THE WORLD BANK

44009

The Rural Investment Climate

Analysis and findings

REPORT NO. 44009-GLB

Agriculture & Rural Development Department
World Bank
1818 H Street, N.W.
Washington, D.C. 20433
http://www.worldbank.org/rural

The Rural Investment Climate

Analysis and Findings

THE WORLD BANK
AGRICULTURE AND RURAL DEVELOPMENT DEPARTMENT

1818 H Street, NW
Washington, DC 20433
Telephone 202-473-1000
Internet www.worldbank.org/rural
E-mail ard@worldbank.org

Cover photos by World Bank staff members: Household in Tanzania by Scott Wallace; Rural Enterprise in Sri Lanka by Yosef Hadar; Community in Nicaragua by Curt Carnemark

Contents

Figures

Tables

Boxes

Foreword

Improving the investment climate constitutes one of the strategic pillars in the World Bank Group's strategy to reduce poverty. Methods for assessing the investment climate have been developed over the past decade, and assessments are being mainstreamed in dozens of countries. Although the concept of investment climate is a generic one, the assessment tools developed have focused primarily on larger formal enterprises in the manufacturing sector. However, the applied sampling strategy for registered enterprises has been ineffective in rural areas, where most enterprises are small and informal.

In 2002, when the World Bank Group was preparing its Agriculture and Rural Strategy "Reaching the Rural Poor," conditions in rural areas, where 70 percent of the world's poor people are living, were viewed as being significantly different from those in urban areas. This view was subsequently endorsed in the World Development Report 2008, which focused on rural development problems in developing nations. Hence, the idea was born to modify investment climate assessment methodology to capture much needed information on rural enterprises and their particular growth constraints. A program of six survey-based pilot assessments was launched, initially in Nicaragua, Sri Lanka, and Tanzania and later, when more resources became available, in Benin, Ethiopia, and Indonesia.

Creating tools for assessing the investment climate is very much a process of learning by doing. It includes three aspects: creating a cost-effective survey methodology, establishing an efficient methodology for analyzing survey results, and applying survey findings in rural development policy analysis and intervention design. The lessons learned from the three first pilots, as well as from two other pilots then in advanced stages, were the focus of the first study in this series—*The Rural Investment Climate: It Differs and It Matters* (2006). The present report details the second comprehensive study on the rural investment climate. It builds on the valuable databases now available from the first three pilot studies as well as their assessment reports and a training course and toolkit for practitioners developed by ARD in 2007. Grounded in the use of advanced methodologies, the study explores the use of these

databases in designing better strategies for rural development. Several recommendations are made to help task managers refine survey designs and methods of analysis in future RIC assessments—a vital aspect of using the results of these important assessments in country and sector dialogues. I hope this work will inspire more and better-targeted efforts to promote private-sector development in rural areas.

Jeurgen Voegele
Director, Agriculture and Rural Development

Acknowledgments

This report was prepared by a team led by Kees van der Meer (consultant) and consisting of Wim Vijverberg and Richard Burcroff (consultants). Valuable support was provided by Yyannu Cruz-Aguayo, Andres Lopez, Isabel Rodriguez Tejedo, and Cateryn Vucina Banjanin (consultants). Peer reviewers were Regina Birner (IFPRI), Mary Hallward-Driemeier (DECRG), Stephen Mink (AFTSN), and Andrew Stone (MNSED). This work had three task-team leaders. It started with Kees van der Meer, and Renate Kloeppinger-Todd led the work following Mr. Van der Meer's retirement from the Bank. The work was concluded under the leadership of John Lamb, who actively participated in the final stages of the report. Mona Sur (ARD), Josef Loening (AFTAR), and Fan Qimiao (WBIFP) provided many helpful comments during the preparation of the report. The task team thanks Jeurgen Voegele (Director, ARD), Mark Cackler (Sector Manager, ARD), Kevin Cleaver (former Director, ARD), and Sushma Ganguly (former Sector Manager, ARD) for their support and guidance.

This study is based on findings from rural investment climate pilot studies led by Mona Sur and Ismail Radwan (Sri Lanka), William Leeds Lane and Josef Loening (Tanzania), and Francisco Pichon and Norman Bentley Piccioni (Nicaragua). Renate Kloeppinger and Ram Ramaswami provided background papers on rural finance. Information about the pilot studies is available at http://intranet.worldbank.org/WBSITE/INTRANET/SECTORS/INTARD/0,,contentMDK:20592746~pagePK:210082~piPK:210098~theSitePK:335808,00.html.

From the beginning of this research project, colleagues from the Bank's Financial and Private Sector Development Network and DECRG, Andrew Stone in particular, regularly provided advice on methodology based on their experiences with generic investment climate assessment tools. The econometric modeling forming the analytical core of this study was developed by Wim Vijverberg (consultant).

Preparation of the report and the further development of RICS methodology were supported by the Bank Netherlands Partnership Program (BNPP); the pilots were supported by BNPP and the Norwegian Trust Fund Private Sector and Infrastructure (NTF-PSI).

Abbreviations and Acronyms

ADB	Asian Development Bank
ANOVA	Analysis of variance
BI	Benchmark indicator
DPR	Development Policy Review
EA	Enumeration Area
EICO	Enterprise investment climate outcomes
EICI	Enterprise investment climate interactions
GDP	Gross domestic product
GVA	Gross value added
IC	Investment climate
IFPRI	International Food Policy Research Institute
LSS	Living standards survey
NFHE	Nonfarm household enterprise
NGO	Nongovernmental organization
NVA	Net value added
PICS	Productivity and Investment Climate Surveys
PRSP	Poverty Reduction Strategy Plans
RIC	Rural investment climate
RICS	Rural Investment Climate Survey
RIC1	The first study in this series: *The Rural Investment Climate: It Differs and It Matters* (2006)
RIC2	The present study: *The Rural Investment Climate: Analysis and Findings*
RNFE	Rural nonfarm enterprise
WBG	World Bank Group
WDR	World Development Report

Executive Summary

CONTEXT AND BACKGROUND

Interest in investment climates has emerged relatively recently. In the 1960s and 1970s, governments in many countries believed they should play a direct role in rural credit, input supply, production, trade, transport, distribution, and even marketing. However, in the 1980s and 1990s, government-dominated systems fell into disgrace because of poor performance.

This led to a push for privatization and liberalization, after which conditions for the development of markets and private enterprises were addressed—in short, the investment climate. Initially, investment climate surveys were developed to determine empirically the constraints inhibiting entrepreneurs from expanding their economic activities. Yet early efforts focused primarily on formal and larger manufacturing enterprises in urban areas.

Only in recent years, and especially after the World Conference on Sustainable Development ratified the Millennium Development Goals, has the attention of the development community broadened to include other sectors and smaller enterprises. Recognizing the cardinal significance of the investment climate to economic growth and poverty reduction, the World Bank devoted to the topic its 2005 World Development Report—*A Better Investment Climate for Everyone* (World Bank 2005).

For the rural sector, the primary focus had traditionally been on agriculture, particularly commercial agriculture and agribusiness, which were perceived to be the main drivers of rural growth. There was not much interest in other rural enterprises, since they were thought to be almost fully dependent on agriculture and unimportant to the dynamics of rural economies. This changed in the 1990s, however, as donors focused more on poverty reduction. Household surveys designed to gather new information on the sources of rural household income were then formulated and conducted. As results came in, more attention was given to the importance of rural nonfarm enterprises (RNFEs) to rural livelihoods. Work by Reardon et al (2001) initiated a stream of research efforts recently summarized in a major study led by IFPRI (Haggblade et al. 2007). Within the World Bank, the 2003 rural strategy "Reaching the Rural Poor" realized the importance of the RNFE, and the 2008 WDR *Agriculture for Development* embraced the view that a sound rural investment climate and a rapidly growing agriculture sector are basic ingredients for a dynamic rural economy.

With the recognition of the importance of RNFE, the direct policy question becomes: what can be done to further stimulate it? In most countries very little information is collected on rural enterprises, their constraints, and their potential. To fill some of these policy-critical gaps, the Bank in 2004 launched pilot studies in Sri Lanka, Nicaragua, Tanzania, Indonesia, Benin, and Ethiopia. The main task was to develop a methodology with three salient features:

(i) design and conduct of cost-effective RIC surveys;
(ii) reliable analysis of survey results and interpretation of policy implications; and
(iii) effective intervention in rural development.

For each RIC pilot study, a country team was charged with conducting the surveys, analyzing the data, and compiling a RIC Assessment report. The assessments add substantive information on several issues relevant to the rural IC that were not appropriately addressed in the urban IC surveys and that were overlooked in agriculture-based studies. Yet further use can still be made of the available data. In these analytical efforts only parts of the collected data were explored. Important avenues of analysis were not fully covered due to limitations in time and budget and to the lack of readily available tools for analysis.

Initial results from this survey-based assessment were published in June 2006 under the title *The Rural Investment Climate: It Differs and It Matters* (World Bank 2006). That report mainly focused on survey methodology[1] and descriptive findings. The present second study was announced as a follow up that would delve mainly into analytical questions and the use of findings in shaping development policies. The present study systematically explores the RIC databases for three of the six pilot countries—Nicaragua, Sri Lanka, and Tanzania—with the following objectives:

- to provide broader and deeper understanding of rural nonfarm activity in rural areas, its constraints, and possible ways to mitigate those constraints;
- to initiate a method of benchmarking the investment climate in rural areas; and
- to advance and sharpen the methodology of such analysis and to provide guidance for analysis conducted by future survey teams and policy analysts.

CONTENTS OF THE RIC2 STUDY

What Has Been Covered?

Four analytical chapters form the core of this study.

Chapter 3 addresses the question of how the investment climate affects ***enterprise performance***, as indicated by both sales and net value added. The investment climate is measured by conditions in the community, as shown in the benchmark indicators and other community characteristics. Community factors, including benchmark indicators, are in general highly influential, but, not surprisingly, they capture only part of the variance.

Chapter 4 investigates household ***entrepreneurial choices***, which are twofold. One set of choices refers to the activities pursued by a household. Of primary interest is the choice to engage in entrepreneurial activity, but this choice interacts with the choices to seek wage employment and to operate a farm. The second set of household choices refers to the decision to start up an enterprise. The models applied explain each of these choices, with reference to household characteristics (structure, skills, ethnicity, non-labor income, and assets), benchmark indicators and their components, and other community characteristics.

Chapter 5 examines ***responses of entrepreneurs concerning their perceptions of the investment climate.*** One set of responses concerns a list of potential barriers to the operation and growth of the respondent's enterprise; the other describes interactions with and observations about government. Explanatory variables are the characteristics of the enterprise, the benchmark indicators and their selected components, and community characteristics. One prominent conclusion of the econometric analysis is that the community random effect is always important: community "unobservables" influence the perceptions of the entrepreneur regarding the investment climate.

Chapter 6 explores ***differences between communities.*** It makes a cross-country comparison of benchmark indicators and explores how benchmark indicators are correlated with community-level characteristics and prices collected at the community level. It also tries to explain community-level indicators of economic activity (outcomes), such as enterprise density, income per capita, nonfarm income shares, and enterprise productivity through simple regression on benchmark indicators and a number of other economic descriptors.

What is New About This Study?

Several features of this study are innovative, and each serves to broaden the methodological frontier available for rural investment climate studies. Innovative features include:

1. A comprehensive analysis of the RICS household data.
2. More systematic use of community data than in previous studies.
3. Development of benchmark indicators and their use in the analyses.
4. Analysis of responses about constraints in the investment climate, using the concepts of Enterprise Investment Climate Outcomes (EICO) and Enterprise Investment Climate Interactions (EICI).
5. Novel Stata programs developed for random effect analysis with weights.
6. Use of estimates of the contribution of groups of variables—enterprise, industry, and community characteristics—in explaining variation of outcome variables.
7. Household and enterprise weights used for Nicaragua and Sri Lanka. (For Nicaragua, household and enterprise weights lost during data processing were reestimated. For Sri Lanka, the enterprise weights used by the Sri Lankan team were readily available, but household weights had to be reestimated.)
8. Collection of community-based data that allowed RIC to introduce a spatial dimension not prominent in PICS. This expansion opens new options for analysis of differential enterprise development in a heterogeneous rural space and along the rural-urban axis. The clustering of observations by community raises complex analytical and methodological questions.

The methods and tools developed herein can be employed by policy analysts for future IC studies.

CONCLUSIONS AND RECOMMENDATIONS

What Has Been Learned About the Three Economies?

The following important findings were made about the economies of Nicaragua, Sri Lanka, and Tanzania:

1. The household data show that simultaneous involvement in RNFE self-employment, wage employment, and farming is common among rural households. Available choices for individual households and household members differ with family structure, human capital, assets, and community characteristics. The RNFE self-employment choice fits within a broader livelihood strategy of pursuing opportunities and managing risks. The view held by some analysts that RNFE self-employment results solely from push factors felt by the poor is much too narrow. This study provides evidence that households engaged in RNFE tend to be better off than farming families.
2. The bulk of rural enterprises are very small in size, with only one or a few workers. The large majority of enterprises (68 to 86 percent) only employ unpaid family workers. Enterprise population is not static, however. By comparing the three countries and differences within the countries, a general pattern emerges showing that, with increased per capita income, both enterprise density per 1,000 inhabitants and average enterprise size increase.
3. Large numbers of small rural enterprises buy and sell mainly locally, which is understandable from the nature of their businesses: services, retail trade, repair, and so on. Yet, evidence from Sri Lanka shows that enterprises selling mainly outside their communities have higher productivity.
4. The share of agro-processing in rural enterprises ranges from only 2 percent in Sri Lanka to 14 percent in Nicaragua. This shows that a development focus on agribusiness as the transformation of raw products represents too narrow a target for private-sector development in rural areas.
5. In all three countries, enterprise productivity appears to differ with enterprise and community characteristics, leading to major differences within and between countries. Generally, enterprise age and entrepreneurial experience matter, indicating the importance for increased productivity of adaptation and innovation through learning by doing.
6. As diverse as are the three countries in this analysis, the list of their most important

investment climate concerns shows remarkable overlap. On the top-ten list of concerns averaged across the countries, six items are found on the individual top-ten list of each country and three more occur among two of the three countries' top-ten concerns. The common concerns involve (i) the cost, availability, and procedures of finance; (ii) electricity and water (access, cost, or reliability); and (iii) road quality and access. The other top-ten concerns are more varied: market demand, economic policy uncertainty, and telecommunications. At the other end of the spectrum, probably because of their enterprises' small scale, few rural entrepreneurs in any country see as obstacles: food safety regulation, regulation on land use, customs rules, or work permits for expatriates, to mention a few.

7. It is often said that access to credit is the factor most constraining to RNFEs. This study found that the large majority of enterprises (58 to 79 percent) reported having no formal or informal debt, and only 1 to 31 percent reported formal debt of more than 50 percent of equity. Although entrepreneurs may genuinely desire inexpensive credit, the benchmark indicator and benchmark components for finance services do not reveal credit as having a strong effect on enterprise productivity.
8. Improved infrastructure provision alleviates perceived investment climate obstacles related to infrastructure services. When more households in a community have access to protected water sources, for example, entrepreneurs complain less about water. The same result was found with respect to electricity, telecommunication, and postal services.
9. Enterprise performance; household choices about entrepreneurship, farming, and wage employment; and perceptions about investment climate obstacles are all affected by the community's economic environment. Observed community characteristics (including benchmarks) matter, but apart from these, regression analysis shows that unmeasured community factors are highly influential in causing similar choices and behavior within a community.

Policy Perspectives

The RIC results can contribute in several ways to preparing effective policy interventions.

1. The RIC survey results provide considerable otherwise unavailable information about the economic activities of rural households, local enterprise populations, and institutional, policy, and infrastructural constraints. These findings can importantly inform the policy dialogue between civil society—especially private-sector stakeholders—and government. The quality of dialogue can be enhanced by good RIC analysis, and the further assessment of options facilitated by the RIC surveys.
2. The various regressions—enterprise performance and entrepreneurial choice—show that, from micro-perspectives, various factors constrain productivity, employment, investment, and income earning. Since many of these constraints can be alleviated (some more readily than others), the RIC analyses can identify and document challenges for policy intervention.
3. In many cases, general RIC analysis as outlined in this report can and should be followed up by further exploration of the RIC databases, with regard, for example, to the specificity of constraints, locations, and groups of enterprises affected. RIC surveys create multipurpose databases of a public-goods nature that for several years hence can be used as information sources to support policy and project design. Moreover, additional information can be obtained from other sources and through targeted interviews and small-scale follow-up surveys. In fact, information from RIC databases will also serve to improve the designs of other kinds of rural surveys.

Statistical and Analytical Limitations

The economic environment of enterprises consists of large numbers of factors that individually and in combination shape the options for entrepreneurial decisions. Many of these factors interrelate in complex ways. This basic situation creates major unavoidable challenges offering analysts no simple solutions. The main challenges are:

- The number of potentially relevant variables that have been collected is generally too large to include simultaneously in econometric models. The pilots cover only 100 to 150 communities, roughly equal to the numbers of variables; hence either information must be condensed or variables must be handpicked.
- The investment climate concept is broad and multifaceted. Because of restrictions on questionnaire length, each facet can only be covered in limited detail if all facets are to be addressed. If one facet is found to be influential, further details always remain to be discovered.
- Many relevant factors cannot be well observed or adequately measured using surveys, and these play a role in the analyses as unobserved background variables.
- Communities often pursue improvements in the economic environment in many ways simultaneously. The interrelationships among variables easily results in problems of multicollinearity, which, together with the problems of unobserved variables, can result in biased and imprecise parameter estimates.

There are no quick solutions to these challenges: analysis requires use of different tools, focused on different questions, and addressing different aspects of the database.

Implications for RIC Methodology

Below are some of the conclusions and recommendations that follow from consideration of the RIC methodology and its further application.

Survey design

1. Stratified sampling is indispensable for drawing a cost-effective sample. This implies that sampling weights are absolutely necessary in the analysis for obtaining unbiased results. With the loss of weights, as in the case of Tanzania, much of the value of the RICS is lost.[2]
2. The length of the standard questionnaire should be substantially reduced through careful revision.

Quality of data collection

Experience shows that the cost-effectiveness of RIC surveys can be greatly improved by the following:

1. *Improvement of the quality of data collection and processing*.[3] Poor quality data collection and processing has a major effect on the outcomes and efficiency of RIC work. It limits potential output and the precision of findings. It also increases the cost of analysis, since the time required for analysis is much greater than for good-quality data sets. If fewer records must be discarded because of data-quality problems, sample sizes can be reduced somewhat or coverage increased.
2. *Reducing the length of questionnaires*. Many data have been collected in the pilot RIC surveys that were not used in the country reports or in the analysis presented in this report. This is understandable because of the pilot character of the three surveys. With the benefit of experience, however, a careful revision leading to a substantially shortened questionnaire would be the logical next step.
3. *Further standardization of questionnaires*. Country teams were generally inclined to design country-specific variations of the survey designs. This contributed to an excessive collection of unused data while reducing inter-country comparability. Similarly, standardization of the structure of the RIC databases across countries will enhance the productivity of empirical investigations.

Further Research

Many potential future research topics present themselves. Among these are:

1. The analytical tools described in this report deserve to be applied with priority to other available data sets, such as those for Bangladesh, Benin, Ethiopia, Indonesia, and Pakistan. This will add to understanding of the similarities and variance in rural economies and to further sharpening of the analytical methods applied.
2. A further build-up of analytical tools is warranted. Future studies should elucidate the interplay between perceptions about the investment climate and the actual conditions in the community and how this determines enterprise performance and entrepreneurship choice, with careful attention to bidirectional causality and self-selection issues.

3. Further exploration of the rural urban continuum and rural urban divides is warranted. One approach would be to conduct comparative studies using the RICS and PICS databases. Another is to make use of the RICS database codes for communities' geographic longitude-latitude information. These fix communities in rural space, but their spatial implications have not yet been explored at all. Spatial regression methods could exploit correlations among nearby communities, an approach that has been widely ignored in applied research to date.
4. The various trade-offs (in terms of statistical efficiency) between number of communities and number of observations per community can be analyzed using existing models and databases. In particular, a few communities in the existing databases contain an unusually small number of households and enterprises, while a few others contain unusually large numbers. This aspect of sampling raises the questions: What effect do these types of communities have on the reliability of results? Given the RICS methodology of using enterprise response aggregates to describe aspects of the investment climate, what is, in fact, the optimal number of observations per community?
5. Just as for other large area surveys, such as governance, cost of doing business, and PICS, benchmark indicators are needed for comparison and analysis. The experience with use of the benchmark indicators in the regressions is encouraging, but further analysis is needed to sharpen and optimize their conceptualization and measurement and to render the indicators more comparable across countries. The present estimation models can be used to test the effect of changes in definitions of the benchmark indicators. The addition of other countries to the sample, say Bangladesh, Benin, Ethiopia, Indonesia, and Pakistan, could greatly enhance the effectiveness of this study's work on benchmarks.

1 Introduction

1.1 ASSESSING THE RURAL INVESTMENT CLIMATE

A sound rural investment climate and rapidly expanding agriculture are basic ingredients of a dynamic rural economy, according to the 2008 World Development Report, *Agriculture for Development* (World Bank 2008b). Recognizing the cardinal significance of investment climate to economic growth and poverty reduction, the World Bank devoted to the topic its 2005 World Development Report, *A Better Investment Climate for Everyone* (World Bank 2005).

Interest in investment climates has emerged relatively recently. In the 1960s and 1970s, governments in many countries believed they should play a direct role in rural credit, input supply, and the production, transport, distribution, marketing and trade of certain products, especially agricultural. In the 1980s and 1990s, however, government-dominated systems fell into disgrace because of their poor performance. Although privatization and liberalization were often necessary to stabilize economies and provide a basis for economic growth, in many cases these did not result in a quick response from private investors. The long legacy of state-controlled and parastatal-managed markets left underdeveloped the institutions and policy frameworks for privately led markets. For those reasons, renewed attention is now being given to the conditions under which markets and private enterprises develop—in short, to the investment climate (Box 1.1).

Market and policy failures can present obstacles to private enterprise development, which governments can help to overcome through direct and generic interventions. Direct interventions target individual enterprises or groups of enterprises, aiming, for example, to improve access to finance, technology, markets, or information. By contrast, generic interventions aim to strengthen the enabling environment for enterprises with, say, improvements to the legal and regulatory framework, fewer administrative burdens, improved infrastructure, and a better functioning financial sector. Direct and generic interventions do share areas of overlap. Even so, with the exception of infrastructure and agriculture policies, the general tendency in rural development has mostly been to support enterprise development through direct interventions while neglecting broader investment climate issues.

Because of very limited information, the largely informal firms in rural areas have been absent from the more analytically oriented

Box 1.1 What Is *Investment Climate*?

The investment climate (IC) consists of the political, administrative, economic, and infrastructural conditions for getting a reasonable return on investment as perceived by potential private investors. It is a subjective appraisal by entrepreneurs of the enabling environment.

Peace is crucial, as are prevailing conditions of law and order conducive to private-sector development. There must be reasonable macroeconomic stability, financial stability, a realistic exchange rate, and low inflation. Supportive business laws, property rights, and bankruptcy laws should be in place and enforced by clearly defined judicial authorities. The tax system and taxation should not be disruptive for business.

Important elements in the regulatory framework are free entry for new enterprises; freedom to operate without the need for difficult-to-obtain permits and without heavy administrative burdens; freedom to trade domestically and internationally without serious administrative obstacles; and functional competition laws, auditing requirements, industrial standards, and market regulations. Policies should enable, but not distort markets. Good governance should also be evident, including transparent policymaking and absence of corruption among and harassment by public servants.

Also of importance are the quality and availability of public services; availability of a healthy, educated, and skilled labor force; availability of finance services; and availability of infrastructure-based services such as transport, telecommunications, post office, power, water, and sanitation.

Source: The Authors.

studies of rural growth and employment. As a consequence of this lack of attention, as well as the almost complete reliance in the literature on household-based samples, the profession until only recently knew very little about the nature and size distribution of rural firms, the constraints they face when trying to expand or even to survive, the significance of their effect on the broader rural economy, and even about policies that could spur investment by rural nonfarm enterprises (RNFEs).

Though still in the pilot stage, the Bank's small portfolio of RIC assessments has already identified concrete actions that can improve the investment climate in rural areas (Box 7.2). To date, the methodology for assessments has been tested, developed, and refined, and directions for overall improvements can now be teased out of the pilots' designs for data analysis.

Improving the rural investment climate likely will involve a differing set of remedial actions, consisting of a tableau of nuanced priorities for the kinds of public-sector support needed for RNFEs, distinct from and in contrast to intervention geared toward stimulating investment by the more formal urban-based enterprise sector.[4] Targeting the rural investment climate is consistent with a rural development policy that stresses the importance of a thriving private sector in rural areas as the best means of achieving growth, creating employment, and reducing poverty—a policy that is, in fact, the focus of the World Bank Group's approach to rural development.

Determining the shortcomings in the rural investment climate poses empirical questions. What hinders enterprises in one country may not bother those in another. What constrains them in one region or community may differ from what holds them back in another. Such questions are not easily answered. Partly because of limits on available sample frames, the universe of largely informal firms in rural areas has been practically absent from any type of analysis.[5] As a consequence of this lack of attention, the profession until only recently knew very little about the nature and size distribution of rural firms, the constraints they face when trying to expand or even to survive, the significance of their effect on the broader rural economy, and the policies that could spur their growth.

To fill some of the policy-critical gaps in information about the nonfarm rural sector, the Bank in 2004 launched a project, the first of its kind, to collect enterprise-level data on the rural investment climate. Initial results from this survey-based assessment were published in June 2006 under the title *The Rural Investment Climate: It Differs and It Matters* (World Bank). The study featured prototype instruments and designs for data analysis grounded in surveys at the community, enterprise, and household levels. Supplementary information about prices and unit costs were also surveyed to gain a handle on factors contributing to the profitability of the surveyed enterprises as well as the transactions costs of doing business in the surveyed areas. With adaptations to satisfy local

concerns, the survey instruments were administered as pilot studies in Sri Lanka, Nicaragua, Tanzania, Indonesia, and Benin. Currently, the final pilot assessment is underway in Ethiopia. The mainstreaming of RIC assessments by the Bank's Regional Offices is underway or in advanced planning stages in Bangladesh, Cameroon, Mozambique, Nigeria, Pakistan, and Zambia. These RIC assessments have added or will add substantive information on several issues relevant to the rural IC that were not appropriately addressed in urban IC surveys and were overlooked in agriculture-based studies.

Developing tools for assessing the investment climate has been very much a process of learning by doing. It includes three aspects: a methodology for cost-effective surveying, a methodology for analyzing survey results, and application of the findings to policy analysis and the design of rural development interventions. The lessons learned from the pilots are most encouraging and have provided the basis for both the present study and for follow-up work to the 2006 study, including a RIC Implementation Manual (World Bank 2007c) summarizing the lessons learned while implementing the pilot surveys and offering revised prototype survey instruments for use in future RIC assessments.[6]

The need remains, however, to improve the RICS by developing more rigorous analytical methodologies than those used in many of the pilots. New analytical tools have been developed for this study for advanced modeling of investment determinants, factors underlying entrepreneurial decisions to open new enterprises, and the analysis of reported IC constraints. This study also develops benchmark indicators that scale community factors (features) along several axes to test their influence as potential facilitators of or exogenous constraints on the local RIC and thus on enterprise performance and entrepreneurial behavior.

The ongoing modifications also aim at standardization, which will make RIC data more comparable and encourage the analysis of RIC data sets as they become available—both inside and outside the Bank. Expanding the base of available RIC assessment results will also help make the sum of all results more usefully inform policymaking. For this purpose, the RIC study team recommended that ten additional RIC assessments be commissioned over the next four years (World Bank 2007c). Within the Bank, the Rural Anchor and its component units are well placed to disseminate the methodology as it develops and to provide expertise and financial resources to the teams implementing projects in the regions.

1.2 IMPLICATIONS FOR THE RURAL AGENDA

The International Food Policy Research Institute led a worldwide, stage-setting review of developments in the rural nonfarm economy, resulting in *Transforming the Rural Nonfarm Economy* (Haggblade et al. 2007). The publication provided comprehensive, in-depth assessments of off-farm employment growth and the expansion of rural enterprises in several developing regions. It also discussed incentives and constraints, including the fostering role played by dynamic supply chains as enabling factors for enterprise start-up and expansion, and it concluded with several general recommendations for development policy. For policy environments "where the basic components of an incentive system favorable to rural business are in place activity," the study identified specific promotional activities that can "accelerate rural growth as well as the participation of poor households in the rural nonfarm economy."[7]

Key promotional activities cited were: (i) identification of potential engines for rural growth, (ii) diagnostic evaluation of supply-chain dynamics to identify points and types of strategic interventions, and (iii) construction of flexible institutional coalitions for implementation. The report then offered detailed guidelines for effectuating promotional activities, inclusive of the role (and limitations) of decentralized implementation, the government-centric role contained in most Poverty Reduction Strategy Plans (PRSPs), and involvement of NGOs. Such promotional activities do not directly address the main elements of RIC reform, however, as it is presumed that the necessary improvements will already be in place. Nonetheless, as RIC improvements are likely to be introduced in phases, subsequent reforms would no doubt be influenced and supported by the above promotional activities, which in turn would likely influence the identification and nature of programs and investments that could further improve the rural investment climate. In sum, IFPRI's report provides a valuable reference for use during the "stage-setting" phase of RIC assessment design.

Generally, mainstreaming improved RIC assessments into World Bank operations and rural development programs will ultimately lead to a more informed policy dialogue, built on rigorous

supporting analysis. The analysis will also provide private-sector actors with improved information on which to base investment decisions—the increased confidence that comes from better information is in itself an important element of the rural investment climate. For the World Bank, the information the assessments provide will inform the design and monitoring of operations that support rural enterprise development. It will also improve our understanding of the structure and functions of the rural economy generally, beyond the parameters of individual economies—providing insight into agricultural and rural economics and their place in the global economy.

1.3 STUDY OBJECTIVES

RIC1, the initial study in this series (World Bank 2006) documented the need to address the rural investment climate separately from on-going Bank surveys of the general investment climate known as Productivity and Investment Climate Surveys (PICS).[8] Its findings and conclusions were presented mostly in the form of cross-tabulations and derived mainly from subjective assessments of the rural investment climate by enterprise owners and managers. The study reviewed technical and methodological issues encountered by the country survey teams in data collection and provided guidance for future surveys. The study indicated that statistical analysis of the data sets would be pursued in a subsequent study—this study, RIC2.

RIC2 provides a comprehensive, largely econometric analysis of the data collected in the pilot surveys with respect to household, enterprise, and community characteristics. Focusing mainly on the data sets for Nicaragua, Sri Lanka, and Tanzania, it analyzes relations between household and community characteristics and enterprise dynamics and performance. The objectives are:

- to provide broader and deeper understanding of nonfarm activity in rural areas, its constraints, and possible ways to mitigate those constraints;
- to initiate a method of benchmarking the investment climate in rural areas; and
- to advance and sharpen analysis methodology and to provide guidance on analysis to future survey teams and policy analysts.

RIC1, the Implementation Manual (World Bank 2007c), and RIC2 together will provide guidance useful for any team designing and conducting surveys, analyzing the data obtained, and using the findings in policy dialogue and preparation of policy interventions.

Audience

The following chapters contain several key lessons learned from the pilots, along with suggestions for robust approaches to data analysis and pointers for improving the rural investment climate. Although it was constructed primarily for the benefit of the development community—including governments, donor agencies, and larger NGOs—this publication constitutes a useful reference for survey managers and staff, for the agencies responsible for the assessments, and for independent policy research entities.

1.4 ORGANIZATION OF THE REPORT

This study's essential findings and policy implications are organized into the following six chapters. Annexes provide supporting material. Chapter 2 presents an overview of related work and of the literature; it also describes the subsequent chapters' methodological framework, including new ways of addressing questions of endogeneity in these kinds of surveys while seeking to isolate cause and effect. Chapter 3, by applying econometric analysis, measurably extends the examination of enterprise performance and investment climate constraints initiated in RIC1. A rigorous examination of enterprise dynamics and entrepreneurial choice is developed in Chapter 4. Aiming to highlight the differing effects on RNFEs, Chapter 5 draws together the main implications RIC2 findings on the rural investment climate in the three country pilots. Community-level influences also matter, and Chapter 6 examines how the local IC and other community characteristics shape the environment for economic activity. Conclusions and recommendations appear in Chapter 7, including suggestions for using RIC results for policy reform and for targeting the rural public expenditures needed to foster improvements in the rural investment climate. The Annexes describe the databases employed and the methodologies used in the study, as well providing detailed regression results.

2 Approach and Methodology

RIC2 is grounded in methodological extensions of the existing literature, but it also offers novel approaches to modeling enterprise performance and dynamics, investor behavior, and the outcomes of entrepreneurial choices in the face of rural investment climate constraints. This chapter first reviews the findings of prior research, both in the general literature and in specific assessments of RIC data, and then explains the twofold methodological contributions of RIC2: (i) the construction of benchmark indicators and the performance testing of these indicators as explanations for enterprise performance, entrepreneurship choice, and community-level influences on the IC faced by RNFEs, and (ii) new empirical models and methodologies addressing previously unexplored features of the RIC database.

2.1 LAYING THE GROUNDWORK

2.1.1 General Literature

Between two-thirds and three-quarters of the estimated 1.2 billion people living below the one dollar a day poverty line are estimated to live in rural areas (World Bank 2008b). While agricultural productivity can make an important contribution to improving rural welfare, recognition is growing that the poor rely on a diversified income portfolio to which the rural nonfarm sector makes an increasingly important contribution (Reardon et al. 2001; Barrett et al. 2001; Lanjouw and Lanjouw 2001).

Along with Haggblade et al. (2007) and the recent World Development Report (World Bank 2008b), Lanjouw and Lanjouw (2001) offer a thoughtful, comprehensive review of the rural nonfarm sector. Among its insights was the observation of the widening recognition in recent years of the contributions of the rural nonfarm sector to economic growth and rural employment. The authors review and document the size and heterogeneity of the sector in developing countries worldwide, pointing to evidence that the sector in many countries is expanding rather than declining and that a positive effect on the poor, while by no means inevitable, can be considerable.

The literature reveals differences, however, concerning the role of geography. Larson and Cruz-Aguayo (2008) point out that research has often emphasized the distinction between rural and urban incomes rather than sectoral differences. This distinction becomes more than a convenient expository device when one recalls that explicit motivations for migration are often tied to urban settings.

Some arguments relate to the accumulation of human capital, viewed as an important determinant of a successful transition from agriculture to employment in other sectors. Also, cities may offer a range of supporting markets and economies of scope that are lacking in rural areas.

Methodologically, most studies of nonfarm income are set up to explain labor participation decisions or indicators of specialization, such as shares of income derived from off-farm sources. Although less typical, a few household studies address the complementary issue of how households faced with productivity constraints allocate existing resources among a portfolio of activities. Such studies take a structural approach that allows comparison of average returns to quasi-fixed and flexible classes of labor and capital assets.[9]

Much recent literature on enterprise performance focuses (understandably) on Sub-Saharan Africa. Poor business environments often mean high costs for certain services important to manufacturers. Eifert, Gelb, and Ramachandran (2005) show that African firms face high costs (for transport, logistics, telecommunications, water, electricity, land and buildings, marketing, accounting, security, and bribes) compared with Asian firms and that African firms suffer substantial losses from power outages, crime, shipment losses, and the like.

African financial markets are the least developed in the world, and development economists have long held that this impedes growth. Bigsten et al. (2003) examined formal credit market participation and credit constraints based on 1991–95 survey data. They found that the demand for formal loans among African manufacturers was quite low and that credit market imperfections translate into binding constraints only if firms have a desire to invest.[10] As uncertainty hampers investment, Bigsten and Soderbom (2006) conclude that governance is likely central to reducing risk. Meanwhile, an important issue is how to accommodate Africa's large informal manufacturing sector, given the high costs associated with formality.

The question of informality has also been studied in the Latin American context. Building on seminal work by de Soto (de Soto 1989), Perry, Maloney, et al. (2007) review the ubiquity and context of enterprise informality in the LAC Region and offer several thought-provoking insights and conclusions. In practice, informality discourages growth in firm size, as the managers of informal enterprises seek to reduce informal operating costs by remaining beneath the radar of the licensing authorities and the scrutiny of tax system operatives. The authors view informality as essentially a transitional phenomenon, a condition expected to fade away as economies develop and state regulation and services transition in parallel from obstructive to more facilitating stances. But this will take time. Thus a more sympathetic understanding of the reasons why so many firms elect to remain informal—coupled with a more supportive system of governance and visible, sustained efforts to discourage rent-seeking by government officials—will likely increase the appeal of registering small enterprises as formal, legal entities.

Vijverberg (2005) reviewed the adequacy of using living standards survey (LSS) databases to examine the rural investment climate and nonfarm entrepreneurship in Ghana, Guatemala, the Kyrgyz Republic, and Vietnam. The study asked whether "a multi-purpose household survey [is] a suitable vehicle in order to measure the effect of investment climate (IC) variables on non-agricultural selection and performance," and it concluded that "while the IC content is informative, it remains limited." Vijverberg found that the main gaps in the LSS community questionnaire concern the topics of credit, business organization, and local governance. Further, the most elaborately designed enterprise module, while recording enterprise outcomes such as size, sales, income, and employment, failed to shed light on the effect of the economic environment on the performance of the enterprise.[11]

Two of the most recent and more comprehensive contributions to the literature are the WDR (2008b) and Haggblade et al. (2007).[12] Focused mainly on the policy and program requisites for stimulating primary agricultural production, the WDR offers conclusions and recommendations for improving the rural investment climate that tend to be thematic. Nonetheless, many are germane to the present study.

The WDR (2008b) stressed the importance of off-farm income generation and employment and promoted the potential for stimulating RNFEs through reforms to improve the rural investment climate. This approach is new—no primary rural development document by the World Bank Group or any other international organization has focused so centrally on the importance of the rural

investment climate. In the overview section, the report posits:

- the basic ingredients of a dynamic rural nonfarm economy are a rapidly expanding agriculture and a good RIC, which is as essential for agriculture as it is for RNFE development; and
- the role of the state in promoting agri-food market development includes improving the investment climate for the private sector.

The report goes on to note that indirect costs stemming from a poor IC appear to be highest on average in Africa.

SME development was also a central focus of the WDR, especially in the agro-processing sectors. Here the WDR argues that an improved RIC can foster contestability and competitiveness and thus should be considered as a key policy instrument for stimulating SME development in rural areas. Examples cited include the seeds industry and the contribution to R&D and to the operation of rural labor markets. Noting the still limited capacity of rural financial markets to serve rural financial needs, the WDR observed that much of the investment needs for RNFE development and growth must continue to be financed from rural savings and the gains from private-sector development. The rural investment climate will be an important determining factor, governing both the robustness of investment demand and the volume of "informal" savings made available to new or expanding enterprises. Lastly, the report signaled the lack of a generic mandate imbued in most Agricultural Ministries to deal with issues and reforms aimed at improving the RIC, while other ministries generally have limited interest in stimulating agricultural value chains.

Although the IFPR policy recommendations (Haggblade 2007) presume that most of the prerequisites for improving the rural investment climate will already be in place (see Chapter 1 above) the study offers many pointers germane for advancing the analysis of the RNFE phenomena in developing countries. These pointers helped to establish the agenda for the analytics presented in the following chapters. Key among these are implications concerning (i) enterprise dynamics, and (ii) factors that stimulate rural household income diversification into rural nonfarm activities.

The review of *enterprise dynamics* found that new firms were being created at a high rate; over 20 percent of the firms examined in the studies recorded in Haggblade (2007), Chapter 5, were new entrants. But large numbers (over 50 percent in some cases) disappeared within three years. The annual growth rate of the surviving RNFEs was high, but in most countries a minority of enterprises fueled this expansion.

Jobs created by the expansion of existing enterprises are more likely than start-ups to reflect increasing efficiency and to demand pull forces in the economy. The expanding enterprises show several common characteristics. They tend to be younger, to have started smaller, to be in the manufacturing or service sector, and to be operated by male entrepreneurs from the home. The IFPRI report thus posits that government interventions should focus more on existing enterprises than on new start-ups and that unblocking market constraints further up the supply chain will be key to expanding opportunities for these growing firms.

An intriguing typology is presented in the study's exposition on *household income diversification.* The authors note that household motives for income diversification, as well as the available opportunities, differ significantly across settings and income groups. This suggests that an important distinction should be made between diversification undertaken for accumulation objectives, driven mainly by pull factors, and diversification undertaken to manage risk or supplement stagnant or even declining incomes from farming activities, driven by push factors.

The observed patterns of diversification in the several studies reviewed in the IFPRI report tend to validate the much larger body of anecdotal evidence now available. In richer areas, household specialization in farm or in nonfarm activities is much more pronounced than in poorer zones. While pluriactive households in income deficient areas may not be able to engage in either pursuit very efficiently, they are able to spread risk, compensate for a limited asset base, and generally survive—albeit at a semisubsistence level. This specialization is also reflected when comparing individual households. While many wealthier households are genuinely pluriactive, the individual members tend to specialize to a far greater extent than do the less well off cohort. Furthermore, the available evidence suggests that rural nonfarm income greatly exceeds the value of farm wage-earning, by factors of 4.5–5 to 1 in India and

Latin America and more than 20 to 1 in some parts of Africa.[13]

The determinants of household capacity to diversify into rural nonfarm activities were examined. Skill-based and financial barriers to entry and expansion typically do not deter asset-rich households, whose members generally appear able to cream off the more lucrative RNFE (and other nonfarm) activities. Conversely, asset-poor households remain confined to the low-return segment from rural nonfarm earnings. Referring to "static and dynamic capital holdings,"[14] the study essentially partitioned the householders' asset base into meso- and micro-level assets. Infrastructure availability and its quality and reliability are probably the single most important meso-level asset. Nearly all studies reviewed showed infrastructure to be a significant correlate of rural nonfarm employment, especially roads, power, and water supply. Proximity to urban marketing centers can be another meso-asset. Few studies have used household data to analyze disaggregated rural nonfarm activities with reference to spatiality, but Haggblade et al. record one instance from Nepal showing that rural nonfarm wage employment fell away quickly from peri-urban areas to the rural hinterland, while a U-shaped pattern for self-employment revealed some rural nonfarm activity serves local needs not met by supply from urban areas.

The micro-assets typically resemble private services or are privately held physical assets (for example, land). Good examples of the former are organizational and social assets. These have been relatively underexplored in the rural nonfarm literature but they deserve much more emphasis, as such social linkages can play an important role in reducing risk and transaction costs for households operating RNFEs. The available evidence, though paltry, points to a strong link between membership in organizations and other connections and successful participation in nonfarm rural activities. Some studies also show that the better endowed households are more likely to benefit from such social capital.

Perhaps most significant, however, is the strong correlation shown in several of the studies between human capital in general and participation in and returns from rural nonfarm employment. Education is key, although there may be significant interactivity differences in skills. When schooling effects are disaggregated by types of rural income generation, that is, returns from nonfarming wage labor versus nonfarm self-employment, farming, or farm wage labor, the pay-off from education was highest in rural nonfarm wage labor, although less so in cash cropping and self-employment. Because of the close relationship between education and land ownership, the contribution of the latter to the above noted patterns usually autocorrelates with the influence of education.

In summary, the IFPRI study infers that

> Because of initial differences in asset endowments, rich and poorer household diversify differently. The rich typically engage in more capital- (including human capital) intensive and more remunerative activities, leaving the poor confined to labor-intensive, highly contested niches with low barriers to entry and low returns.[15]

Haggblade (2007) also discusses the potentially transformative contributions of agricultural development to RNFE development via intersectoral growth linkages. From several cross-sectional studies, the authors conclude that sustained RNFE growth and development is most likely to be centered in rural locales experiencing a vibrant agricultural sector. Backstopped by evidence from East Asia and from several locales in India, Africa, and Latin America having a more dynamic agriculture, Haggblade et al. also note that in time the growing strength of urban demand linkages for rural manufacturing will eventually replace stimuli emanating from agriculture and may even eventually absorb or displace RNFEs, as the larger urban firms relocate to the urban periphery or deeper into rural areas.

2.1.2 Prior Studies Using Data from the RIC Pilots

RIC2 provides comprehensive analysis of the data collected in the pilot surveys with respect to household, enterprise, and community characteristics. Deininger, Jin, and Sur (2007) analyzed data sets from the Sri Lanka pilot, and Sundaram-Stukel, Deininger, and Jin (2006) did similar analysis for the Tanzania data set. Both studies explore characteristics of the nonfarm sector and demonstrate quantitatively the sector's economic importance, identifying enterprise characteristics and IC obstacles to sector expansion and productivity. Deininger et al. estimate that RNF value added

amounts to 80 percent of agricultural GDP. The incomes of participating households were found to be significantly higher than those of nonparticipating households. Barriers to entry are low, and nonfarm development only modestly affects inequality, implying a large potential contribution to growth and poverty reduction. Infrastructure constraints negatively affect new start-ups, as well as investment in and productivity of existing enterprises, with small enterprises being especially affected.

The Sri Lanka and Tanzania studies employed probit regression models to estimate and explain factors behind the decision to start up an enterprise and probit and tobit models to explain determinants of new investment in existing enterprises. Explanatory factors included both objective measures from community-level surveys (for example, availability and adequacy of physical infrastructure) and subjective perceptions from enterprise managers about the relative importance of several investment climate constraints. Similarly, estimates of total factor productivity were made to portray the extent to which investment climate constraints reduce the efficiency of resource use in the rural nonfarm economy.

In stark contrast to most of the findings for the formal sector in most countries, where taxation and other regulatory constraints were identified as key constraints, both IC studies found that infrastructure constraints (but not regulatory obstacles) pose a formidable barrier to rural households' participation in rural nonfarm enterprises and to investment and increased productivity by existing ones. These barriers being particularly harmful for small enterprises, policies to improve delivery of the public services in question will be important foundations for a flourishing rural nonfarm sector, which in turn will have an important role in poverty reduction. Lacking proper sampling weights, the reliability of the estimates for Tanzania is open to question; however, the general econometric findings accord well with more anecdotal evidence.

Motivated by the question of whether policies supporting nonfarm businesses can reduce rural poverty, Larson and Cruz-Aguayo (2008) examined the consequences for income of household decisions about family labor, farms, and businesses. Using survey data from the Nicaragua RIC pilot, the authors identified sector-related differences in returns on allocated household assets indicative of market constraints. On the basis of both parametric and nonparametric conditional productivity measures, entrepreneurial activity in rural Nicaragua was found to have a dual character: sometimes representing opportunity and sometimes serving as a refuge for family labor. These results are consistent with the scale-related constraints on asset-poor households affecting rural incomes across all levels of specialization. For most rural households, labor appears to be better rewarded by formal wage markets than by family businesses or farms. In contrast, for wealthier households, differences in productivity do not relate significantly to sector or self-employment choices.

2.2 CONTRIBUTIONS OF RIC2

RIC2 offers several methodological extensions of the existing literature and suggests novel approaches to modeling enterprise performance and dynamics, investor behavior, and the outcomes of entrepreneurial choices in the face of perceived rural investment climate constraints. It also develops a study of perceived investment climate obstacles in light of objectively measured community characteristics. The findings are presented so as to provide methodological guidance for future survey teams and policy analysts.

As a first step, benchmark indicators were developed to summarize the many community-level factors commonly associated with the investment climate. Benchmarks serve two purposes: (i), they may be used to compare countries, and (ii) they are potentially useful devices for explaining enterprise performance, entrepreneurial choices, and patterns in entrepreneurial complaints about IC constraints. Benchmark indicators are described in subsection 2.2.1.

RIC data are gathered in such a way that assumptions behind standard econometric models do not apply. The peculiarities, owing to endogeneity and fixed effects, are shared by other common surveys, such as Living Standards and agricultural surveys. Accommodating these peculiarities requires important modifications to standard econometric techniques, one of the signal features (outcomes) of this report. Section 2.2.2 below discusses these, as well as introducing in greater detail the topics to which new models are applied: enterprise performance, entrepreneurship choice and enterprise start-up, perceived

investment climate obstacles and interactions between entrepreneurs and government, and differences in the rural investment climate between communities.

2.2.1 Benchmarking

The RIC household, enterprise, and community questionnaires ask several hundred questions about various aspects of the geographic, institutional, economic, and social environment faced by RNFEs. The variables that measure these aspects are far too numerous for direct comparison between countries, among regions within a country, or even among communities in the data set. They are also far too numerous for simultaneous inclusion in the study's regressions. A common approach in policy analysis is to compute benchmark indicators for which information from many variables—often covariant ones—is scaled and synthesized.[16]

Of necessity, developing benchmark indicators is an iterative process. It starts with the choice of potential subindicators and indicators based on theoretical and empirical considerations. Subindicators must combine related variables that point in the same direction. Also required is sufficient good quality empirical information that can be scaled to similar quantitative ranges and aggregated into subindicators, which in turn should allow for meaningful aggregation into main benchmark indicators. The subindicators and indicators should be scaled so they can be compared between geographic areas and countries. To be usable, indicators must provide consistent, robust information; cover sufficiently wide ranges; be empirically meaningful; allow for unbiased comparisons; and serve as useful input variables in a broader analytical analysis. Missing observations and unreliable answers will jeopardize the reliability of indicators.

During data exploration for this study, sets of variables were subjected to a factor analysis to help identify patterns that could help define benchmarks. Owing to differences in the pilot surveys' designs and definitions, factors extracted from the data were unusable for benchmarking, as they were country-specific and not replicable across countries. But had essentially uniform survey instruments been employed, at least for the core variables, a common factor analysis could have been applied to each pilot's data set and benchmarks synthesized using common algorithms. Benchmarking was nonetheless undertaken to exploit the information contained in the large number of community-level variables to achieve (i) cross-country comparison, and (ii) analysis of enterprise performance, entrepreneurial choices, and patterns in entrepreneurial complaints about IC constraints.

The RIC assessments yielded six benchmark indicators:

(i) ***connectivity,*** that is, indexes measuring rural-urban connectivity and regional economic integration, including the quality and capacity of communications and times required to travel to main markets and nearby cities or to access postal and other government services;
(ii) the quality and availability to community residents of ***infrastructure services,*** including power, water, sewerage, and other household services, and of local roads;
(iii) access to ***business services,*** including management, marketing, accounting, legal, insurance, technical, and information technology;
(iv) the quality, reliability, and transparency of ***governance;***
(v) the level of ***human capital,*** assessed via an index that scales education and experience levels in the sampled communities; and
(vi) availability, access, and costs of ***finance services,*** including the number and kinds of formal financial and insurance service institutions and services provided, access to commodity futures and options, and the general availability of formal lending services.

Definitions for each benchmark indicator and subindicator indexes are provided in Annex E of this report, with illustrations from the Sri Lanka pilot. The indicators are mainly derived from data collected by the community questionnaires, but also in part from aggregations at the community level of data collected through the household and enterprise questionnaires. Although the RICS community surveys have no parallel in the PICS, the benchmark indicators ii, iii, iv, and vi for RIC assessments resemble to some extent four of the indicators developed by the World Bank Group for PIC assessments:

1) access to infrastructural services and service delivery;
2) availability of business development services;

3) governance and corruption, covering quality of public-service delivery and aspects of corruption; and
4) access to financial services and service delivery.

In this study, benchmarks iv and v and one subindicator of vi are aggregates of household and enterprise data. Provided communities are large enough and contain sufficient sampling units, the benchmark indicators will be mainly exogenous and, in principle, useful as explanatory variables in regressions on entrepreneurship and enterprise performance. The use of benchmark indicators thus provides an alternative to the ad hoc selection of variables from the hundreds of potential variables that could be chosen for the regressions. The subindicators also form a systematic set of specific indicators.

An accurate presentation of countries' rural investment climate and their comparison across countries are crucial aspects of policy analysis and RIC reform. From the perspective of the donor community (and the Bank), international benchmarking constitutes one of the most important contributions made by a country-based RICA, as international comparison can contribute significantly to the continuing policy dialogue over priorities for reform and investment. The benchmarks defined in this study have already been applied in RIC studies for Bangladesh and Pakistan. Table 2.1 provides a comparison.

RIC2 forms a first step in the development of RICA benchmark indicators, used in the empirical analysis discussed in Chapters 3, 4, and 5. Chapter 6 assesses the strengths and weaknesses of the indicators for descriptive comparison and analysis.

2.2.2 Econometric Methods and RIC Applications

An econometric study of enterprise performance and entrepreneurship choices must accommodate the structure of the sample data. Two main features are prominent. First, survey data should be collected using a two-stage sampling process. In the first stage, communities are selected; in some countries they may be stratified by region to ensure adequate regional representation. Then, a list of buildings in each of the selected communities is made, and a random sample of enterprises and households is determined for each community.[17] In most countries, enterprises may usefully be stratified by enterprise size and/or activity sector. Most of the selected communities will normally include very many small one- or two-employee units. These can easily overwhelm the sample. It is therefore desirable, indeed necessary, to also sample as many of the larger enterprises as can be found—perhaps by designing a separate, purposive sample. To correct for oversampling any segment of the enterprise population, sampling weights must be computed when the survey is completed and then used in any statistical evaluation of the data.[18] Sampling weights also adjust for the relative size of the selected communities, in comparison to the population of the stratified regions, to maintain representation at the national rural level.

In every country, the sample of households is stratified by entrepreneurship status. The rate of entrepreneurship falls around 25 percent of households in most countries, but it will vary substantially between rural communities. Some remain primarily agricultural; in others, agriculture has become secondary to nonfarm sectors. Without

Table 2.1 International Comparison of RIC Benchmark Indicators

	Bangladesh	Nicaragua	Pakistan	Sri Lanka	Tanzania
	Index	Index	Index	Index	Index
Connectivity	0.45	0.34	0.38	0.40	0.20
Infrastructure services	0.44	0.51	0.58	0.35	0.19
Business services	0.05	0.24	0.40	0.15	0.07
Governance	0.62	0.67	0.72	0.67	0.50
Human capital	0.24	0.21	0.21	0.33	0.21
Finance services	0.19	0.17	0.39	0.50	0.18
Average	***0.33***	***0.36***	***0.45***	***0.40***	***0.22***

Source: Annex K, Table K.1.

stratification, visits to mainly agricultural communities will yield an abundance of nonentrepreneurial households and few observations for the enterprise sample. RICS surveys typically aim for a better than 50-50 proportion of entrepreneurial households, using sampling weights to correct for oversampling households with an enterprise.

Second, standard econometric models assume that observations (enterprises, households) are independent of each other, that random disturbances in their behavior are uncorrelated. Study of the effect of the rural investment climate, however, is founded on the notion that the economic environment influences individual (that is, enterprise or household) behavior. Some aspects of the economic environment are known: benchmark indicators, indicator components, and other community characteristics such as size and seasonality.

But it would be presumptuous to claim that observations made about a community's economic environment are the sum total of that environment. Rather, it must be assumed that relevant aspects of the investment climate are unobserved—and indeed probably unobservable, even with the best-designed questionnaire. The unobserved aspects enter the econometric model as community-level "disturbances." By the nature of the sample data design, where enterprises and households are clustered into communities, all observations in a given community share the same community-level disturbance. Thus, the overall random influences on an individual's behavior are in part idiosyncratic and in part communal. This violates the independence assumption underlying standard econometric models.

In the econometrics literature, the community disturbance is called a *random effect* (sometimes referred to as a "fixed effect"). The literature, specifically the subfield of panel econometrics, offers many techniques for dealing with random effects. Panel data are repeatedly observed cross-sections: a random effect that stays with a person (or household or enterprise) over time would be an unobservable. Econometrically, a community random effect that attaches itself to all enterprises in a community is similar to a random effect that attaches itself to an enterprise over time. The panel econometrics literature, however, never concerns itself with sampling weights; common statistical software does not permit estimation of panel econometric models with sampling weights. Annex I outlines econometric models appropriate for RICS data,[19] inserting sampling weights into panel-econometric models; the Annex also offers an Appendix with programs in Stata that permit estimation in the presence of random effects and sampling weights. The development and application of these models is a major contribution of RIC2, as typical studies of rural investment climate using similar kinds of data ignore the issues addressed here.

Chapter 3 addresses the question of how the investment climate affects enterprise performance, indicated by both sales and net value added. The investment climate is measured by conditions in the community, as given in the benchmark indicators and other community characteristics. Community factors in general are highly influential, but, not surprisingly, the benchmark indicators or their subindicators capture only a portion.

Chapter 4 investigates household choices. These are twofold. One set of choices refers to the activities the household pursues. Of primary interest, of course, is the entrepreneurship choice: are members of the household operating an enterprise? But because of specific opportunities, a desire to diversify, the skill mix in the household, or other factors, the entrepreneurship choice interacts with the choice to seek wage employment or to operate a farm. Thus, the econometric analysis expands to cover those choices as well, not in the least because some development economists view the distinction between wage-employment and nonfarm self-employment as less relevant than the distinction between farm and nonfarm activities. As the analysis will show, such a two-way division is less informative than a three-way one. The second household choice refers to the decision to start up an enterprise. Whereas entrepreneurship status may be viewed as a *stock,* enterprise start-up is a *flow.* Successful policy changes to the investment climate will quickly be reflected in the number of enterprise start-ups; only over time will they affect the rate of entrepreneurship.[20]

Explanatory variables in the analysis of entrepreneurial household behavior include household characteristics (structure, skills, ethnicity, nonlabor income, assets), benchmark indicators and their components, and other community characteristics. Household assets may be endogenous to this behavior: past entrepreneurship may have generated incomes that led to increased savings and greater assets at present. Among the community characteristics, a higher per capita income

may attract households to entrepreneurship, but their success in turn leads to higher incomes in the community. The potential endogeneity of these two variables could bias the estimated entrepreneurship relationship. Thus, the econometric model of entrepreneurship choice is augmented with an allowance for endogeneity of one household-level variable and one community-level variable (see Annex I). Empirically, however, endogeneity does not appear to be a relevant issue.

Chapter 5 examines two sets of responses from entrepreneurs about the investment climate. One concerns a list of potential barriers to the operation and growth of the enterprise; the other describes interactions with and observations about government. These responses are in part subjective and in part factual; they represent the entrepreneur's perspectives about the investment climate. Explanatory variables are the characteristics of the enterprise, the benchmark indicators and selected components, and community characteristics. One prominent conclusion of the econometric analysis is that the community random effect is always important: community unobservables influence entrepreneurs' perceptions about the investment climate.

Chapter 6 delves into differences between communities. It makes a cross-country comparison of benchmark indicators and explores how benchmark indicators are correlated with community-level characteristics and with prices collected at the community level. It also tries to explain community-level indicators of economic activity (outcomes), such as enterprise density, income per capita, nonfarm income shares, and enterprise productivity, through simple regression on benchmark indicators and a number of other economic descriptors. This analysis plays itself out at the community level, and the econometric model must only account for sampling weights that represent the relative importance of each community.

No claim is made here that the analysis in the following four chapters is definitive. Much work of great interest remains to be done. In principle, for example, it is arguable that enterprise performance and entrepreneurship behavior is driven by perceptions about the economic conditions in the community rather than by actual, factual conditions—and that perceptions are driven in part by the actual conditions and in part by education, networks, and opportunities. Thus, models estimated in Chapters 3 and 4 substitute actual conditions for entrepreneurial perceptions: they constitute reduced-form rather than structural models. Chapter 5 develops this issue further and outlines the next steps that beckon the analyst. In brief, perceptions about the investment climate may well be endogenous to the performance of the enterprise, as growth may reveal constraints that are not apparent to smaller-scale entrepreneurs, while entrepreneurs' evaluation of the investment climate may differ from that common among those chosing *not* to operate an enterprise, suggesting that a significant problem of self-selection must be addressed before investment climate perceptions can be considered as potential determinants of entrepreneurial choice.

3 Enterprise Performance and Investment Climate Constraints

This chapter analyzes the economic performance of the enterprises sampled in the Nicaraguan, Sri Lankan, and Tanzanian RIC pilot surveys. Four performance measures are used: gross productivity, net factor productivity, employment generation, and capital generation. A brief review of enterprise characteristics is followed by summaries of the methods used and the main findings. The chapter concludes with an economic interpretation of the regression results and a discussion of the general findings as they relate to RIC survey methods.

3.1 THE ENTERPRISES AND THEIR ENVIRONMENTS

Nicaragua, Sri Lanka, and Tanzania offer rural enterprises different economic environments. Over the past five years, Nicaragua's economy grew by about three percent annually, Sri Lanka's by almost 4 percent, and Tanzania's by over 6 percent. Of the three countries, Sri Lanka had the highest GDP per capita in 2006. Sri Lanka's non-agricultural value added per worker was also about 30 percent higher than Nicaragua's and more than 60 percent higher than Tanzania's (Table 3.1). Agricultural value added per worker in Nicaragua, however, is more than 2.5 times higher than in Sri Lanka. Tanzania scores lowest in both indicators. Sri Lanka's population density is relatively high, and a proportionately higher percentage of its roads are paved.

Among enterprises for which sufficient data was available for analysis, the countries show different sectoral compositions (Table 3.2). At 50 percent, Tanzania has the most trading enterprises, while Sri Lanka has the most production enterprises. The share of service enterprises is about the same for the three countries. Contrary to common belief, agro-processing enterprises form only 2 to 14 percent of the rural enterprises for these countries. The number of enterprises per 1,000 people is about the same in Nicaragua and Sri Lanka.

More than half of Sri Lanka's rural enterprises are registered, representing a significantly larger share than in the other two countries (Table 3.3). In Tanzania only 20 percent of enterprises report being registered. The average age of enterprises in Nicaragua is 10.7 years, more than 2 years older than RNFEs in Sri Lanka and Nicaragua.

The large majority of enterprises in Nicaragua and Tanzania operate without paid labor and without debt. In Sri Lanka, 32 percent have paid labor, in Nicaragua and Tanzania only 17 and 14 percent respectively. Although the absolute numbers are small, twice as

Table 3.1 Basic Economic Characteristics of Selected Countries

	Nicaragua	Sri Lanka	Tanzania
GDP Growth Rate (%) (annual average 2000–04)	3.1	3.9	6.2
GDP / cap (2006) (PPP constant 2000 US$)	3340	4034	620
Inflation (GDP deflator 2000–04)	6.9	8.6	7.3
Agric. Value Added (AVA) (% of GDP)	19.2	17.8	44.8
AVA / worker (constant 2000 US$)	1946	746	290
Non-Agric. VA / worker (constant 2000 US$)	2280	3048	1885
Rural population density (persons / sq. km)	117	1659	596
Population density (people / sq. km)	43	298	42
Roads paved (% of total roads)	11	81	9

Source: World Bank development indicator database (World Bank 2007d).

Table 3.2 Sector of Operations, Sales, Net Value Added, and Productivity of Enterprises

	Nicaragua	Sri Lanka	Tanzania
Observations	846	1018	947
Of which (%)			
Trade	40	39	50
Service	19	20	18
Agricultural processing	14	2	4
Other production	11	32	7
Mixed	16	7	21
Enterprise density (per 1,000 people)	50	52	NA

Source: Annex D, Annex B.

many enterprises in Sri Lanka operate with debt than is the case in Tanzania and Nicaragua. This may reflect the greater density of formal financial services in rural Sri Lanka, where more than a quarter of Sri Lankan RNFEs have obtained formal-sector loans equivalent to 50 percent or more of owner's equity. Only 7 percent of the enterprises surveyed in Nicaragua relied substantially on loans from formal sources; in Tanzania the number is negligible.

Data on the experience, education, and gender of enterprise managers was collected by the Sri Lankan and Tanzanian pilots. While the formal education of managers in terms of years of schooling is similar in both countries, Sri Lankan managers, on average, have almost two more years of experience than do their Tanzanian counterparts. In both countries more than three-quarters of the managers are male.

The three countries' enterprises clearly vary in size. Using sales as an indicator, Sri Lanka's enterprises are much larger than Nicaragua's; enterprises in Tanzania are by far the smallest (Table 3.4). Using net value added (NVA), perhaps a more useful indicator of enterprise size, the differences between Nicaragua and Sri Lanka were smaller, but those for Tanzania were larger. It is likely that the differences in size reflect levels of income per capita.

Table 3.3 Characteristics of Surveyed Enterprises

	Nicaragua	Sri Lanka	Tanzania
Number in sample	846	1018	947
Registered/formal (%)	30	58	20
Age of enterprise (years)	10.7	8.7	8.5
Stand-alone (%)	7	65	37
Enterprises with paid labor (%)	17	32	14
Enterprises with debt (%)	21	42	23
Formal loan > 50% of equity	7	31	1
Experience of manager (years)	NA	7.0	5.0
Education of manager (years)	NA	9.5	7.8
Female manager (%)	NA	23	22

Source: Annex D, Annex B.

The difference between the countries suggests that enterprise size increases with the level of economic development measured by GDP per capita. In Nicaragua and Sri Lanka, larger communities had larger median enterprise sizes. Assuming that income per capita tends to be higher in more urbanized areas, enterprise size would seem to increase with level of income. Tanzania, however, shows no difference in enterprise size between smaller and larger communities.

Finally, Table 3.4 shows the size distributions of RNFE sales and net value added for the three countries, and Table 3.5 contrasts these two indicators of enterprise size with indications of community size. Average enterprise sizes measured by sales or NVA are almost six times larger in Sri Lanka than in Tanzania and almost three times larger than in Nicaragua. The maximum enterprise size in the Sri Lankan sample affects the average.

Table 3.4 Enterprise Size Distribution in Terms of Sales and Net Value Added

	Nicaragua	Sri Lanka	Tanzania
Observations	846	1018	947
Sales (US$)			
Average	2438	6588	1186
Standard deviation	5980	41584	2153
Distribution points:			
Minimum	65	83	73
20 percent	452	456	193
40 percent	753	984	367
60 percent	1369	2072	576
80 percent	2881	4973	1378
Maximum	102954	1864962	18368
NVA (US$)			
Average	1975	2458	381
Standard deviation	4688	28636	2230
Distribution points:			
Minimum	−11752	−22094	−37382
20 percent	324	75	28
40 percent	646	262	110
60 percent	1140	613	241
80 percent	2275	1579	551
Maximum	88378	1388692	15732

Source: Annex D, Annex B.

Table 3.5 Median Sales and Net Value Added, by Size of Community (US $)

	Nicaragua (n = 98)		Sri Lanka (n = 147)		Tanzania (n = 154)	
Community:	Smaller	Larger	Smaller	Larger	Smaller	Larger
Sales	753	1157	881	1554	459	459
NVA	734	1004	406	477	166	165

Source: RIC Surveys.

The comparison in Table 3.6 of indicators of average productivity suggests that rural enterprises in Nicaragua are considerably more productive than are enterprises in Sri Lanka and Tanzania. These results must be treated with caution, however, as the differences between countries may reflect structural differences in supply chains and dissimilarities in the survey instruments.

3.2 ENTERPRISES AND VARIABLES

The database for the enterprise sample and related variables is described in Annex A. RIC surveys collected data for about 1350 enterprises in each country, although as explained in the Annex, the actual number of enterprises is smaller because of survey and data deficiencies (Table A.1). For Nicaragua, data for households with more than one enterprise were consolidated during data processing. For Sri Lanka, essential community variables in war-torn areas are incomplete, leading to a significant loss in the number of observations. In Nicaragua, a data-entry failure resulted in loss of observations. In a few cases in all three countries, some explanatory enterprise-level variables are missing. In addition, a threshold was applied: enterprises reporting negative or very low NVA, often presumably because of incorrect reporting, were omitted.[21] Although the loss of potentially good observations reduces efficiency in the regression analysis, there is no reason to assume that the data losses caused biases in the estimates.

From the data collected in the RIC surveys, some variables were taken and others calculated for an analysis of enterprise performance. Table A.2 in Annex A provides definitions of data grouped as financial data, other enterprise characteristics, and community characteristics. The numerical values of the variables entered in the regressions are provided in Annex B.

3.3 SPECIFICATION OF ENTERPRISE PERFORMANCE REGRESSION MODELS

Four enterprise performance measures are used: total productivity, net factor productivity, labor generation, and capital generation. Total productivity is measured at the sales level (gross output) and at the net value added level (net enterprise income).

Table 3.6 Enterprise Productivity in Terms of Gross Sales and Net Value Added

	Nicaragua	Sri Lanka	Tanzania
Observations	846	1018	947
Average Profitability (US$)			
Ratio: Sales/Total Cost	2.59	1.09	1.01
Ratio: NVA/ Total Factor Cost	2.79	1.41	1.48
Average Productivity (US$)			
Ratio: Sales/day of labor	3.55	9.73	4.59
Ratio: NVA/day of labor	2.88	3.63	1.47

Source: Annex B.

Productivity

The primary empirical specification for measuring productivity uses a Cobb-Douglas production function, which is loglinear in output and inputs, augmented with variables hypothesized to enhance enterprise productivity. Throughout this report reference will be made to *sales variants* and *net value added variants*.

The choice of empirical variants for the regressions is discussed in detail in Annex C. Because of inherent interest in the contribution of enterprises of different sizes to the rural economy, quadratic and interactive specifications of the factor input variables *labor* and *capital* can better describe differences related to size than loglinear forms do.

The aim in formulating the regression models is to choose variables that describe enterprise characteristics and the community environment and to avoid variables that could be seen as largely endogenous. Since the main objective of the analysis is to shed light on the contribution of the investment climate, three stepwise regressions are carried out. This reveals the structure of the data sets and illustrates the variance captured by enterprise variables, general community variables, and benchmark indicators. In addition, a fourth variant is estimated in which the benchmark indicators are replaced by those of their components providing the highest individual explanation.

Underemployment of household labor may lead to a lower marginal productivity for household labor than for paid labor. Such differences may also exist for household-owned assets and borrowed assets. One way to deal with this is to utilize separate specifications in the regression model for self-owned and for borrowed inputs. A fifth regression variant separates imputed and paid labor and capital. Table C.1 in Annex C provides an overview of the regression variants for productivity analysis. Table A.2 in Annex A fully specifies the variables used. Table B.1 to B.6 in Annex B list descriptive statistics for the variables used.

The selection of variables includes the factor inputs—labor and capital—for the NVA variants and the factor and nonfactor inputs and depreciation for the sales variants. Also considered are a number of enterprise characteristics that could add to gross and net productivity. They include enterprise age, registration status, and line of business and entrepreneur experience, gender, and education.

Community characteristics present direct incentives and disincentives for enterprises and hence may affect their productivity. For Nicaragua and Sri Lanka, these include data about seasonality in agricultural labor use and the number of enterprises per 1,000 inhabitants (enterprise density). In all three countries, community population size is included as an indicator of the degree of urbanization within which the enterprise operates, and agricultural land cultivated per capita and the rate of illiteracy are also noted. In Sri Lanka, information about where households from the community buy and sell most of their goods was used to construct dummies that to some extent indicate the community's degree of commercial openness. In Sri Lanka and Tanzania, data about the community's main source of income, that is, agriculture, wages, or nonfarm enterprises, were used to create dummies that help capture differences in markets and income between communities.

All regression models incorporate benchmark indicators, which are community-level composite indexes of characteristics of location, infrastructure availability, utilities, public and private services, governance and corruption, human capital, and finance services, as described in Chapter 2 above. Benchmark indicators and their components describe both the community-level environment and the investment climate.

Employment and Capital Formation

What conditions generate enterprises with more employment and what conditions generate enterprises with more capital? Or, expressed differently, what explains differences in enterprise size, measured by labor and by capital input respectively? Regression models, using the same enterprise and community data used for productivity measurement, explain employment and capital generation from industry mix, enterprise characteristics, benchmark indicators, and other community characteristics. The parameter estimates appear in Annex C.

3.4 REGRESSION RESULTS

3.4.1 Variance Explained

Productivity analysis. A dominant characteristic of the enterprise data sets is that enterprise characteristics explain by far most of the variance in the

data set; industry and community levels explain only a small part.[22] Table C.2 in Annex C illustrates this for productivity. The regressions on benchmark indicators and enterprise dummies for the control of industry differences explain only between 3.6 and 9.3 percent of the variance. When enterprise variables are added, the explanation jumps by a factor of five to eight. General community variables explain 0.2 to 1.8 percent. If the benchmark indicators are replaced by those of their components that add most, the explanation of the regression increases by another 0.5 to 2.0 percent.

The NVA regressions have generally a one-third lower explanation in Sri Lanka and Tanzania; in Nicaragua the difference is smaller. When imputed and paid labor and capital are separated (regression variant 5), the explained variation increases by a few percent only, and for Sri Lanka's NVA and Tanzania's sales model it even decreases.

Although no comparable data are available for urban enterprises, these findings may not be unique to rural enterprises. Most of the variance will relate to enterprise size, and IC variables will explain a modest percentage in productivity differences. Yet the addition of these few percentage points may represent an important contribution to profitability.

Employment and capital. The findings for employment generation and capital formation show a similar pattern. Industry mix and benchmark indicators together can explain 4 to 5 percent of the variation in employment generation. (For details see Annex C, Table C.3.) Addition of enterprise characteristics doubles the explanation in Nicaragua, triples it in Sri Lanka, and increases it by half in Tanzania. Other community variables add 1.5 percent in Nicaragua, almost 2 percent in Tanzania, and almost 4 percent in Sri Lanka. The higher percentage of explanation in Sri Lanka could well be the result of its larger number of enterprise and community variables as compared to the other countries. Differences in capital generation can be better explained than can differences in employment generation. Partly this could be caused by greater weaknesses in the data measuring labor input. Enterprise characteristics contribute most to differences; other community variables contribute much less. In Nicaragua adding community variables results in a lower percentage of explanation; for Sri Lanka and Tanzania the explanation increases by 1.2 and 2.4 percent.

3.4.2 Productivity Analysis

Production elasticities and economies of scale. The Cobb-Douglas model augmented with quadratic and interactive terms (log of labor and log of capital) has greater overall explanatory power than do models with linear terms only. Results from this specification are recorded in Table 3.7 Annex C, as variant (4).

The economies of scale vary with enterprise size. For the sales variant (4) in Sri Lanka, the implied returns to scale at mean input levels equals 0.911 (Table 3.7) and rises with increasing levels of labor and capital. For Nicaragua and Tanzania it is lower, at 0.677 and 0.355 respectively, but unlike the other countries, in Tanzania it decreases with factor input. For the NVA variants (4), the returns to scale are about the same for Nicaragua and Tanzania and lower for Sri Lanka (0.733), in all cases increasing with the level of factor input.

Separate specifications (referred to as variant 5) for paid and imputed labor and capital input lead

Table 3.7 Economies of Scale and Production Elasticities for Variant (4)

	Nicaragua		Sri Lanka		Tanzania	
Variant	Sales (4)	NVA (4)	Sales (4)	NVA (4)	Sales (4)	NVA (4)
Production elasticity labor	0.422	0.432	0.220	0.369	−0.013	0.055
Production elasticity capital	0.193	0.269	0.255	0.364	0.148	0.323
Economies of scale	0.677	0.700	0.911	0.733	0.355	0.379

Source: Annex C, Table C.4 to Table C.9.

Notes: Elasticities are evaluated at the log of the mean input values, where the mean is computed under the assumption that the inputs are lognormally distributed in order to reduce the effect of outliers.

Table 3.8 Economies of Scale and Production Elasticities for Variant (5)

	Nicaragua		Sri Lanka		Tanzania	
Variant	**Sales (5)**	**NVA (5)**	**Sales (5)**	**NVA (5)**	**Sales (5)**	**NVA (5)**
Production elasticity labor						
Total labor input	0.386	0.391	0.216	0.400	0.239	0.142
Imputed labor input	0.316	0.312	−0.015	−0.032	0.215	0.084
Paid labor input	0.070	0.079	0.231	0.432	0.024	0.058
Production elasticity capital						
Total labor input	0.154	0.196	0.202	0.258	0.096	0.352
Imputed labor input	0.201	0.252	0.190	0.225	0.072	0.225
Paid labor input	−0.046	−0.055	0.012	0.033	0.025	0.127
Economies of scale	0.687	0.587	0.865	0.646	0.590	0.493

Source: Annex C, Table C.10 to Table C.15.

Notes: Elasticities are evaluated at the log of the mean input values, where the mean is computed under the assumption that the inputs are lognormally distributed in order to reduce the effect of outliers.

to better explanations, except for sales in Tanzania and NVA in Sri Lanka. The models allow for differences in production elasticities for imputed and paid labor and capital costs. Evaluation of the parameter estimates shows that economies of scale are roughly similar for variants (4) and (5) for Nicaragua and Sri Lanka (Table 3.8). Similarly, total labor and capital elasticities do not differ much either, but differences appear in the component effects: in Nicaragua, an increase in imputed (household) labor contributes more to output than does an increase in paid labor; in Sri Lanka, an increase in imputed labor does nothing for output, whereas paid labor substantially raises output; and in both countries output responds more strongly to variations in imputed capital than in paid capital. The Tanzania results differ from the other two countries in that the labor elasticities in variant (5) are substantially higher, causing a higher estimate of the economies of scale (though still much less than 1); and while imputed labor and capital inputs are more influential than paid inputs, the difference is less pronounced than in Nicaragua and Sri Lanka. While Table 3.8 focuses only on the implications of the alternative specification of inputs, it is important to note that this modeling exercise can strongly afect parameter estimates for other variables.

Contribution of enterprise and community variables. Table 3.9 provides an overview of variables with statistically significant contributions to enterprise productivity. Enterprise age makes an important contribution to productivity in Nicaragua and Sri Lanka, and in Sri Lanka manager education is important as well. In Tanzania manager experience contributes to total productivity and manager education to net factor productivity. These findings suggest that learning by doing and innovation are important for survival and increased productivity.

Manager gender is known for Tanzania and Sri Lanka. In Tanzania male managers appear to induce higher total productivity, but not net productivity. Sri Lanka shows no significant productivity differences. Registered enterprises have higher productivity in Tanzania and higher net factor productivity in Nicaragua, but for Sri Lanka no difference appears. Registration may encourage higher productivity in two ways. Enterprises with higher net income or sales may be unable to escape registration, but registration may also confer particular benefits that encourage growth. Small, low-income enterprises may simply escape registration because their officers and entrepreneurs don't bother with it.

Line of business explains few productivity differences. In Tanzania no differences were found, and in Sri Lanka only service enterprises had higher total productivities than trading enterprises (serving as the comparison). Nicaragua shows more differences: service enterprises have significantly lower productivity and mixed enterprises have higher productivity; total productivity is higher for manufacturing enterprises; and net factor productivity is lower for other enterprises (mainly mining and construction).

Table 3.9 Significant Contributions to Productivity by Enterprise and Community Characteristics

	Nicaragua		Sri Lanka		Tanzania	
Variant	**Sales (4)**	**NVA (4)**	**Sales (4)**	**NVA (4)**	**Sales (4)**	**NVA (4)**
Enterprise characteristics						
Age of enterprise	+	+	+	+		
Experience of manager	NA	NA			+	
Male manager	NA	NA			+	
Education of manager	NA	NA	+	+		+
Registration		+			+	+
Line of business						
Service enterprise	−	−	+			
Manufacturing of nonagricultural products	+					
Agricultural processing						
Other production		−				
Mixed enterprise	+	+				
Community characteristics						
Agricultural seasonality					NA	NA
Enterprise density					NA	NA
Community population size						
Agricultural land per capita						
Illiteracy		−				
Main market in						
Neighboring communities	NA	NA	+	+	NA	NA
Commercial center	NA	NA	+	+	NA	NA
Nearest city	NA	NA			NA	NA
Main source of income						
Wages	NA	NA				
Self-employment	NA	NA				

Source: Annex C, Table C.4 to Table C.9.

Note: Statistically significant at 10 percent or higher with positive (+) or negative (−) parameter estimate. NA = Not available.

Most of the community characteristics make no significant contributions to productivity differences. The only exceptions are communities where the enterprises' main markets are in neighboring communities and commercial centers. This suggests that openness and access to more distant markets plays a positive role. The general finding of the limited contribution of community characteristics also suggests that the benchmarks and benchmark components effectively describe differences between communities.

Nicaragua shows relatively many significant contributions to productivity by benchmark indicators and benchmark components (Table 3.10). Connectivity is important in Nicaragua and Tanzania, business services in Sri Lanka, and governance in Nicaragua. The role of finance services in Sri Lanka is limited to loan access. The benchmark indicators and their components have been constructed such that a higher value between (0 and 1) indicates an improvement for business. Hence, the expectation is that all parameters would have positive signs. Yet in Tanzania access has a negative value, and in Nicaragua human capital does. This is most likely due to unobserved background variables and multicollinearity. Multicollinearity can explain the contradictory signs in Nicaragua for "time to near city" and "cost to near city," for example. The negative sign for "sewage system" in Tanzania is most likely caused by unobserved variables in the few communities having sewage systems.

3.4.3 Explaining Employment and Capital Generation

Many factors contribute to differences in employment and capital generation (Table 3.11). Enterprise age matters only in Nicaragua. Manager

Table 3.10 Significant Contributions to Productivity of Benchmark Indicators and Components

Nicaragua		Sri Lanka		Tanzania	
Sales	**NVA**	**Sales**	**NVA**	**Sales**	**NVA**
Connectivity					
Connectivity	**Connectivity**		Cost of Transportation	**Connectivity**	**Connectivity**
Time to near city	Time to near city		Time to main market (negative)	Time to main market	Rail stop
Cost to near city (negative)	Cost to near city (negative)				
Cost to main market	Cost to main market				
	Distance to post office				
Infrastructure services					
Percent with electricity	Percent with electricity	Concrete/ asphalt road (negative)		**Access (negative)**	
Cost of transportation (negative)	Availability electricity (negative)			Sewage system	
Garbage collection	Percent with protected water (negative)				
Sewage system	Garbage collection				
Business services					
Technology services		**Business services**	**Business services**		
		Management consulting			
Governance					
Governance	**Governance**	Public services / institutions			Insurance service
Human capital					
Human capital (negative)	**Human capital (negative)**				
Finance Services		Access to loans		Access to loans	

Source: Annex C, Table C.4 to Table C.9.

Notes: The table lists findings from regression variants (3) and (4) with a significance level of at least 10 percent. The benchmarks are printed in boldface.

experience matters for capital generation, but not for employment generation. Male managers generate more employment and capital. Manager education makes no difference. Registration affects the amount of employment and capital generated for two reasons: registered enterprises have better access to services, and licensing officials may make special efforts to register enterprises with significant factor input. Generation of employment and capital shows more differences among sectors than does productivity. Community characteristics play varying roles. Enterprise density relates especially to employment generation and in Sri Lanka to capital generation as well. Interestingly, in Sri Lanka openness (that is, marketing outside the community) has a positive effect on employment generation but not on capital generation.

The benchmark indicators and their components contribute differently to employment and capital generation (Table 3.12). In Nicaragua and Tanzania, connectivity contributes significantly to employment generation and human capital to

Table 3.11 Significant Contributions to Employment and Capital Generation by Enterprise and Community Characteristics

	Nicaragua		Sri Lanka		Tanzania	
Variant	Employment (4)	Capital (4)	Employment (4)	Capital (4)	Employment (4)	Capital (4)
Enterprise characteristics						
Age of enterprise	+	+				
Experience of manager	NA	NA		+		+
Male manager	NA	NA	+	+	+	+
Education of manager	NA	NA	+	+	+	
Registration	+	+	+	+		+
Line of business/industry dummies						
Service enterprise	–		–	–		+
Manufacturing nonagricultural products	–					
Agricultural processing	–	–	+	+	–	–
Other production	+		+		+	
Mixed enterprise	+	+				+
Community characteristics						
Agricultural seasonality					NA	NA
Enterprise density	+		+	+	NA	NA
Community population size	+	–		+	–	
Agricultural land per capita	+	–				+
Illiteracy						
Main market in						
Neighboring communities	NA	NA	+		NA	NA
Commercial center	NA	NA	+		NA	NA
Nearest city	NA	NA	+		NA	NA
Main source of income						
Wages	NA	NA				
Self-employment	NA	NA				

Source: Annex C, Table C.10 to Table C.15.

Note: Statistically significant at 10 percent or higher with positive (+) or negative (−) parameter estimate. NA = Not available.

capital generation. In Sri Lanka, similar to findings for productivity, business services are important for employment and capital generation. Finance services in Nicaragua contribute only to capital generation. As with their contribution to productivity, benchmarks and their components at times yield counterintuitive outcomes due to unobserved background variables or multicollinearity. Governance, time to market, and time to post office have negative signs in Nicaragua. Concrete or asphalt roads have a positive sign in Tanzania and a negative one in Sri Lanka. Obviously, the variables do not include other powerful location choice factors. Remarkably, finance services play no role in explaining enterprise size in terms of employment and amount of investment, with capital generated in Nicaragua the only exception.

3.5 GENERAL DISCUSSION

3.5.1 Data Quality

As with any empirical study, the value of the results depends on the quality of the data. In common with other surveys, RICS data quality suffers from limitations of human memory, respondent errors, and, at times, respondents' unwillingness to disclose full financial data. Moreover, some computed variables have conceptual weaknesses, and perhaps biases, owing to insufficient detail in the collected data (see Annex A). This may hold especially for computed labor input, for which a flawed stock concept had to be used. For capital, separation of capital used for consumption from capital used for production can be arbitrary, including as it does such elements as housing,

Table 3.12 Significant Contributions to Employment and Capital Generation of Benchmark Indicators and Components

Nicaragua		Sri Lanka		Tanzania	
Employment	**Capital**	**Employment**	**Capital**	**Employment**	**Capital**
Connectivity					
Connectivity	Time to market (negative) Time to post office (negative)			**Connectivity** Distance to post office	
Infrastructure services					
Infrastructure services (negative) Percent with electricity (negative) Percent with cell phones Sewage system (negative)	Cost of transportation (negative) Percent with fixed phone Garbage collection	Availability electricity Percent with fixed phone lines	Percent with protected water Percent with fixed phone lines Concrete / asphalt road (negative)		Percent with cell phone Concrete / asphalt road
Business services					
Technology services		**Business services**	**Business services**	Marketing service	
Governance					
Governance (negative) Public service (negative)	Public services (negative) Public institutions			Public services / institutions	
Human capital					
	Human capital				**Human capital**
Finance services					
	Finance services Number of bank services				Number of bank services (negative)

Source: Annex C, Table C.10 to Table C.15.

Notes: The table lists findings from regression variants (3) and (4) with a significance level of at least 10 percent. The benchmarks, if significant, are printed in boldface.

equipment, and vehicles; evaluation too can present difficulties. Furthermore, as already noted, missing information and data entry errors cause loss of observations and possibly loss of representativeness in the analysis subsample.

It is fair to say, therefore, that the observations and coded data are likely to contain errors, in part because of the surveys' pilot character. This of course affects the precision and efficiency of parameters estimated in the regressions. Generally, biases of estimates presented here are considered to be modest, but the usual caution should be applied when interpreting the empirical results.

3.5.2 Statistical Importance of RIC

The regressions provide many useful results on factors affecting differences in enterprise performance, which in this chapter are measured as productivity and generation of labor and capital. These vary in response to many factors, and the pilots' findings can provide input to all kinds of further analysis to help establish the policy agenda and support the policy dialogue. Work on the policy agenda will be discussed in Chapter 7, following further discussion in Chapters 4, 5, and 6 of the analytical work involved. Here some attention will

be given to statistical limitations and counterintuitive results.

Variation depends on many factors, some of which the data set may not explicitly include but which may still be influential in the background as unobserved variables. Indeed some factors may pertain to only a few enterprises or locations. In principle, variation in dependent variables can be expected to derive from (i) observed enterprise variables, (ii) unobserved enterprise variables, (iii) observed community variables, and (iv) unobserved community variables. In the pilots, community-level factors explain a relatively small part of enterprise productivity variation. Box C.1, Annex C, shows that, at best, community factors explain around 12 percent of the total variation in productivity in Nicaragua, between 7 and 14 percent in Sri Lanka, and about 27 percent in Tanzania. In practice, the explanation will be much lower.

The low explanatory contribution of community-level factors, including investment climate variables, indicates great heterogeneity among enterprises, hindering analysis of the effect of the rural investment climate. Analysts will simply have to cope with this fact. Although efforts were made in the pilots to collect data from communities ranging from typical rural hamlets to larger market towns, future RIC assessments could perhaps capture an even larger and more diverse mix of settlements. Also the scant number of communities was limiting: the RIC pilots sampled in only 99 communities in Nicaragua, 151 in Sri Lanka, and 149 in Tanzania. Because of missing data, the enterprise performance regression models were estimated with enterprise data from even fewer communities (93 in Nicaragua, 118 in Sri Lanka, and 136 and 127 respectively for the sales and NVA variants in Tanzania [see Annex C, Box C.1]). Thus while productivity variation may have many potential determinants across communities, only a limited number can be explored. In sum, although future surveys may be successful in capturing more variation, analysts must cope with data sets with many interrelated variables, multicollinearity, unobserved variables, and a relatively small contribution of community variables.

3.5.3 Benchmark Indicators

The conclusion from the previous section is that benchmark indicators, which together combine many of the community-level economic and investment climate variables, can explain only a modest part of the total variation. This section focuses on productivity analysis. Some benchmark indicators contribute well to explanations of productivity. In Nicaragua and Tanzania, for example, connectivity contributes significantly, but not in Sri Lanka: connectivity captures many of the infrastructure deficiencies related to distance, which are more prominent in Nicaragua and Tanzania.

The infrastructure services indicator did not perform as expected; it has negative signs in many cases, although these are significant only for Tanzania's sales regressions. Reasons could be multicollinearity—infrastructure services are correlated with connectivity—and unobserved background variables.

The business services indicator does well in Sri Lanka, but it does not appear to matter in Nicaragua and Tanzania. Business services may play a limited role, if any, for many enterprises, in particular in less developed situations. For that reason the revealed contribution may reflect Sri Lanka's relatively high level of development.

Although the governance indicator shows up in the tables with more positive than negative estimated effects, the regression analysis did not yield enough evidence to suggest it really matters. Governance has perhaps the poorest data reliability of all indexces, in part because it is an aggregate of more than 20 pieces of information gathered from sample enterprises in the community; for that reason alone it may be a weak indicator. Some enterprises, by self-selection or active participation, may even profit from weak governance and corruption.

Human capital does not add much to explain productivity; it exhibits a few significant but counterintuitive negative signs in Nicaragua, and only hints at a positive effect in Tanzania's sales regressions. Community human capital may offer only a weak additional contribution to the productivity of individual enterprises in models that already include entrepreneur experience and education as characteristics of a successful enterprise. Another possible reason is that the measured human capital index is a weak indicator of human capital at the community level because it is based on an aggregation of data from a limited number of households. Finally, many enterprises may not be conducting businesses sensitive to community-level human capital.[23]

The finance services indicator contributed effectively in Nicaragua and to a lesser extent in Sri Lanka. In Tanzania finance services had a significantly negative sign, which unfortunately makes little economic sense.

Replacing benchmark indicators with the subcomponents with greatest explanatory power provides interesting results. As expected, in general the added components increase the percentage of variation explained. Many components have significant positive parameters, helping to signal which among the investment climate variables are most important. But another finding is that, in several cases, underlying components are significant with opposite signs. Examples include connectivity components in the NVA regressions in Sri Lanka, infrastructure services components in both regressions for Nicaragua, business services in both regressions in Tanzania, governance in the sales regression in Sri Lanka, and finance services in the sales model in Tanzania. Understandably, for all cases, the aggregate benchmark did not have a significant effect one way or the other.

3.5.4 Comparative Observations on the Economies

The findings on productivity and generation of employment and capital show a wide variety of factors contributing to differences among the three countries. This is as expected, as it reflects a range of factors related to differences in natural conditions, policies, institutions, infrastructure, and level of development.

Tanzania has the lowest level of development, with weak commercialization and small enterprises. Many of the enterprises have a local focus and function in communities with large subsistence components. The finding that connectivity most affects productivity indicates priority should be placed on infrastructure and services to open rural areas to further commercialization. Connectivity also contributes to employment generation. Human capital is important for capital generation.

Nicaragua's economy is more developed but shows major differences between modern enterprises and commercialization in the country's advanced areas and low levels of development and commercialization in less advanced areas. Improved connectivity adds most to explaining productivity and employment generation differences. Some infrastructure and business services also make positive contributions. Less developed areas require strengthened commercialization; more developed areas require a more mature economic and institutional infrastructure. Human capital is a significant factor in capital generation.

In Sri Lanka, the most developed of the three countries, business services contribute most to explaining productivity and employment and capital generation. Perhaps because of the much higher population density and higher level of road infrastructure, connectivity contributes less to explaining productivity differences. Some connectivity and infrastructure services subcomponents, however, may play a role in specific situations. Further commercialization emphasizing private-sector service provision seems to be priority for Sri Lanka.

3.5.5 Comparison with Other Findings

The three pilot studies were conducted by country teams, and the results were published in country reports for Nicaragua (World Bank/RUTA 2008 forthcoming), Sri Lanka (World Bank and Asian Development Bank 2005), and Tanzania (World Bank 2007e). Each of these reports includes some effort to describe enterprise performance in the light of investment climate constraints. In addition, econometric research was carried out by Deininger, Jin, and Sur (2007) for Sri Lanka and Sundaram-Stukel, Deininger, and Jin (2006) for Tanzania.

All teams used a Cobb-Douglas–based model for analysis of factor productivity (gross value added) and investment climate constraints. The choice of IC variables was guided by the IC constraints reported by entrepreneurs in the enterprise survey and by availability of corresponding IC variables reported in the community surveys. A practical concern with this approach is the large numbers of IC variables covered in the enterprise and community surveys, which due to sheer numbers cannot all be included in a model. Given the limited number of community observations, the number of community variables should not exceed 10 to 15. This requires that researchers use subjective criteria to select variables.

RIC2 offers a unique parallel analysis of three countries using a framework of benchmark variables and subcomponents based on theoretical considerations and factor analysis to explore the structure of community data. These variables are

more robust than those used in other studies. Moreover, this framework allows control of many aspects of the investment climate while simultaneously exploring the contribution of components in one benchmark dimension. This reduces the bias in the representation of components' influence. Furthermore, several other community characteristics were added that appear useful for describing the enterprises' local environments. In contrast with previous econometric studies, regression models specifically incorporate the existence of unobservable community-level factors (as a so-called random effect), as elaborated in Chapter 2 and Annex I. The analysis is carried out for gross output and net value added, with a regression model empirically more flexible than the basic Cobb-Douglas form. To improve the quality of the analysis, the model reconstructs lost sampling weights for Nicaragua and adds an econometric analysis for Nicaragua that includes both household-based and stand-alone enterprises.

RIC2 provides a comparative perspective on the three countries. Although direct comparison of results is not possible owing to the surveys' differing model specifications and variable selections, together they provide a broad overview of data, methods of analysis, and findings that will help further research on investment climate in these three and other countries.

4 Entrepreneurship Choice and Enterprise Start-Up

Across the developing world, rural areas exhibit richly varied patterns of activities. Many households, of course, derive their livelihoods from farming, but people are also active as traders, service providers, and manufacturers. In fact, rural nonfarm enterprises provide 30 to 45 percent of rural incomes across the developing world and an even higher share among the rural poor (Haggblade et al. 2007).

The level of RNFE activity not only depends on growth in agricultural production and rural consumption, it also stimulates it: the relation is one of interaction and synergy. RNFE efficiency translates directly into lower transaction costs for bringing farm products to consumers and distant markets and for bringing inputs and tradable consumer goods and services from distant markets to rural residents. Hence, RNFEs affect the terms of trade for farmers, their real income, and their incentives to produce.

Yet it would be a mistake to think of RNFEs only in terms of the loop between agricultural production and rural consumption. The low cost of rural labor may induce some urban entrepreneurs to establish rural branches, and it may also lead some rural entrepreneurs to produce for urban markets (Hayami 1998; Haggblade et al. 2007). In such a decision, the cost of transport and communication plays an obvious role. But other relative improvements in the rural investment climate also offer scope for stimulating rural incomes and employment by attracting economic activities.

This chapter explores why households engage in the operation of a RNFE and which households choose to do so. Is it only a matter of the skills and resources available within the household? Or do favorable investment climates attract households to entrepreneurship?

Entrepreneurship, defined as the current operation of an enterprise, is a stock concept (Van der Sluis, Van Praag, and Vijverberg, 2005). But entrepreneurship is more often thought of in dynamic terms: finding market niches, organizing efficient methods of production, exploring new opportunities. If entrepreneurship means the pursuit of success, some ventures must inevitably fail. Indeed, turnover among small enterprises in most countries around the world is often high; 15 to 20 percent of business activities are abandoned each year—but about an equivalent number start up (see Vijverberg 2005).[24] The RIC survey does not permit analysis of enterprise deaths, but analysis of enterprise start-ups is feasible. The obvious policy questions about the "flow" concept of entrepreneurship are: Which households start up new enterprises? What

influence, if any, does investment climate have on these decisions?

Entrepreneurship is just one of several forms of employment and income generation. Households may also pursue wage employment or farm activities. It is useful to analyze the three together, since transition from an agrarian society to an industrial economy moves through a phase in which nonfarm self-employment and small-scale entrepreneurship are important. As enterprises become successful, they expand and hire workers who in an earlier phase of the transition might have set up their own businesses. Aspects of the investment climate may encourage entrepreneurship in one society by drawing people away from farming, but discourage it in another by drawing people into wage employment at larger enterprises.

The next section summarizes patterns of employment, income, and assets among rural households surveyed by the three pilots. Section 4.2 examines household involvement in nonfarm enterprises through regression analysis; Section 4.3 broadens the perspective to include other forms of income earning (wage employment and farming); Section 4.4 addresses the issue of enterprise start-up; and Section 4.5 summarizes findings. For reasons explained in Section 4.1, the multivariate analysis in Sections 4.2 to 4.4 utilizes RICS data from Nicaragua and Sri Lanka only, omitting Tanzania.

4.1 HOUSEHOLDS

Demographics

The age distribution of the sample households in the three pilot countries is skewed to the left, as shown in Annex D, Table D.2. The percentage share of the male population of the sample households is consistently lower than that of their female counterparts, likely reflecting emigration by the male cohort in search of more remunerative employment.

The rural population of Sri Lanka has completed more education (7.4 years) than have their counterparts in Tanzania (5.0 years) and Nicaragua (4.3 years). The male-female gap is 0.4 years in Sri Lanka and 0.6 years in Tanzania, while in Nicaragua women have received more (0.6 years) education than men. These statistics obviously average only those residing in rural areas; they exclude relatives who may have temporarily or permanently migrated but still belong to the household in a social or economic (breadwinner) sense.

Employment

Comprehensive data for the rural labor force is portrayed in Annex D, Table D.8. The table indicates that the average number of working-age adults (16 to 65 years of age) is higher in Sri Lanka (2.99) than in Nicaragua (2.80) and Tanzania (2.76). The number of working adults, however, is lowest in Sri Lanka (1.78) and highest in Tanzania (2.48): the labor force participation rate in Tanzania is 90 percent, as opposed to 57 percent in Sri Lanka and 65 percent in Nicaragua. The male-female gap is trivial in Tanzania (2 percent) but quite large in Sri Lanka (37 percent) and Nicaragua (34 percent).

Table 4.1 provides information on household members' involvement in major economic activities such as self-employment, wage employment, and farming. It is striking that 46 percent of households in Sri Lanka, 37 percent in Tanzania, and 40 percent in Nicaragua have multiple sources of income. Considering all households together, 50 percent of households in Sri Lanka, 90 percent in Tanzania, and only 29 percent in Nicaragua have at least one member engaged in farming,[25] and 61 percent, 25 percent, and 60 percent of households in Sri Lanka, Tanzania, and Nicaragua, respectively, have at least one member involved in wage employment.

Income

Total household income is estimated by aggregating all sources of income minus expenses of productive activities.[26] Table 4.2 provides total household income in the selected countries. Average household income in Sri Lanka (US$2,693) is more than 25 percent higher than that of Nicaragua (US$2,016) and four times that of Tanzania (US$671). All of the pilots' household surveys show an unequal distribution of income, although less so in Nicaragua (Annex D, Table D.23). In Sri Lanka almost 21 percent of households' average income is less than US$500, while this figure is only 14 percent in Nicaragua.

Households tend to engage in several activities for their livelihoods and subsistence. Household members earn income from wages and salaries; from self-employment in RNFEs; and from agriculture, aquaculture, and animal husbandry.

Table 4.1 Economic Activities of Households and Household Members

	Any income[a] from stated source		
	Nicaragua	Sri Lanka	Tanzania
Enterprise only	7.7	0.0	14.6
Enterprise and wage	8.3	9.9	2.2
Enterprise and farming	3.5	1.5	38.9
Enterprise, wage, and farming	2.7	6.6	6.2
Wage only	45.5	36.0	1.5
Farming only	10.0	18.3	26.6
Wage and farming	12.8	26.4	2.7
Not employed	9.5	1.4	7.3
Total	100.0	100.0	100.0

Source: RIC Surveys.

Note: a. For Sri Lanka, adults living in 149 households with zero or negative total incomes are excluded.

Table 4.2 Average Total Household Income and Its Components

	Wage	Enterprise	Farm	Remittance	Other	Total
Nicaragua	988	434	228	230	136	2016
Sri Lanka	835	162	343	1088	268	2693
Tanzania[a]	81	261	203	n.a.	126	671

Source: RIC Survey Data.

Note: a. Remittance incomes were not recorded by the Tanzanian pilot.

Table 4.3 Measures of Inequality in per Capita Total Income

	Nicaragua	Sri Lanka	Tanzania
Coefficient of variation	2.347	2.942	3.468
Gini	0.568	0.769	0.757

Source: RIC Surveys.

Note: Households with negative total income were omitted from these calculations.

In Tanzania and Nicaragua, the sum of wage and enterprise incomes—earned largely off-farm—constitutes the largest single source of household income.[27] The relative importance of the several income sources in total household income is portrayed schematically in Annex D, Figure D.2.

In all three pilots, the distribution of rural per capita incomes appears to be highly skewed, more so in Nicaragua. In Table 4.3 this is reflected both in the large coefficients of variations for household income estimates from the RIC surveys and in the high Gini coefficients.

Assets and Investments

Sri Lanka has the highest average total asset holding, followed by Nicaragua (Table 4.4). The average asset holding is lowest in Tanzania, where 43 percent of the country's households own less than US$500 in assets. In contrast, only 27 percent of households in Nicaragua own less than US$1,000 worth of assets (Annex D, Table D.19), while 47 percent own assets worth between US$1,000 and US$4,000. In Sri Lanka only 11 percent of households own US$1,000 in assets or less, while

Table 4.4 Cross-Pilot Comparison of Household Assets by Type of Asset (US$)[a]

	Housing	Household Durables	Land	Other Agricultural Assets
Nicaragua	3,162	78	1,869	2,292
Sri Lanka	4,400	1,098	2,900	244
Tanzania	1,683	17	667	96

Source: RIC Surveys.

Note: a. excludes financial assets.

30 percent own between US$1,000 and US$4,000. Average holdings, by type of asset, are summarized in the following table.

Land remains a prominent asset as it strongly influences rural households' economic position. Table D.14 summarizes land holdings per household. Table D.15 provides a distribution of households by land value. The valuation of other agricultural assets extends to production-related tools and equipment, fruit bearing trees, and animal inventories. Table 4.4 above provides estimates of such other agricultural assets. Although land may be the most prominent asset, housing remains the most valuable in all three countries. Durable assets include durable consumables, such as electronics, furniture, transport equipment, and so on. Sri Lanka has the highest durable asset base.

As with per capita incomes, the value of the asset base (especially housing) is most skewed in Nicaragua (see Table 4.5 and Annex D, Table D.19).

Household investment is defined as the net addition to total assets (Annex D, Table D.21). An average household increased its assets by US$136 in Sri Lanka and by US$44 in Tanzania; large standard deviations indicate that household investment varies widely among households. Unfortunately, the Nicaragua survey data do not permit an estimate of household investment.

Tanzania's Lack of Sampling Weights

In contrast to the databases for Nicaragua and Sri Lanka, the Tanzania database lacks household sampling weights. This is an unfortunate omission. In gathering RICS data, the sample of households operating enterprises is augmented with a random sample of households not operating enterprises. To save on costs, this latter sample constitutes a smaller proportion of the total sample than is the case in the overall population. The tables above, containing weighted statistics on Nicaragua and Sri Lanka and unweighted statistics on Tanzania, suggest that nonfarm self-employment and income are more prevalent in Tanzania; the results should be carefully interpreted, however. Because of the sampling design, the rate of entrepreneurship is substantially higher in the RICS household sample than in the population: in the Sri Lanka database, 57 percent of sample households operate an enterprise, whereas the rate of entrepreneurship in the rural population is estimated to be 29 percent; the two Nicaragua percentages are 68 percent and 22 percent, respectively.[28] By contrast, in the Tanzania sample, 62 percent of households derive income from an enterprise, a figure probably so much higher than the population rate that any analysis of entrepreneurship determinants likely suffers from a strong

Table 4.5 Measures of Inequality in Total Assets

	Nicaragua	Sri Lanka	Tanzania
Coefficient of variation	3.610	1.103	3.034
Gini	0.821	0.590	0.838

Source: RIC Surveys.

bias. Especially in an analysis of the entrepreneurship determinants, sampling weights are important in maintaining proper sample representation of the population. Thus, whereas most of this report examines data from the RICS in Nicaragua, Sri Lanka, and Tanzania in parallel, the balance of this chapter drops the Tanzania data from the multivariate analysis.

4.2 ENTREPRENEURSHIP AMONG RURAL HOUSEHOLDS

Definition of the Variables

Precise definitions and descriptive statistics of all variables are provided in Annex F, Table F.1. In Nicaragua, household entrepreneurship is defined as households deriving a positive income from an enterprise. Because of an error in the Sri Lanka questionnaire design, this definition would understate the incidence of entrepreneurship; thus the definition used for Sri Lanka is household members working on their own account.

The explanatory variables are divided into three sets (for more detail on expected effects on household entrepreneurship, see 7Annex F). The first pertains to characteristics of the household and includes demographics and ethnicity, evidence of entrepreneurship by the household head's parents, and access to finance (remittance income, other income, and household assets).[29]

The second set of variables describes the investment climate in the household's community as summarized by the six benchmark indicators. These indicators and their components are used in the econometric analysis to measure the effect of investment climate on entrepreneurship, with a higher value expected to favor enterprise performance. As mentioned above, however, a robust investment climate may expand the nonfarm economy while at the same time creating opportunities more lucrative than household-owned and operated enterprises, thus leading to enterprise closures. In other words, the effect of investment climate on entrepreneurship may be ambiguous.

The third set of variables describes the community apart from immediate associations with the policy-related investment climate: community size, level of per capita income, typical male wage-rate, enterprise openness, and agricultural seasonality. As for the BIs, the a priori effect of these variables on entrepreneurship is ambiguous.

Regression Results

Entrepreneurship is a discrete 0/1 outcome and is therefore analyzed with a probit regression model. Sampling weights are applied. The clustering of households in communities makes it plausible that, aside from observable community characteristics, unobservable community factors may be common to all households, causing correlation among the model's disturbance terms. This renders the standard probit model inadequate: a random effect must be incorporated. In cases where the estimated standard deviation of the random effect proved to be exceedingly small, however, the estimation strategy was simplified to a weighted probit model (without a random effect).[30]

Table F.2, Annex F, reports the estimated models. For Nicaragua, the random effect always vanishes in estimation. For Sri Lanka, a random effect model is needed, since the standard deviation of community random effect is estimated at about 0.35, one-third the standard deviation of household-level random disturbance. Related to this, compared with Sri Lanka, the Nicaragua regressions find more community variables with statistically significant effects—similar to the regression results for enterprise performance discussed in Chapter 3.

The first column of Table F.2 contains only the six benchmark indicators; the second column adds the household characteristics; the third column fills in the remaining community variables. The fourth column selects variables from among the set of benchmark indicators, their components, and other community variables making the largest contribution to explanations of entrepreneurship choice.

Since parameter estimates of the probit model are tedious to interpret and the scales of the variables differ, the table shows the percentage point effect of a one-standard-deviation change in each explanatory variable, evaluated for an average household. For explanatory dummy variables such as "Head female," "Head parents entrepreneur," and "Not Sinhalese," the change in the explanatory variable is one unit, which is clearly more appropriate. It is useful to keep in mind that the rate of entrepreneurship in Nicaragua is 22 percent. Thus, the second column indicates that a household with 0.99 additional male adults (the standard deviation of this variable; Table F.1) is 3.08 percentage points more likely to operate an

enterprise. Also, when parents of the household head were entrepreneurs, the probability that the household operates an enterprise rises by 2.27 percentage points, although this effect is not estimated precisely as the t-statistic equals only 0.86 (see Table F.2, part A).

In Nicaragua, entrepreneurship is more likely in households with more adults, whether male or female, and in households with older members. Human capital does not matter; skilled workers may be more productive in the enterprise but, as the literature on wage earnings shows, they are also capable of earning higher wages when working elsewhere. Female-headed households are 6 percentage points more likely to operate an enterprise.

In Sri Lanka, where 25 percent of households operate an enterprise, the effect of household size, age structure, and human capital is smaller and not statistically significant. Female headship has a numerically strong negative effect, equal to 9 percentage points, but it, too, is not precisely enough estimated to attain statistical significance. Members of the ethnic (non-Sinhalese) minorities tend not to operate enterprises, although the effect is also not precisely estimated. Surprisingly, the occupation of the head's parents proves to be irrelevant in both countries. Elsewhere, such as in Vietnam (see, for example, Vijverberg and Haughton 2004), the intergenerational effect is rather pronounced.

In regard to household financial variables, the specification of the entrepreneurship model varies between the two countries owing to the bias in measured income components referred to above for Sri Lanka. In Nicaragua, access to remittance income and other income sources reduces the incentives for self-employment.

On the other hand, assets improve the chances. Specifically, the standardized effect of log-assets raises the probability of entrepreneurship by 5 percentage points in Nicaragua and 10 percentage points in Sri Lanka. To interpret this for Nicaragua, where the average of log-assets amounts to US$1,419, a one-standard-deviation change in log-assets equals US$6,485, which would take a household from, say, the twenty-fifth percentile to about the eighty-fifth percentile of the distribution of assets. For Sri Lanka, the 10 percentage point increase in the likelihood of entrepreneurship comes about in response to a rise of US$11,621 in assets, relative to a base of US$5,405; this is equivalent to moving the household from the twenty-fifth percentile to the eightieth percentile of the distribution of assets.[31] These computations show that household assets are important determinants of entrepreneurship but also that establishing and operating an enterprise demands large sums of resources.[32]

Among the benchmark indicators, community human capital stands out in Nicaragua, where the standardized effect is about 5 percentage points. Infrastructure services have a positive effect of about 5.5 percentage points as well, but with the introduction of other community variables the effect diminishes to about 3. Business services may reduce the likelihood of entrepreneurship, and connectivity, governance, and finance have no effect. The last column of Table F.2, which contains only those community variables with the greatest explanatory power, lists positive effects of proximity to the post office and access to protected water: for both, the effect is 3 percentage points per standard deviation increase in the benchmark component.

The Sri Lanka results list a positive effect for connectivity, with the proximity to the post office highlighted in this country as well; the effect of 5 percentage points is quite large. The indexes for infrastructure services,, business services, governance, human capital, and finance do not appear to matter much. The last column highlights sewage systems, which are found to have a strongly negative effect and thus contribute to the overall negative effect of the infrastructure services indicator; engineering services, with a negative effect; and information technology services, with a positive effect. The negative effects are counterintuitive, but these variables are all simple 0/1 indicators. As only 13 communities have sewage systems and merely 5 have engineering services available, the estimation results may simply be the luck of the draw.

The last group of variables consists of various community characteristics. Here, Nicaragua and Sri Lanka yield opposite results. In Nicaragua, enterprise openness (that is, selling and buying outside the community) raises the likelihood of entrepreneurship in a household by about 5 percentage points. The community income level may have a positive effect, but its effect is collinear with community human capital, the effect of which is more robust and therefore enters the last column. Agricultural seasonality, community size, and the level of male wages do not matter. In Sri Lanka, on the other hand, the likelihood of entrepreneurship

is higher in larger communities with higher wages, although income per capita, seasonality, and enterprise openness had no effect.

As mentioned before, Nicaragua's model is weighted probit without a random community effect. In Sri Lanka, the random community effect is statistically important. A one-standard deviation in the random effect moves the likelihood of entrepreneurship in a household up by almost 10 percentage points in the model with benchmark indicators and 7 percentage points in the model with selected benchmark components.

To conclude this examination of household entrepreneurship choice, it is worthwhile to reflect on the explanatory power of the model. The evidence is summarized in Table F.3. The overall explanatory power of the model with household variables, benchmark indicators, and community characteristics (Panel A) is 12.42 percent in Nicaragua and 7.85 percent in Sri Lanka. Household characteristics explain (that is, improve the criterion function by) 8.09 percent in Nicaragua and 4.58 percent in Sri Lanka, where, it should be noted, the model does not include remittance and other income, which are influential variables in Nicaragua. The benchmark indicators help to explain 2.98 percent in Nicaragua and 1.46 percent in Sri Lanka, and community characteristics further add about half as much. In all, benchmark indicators and community characteristics contribute 35 percent (= (2.98 + 1.35)/12.42) of the explanation of entrepreneurship choice in Nicaragua and 28 percent (= (1.46 + 0.72)/7.85) in Sri Lanka. Further checks confirmed that benchmark indicators carry relatively less information in Sri Lanka than in Nicaragua.

4.3 ACTIVITIES OF HOUSEHOLDS

Typically, two-thirds of households with enterprises are also involved in other activities; even among individuals, many engage in several modes of employment over the course of a twelve-month period (see Table 4.1). This parallels findings in Cote d'Ivoire, Ghana, Guatemala, Kyrgyz Republic, Peru, and Vietnam (Moock, Musgrove, and Stelcner 1990; Lanjouw 2001; Vijverberg 1990, 1998, 1999, 2005). Even if some activities—wage, enterprise, or farm—are rather insignificant, earning less than $60 per year, the level of participation demonstrates that employment diversity is the norm.

The top four lines of Table F.1, Annex F, list the percentages of households in Nicaragua and Sri Lanka participating in the various activities. The rate of wage employment is surprisingly high, given that these are rural samples; the percentage of households that do any farming at all is only 27 percent in Nicaragua, which has large numbers of farm laborers, and 48 percent in Sri Lanka.[33] This raises the question of how rural these communities are.

In any case, just as households may choose to engage in nonfarm entrepreneurship, they may also choose to pursue wage employment or farming. Therefore, this section extends the analysis of entrepreneurship choice to the other wage employment and farming choices households make. Moreover, as mentioned above, since during the course of economic development people may switch from self-employed farming into rural nonfarm self-employment and wage employment, the percentage of households that pursue wage employment or nonfarm entrepreneurship is of interest: it integrates the development of the nonfarm economy into a single concept.

Estimation results are reported in Table F.4, showing only the specifications with selected components. Each equation is estimated separately with a random effect probit model.[34] The first column refers to engagement in nonfarm enterprises and is therefore identical to the fourth column of Table F.2.

In Nicaragua, households with more men diversify into more activities; those with more women pursue only nonfarm entrepreneurship. In Sri Lanka, diversification is stronger yet when there are more adult males; specialization is stronger when there are more adult women. The age pattern is similar in the two countries: households with older members are more likely to farm and perhaps operate an enterprise; younger households seek wage employment. In Nicaragua younger households also engage more in entrepreneurship. Human capital draws households into wage employment, perhaps out of nonfarm entrepreneurship and likely out of farming. The entrepreneurial experience of the head's parents, however, leaves no sorting imprint on the head's household.

Nicaragua and Sri Lanka differ in the sorting patterns according to the household head's gender. Female-headed households in Nicaragua are more likely to operate an enterprise, whereas those in Sri Lanka are significantly more likely to have

wage jobs rather than a household enterprise. In both countries, however, farming is less likely for female-headed households; the effect is large but imprecisely estimated. In Sri Lanka, non-Sinhalese are less likely to participate in the nonfarm economy. In particular, they are found in farming rather than in nonfarm businesses.

Extraneous income sources reduce the incentive for households to participate not only in a business, as seen above, but actually in any kind of work. This is consistent with the notion that, in standard labor economics terms, leisure is a normal good. Assets strongly sort households into both enterprises and farming; households with few assets seek wage employment.

The third column of Table F.4 focuses on nonfarm economic activity, combining (both farm and nonfarm) wage employment and nonfarm self-employment. For some explanatory variables, the third column sums up contradictory signals from the first and second columns of the table; for example, see the effect of age, female headship, ethnicity, or asset ownership. Among the strongest results, human capital pulls people into nonfarm economic activity. In Sri Lanka, this is supplemented by a positive effect for adult women and negative effects for age and non-Sinhalese ethnicity. But balancing these insights is an equally valid inference that an exclusive study of nonfarm economic activity hides interesting sorting patterns in the economy among wage and self-employment.

Searching for the best explanation through community variables, benchmark indicators, and benchmark components yields various results. In Nicaraguan communities with higher incomes, households are more likely to engage in wage employment and less likely to be found in self-employed farming, and, when seasonality is more pronounced, farming becomes more likely and wage employment less. Wage employment diminishes when more households have access to electricity or when government agencies are more transparent in their dealings with entrepreneurs in the community: these effects are counterintuitive, since one would expect the demand for wage labor to flourish under those conditions.[35] The weaker negative effect of human capital in the community should be seen in the light of the positive private effect of human capital: a comparative advantage in skills disappears when other members in the community are skilled as well. For farming, the data suggest that the more rural and remote the community, the more likely households will be to earn their living by farming, although the estimates are not precise enough to imply statistical significance. Finally, the community random effect may be absent in the determination of entrepreneurship choice, but it is highly significant in households' choice of wage employment and farming: the standardized effect ranges from 15 to 19 percentage points.

In Sri Lanka, the list of community variables and benchmark components is long and varied. Households in larger communities are more likely to work for a wage and less likely to farm. Greater seasonality is associated with farming choice. Turning to benchmark components, chances of wage employment seem higher in communities farther from a major city or closer to a main market. Overall access to infrastructure helps; more detailed specifications show positive effects for the percentage of households with electricity and the presence of concrete or asphalt roads and a negative estimate for the percentage of households with cell phones. Furthermore, according to the estimates, wage employment is more likely with the presence of marketing services available to businesses in the community and less likely with higher "financial penetration," that is, the percentage of entrepreneurs in the community that have considered asking a financial institution for a loan in years prior to the RIC survey. Farming is more prevalent in communities with poorer access to electricity or phone service. As the last variable, financial penetration has a positive effect, as farming may also benefit from access to credit.[36] Across all outcome variables in Table F.4B, the random community effect is influential. In particular, the prevalence of farming varies greatly from community to community in ways not captured in the observable information about communities. This is consistent with the notion that farming, more than other enterprises, depends more on land and water availability in the location than on investment climate variables.

4.4 ENTERPRISE START-UP

Measuring Start-Up

This section examines the effect of investment climate on enterprise start-up. Households are considered to have started up an enterprise if at the time of the survey the enterprise they operate is

less than two years old and is based on the premises of the household's residence. This understates the true two-year start-up rate, since these enterprises must have survived for up to two years to be included; the survey cannot count recently started but already defunct enterprises or enterprises started as or converted into stand-alone units.

In Sri Lanka, therefore, where many enterprises are stand-alones, the measured start-up rate is only 1.2 percent, compared to an estimated true start-up rate that might be as high as 6.05 percent (see Annex F, Box F.1). In Nicaragua, the situation is quite different: the enterprise start-up rate is measured to be 3.0 percent, probably a fairly accurate number, because stand-alone enterprises are only 7.5 percent of the sample.[37]

If it is assumed that the enterprise population is stationary, such that new start-ups merely replace exiting enterprises, enterprise start-up information may be translated into enterprise survival estimates. If so, the annual enterprise survival rate[38] among household-based and stand-alone enterprises in Sri Lanka equals 86.6 percent and 87.2 percent, respectively, and in Nicaragua, where it is assumed that the data captured only half the start-ups in 2004, 90.7 percent and 84.7 percent. By comparison, the enterprise survival rate is about 83 percent in Vietnam and similar elsewhere (Vijverberg 2005).

Regression Results

Table F.5, Annex F, examines enterprise start-up determinants. These are estimates of weighted probit models, expressing the effect of explanatory variables, again in standardized form. The community random effect vanished in both countries as its estimated standard deviation converged to 0. Moreover, because the number of start-ups is so small, the model could not be populated with a full set of benchmarks and community characteristics. This very fact implies that estimates of community variable effects should be interpreted carefully.

Actually, few explanatory variables matter. Moreover, as the average start-up rate is only 3.0 percent in Nicaragua and 1.2 percent in Sri Lanka, the estimated effects are small, certainly in comparison with Table F.2, where the average rate of entrepreneurship was 22 percent and 25 percent respectively. Among the statistically significant estimates, in Nicaragua only households receiving income from remittances and other sources are less likely to start up a new enterprise. In Sri Lanka, assets had a positive effect; the number of female adults and the entrepreneurial experience of the head's parents may raise the likelihood of start-up; chances are lower among older or non-Sinhalese households.

Among benchmark indicators, connectivity hurts entrepreneurship chances in Nicaragua but has no effect in Sri Lanka. Proximity to a major city has a negative effect in Nicaragua but a positive one in Sri Lanka. Infrastructure services has a positive effect in Nicaragua but a negative one in Sri Lanka. Community human capital raises chances in Sri Lanka. The remaining indicators do not matter.

Other community characteristics show a similar disagreement between the two countries. Enterprise start-up is more likely in Nicaragua in larger communities where enterprises are accustomed to trading with clients outside the community and where agricultural seasonality is more pronounced. Of these three variables in Sri Lanka, enterprise openness has an opposite effect, and the other two do not matter.

Overall, therefore, the effect of investment climate varies considerably between the two countries. It is possible that behavior differs; as mentioned above, the effect of household and community variables is a priori ambiguous. Alternatively, these may simply be random patterns that appear when communities are few and variation in the dependent variable is limited. Moreover, one should expect differences between the two countries since Sri Lanka has a much greater proportion of stand-alone entities: the analysis concerns start-up of household-based enterprises only, omitting start-up of stand-alone enterprises.

4.5 GENERAL DISCUSSION

Having completed this empirical analysis, it is useful to ask whether a study such as this accomplishes important objectives. What light does it shed on entrepreneurship choice? Is the research vehicle—the econometric model—appropriate? Is the investment climate properly captured in the analysis? Questions of whether the data are appropriate for the objective will be addressed in the following section.

Entrepreneurship Choice

This chapter looks at only one facet of entrepreneurship choice: whether to set up a business. Entrepreneurs make other choices: which sector to enter, how to arrange for financing, which technology to use, where to locate, and how to seek out clients, to mention a few. In this sense, this chapter's scope is limited. The RICS data, however, do offer opportunities for further analysis to better understand the full dimension of entrepreneurship choice.

Entrepreneurship choice must be placed in context. If farming is more profitable than anything else, rational people will operate a farm. If a wage job pays better, many people will elect to work for an employer. If one's own enterprise is the most profitable option, it makes sense to start a business. Income may not be the only consideration in this comparison, as each option has its own nonmonetary compensation and imposes its own income risk. Households may pursue diversification strategies. In all, the results obtained in this chapter illustrate how a focus on the 0/1 outcome of enterprise operation fails to uncover a rich sorting pattern across the different economic activities.

This chapter also illustrates that adopting a dichotomy of farm versus nonfarm economic activities, where the latter aggregates wage and nonfarm self-employment, can lead to an inadequate insight into the process of development. Self-employment sorting patterns are not the same as those for wage employment. At the same time, the RICS data examined here do not distinguish wage employment by sector: in the future, it would be useful to separate agricultural and nonagricultural laborers.

Econometric Modeling

Participation in farming, wage employment, and a nonfarm enterprise was modeled with three separate regression equations, each estimated independently. This actually hides the interdependence. Greater efficiency is feasible by simultaneously estimating the three equations, but this comes at a cost of complexity. The structure of the data has households clustered in communities. In consequence, the model must allow for unobservable community factors (that is, random effects). Estimation of a single random effect probit equation is already a more laborious procedure than the simple, common probit model; simultaneously estimating three equations will be time consuming.

An alternative approach to the simple 0/1 modeling strategy would be to describe how many of the household members and even how much of their time is devoted to each economic activity. This converts the entrepreneurship decision into a continuous variable with a lower bound at 0. It also extracts more information from the RICS data, placing more demands on their collection.

Application of the random effect probit model is complicated by the stratification of the household sample and the need to use sampling weights. A typical RICS database draws more heavily from households with enterprises than from households without them, whereas the population consists of more households without enterprises than with them. Thus, an unweighted analysis of economic activities yields biased estimates. Stata software, however, does not allow sampling weights to be specified for the panel data procedures used when community random or fixed effects enter the model. In experimentation with different estimation methods, estimates sometimes varied substantially among weighted probit, unweighted random effect probit, and weighted random effect probit.

Estimation is further complicated by potential endogeneity of explanatory variables, such as income levels in the community and the household's assets and nonlabor income. The analysis above failed to disprove the exogeneity of the first two variables and did not examine the potential endogeneity of nonlabor income. But more research is needed to successfully incorporate endogenous explanatory variables.

Capturing the Community's Investment Climate

The estimates in this chapter show that the community random effect is frequently important. Its standardized effect is usually as large as, and sometimes much larger than, any of the observed explanatory variables. This shows that community factors matter. The question is whether the survey data have captured them in meaningful ways.

Yet, in Nicaragua's entrepreneurship choice model and in both countries' enterprise start-up model, the estimated standard deviation of the community random effect was zero, and the random effect vanished. As Table F.3, Annex F, shows that in Sri Lanka community benchmark indicators and community characteristics contributed substantial explanation after allowance was made

for a community random effect. Nevertheless, the effect of benchmark indicators is mostly muted and varies between the two countries.

The investment climate is a multidimensional concept. Defining benchmarks does not prove their usefulness, and it is as yet an empirical question whether the investment climate actually matters. Theory may suggest contrary directions of influence. Connectivity, for example, opens up distant markets for local entrepreneurs and local markets for distant entrepreneurs. Better governance may be good both for small-scale entrepreneurs and for large-scale employers that hire many locals for a wage. But a number of benchmark indicators clearly relate to the likelihood of entrepreneurship, whether positively or negatively, which may be viewed as evidence that benchmarks matter.

The aggregative nature of benchmarks may also be their weakness. If only one component matters, its effect is weighed down by its pairing with a number of other ineffective components. It is therefore advisable to search among benchmark components for variables that provide strong explanatory power. The drawback is that there are so many components. Each is listed as a candidate variable because it reflects an aspect of the investment climate and might, on economic grounds, matter for entrepreneurship, enterprise start-up, or enterprise performance. The eventual selection is the result of, essentially, a statistical rather than an economic comparison, that is, a stepwise regression procedure. The selection is then vulnerable to accidental statistical significance, as probably illustrated by the effect of management consulting and marketing services on farming in Sri Lanka (Table F.4B). In retrospect, some aspects of the investment climate in play in only a few (say, less than 10) communities should perhaps have been discarded from this selection procedure.

Nevertheless, experimentation has shown that a few well-selected components and community characteristics can provide equal or better explanatory power than a full list of benchmark indicators and community characteristics. Proximity to the post office, for example, matters for entrepreneurship choice in both countries. Access to protected water raises chances in Nicaragua, as do information technology services in Sri Lanka. Harder to explain are the negative effects of the presence of a sewage system and the availability of engineering services in Sri Lanka.

Implications for RIC Methodology

The data from Nicaragua, Sri Lanka, and Tanzania were gathered during the pilot phase of the RIC project. The analysis in this chapter uncovers at least four important lessons for the RIC methodology. First, sampling weights must be calculated when data are collected: since the sampling strategy includes clustering by community and stratification by entrepreneurship status, implementation of the survey implies that sampling weights must be carefully constructed and constitute an essential element of the survey data.

Second, in a survey database that includes both household and enterprise data, one would expect that entrepreneurship in a household would be defined by (i) whether a linked enterprise was present, (ii) whether any members of the household were employed on their own accounts, and (iii) whether the household earned income from work on its own account—and that these three aspects would be mutually consistent in the database. The three countries' databases exhibited too many inconsistencies. In future RIC surveys, inconsistencies should be flagged and resolved while teams of enumerators remain in the field.

Third, the investigation of enterprise start-up among rural households was greatly hindered by the design of the RIC survey as implemented in Sri Lanka and Nicaragua, which treated stand-alone enterprises as entities without a household connection, whereas the data (particularly in Sri Lanka) show that stand-alone enterprises are under household ownership. As a result, enterprise start-up measures refer only to household-based enterprises, thus weakening the analysis. The survey design should never view an enterprise separately from its owner but should, rather, interview the associated household whenever possible.

Finally, the households in the research samples reside in 93 communities in Nicaragua and in 117 communities in Sri Lanka. Community-level variables can explain only the between-community variation in the dependent variable. This means that the investigation of investment climate effects works with roughly 100 data points. Yet, the investment climate has many dimensions: the six benchmark indicators alone consist of 30 components. This makes reliable detection of investment climate effects quite challenging.

5 Perceptions About Investment Climate Constraints

An important objective of the Rural Investment Climate Survey program is to give entrepreneurs the opportunity to voice their interests and concerns regarding the investment climate. How do they view the environment in which they operate their businesses? How do these perceptions vary by sector and by enterprise type? Does the perception correspond to the actual conditions in the community, as assessed by objective measures? Questions of perception are relevant, since it may be argued that entrepreneurs base their decisions on their perceptions. They are more likely to invest when they view the climate as favorable; they are more likely to hire workers if they anticipate profitability for their operations.

The RIC questionnaire lists features ranging from availability and cost of electricity to agricultural issues to concerns about crime and civil unrest: for each of these features, respondents are asked whether it constitutes a barrier to the operation and growth of their enterprise. In essence, these responses supply subjective measures of the investment climate: they reflect perceptions colored by the respondent's frame of reference and, presumably, by actual conditions in the community and the region. These variables are dubbed Enterprise Investment Climate Outcomes (EICOs).

Apart from these questions, entrepreneur respondents are asked about their interactions with government agencies and their views on the legal system as it relates to operation of an enterprise. Some of the responses are objective, such as the number of inspections or the time spent dealing with government officials; others are subjective, such as level of trust. Entrepreneurs do not rate these conditions as barriers; the questionnaire makes no attempt to link these facets of the investment climate to enterprise operation and growth. Rather, whether essentially objective or subjective, factual responses are sought. These variables are dubbed Enterprise Investment Climate Interactions (EICIs).

This chapter examines determinants of EICO and EICI perceptions. Section 5.1examines which investment climate elements rank high as barriers on the lists of entrepreneurs' concerns; Section 5.2 describes the model; and Sections 5.3 and 5.4 present estimates of regression models for EICO and EICI variables. Section 5.5 offers suggestions for further research, and Section 5.6 discusses implications for the RIC methodology.

5.1 ENTREPRENEURS' INVESTMENT CLIMATE CONCERNS

5.1.1 EICOs

The questionnaires list nearly 40 investment climate features entrepreneurs may indicate as barriers to the operation and growth of their enterprises. This list is wide-ranging and may be divided into the following topics:

- Public utilities
- Transportation
- Financing
- Marketing
- Registration, license, and permits
- Taxation
- Labor
- Land
- Agricultural policy
- Nonagricultural trade policy
- Environmental policy
- Governance[39]

Across the three pilots, the detailed questions were broadly similar; any differences are inconsequential for a comparative analysis.[40] In Sri Lanka and Tanzania, the responses are coded as *0 = No obstacle* to *4 = Very severe obstacle,* whereas in Nicaragua the variation in the responses is limited to *0 = No obstacle, 1 = Moderate obstacle,* and *2 = Severe obstacle.* This hinders direct numerical comparison between countries, but it does not affect qualitative comparison of the types of barriers entrepreneurs highlight most prominently.

Figure 5.1 ranks the more important EICOs, treating the categorical responses as true ordinal numbers; Table H.1, Annex H, reports detailed frequencies by category for all variables. In Nicaragua and Sri Lanka, entrepreneurs highlight some 10 to 12 barriers; the other EICOs fail to elicit much of a response. In Tanzania, concerns are more wide-ranging and, compared to Sri Lanka, which uses the same coding scheme, more intense.[41]

To get a better sense of the barriers generally important in rural areas, Table 5.1 lists the top ten EICOs in each country as well as the top ten overall for these countries on the basis of their average ranks. Six of the ten most frequently cited EICO barriers are common across Nicaragua, Sri Lanka, and Tanzania. On the individual country lists, no more than two important concerns (that is, listed among the top ten) are unique to that country.

Overall, finance is viewed as the most important barrier: high interest rates, difficult loan procedures, and problems finding lenders rank 1, 2, and 4, respectively. Electricity ranks third to fifth in each country and third overall. Issues with road quality, road access, and water supply rank 5, 7, and 8, respectively, and appear on each country's list. Opinions about market demand, economic policy uncertainty, and corruption vary sharply across countries, ranking near the top in one country but being rather unimportant in another. Telecommunication is ranked more evenly and thus makes the top-ten list even though no country ranks it higher than seventh. Because of inherent interest, Tables H.8 and H.10 elaborate on the measurement of the corruption EICO.

This brief summary indicates both the breadth of topics covered in the questionnaire and their lack of detail. Concerns about "electricity," for example, can take many forms: obtaining a hookup, experiencing frequent blackouts, high tariffs, and the like. "Water supply" and "telecommunication" are equally vague. But for this study, detailed inquiry into every EICO dimension was simply not feasible.

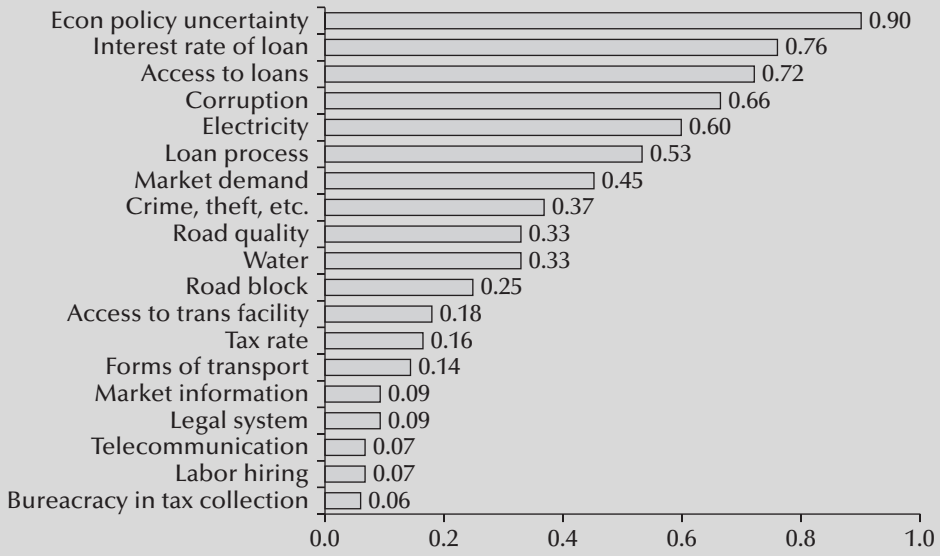
5.1.2 EICIs

The RIC survey instruments contained questions on numerous aspects of respondents' business experience: applications for and renewals of licenses and permits; time and monetary outlays (both formal and illicit) when dealing with government agencies; utilities issues; efficiency of government services, including whether government officials can be influenced regarding enforcing and drafting laws and regulations; and the legal system and conflict resolution methods. Disentangling the many patterns in all this information is beyond the scope of this report, but it is instructive to examine a few of the more accessible and comparable items.

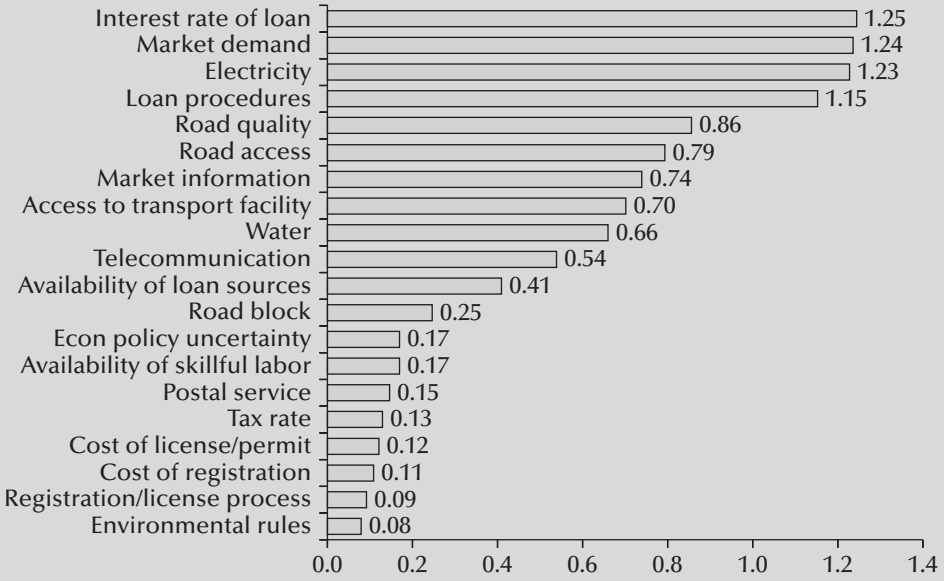
Table 5.2 pulls together three elements of government efficiency: the predictability of laws and regulations, the willful or unintentional misinterpretation of laws and regulations, and influence on government officials. Many entrepreneurs feel insecure about laws and regulations: often, enforcement of these is perceived to reflect government officials' whims. With regard to

Figure 5.1 Enterprise Investment Climate Outcomes

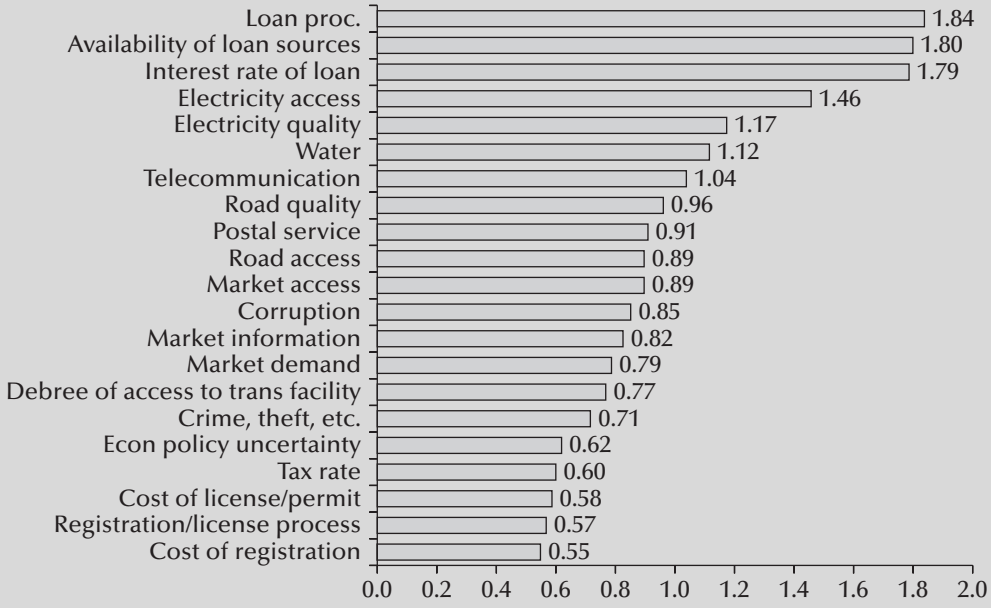
A: Nicaragua



B: Sri Lanka



C: Tanzania



Source: RIC Surveys.

Table 5.1 Ranking EICOs Across Countries

Rank	Nicaragua	Sri Lanka	Tanzania	Overall	Average Rank[d]
1	economic policy uncertainty	interest rate of loan	loan procedures	**interest rate of loan**	2.00
2	interest rate of loan	market demand	availability of loan sources	**loan procedures**	3.67
3	access to loans[a]	electricity	interest rate of loan	**electricity[b]**	4.17
4	*Corruption*	loan procedures	electricity access	availability of loan sources	5.33
5	Electricity	road quality	electricity quality	**road quality**	7.33
6	loan process	road access	water	market demand	7.67
7	market demand	*market information*	telecommunication	**road access[c]**	8.00
8	*crime, theft, etc.*	*access to transportation facility*	road quality	**water**	8.33
9	road quality	water	*postal service*	economic policy uncertainty	10.33
10	Water	telecommunication	road access	telecommunication	11.33

Source: RIC Surveys.

Notes: Items in italics do not appear in the overall list. Bold-faced items occur in the top-ten list for each country.
a. Interpreted to be similar to "availability of loan sources."
b. In Tanzania, electricity concerns distinguish between access and quality, which are given equal weights in computing the overall rank.
c. Measured in Sri Lanka and Tanzania only.
d. Simple average of the country-specific rankings.

influencing government officials in Nicaragua, the question referred to the effect of informal payments by the entrepreneur or others,[42] whereas in Sri Lanka and Tanzania respondents were asked whether they or other entrepreneurs could influence government officials. In Nicaragua a similar follow-up question probed the degree of influence, but responses were few.[43] In any case, about 15 percent of Nicaraguan entrepreneurs consider themselves affected—whether favorably or not—a considerably higher response than in Sri Lanka or Tanzania, where bribes are considered effective regarding the enforcement of laws and regulations but not their drafting.

Entrepreneurial perspectives on the legal system are summarized in three measures: whether the entrepreneur must rely on the reputation of those with whom they enter into agreements; whether entrepreneurs expect a contract to protect them from being cheated by others; and whether they expect that the legal system will uphold their contract and property rights in business disputes. Table 5.3 summarizes the responses. The coding of the responses varied a little across countries: wheres Sri Lanka and Tanzania used codes ranging from *Strongly agree* to *Strongly disagree,* the Nicaraguan questionnaire allowed only *Agree, Neither agree nor disagree,* and *Disagree*. Relative intensities are expressed in the table formatting. Moreover, the numerical summary uses values of 0/1/2/3 for Sri Lanka and 0.5/1.5/2.5 for Nicaragua. With this proviso, the countries' responses may be compared.

A need to rely on reputation illustrates an incomplete market mechanism: a low average score indicates a more problematic investment climate. The other two items, protection by means of contracts and support from the legal system, reveal a better-developed market mechanism: in this case, higher scores are evidence of a poorer investment climate.

One might expect the need to rely on reputation to be lessened by contracts and legal system support and thus, to the degree that correlation coefficients of these categorical responses are appropriate, that correlations would be negative.

Table 5.2 Investment Climate Evidenced in Government Efficiency

A: Predictability of laws and regulations

	Completely predictable	Highly predictable	Fairly predictable	Fairly unpredictable	Highly unpredictable	Completely unpredictable	Total	N
Nicaragua	→	55.3	←	→	44.7	←	100	1139
Sri Lanka	6.5	27.6	35.6	20.9	9.4	←	100	1324
Tanzania	19.1	16.1	20.7	23.2	11.5	9.4	100	747

B: Laws and regulations can be misinterpreted or manipulated

	Strongly agree	Agree	Disagree	Strongly disagree	Total	N
Sri Lanka	9.7	33.0	44.6	12.7	100	730
Tanzania	18.6	24.9	44.1	12.5	100	1051

C1: Your establishment or others can influence the contents of laws and regulations

	No	Yes	Total	N
Sri Lanka	90.4	9.6	100	432
Tanzania	88.4	11.6	100	1211

C2: Your establishment or others can influence officials in drafting of laws, decrees, regulations, etc.

	No	Yes	Total	N
Sri Lanka	97.9	2.1	100	472
Tanzania	93.1	6.9	100	1216

C3: Nicaragua: Impact on your business from kickbacks

	No impact	Little impact	Lot of impact	Total	N
. . . laws	84.4	9.7	5.9	100	904
. . . enforcement	83.8	10.0	6.2	100	904
. . . judge	84.8	9.8	5.4	100	894
. . . credit	87.4	9.0	3.7	100	880
. . . politician	86.1	10.0	3.9	100	875

Source: RIC Surveys.

Note: Arrows show contractions in the response coding that led to combining cells.

The correlation between reliance on reputation on the one hand and contract protection and legal system support on the other is, however, around 0.6 in Sri Lanka, 0.5 in Tanzania, and 0.55 in Nicaragua. In comparison, the correlation between contract protection and legal system support is 0.4 in Tanzania and 0.74 in Sri Lanka. It is difficult to reconcile thes strong positive correlations with the implications for the measured RIC.

5.2 MODELING ISSUES

EICOs and EICIs express the investment climate as seen by the entrepreneur. Perceptions are founded on the actual environment, on the entrepreneur's frame of reference, and on encounters occasioned by business operations. To elaborate, the economic, social, and political environment in which the enterprise operates should help determine perceptions: economists often assume that entrepreneurs

Table 5.3 Investment Climate Evidenced in the Legal System

	Strongly agree		Agree		Disagree		Strongly disagree	Total	N	Mean
A: Sri Lanka										
Need to rely on reputation	38.6		54.7		5.9		0.8	100	1021	0.69
Contract will protect	28.3		58.1		11.1		2.5	100	953	0.88
Legal system will support	26.6		56.0		15.2		2.3	100	829	0.93
B: Tanzania										
Need to rely on reputation	32.5		44.2		16.8		6.5	100	1194	1.97
Contract will protect	29.9		51.1		12.9		6.1	100	1193	1.95
Legal system will support	22.2		51.6		18.3		7.9	100	1195	2.12
C: Nicaragua										
Need to rely on reputation	_	54.0	_//_	19.5	_//_	26.5	_//	100	1110	1.22
Contract will protect		52.9		18.6		28.5		100	1111	1.26
Legal system will support		50.0		21.0		29.1		100	1091	1.29

Source: RIC Surveys.

have perfect information in this regard, but reality may fall well short of that. Experience, education, and network relationships build a mindset about the external world. Business operations may bring the entrepreneur face to face with regulators and government officials. One type of business (for example, transport) may encounter barriers that other types (for example, personal services) never deal with. The econometric model developed below attempts to capture some of these features as an explanation of the variations in entrepreneurial perceptions of the investment climate.

This enterprise-level information is complemented by community-level information. Much of the latter reflects objective, factual conditions, information on which was gathered from knowledgeable community respondents including community leaders, teachers, business leaders, and local bank officials. This information was used in two ways. First, one set of models employs five of the six benchmark indicators.[44] This approach is compact, as benchmark indicators combine a host of community characteristics, but its drawback is that benchmark indicators sometimes mask interesting detail.[45] For this reason, a second set of models contains the selected benchmark indicator components that are a priori the most relevant community characteristics for the analysis of a given EICO. This second model therefore varies between EICOs. A search for the best predictive model, as was done for enterprise performance in Chapter 3, was not attempted since, with ten EICOs under study in each country, it would be a Herculean task. The descriptive statistics of the explanatory variables are provided in Table H.2, Annex H.

EICO responses are discrete categories on an integer scale from 0 to 4 for Sri Lanka and Tanzania and from 0 to 2 for Nicaragua; EICIs are measured on various scales, including 0/1. This calls for estimation with an ordered probit model (or probit if the outcome is 0/1). Since enterprises cluster in communities, responses are determined by (observed) community-level information and by community-level unobservables representing unmeasured and unmeasurable community characteristics, that is, by a community-level random effect.

Annex I outlines the econometric details of the estimation of the weighted random effect ordered probit model. Annex H contains a further note about the computation of standard errors of the estimate. In order to facilitate the interpretation of the results, the parameter estimates are converted to reflect the effect of a one-standard-deviation change of a continuous variable[46] or a one-unit (0 to 1) change of a dummy variable on a scale that corresponds with the original discrete coding. Technically, the ordered probit model estimates a "tendency to report obstacles" equation that is subject to an arbitrary scaling that disconnects the parameter estimates from the coding of the EICOs and renders comparison between different EICOs

and different countries difficult. For the purpose of this report, the average tendency of a typical enterprise reporting no obstacles is set at 0, and the average tendency of a typical enterprise reporting very severe obstacles is set at 4 (or 2 in the case of Nicaragua). These averages take into account both the randomness in the responses at the enterprise level and also the community-level random effect.

This scaling facilitates interpretation of the estimation results. The effect of connectivity on electricity in Nicaragua, for example, is reported to be −0.194: this implies that a one-standard-deviation increase in connectivity—an improvement in the investment climate—reduces the tendency to view electricity as an obstacle by 0.194 on a 2-point scale.[47] Consider this dummy variable example: in Sri Lanka the effect on "electricity as an obstacle" of being a female manager is reported to be −0.361: on a 4-point scale, by 0.361 points female entrepreneurs view electricity as less of an obstacle than do their male counterparts.

Expressing all EICO results on the same scale also facilitates comparison across EICOs. Estimates for Tanzania, for example, indicate that improving connectivity actually matters little for the problems entrepreneurs report having with road access (−0.017) or road quality (−0.150), but it helps in locating sources of financial assistance (availability of loan sources) (−0.358).[48] Similarly, manufacturers in Sri Lanka find electricity to be much more problematic (0.989) than do traders, but the gap is much smaller in regard to costs of finance (0.321) or loan procedures (0.231).[49]

5.3 ESTIMATION RESULTS: EICOs

5.3.1 Facets of the Investment Climate

Estimation results from the analysis with Enterprise Investment Climate Outcomes are presented in Table H.3 for the model with benchmark indicators and in Table H.4 for the model with selected benchmark components. Table H.3 groups EICOs by country; Table H.4 tabulates results across countries by EICO. Both comparisons offer useful information. For example, Table H.3 permits quick insight into the nature of obstacles perceived by female entrepreneurs or household-based enterprises, whereas Table H.4 helps establish whether patterns for a given EICO hold across countries.

A summary of the results is provided in Table 5.4 in regard to enterprise characteristics and Table 5.5 with respect to community characteristics. Consider first the effect of enterprise characteristics. Although Tanzanian entrepreneurs raised more complaints (Figure 5.1), the nature of the enterprise or the type of entrepreneur rarely relates to the EICO value. In Nicaragua, entrepreneurs operating household-based enterprises perceive fewer obstacles than do their peers with stand-alone enterprises. In Sri Lanka, Sinhalese managers are more vocal about obstacles, especially in the area of finance. Female entrepreneurs complain less about electricity and water, and they also see fewer problems in the cost of loans. Substantial differences emerge among industries, with service enterprises registering more obstacles related to electricity and water compared to traders and fewer in most other aspects. Younger and household-based enterprises view road quality and access as problematic; older enterprises may have learned to cope or may have survived despite this barrier.

Across countries, **Table 5.4** reveals few similarities. Mixed enterprises[50] register more obstacles related to electricity and water in both Nicaragua and Sri Lanka, and household-based enterprises have less trouble with electricity, but that sums up the degree of similarity among significant enterprise characteristics. Inspection of Table H.3 reveals additional patterns, even if the individual estimates are not statistically significant. Household-based enterprises experience fewer obstacles with the loan cost, procedures, and availability as well as with electricity and water; the effect on other EICOs is mixed. Female entrepreneurs in Sri Lanka mention obstacles with loan cost, procedure**s,** and availability less often**,** but the effect for Tanzanian women is mildly positive. The effect of enterprise age is generally minor. For brevity's sake, Table 5.4 skips a small number of EICOs unique to each country.

Not included in the list of enterprise characteristics are measures of size and productivity. Therefore, the estimated models represent reduced form equations. The estimated effect of, say, female entrepreneurship on the EICO response is a combination of a direct effect of gender on the EICO and an indirect effect through size and productivity. For further discussion on this issue, see 7Annex H.

It is easy to argue that size matters. A large enterprise, for example, may be severely disrupted by poor-quality electricity delivery; it may have an

Table 5.4 Significant Enterprise Characteristics in EICO Equations[a]

EICO variable	Nicaragua	Sri Lanka	Tanzania
Interest rate of loan			
+		Sinhalese manager Manufacturing, non-agricultrura; enterprise	Mixed enterprise
−	Household-based enterprise	Services enterprise *Female manager*	
Loan procedures			
+		Sinhalese manager	Female manager
−	*Household-based enterprise*	Services enterprise	
Electricity			
+	Mixed enterprise	Services enterprise Manufacturing, nonagricultural enterprise Agricultural processing enterprise Mixed enterprise *Education of manager*	
−	Household-based enterprise	Female manager Household-based enterprise	
Availability of loan sources			
+		n.e.	Agricultural processing enterprise
−	Age of enterprise	n.e.	
Road quality			
+		Education of manager Sinhalese manager Household-based enterprise Manufacturing, nonagricultural enterprise	
−		Age of enterprise	
Low market demand			
+	*Age of enterprise*		n.e.
−		Services enterprise	n.e.
Road access			
+	n.a.	Household-based enterprise	
−	n.a.	*Age of enterprise* *Services enterprise*	Other production enterprise
Water			
+	Mixed enterprise	Services enterprise Agricultural processing enterprise Mixed enterprise	
−	Household-based enterprise	Female manager	
Telecommunication			
+	n.e.	Education of manager	
−	n.e.		

Source: Annex H, Table H.3 and Table H.4.

Note: a. Variables with at least 10 percent significance level. Variables in models with benchmarks are in regular font, variables in models with benchmark components are in italic font. Variables that are significant in models with benchmarks are not listed twice if they are also significant in models with benchmark components.

n.a. = Not applicable; n.e. = Not estimated.

inside track to bank loans and other financial instruments; it may face more harassment from government agencies. But if EICO obstacles hinder the enterprise's operation and growth, as is indeed the specific question posed to entrepreneurs, one would expect to find that a greater obstacles reduce the enterprise size and productivity. In other words, the direction of causality goes both ways, and estimated effects of size and productivity on EICO responses would be negatively biased. This resulting endogeneity bias could be solved through instrumental variables. Alternatively, size and productivity measures may be omitted and a reduced form equation may be estimated—which indeed was the strategy followed here. The drawback of this approach is that it remains unclear whether the estimated effect of, say, female entrepreneurship on the EICO response is direct or is manifested indirectly through size and productivity.[51]

Table 5.5 describes the findings with regard to the effect of community characteristics as Table 5.4 did for enterprise characteristics. The column for Tanzania registers many significant effects; in Sri Lanka and Nicaragua the effects are not quite as strong. With respect to loans, a strong pattern emerges: in Sri Lanka and Tanzania, the number of banks has a positive effect on the number of obstacles regarding loan cost, procedures, and availability, while the number of bank services has a negative effect and alleviates obstacles. Yet, surprisingly, the effects in Nicaragua are always just the opposite, even though the Nicaraguan effects are smaller and not statistically significant.

Most often, estimation results indicate entrepreneurs report fewer obstacles with infrastructure whenever infrastructure provision at the community level improves. Reported obstacles regarding electricity decline in every country when the percentage of households with electricity rises. In every country, road quality is better in communities where the travel time to the nearest city is lower. A concrete or asphalt road (recorded only in Tanzania and Sri Lanka) alleviated obstacles of road access and quality, even if insignificantly so in Sri Lanka. In Nicaragua and Tanzania, complaints about water supply decline when more households obtain drinking water from a piped source or from a protected well, but in Sri Lanka this variable has no effect.[52] Telecommunications are better in communities having many households with fixed-line phones. In Sri Lanka, access to transport facilities improves as scores improve on the connectivity benchmark indicator. In Tanzania, proximity to the post office reduces reported problems with the postal service. Improvements in the aggregated infrastructure services benchmark indicate reduced obstacles of electricity quality and access, telecommunications, water, road access, and postal service.

Reports on obstacles presented by weak market demand proved difficult to explain using community characteristics. Higher connectivity scores in Nicaragua create a greater obstacle (though not strongly so), but travel time to the nearest city in Sri Lanka reduces it.[53] In Nicaragua and Sri Lanka, the availability of business services in the community did nothing to alleviate perceived low market demand or, in Sri Lanka, the lack of market information.

In Nicaragua, the EICOs of economic policy uncertainty, corruption, and crime/theft rose to the top ten. The models with benchmark indicators assign a significant role to connectivity (negative), infrastructure services (positive), and business services (positive). Probing with specific benchmark components yields less conclusive results: proximity (inverse time) to the nearest city has a negative effect but only significantly so for crime/theft, and the share of households with telephone services has no effect.

The human capital benchmark indicator enters all models with selected components because of its presumed cognitive insight into the social and economic working of society and the effect of obstacles. In Tanzania, six of the ten EICO regressions find a significant negative human capital effect, and twice it is positive (significant for electricity access and nearly significant for electricity quality). In Nicaragua and Sri Lanka, the effect is more variable. In Nicaragua, higher levels of human capital in the community raise economic policy uncertainty and concerns about electricity but somewhat decrease obstacles of low market demand. All other effects are not only statistically insignificant but also economically small. In Sri Lanka, the human capital benchmark reduces obstacles with electricity and water. Even if statistically insignificant, however, it raises each finance obstacle—cost, procedure, and availability.

Table H.3 and Table H.4 report the effect of unmeasured community factors, that is, the community random effect. Across all EICO models with benchmark indicators, the average contribution is 0.45 in Nicaragua on a scale of 0/2, 0.77 in Sri Lanka

Table 5.5 Significant Enterprise Characteristics in EICO Equations[a]

EICO variable	Nicaragua	Sri Lanka	Tanzania
Interest rate of loan			
+		*Number of banks*	*Number of banks*
−		*Number of bank services*	Share income from agriculture Connectivity Finance services *Time to near city* *Number bank services*
Loan procedures			
+	Business services *Insurance service*	*Number of banks*	Income per capita *Number of banks*
−			Share income from agriculture Connectivity *Human capital*
Electricity			
+	Community population size *Human capital*		Community population size Income per capita
−	Connectivity Infrastructure services *Percent households with electricity*	Human capital *Percent households with electricity*	Share income from agriculture Infrastructure services *Percent households with electricity* *Availability electricity*
Availability of loan sources			
+	Business services	n.e.	Income per capita
−		n.e.	Share income from agriculture Connectivity Finance services *Time to nearest city* *Human capital* *Number of bank services*
Road quality			
+			
−	Connectivity *Time to nearest city*	Agricultual seasonality Share income from agriculture Connectivity Business services Finance services *Time to nearest city*	Business services Human capital *Time to nearest city* *Concrete/asphalt road*
Low market demand			
+	Connectivity		n.e.
−	Human capital	*Time to nearest city*	n.e.
Road access			
+	n.a.		Income per capita
−	n.a.	Finance services	Infrastructure services Business services Human capital *Time to nearest city* *Concrete/asphalt road*

Table 5.5 (*Continued*)

EICO variable	Nicaragua	Sri Lanka	Tanzania
Water			
+		*Sewage system*	Income per capita
−	Connectivity *Percent households with protected water*	Income per capita *Human capital*	Connectivity Infrastructure services *Time to main market* *Percent households with protected water* *Human capital*
Telecommunication			
+	n.e.		Income per capita
−	n.e.	Infrastructure services Business services *Percent households with fixed phone line*	Infrastructure services Business services Finance services *Share income from agriculture* *Percent households with fixed phone line* *Availability electricity*

Source: Annex H, Table H.3 and Table H.4.

Note: a. Variables with at least 10 percent significance level. Variables in models with benchmarks are in regular font, variables in models with benchmark components are in italic font. Variables that are significant in models with benchmarks are not listed twice if they are also significant in models with benchmark components.

n.a. = Not applicable; n.e. = Not estimated.

on a scale of 0/4, and 1.13 in Tanzania on a scale of 0/4; all of the estimated effects are statistically highly significant. Thus, first of all, significant similarities in the responses of entrepreneurs in the same community are not explained by the community characteristics. Second, the magnitude of the effect of unmeasured community factors is relatively smaller in Sri Lanka and larger in Tanzania, with Nicaragua falling between—even if in Tanzania more measured community characteristics showed up as statistically significant. Third, the magnitude of the effect of unmeasured community factors is large: the largest effect of any measured community variable is only one-half as large. Thus, the EICO responses reflect much information about the community investment climate that is not obvious from the community questionnaire.

5.3.2 Total Burden of the Investment Climate

The sheer number of investment climate facets covered in the RICS questionnaire may easily overwhelm the analyst who tries to discern patterns in the responses. Is a simple summary measure informative? Consider two summary statistics: (i) *EICO Burden,* defined as the simple sum of all EICO responses, thus treating the categorical values as ordinal scales; and (ii) *EICO Obstacle,* defined as a count of EICO factors rated *major* or *very severe* obstacles (or in Nicaragua merely *severe*).

Table 5.6 reports results of a simple analysis of variance of these summary statistics across communities. Enterprise characteristics explain at most 6.6 percent of the variation in the EICO Burden or EICO Obstacle. Community dummies, which capture all observable and unobservable features of a community's investment climate, explain between 27 and 64 percent of the variation. This is further evidence that entrepreneurs in a given community tend to share similar perceptions about their community's investment climate. In other words, the investment climate is indeed a community factor.

A logical next question is to what degree observable community variables can explain the Burden and Obstacle variables. Table H.6 in the Annex provides estimates. Household-based entrepreneurs tend to perceive fewer obstacles, as do female entrepreneurs, especially in Sri Lanka.

Table 5.6 Analysis of Variance of Total Brden of the Investment Climate

Proportion of variation explained by	Nicaragua		Sri Lanka		Tanzania	
	EICO Burden	EICO Obstacle	EICO Burden	EICO Obstacle	EICO Burden	EICO Obstacle
Enterprise characteristics	0.026	0.015	0.064	0.066	0.019	0.019
Community dummies	0.313	0.272	0.402	0.365	0.639	0.599
Enterprise characteristics + community dummies	0.326	0.279	0.450	0.411	0.642	0.604
Enterprise characteristics + benchmark indicators + community variables	0.108	0.066	0.089	0.095	0.081	0.073

Source: RIC Surveys and, for the fourth row, Annex H, Table H.6.

Sinhalese entrepreneurs mention more obstacles. Nicaragua shows no significant variations by industry, but in Sri Lanka and Tanzania service enterprises face fewer obstacles than traders, while in Sri Lanka manufacturing and other enterprises list more obstacles.

Benchmark indicators as a group are jointly significant in Nicaragua and Tanzania. In Nicaragua, better connectivity lowers the total EICO burden, but business services raise it; in Tanzania, community human capital is associated with lower total EICO burden. Overall, most of the estimated benchmark effects are negative, even if not statistically significant.

Other community variables do not have a significant effect on the reported burden of the investment climate, even though the effect of population is positive in every country and the effect of the importance of agriculture in the economy (measured by seasonality and the share of income derived from agricultural activities) is negative.

Together, the community variables and benchmark indicators explain between 2.5 percent and 8.2 percent of the variation in the Burden and Obstacle variables, only a rather small portion of the variation between the communities (Table 5.6). A large portion of the investment climate at the community level is still shrouded from direct measurement.

5.4 ESTIMATION RESULTS: EICIs

Where Enterprise Investment Climate Outcome (EICO) variables consider obstacles to the operation and growth of enterprises, Enterprise Investment Climate Interactions (EICI) variables reflect the interaction between entrepreneurs and government bodies. This section relates three types of EICIs to the standard set of enterprise characteristics and selected community characteristics: (i) two aspects of government reliability: predictability or variance in the application of laws and the misinterpretation and manipulation of laws by government officials; (ii) the perceived effect of kickbacks on businesses; and (iii) perceptions about the support received from the legal system.

Five community characteristics are inserted into the empirical models.[54] Population size denotes the size of networks, the degree to which people know each other, the extent to which government officials can hide their discretionary behavior, and so forth. Income per capita represents the level of development in the community. The connectivity benchmark measures the community's degree of integration with other parts of the country, which may alternately raise or reduce the scope of government discretionary behavior. The business services benchmark is inserted to mark entrepreneurs' potential to plug into commercial networks, receive information, and learn about common practices. Whether this raises or lowers government officials' perceived discretionary activity is an empirical question; arguments could be made both ways. The community's level of human capital denotes cognitive development that on the one hand prevents rogue exploitation of illiterate members of the population but on the other hand allows creativity in bending rules to one's advantage.

5.4.1 Government Reliability: Predictability and Manipulation of Laws

Table H.7 relates enterprise and community characteristics to perceptions of laws as predictable and manipulable. Only a few enterprise characteristics matter. Educated Sri Lankan entrepreneurs do not think laws are unpredictable: per year of schooling, the probability diminishes by 5.3 percentage points, which is a strong effect indeed. Similarly, the gender effect in Tanzania is negative: women view laws as more predictable than do men. Tanzania lacks the education effect, however, and Sri Lanka gives no evidence of the gender effect; in Nicaragua, manager characteristics were not recorded in the survey data.

The effect of community environment is hardly uniform across the three countries either. The only variable that carries the same sign in all three columns is per capita income, but its effect is statistically insignificant. Connectivity creates predictability in Nicaragua but has no effect elsewhere. Business services associate with predictability in Tanzania, but Nicaragua shows an opposite, although statistically insignificant, effect. In Tanzania, human capital raises unpredictability sharply, but this variable has a minimal effect elsewhere.

One element is common to all these models, however: the effect of unmeasured community factors is large in every country. A one-standard-deviation swing in these factors changes by 20 percentage points the probability that entrepreneurs believe laws to be unpredictable. This commonality in the response by community suggests strongly that local factors play a highly important role in how laws are applied. Unfortunately, these factors are not captured well in the community characteristics employed here.

5.4.2 Kickbacks

In Nicaragua, a sufficient number of entrepreneurs provided information about the effect of kickbacks on their businesses that an analysis with an ordered probit model is warranted. Table H.8 reports few distinctions among enterprises in their views on bribes. There is no age effect, and service and mixed enterprises may experience less of an obstacle than do trading and production enterprises.

The real action in this model appears, once again, in the community variables. The effect of kickbacks is felt more strongly in larger communities with more business services and poorer connectivity. Income per capita has a borderline-significant negative effect. Community human capital has no effect. A significant unmeasured community factor is larger than any measured community variable but not dramatically so.

The similarity between the five equations may suggest that entrepreneurs tend to respond in the same way to all five questions. That is quite correct: 81.2 percent of the entrepreneurs who responded to all questions never felt an effect from kickbacks in any dimension, and half of the rest gave the same answer to all questions. But the variations are large enough to allow the effect of business services to evaporate from the credit kickback equation and to raise the effect of population size and connectivity in the politician kickback equation.

5.4.3 Efficiency of the Legal System

Business dealings take place within an institutional framework of rules and regulations. The two sides engaging in any transaction implicitly assume that any disagreement or dispute that might arise can be resolved in a predictable manner as established by that framework. But if the set of rules and regulations leaves ambiguity, has loopholes, or is inconsistently enforced, transaction costs increase.

The RIC database includes several items that reflect on the institutional framework; three of them are examined in Table H.9: reliance on reputation, protection through a contract, and support by the legal system. Considering patterns among enterprise characteristics across the three sets of estimates, it is striking that older businesses in Nicaragua rely less on reputation yet do not expect that contracts will protect them from cheating or that the legal system will uphold that contract. In Sri Lanka and Tanzania, the signs of the parameter estimates of education, gender, and ethnicity exhibit similar patterns, although none of the estimates (with one exception) is statistically significant. Differences between sectors are also more or less similar among the three sets of estimates. As mentioned before, reputation might be expected to count more heavily in economic environments where the rod of the legal system is short, but neither the descriptive statistics in Table 5.3 nor the regression results in Table H.9 bear out this expectation.

Overall, statistical significance is weak, but three patterns are notable in Table H.7, H.8, and

H.9. First, the gender effect is fairly similar in Sri Lanka and Tanzania, even though it was weakly different for opinions about the predictability and manipulability of laws. Of course, regardless of what one thinks about the manner in which laws and rules are written and enforced, their actually ability to provide legal protection is an entirely different matter. Women appear to feel more vulnerable in this aspect. Second, if the statistically insignificant coefficients may be interpreted this way, Sinhalese entrepreneurs both rely more on reputation and expect contract and legal system support—even though they view laws as manipulable by government officials. Third, Nicaragua and Sri Lanka show a pronounced difference in the way manufacturers as opposed to traders respond to these three questions.[55] In Nicaragua, traders perceive a greater lack of support from the legal system than do manufacturers; in Sri Lanka, manufacturers feel disadvantaged relative to traders. It will be interesting to relate this difference to a comparison of the two countries' commercial laws.

In Nicaragua, entrepreneurs in larger communities expect more protection from contracts; in Sri Lanka and Tanzania, reputation is less important in larger communities—as indeed transactions become less personalized—but in Sri Lanka largetown entrepreneurs do not expect as much support from the legal system. Other community variables matter hardly at all. Unmeasured community factors are important, but the magnitude of the effect is smaller than was found in other tables: the predictability and manipulability of laws is a bit more discretionary than are reputation and legal protection and appear to vary more systematically between communities.

5.5 PUSHING THE ANALYSIS FORWARD: NEXT STEPS

5.5.1 Analyzing Discrete Outcomes

In this chapter, the EICO and EICI variables are obviously discrete, which necessitates the use of probit models for simple 0/1 outcome variables and ordered probit if the outcome has multiple categories. When applying these models, it is important to account for sampling weights and community random effects. Because of the sampling design of the survey data, sampling weights are necessary if empirical results are to be representative of the population at large. From smallest to largest, weights vary by a factor of 13.1 in Nicaragua and a factor of 619.5 in Sri Lanka; sampling weights are not available for Tanzania. Community random effects capture unobservable factors affecting the perceptions of all entrepreneurs in a given community. Conceptually, it is plausible that they exist; empirically, they prove to be important.[56] Unfortunately, as demonstrated in Annex H on the econometric modeling of RIC data, the computation of the standard errors of the parameter estimates is not a fully settled issue and warrants more research.

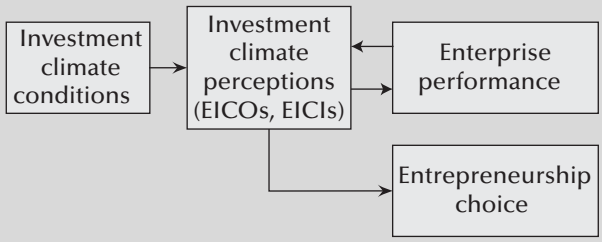
5.5.2 Perceptions, Conditions, and Behavioral Outcomes

The basic premise of investment climate survey research in urban and rural areas alike is that the investment climate determines enterprise performance and therefore entrepreneurship incentives. In this regard, it is useful to distinguish investment climate conditions and investment climate perceptions. Conditions are factual; perceptions are subjective. Conditions impose hard constraints; perceptions create constraints that affect behavior. In a world of perfect information, perceptions are identical to actual conditions. In a more realistic world of imperfect information, perceptions are imperfect reflections of conditions. This raises interesting questions: How is enterprise performance conditioned by investment climate perceptions? How do investment climate perceptions influence entrepreneurship choice? How do investment climate conditions determine investment climate perceptions?

Figure 5.2 describes these posited relationships in a schematic diagram. It also adds one feedback loop: perceptions may be influenced by enterprise performance. Growth in the enterprise may expose entrepreneurs to limitations in their economic environments (for example, limits on credit, poor infrastructure, low-quality labor) that were not obvious before. Not expressed in this diagram is the policy feedback loop in which policymakers act to improve the investment climate in communities with successful enterprises and entrepreneurs who lobby for change.

RIC surveys gather information from entrepreneurs about their investment climate perceptions. The surveys also collect information about investment climate conditions through community

Figure 5.2 Investment Climate and Entrepreneurship



Source: The Authors.

questionnaires and, to some degree, from factual information provided by entrepreneurs. The EICO and EICI models in this chapter target the leftmost relationship in Figure 5.2. They sidestep the interactive relationship with enterprise performance.[57] Instead, they estimate a "reduced form" set of equations in which determinants such as enterprise characteristics and investment climate conditions (through community characteristics) capture both the direct effect on the investment climate and the indirect effect that works through the feedback loop on enterprise performance.

A next step in this research project is to disentangle this relationship. This presents a two-part challenge:

- First, find variables that determine enterprise performance but not investment climate perceptions and find other variables that determine investment climate perceptions but not enterprise performance (that is, in econometric jargon, find variables to identify the investment climate perception and enterprise performance equations); and
- Second, develop an econometric model that incorporates sampling weights and community random effects while accounting for the discrete nature of the measured investment climate perceptions and continuous nature of enterprise performance.

Investment climate perceptions also play a role in the decision to become an entrepreneur and set up a business. But these are the perceptions held by would-be entrepreneurs, whereas the perceptions measured in the survey are those expressed by current entrepreneurs. This creates a self-selection issue: current entrepreneurs may have a more positive view of the investment climate than do those who have decided not to operate a business. Correcting the EICO and EICI measures for the self-selection effect and then applying the corrected measure as a determinant of entrepreneurship choice is a useful next step in the study of entrepreneurship.

In one respect, Figure 5.2 makes a clinical simplification: both investment climate conditions and investment climate perceptions have many dimensions. This chapter grapples with the methodology of how to relate these two multidimensional concepts together. Chapters 3 and 4 faced a similar challenge in relating investment climate conditions to enterprise performance and entrepreneurship choice—in terms of Figure 5.2, bypassing the center box. The next step in investment climate research will be to come to terms with the multidimensionality of perceptions: Should they be summarized into "benchmark perceptions"? How important is each facet of investment climate perception?

5.6 IMPLICATIONS FOR RIC METHODOLOGY

As diverse as are the three countries in this analysis, the lists of their most important investment climate concerns show a remarkable overlap: six items appear in each country's top ten, and three more occur in two of the three. The common concerns are

(i) finance cost, availability, and procedures;
(ii) electricity and water (access, cost, or reliability); and
(iii) road quality and access.

The other top-ten concerns are more varied: market demand, economic policy uncertainty, and telecommunications.

In a number of investment climate facets, countries differ greatly: ranks vary sharply for corruption, economic policy uncertainty, and postal service and road blocks. It is possible that the use

of rather general screen questions in Sri Lanka and Tanzania reduced responses. A question such as "Is governance an obstacle?" may not have triggered an affirmative response, for example, whereas follow-up questions about corruption, economic policy uncertainty, crime, war, and legal systems might have if they had been asked.

Many EICO factors are not really relevant to rural enterprises. In Nicaragua and Sri Lanka, more than 95 percent of respondents rated many variables *No obstacle.* Tanzania's entrepreneurs, who are either more vocal or suffer from a large array of obstacles, include many items of lesser concern, such as import/export, food safety, and labor issues. It is unclear whether all of the almost 40 EICO variables should be collected—but they are easy to collect, as opposed to specific questions about enterprise performance, labor employment, and asset investment.

Throughout the econometric analysis, the magnitude of the effect of unmeasured community factors proved to be large. Thus, the EICO and EICI responses produced much information about the community investment climate not gleaned from the community questionnaires.

RIC2 may not yet have explored in sufficient depth the range of information available in the community questionnaires. Extracting this information is a time-intensive exercise, especially since the three RIC databases are constructed in completely different ways and the questionnaires contain significant variations. Moreover, the relevant set of community characteristics may well vary between the various obstacle and interaction variables studied here. Community characteristics, too, are likely intercorrelated, making it difficult to disentangle the precise reason for variation in EICO and EICI responses between communities. On the other hand, community characteristics may suffer from some degree of measurement error, which could limit their statistically significant contributions to the explanations pursued here. The actual contribution then becomes part of the community level random effect.

More likely, the community questionnaire has not yet captured the idiosyncrasies that characterize local investment climates. One could point to factors that may have been overlooked: among them, the influence of major local personalities, the role of elites, and the importance of networks. RIC2 shows that something operating at the community level has yet to be captured. Future RIC practitioners might profitably strive to improve the measurement of that "something."

6 Differences Among Communities

6.1 COMMUNITY CHARACTERISTICS AND THE COMMUNITY ENVIRONMENT

The communities within which rural enterprises operate vary greatly. This is true between economies, but also within the pilot countries. To determine the specific influences these differences might impose on the local rural investment climate, community leaders were questioned about the status of several community characteristics that *ex hypothesis* might support or impede enterprise performance and growth. As shown in the following table, the community descriptors scale community size (population), the presence (or absence) of important infrastructure such as transport and communications; the levels of education, health, and other social or public services; the availability of financial and business services; proximity to urban areas and marketing centers; and the number of taxes paid by local residents. Information was also gathered about the communities' endowments of dry and irrigated agricultural lands. In Nicaragua, an attempt was made to determine gender bias in seasonal agricultural employment, while in Sri Lanka surveyors queried the variability of male and female participation in the agricultural labor force.

A cross-country summary comparison of these descriptors appears in Table 6.1. Differences in community population size make comparison difficult. Sri Lanka's communities are about 11 percent the size of those in Nicaragua and less than a third the size of those in Tanzania.[58] Taking account of these differences, communities in Sri Lanka areseem relatively well supplied with economic and social infrastructure. The concept of market probably differs among the countries: the distance to market is highest in Nicaragua, despite (or perhaps because of) the large community size.

At the community level, the other main differences appear in the proximity of villages to markets and urban demand centers and in the availability of household infrastructure services. Here, residents in Nicaraguan and Sri Lankan communities appear much better provided with electricity than their Tanzanian counterparts.

Community-level differences are also evident in aggregated enterprise performance results. As shown in Annex J, the significance of nonfarm rural employment and of earnings per capita and per worker appears to be greater in larger communities than in smaller ones. In Nicaragua and Sri Lanka, employment share in

Table 6.1 Comparison of Community Descriptors (Community Averages)

Descriptor	Nicaragua	Sri Lanka	Tanzania
Population	16887	1837	6414
Distance (km) to market	25	16	9
Distance (km) to city	35	15	19
Road surface* (max = 4)	2.4	2.4	1.9
Percent households with electricity	74	95	6
Percent households with gas	31	16	NA
School type (highest level)* (max = 3)	1.8	1.6	1.5
Health center type (highest level)* (max = 5)	3.4	2.9**	3.1
Local availability of banking services* (max = 7)	1.0	0.9	0.9
Local availability of business services* (max = 7)	1.7	2.3	0.5

Source: RICS data.

Notes: * scaled; ** not including hospitals; max = 7.

agriculture falls dramatically in the larger communities. Absolute magnitudes of both sets of indicators vary considerably across countries, however, probably reflecting differing levels of development and degrees of urbanization.

Benchmark indicators, introduced in Chapter 2, are defined in Annex E, which outlines their use in the regressions. The following sections show how benchmark indicators vary among the regressions, starting in Section 6.2 with the variation among communities and between countries. Correlations among community variables are explored in some detail, especially with the benchmark indicators (BI). Section 6.3 examines relations between the BI and prices and Section 6.4 those between the BI and other indicators of community characteristics. Regression results for Sri Lanka are presented in Section 6.5 to help explain community economic performance through a set of explanatory variables, as measured by outcome variables at the community level. The chapter concludes with Section 6.6, an assessment of the performance of benchmark indicators in the three data sets.

6.2 CROSS-COUNTRY COMPARISONS OF BENCHMARK INDICATORS

Comparisons

One of the purposes of benchmark indicators is to facilitate comparisons between countries. Table 6.2 reveals that the benchmark indicators for Nicaragua and Sri Lanka tend to average slightly higher, while Tanzania averages much lower. Connectivity, human capital, and finance services score higher in Sri Lanka than in Nicaragua, which is as expected because of Sri Lanka's high population density and traditionally high level of education. In Tanzania all indicators are lower than or at best equal to the other countries', with the exception of finance services.[59] Sri Lanka's relatively low index for business services (compared to Nicaragua's) may explain why this factor scored as significant in the Sri Lanka regressions in Chapter 3.

The breakdown between communities with smaller and larger populations shows, as expected, that the indexes in larger communities are generally higher. The differences in Nicaragua are much greater than those in Sri Lanka and Tanzania, probably reflecting the greater income disparity in there. The BIs with the relatively highest difference between smaller and larger communities are connectivity, infrastructure services, and business services indexes. In Nicaragua, the range in the human capital and finance services indexes is also substantial. It is noteworthy that the governance index in Tanzania and Sri Lanka does not differ between small and large communities, while in Nicaragua the difference is notable.

Correlations Between Benchmark Indicators

The BIs for Nicaragua show nearly all statistically significant correlations with each other at the 10 percent level (Table 6.3). Governance is the only exception, as it is not related to infrastructure

Table 6.2 Comparison of Benchmark Indicators by Country and Community Population Size

	Nicaragua	Sri Lanka	Tanzania
Population size: All			
	Index	Index	Index
Connectivity	0.34	0.40	0.20
Infrastructure services	0.51	0.35	0.19
Business services	0.24	0.15	0.07
Governance	0.67	0.67	0.50
Human capital	0.21	0.33	0.21
Finance services	0.17	0.50	0.18
Average	0.36	0.40	0.22
*Population size: Small**	=<4800	=<1350	=<3000
	Index	Index	Index
Connectivity	0.24	0.36	0.17
Infrastructure services	0.38	0.31	0.14
Business services	0.11	0.12	0.04
Governance	0.64	0.67	0.50
Human capital	0.17	0.33	0.20
Finance services	0.02	0.48	0.18
Average	0.26	0.38	0.21
*Population size: Large**	>4800	>1350	>3000
	Index	Index	Index
Connectivity	0.44	0.44	0.22
Infrastructure services	0.65	0.38	0.23
Business services	0.37	0.18	0.08
Governance	0.71	0.67	0.50
Human capital	0.24	0.33	0.22
Finance services	0.33	0.52	0.19
Average	0.46	0.42	0.24

Source: Annex E.

Note: * the indicated size is about the median for each country.

services or business services. All significant correlations have a positive sign, thereby pointing towards a better rural investment climate.

In Sri Lanka most benchmark indicators are also positively correlated at a significance level of 10 percent, but it shows more insignificant correlations than does Nicaragua. Governance is the exception, with no significant correlation with any other BI. Human capital and business services are unrelated as well.

In Tanzania fewer BIs are significantly correlated than in the other two countries, but the significant ones all have the correct signs. Governance is not correlated at all, and finance services correlates only with connectivity and infrastructure services. The greatest interrelation exists with connectivity and infrastructure services.

Completeness of Information for Benchmark Indicators

Because the BIs and their components are constructed from raw data, the problem of missing observations had to be addressed. Following the procedure outlined in Annex E, synthetic values were substituted for the missing data, computed as averages from available data. This can result in smaller standard deviations and perhaps bias in the estimates. Annex G, Table G.1, provides an overview of communities and subcomponents with missing observations. As missing observations for communities occurred in only 0.3, 4.6, and 1.7 percent of the observations, respectively, in Nicaragua, Sri Lanka, and Tanzania, and in the respective subindicators by 2.6, 3.9, and 2.1 percent, the resulting biases were likely very small.

Table 6.3 Correlation Coefficients Between Benchmark Indicators

	Connectivity	Infrastructure services	Business services	Governance	Human capital	Finance services
Nicaragua						
Connectivity	1.00					
Infrastructure services	0.73*	1.00				
Business services	0.34*	0.49*	1.00			
Governance	0.17*	0.12	−0.09	1.00		
Human capital	0.41*	0.37*	0.26*	0.25*	1.00	
Finance services	0.41*	0.45*	0.33*	0.23*	0.35*	1.00
Sri Lanka						
Connectivity	1.00					
Infrastructure services	0.48*	1.00				
Business services	0.25*	0.24*	1.00			
Governance	−0.09	−0.12	0.06	1.00		
Human capital	0.35*	0.32*	−0.04	−0.07	1.00	
Finance services	0.53*	0.45*	0.22*	0.06	0.31*	1.00
Tanzania						
Connectivity	1.00					
Infrastructure services	0.55*	1.00				
Business services	0.20*	0.29*	1.00			
Governance	−0.04	0.04	0.09	1.00		
Human capital	0.27*	0.49*	0.15*	0.03	1.00	
Finance services	0.21*	0.21*	0.11	−0.04	0.05	1.00

Source: RICS data.

The most missing observations occurred in the governance indicator in Tanzania and Sri Lanka.

Variation Among Communities

As is shown in Annex E, the BIs are generally well distributed over the range between 0 and 1, the main exception being governance, which has relatively high minimum and average values. Human capital has low maximum values, but this is consistent with the still relatively low levels of education in the three countries' rural areas. The standard deviations shown in Annex E also provide information about the range of index values. The low standard deviations for governance and for human capital suggest that the scaling of responses did not distribute the index values over the full unit interval and so was not optimal.

6.3 BENCHMARK INDICATORS AND PRICES

Apart from production technology, which was examined with the production function estimates in Chapter 3, prices are important determinants of enterprise viability and entrepreneurial choice. In remote or not easily accessible villages, prices may differ substantially from those in urban centers. Inadequate infrastructure can substantially raise costs of products brought from elsewhere and, because of high transportation costs, lower RNFE "producer" prices for products destined for distant markets. Thus improved infrastructure and finance services may encourage private investment in modern technologies, thereby lowering production costs. By implication, improvements in the investment climate reflected in the BI should also correlate with price movements in the community.

The set of prices considered below differs among the RIC pilots, in part because the lists of prices collected were not identical. But more importantly, values were missing for some commodities in far too many communities to support a proper correlation analysis.

Among the BI, in Nicaragua, connectivity, governance, and human capital relate most to the different price variables (Annex G, Table G.2). Business and finance services indicators were not at all related to any of the prices collected,

however. Infrastructure services related only to the price of diesel.

The prices of staple foods (rice, maize) are significantly related to connectivity. This probably reflects higher food prices in more urbanized communities. The price of diesel is negatively related to connectivity and infrastructure services, which suggests higher prices in more remote areas. Finance services are not correlated. Amongst subindicators, the costs of phone calls and salt are negatively related to connectivity. The price of transportation, Coca-Cola (included as a representative mass-produced consumer good), and soap are positively related to human capital. In Nicaragua, when governance is supportive, wages and the price of phone calls tend to be low and the costs of transportation high, possibly indicating that governance is less of an issue in more distant rural areas.

Table G.3 shows relations between benchmark indicators and prices in Sri Lanka. As expected, the connectivity and infrastructure services index are negatively related to many prices and positively related to wages and the price of food. Business services are positively related to wage levels and the price of wheat flour, which may reflect the situation in more urban commercial areas. The governance index does not relate to many prices. The relation with rice is positive and that with wheat flour is negative. Human capital is positively related to wages. The finance services index shows much similarity with connectivity: it is negatively related to many price items and positively related to food prices. Distance indicators are nearly all negatively related to connectivity, infrastructure services, human capital, and finance services.

In Tanzania relatively little significant correlation is found between price levels and the BIs (Table G.4). A negative relation exists between phone calls to town and connectivity and human capital; a mobile phone call is negatively related to connectivity, infrastructure services, and business services. The price of cement correlates weakly only with finance services; rice prices are weakly related only to business services; and wage levels are generally unrelated to the benchmark indexes, except that payments for rural works appear to be lower in areas with good governance. Coca-Cola and kerosene prices are not related to the Tanzanian BIs, whereas the price for diesel is negatively related to governance and finance services. In general, patterns are not clear in Tanzania.

6.4 BENCHMARKS AND COMMUNITY CHARACTERISTICS

Benchmark indicators may affect other characteristics of communities as well. Benchmarked variables may affect the provision of schooling, for example, due both to demand for education on the part of households wishing to connect with the urban world and to supply generated by forward-looking community leaders who foresee community integration into the region. These variables also change the incentives for households to engage in farming, in a nonfarm business, or wage employment. Effective governance reduces the cost of conflict. Thus, the annex tables referred to below examine BI correlations with a broad range of community characteristics. As with price variables, the number of characteristics imputed for each pilot varies between countries because of data availability.

The BIs for Nicaragua show several statistically significant correlation characteristics (Table G.5). The data set shows much communality, although many of the relations are quantitatively weak. Most characteristics representing an endowment show positive relations with the benchmark indicators. Negative correlations are found with immigrants as percentage of population and with distance to market. In an important finding, community population size has positive significant correlation with all benchmarks, except governance. This shows, as expected, that the IC is generally more supportive in urban than in rural areas. The highest correlation for connectivity is with the number of public services (0.69); for infrastructure services with number of public services (0.83); for business services with number of services available for businesses (number of public services [0.46] and number of banks [0.43]); and for human capital with percentage of households with gas (0.59). Governance only shows a few low correlations, and finance services correlate positively with number of banks (0.72, which is almost deterministic) and highest level of school type (0.53).

In Sri Lanka the number of significant correlations with the BIs tends to be less than in Nicaragua (Table G.6). Perhaps the differences among communities in Nicaragua are greater than in Sri Lanka. The highest significant correlations are: with connectivity, the number of public services (0.61) and distance to city (−0.56); with

infrastructure services, the number of public services (0.57) and the percentage of households with electricity (0.55); with business services, the percentage of conflicts solved by labor court (0.30) and the number of banks (0.28); with governance, the number of business services (0.26); with human capital, the percentage of households with gas (0.40) and electricity (0.38); and with finance services, the percentage of households with electricity (0.51) and number of public services (0.43). The generally negative relationships with connectivity include seasonality (as reflected in the standard deviation of male and female agricultural labor use), distances to markets, distance to the nearest city, and agricultural land per capita. These indicators seem associated primarily with the more remote agricultural areas. Infrastructural services also show roughly the same negative relations. Interestingly, the finance services index is negatively related to agricultural land per capita.

In Tanzania the BIs are less significantly correlated with community characteristics (Table G.7). The highest significant correlation is connectivity with the number of public services (0.67), followed by infrastructure services also with the number of public services (0.72) and with percentage of households having a cell phone (0.70); business services with the number of banks (0.43); and human capital with the number of public services (0.38). The largest correlation with governance is found with road surface (−0.29, a negative). Finance services shows the highest correlation with the number of banks (0.38) and the number of public services (0.26) and correlates negatively with distance to banking centers (−0.29). Generally, negative BI relationships in Tanzania include distance to market, distance to city, and agricultural land per capita.

In conclusion, the BIs and community characteristics show many significant and sometimes quite high correlations. This demonstrates that a small number of well-selected indicators can represent many facets of community-level influences on the rural investment climate, although they do not and cannot tell the whole story. Using benchmark variables in regression analysis, especially in conjunction with other community variables, may lead to problems of multicollinearity: too many variables, not dissimilar enough from each other, compete in the explanation. This can undermine the precision of the estimated effects and make it difficult to attribute between-community variation in the dependent variable to any particular BI or community variable. But this econometric problem originates more with the multidimensional nature of communities and less with the manner in which the communities are described. Empirically, connectivity, infrastructure services, business services, human capital, measures of governance and corruption, and finance services may be correlated, but conceptually they remain distinct.

6.5 REGRESSIONS AT THE COMMUNITY LEVEL FOR SRI LANKA

An interesting point for analysis is whether the communities' economic performance can be explained by community data. (It should be stressed upfront that this kind of analysis is hampered by the rather small number of communities used in cross-sectional analysis.)[60] The analysis in this section of a set of explanatory variables for Sri Lanka provides an example of regression results to explain community economic performance as measured by community-level outcome variables.

Table G.8, Annex G, provides information on the dependent and explanatory variables used in the community-level regressions. The list of dependent variables starts with enterprise density as a measure of the degree of entrepreneurship in the community. It continues with seven variables describing labor market outcomes, such as income level, structure of the labor force, labor force participation, and labor productivity measures; two variables measuring output generated by enterprises relative to their production costs (focusing on community-level enterprise productivity as far as this can be captured in the sample); two variables focusing on gross value added (GVA) to examine size and variation in size among enterprises; and two variables considering seasonality in the community from the perspective of enterprise production cycles and agricultural labor demand cycles.

Regression Results

Table G.9 displays the regression results for all 14 outcome variables in Sri Lanka. The model shown in Panel A is exploratory: the same set of explanatory variables is used in every regression

model. The risk of such a specification is that irrelevant variables reduce the precision of the parameter estimates of relevant variables. For that reason, Panel B limits the specification to variables with a t-statistic of at least 0.5 in absolute value.

Enterprise density per household is positively related to the share of Sinhalese in the population, who appear to operate relatively more of the enterprises. Proximity to schools and hospitals is beneficial; these are higher in more densely populated areas. Among the benchmarks, connectivity, business services, and human capital raise enterprise density. After deleting three insignificant variables, food prices are found to have a significant positive effect as well, but the finance services index has a negative effect.

Total per capita income from work (farms, RNFEs) is positively related with community population size, business services, and human capital and negatively with enterprise openness, infrastructure services, governance, and the price of rice. Since the price of rice was generally low when the pilot surveys were administered, the relationship suggests that income from work is highest in larger communities having higher average levels of education. The negative infrastructure services effect, however, is more difficult to understand.

The share of rural nonfarm income falls with the share of Sinhalese and proximity to the city. Improvement in the finance services index is beneficial, however. When six insignificant variables are removed, the effect of the remaining variables becomes more clearly defined, but no other variable rises to a 10 percent significance level. It should be noted, however, that a flaw in the Sri Lanka questionnaire has led to an understatement of household nonfarm enterprise income.

The next dependent variable—the ratio of RNFE workers—is closely related and, because of the problem in measuring enterprise income, more appropriate: employment in rural nonfarm enterprises as a share of working adults. This share rises in those larger communities having a lower availability of agricultural lowlands. Dropping ten insignificant variables leaves some statistically significant variables: relatively more people work in their own nonfarm enterprise when illiteracy is lower, schools are nearby, the main city is farther away, and the price of kerosene is lower. The latter probably reflects lower kerosene prices in commercial areas. The other variables reflect the set of opportunities in the community.

The next two variables, rural nonfarm income per capita and per worker, are both affected by the measurement problem in income, so results must be cautiously interpreted. Both variables rise with human capital, which is entirely plausible. They also fall with the governance benchmark, however. Openness of communities[61] is not really helpful; and food price is likewise negatively related.

Labor force participation relates positively to community population size, upland per capita and the benchmark indicators of business services, human capital, and governance. The relationship also suggests that labor force participation is lower in communities with a high share of Singhalese located in the more productive lowlands. The infrastructure services and finance services indicators tend to reduce the labor force participation rate.

The male wage rises rather significantly in communities with more uplands. The human capital index also exerts a positive effect on male wage rates, as one would expect. Food prices are also positively related, but the price of kerosene shows a negative effect, which may hint at the role of commercial areas. These two price variables correlate only weakly, so this contrasting result should not be attributed to multicollinearity. After removing a number of insignificant variables, a high share of Sinhalese, proximity to a city, and the business services benchmark all serve to raise the male wage.

The average ratio of sales to cost is a proxy for average (total) productivity of community enterprises. Enterprise productivity shows weak negative association with connectivity and food prices and weak positive association with business services and infrastructure services.[62] Removing redundant variables suggests a role for several other variables: productivity may be lower in larger communities with a larger share of Sinhalese. The same variables affect net factor productivity (the next dependent variable), but the relationship is less well defined.

Gross value added per enterprise, a size measure, relates positively to community openness and a high finance index. In communities with more Sinhalese and a higher governance benchmark, enterprises are smaller. The business services benchmark comes forward as well with a positive effect when insignificant variables are removed. The standard deviation of gross value added, a measure of size differences, relates to the same variables.

A lower value of the enterprise seasonal adjustment factor indicates more seasonality. Thus, RNFE seasonality is more severe in communities with a greater share of Sinhalese and more widely available business services. Where finance services are more available, seasonality is less. Removing redundant variables changes little in this picture.

Agricultural seasonality is characterized by a high share of Sinhalese, a high share of illiteracy, and low food prices, and yet, at the same time, also by proximity to a hospital and a main city. Among the benchmark indicators, better governance is associated with greater agricultural seasonality, but connectivity, human capital, and finance services are not. Several of these are characteristic of rural areas with paddy land.

Using the R^2 values to represent suitable measures of fit, it appears that community-level analysis best explains agricultural seasonality, enterprise density, and the various measures of enterprise productivity and size. The community analysis corresponds with the high R^2 values found in the enterprise performance analysis for Sri Lanka in Chapter 3.

Factor Analysis

Factor analyses conducted to explore the full set of community variables revealed a considerable number of factors and mostly low factor loadings. Clearly, communities are multifaceted in their characteristics, and the structure of the variance in community variables cannot well be explained with only a limited number of factors or benchmark indicators. It is of course possible to "fish" for variables with significant explanation in regression analysis, but still it must be expected that only a small part of the variance at the community level can be explained using the data sets in hand.

6.6 PERFORMANCE OF BENCHMARK INDICATORS AS DESCRIPTORS FOR THE INVESTMENT CLIMATE

The economic structure of rural communities, including geographic and climatic characteristics and the enabling environment, has many dimensions. This makes it difficult to compare investment climate conditions and to explain economic performance. The BIs are aggregates developed in the RIC study to facilitate comparison of the investment climate among communities and between countries. This study demonstrates that the BIs are useful aggregates for this purpose, although refinement is needed. The evidence is threefold:

1. In binary correlations the benchmark indicators generally behaved as expected from a theoretical perspective: positive relations with outcome variables and negative relations with obstacles, some with strong statistical significance, even though many of the relations were weak.

2. Together in multiple regressions, some of the parameter estimates of benchmark indicators give counterintuitive relations, probably because of their intercorrelations and unobserved background variables.

3. Replacement of a benchmark by its components, as practiced in Chapters 3, 4, 5, and above in this chapter, can be a useful strategy to see how much explanatory power was lost by aggregation. It can also reveal which are the relatively weaker subcomponents. Because some components correlate highly with other benchmarks, however, the potential gains of replacement are reduced.

7 CONCLUSIONS AND RECOMMENDATIONS

THE AIMS OF THE STUDY

This study explored RIC databases for three countries—Nicaragua, Sri Lanka, and Tanzania—in a wide-ranging look at the effect of investment climate on enterprise performance and entrepreneurship. This is not the first report written on the basis of these data. For each RIC pilot study, country teams were in charge of conducting the surveys, analyzing the data, and writing a RICA report. These reports contain important findings about the investment climate in the pilot countries. RIC1, the first comparative study of the surveys, formally titled *The Rural Investment Climate: It Differs and It Matters,* appeared in 2006. Yet these analytical efforts explored only parts of the collected data, focused on different issues, and used dissimilar methodological tools. Important topics of analysis were not fully covered because of limitations in time and budget and lack of readily available tools for analysis; other topics of great interest were not addressed at all. This study augments and extends earlier studies with further systematic exploration of the three countries' RIC data.

Specifically, this second comprehensive study, RIC2, was undertaken to:

- Provide a broader and deeper understanding of nonfarm activity in rural areas, its constraints, and possible ways to mitigate those constraints;
- Initiate and test a method of benchmarking the investment climate in rural areas; and
- Advance and sharpen the methodologies for RIC assessments and provide analytical guidance for future survey teams and policy analysts.

Developing and refining methodologies for RIC assessments is in part a process of learning by doing. Lessons learned in the first pilots and RIC1 have already been taken into account in the draft Implementation Manual (World Bank 2007c). The findings of the present study, currently being supplemented with ongoing work in other country assessments, provide important information and pointers for further sharpening survey designs and analytical methodologies, as well for updating the Implementation Manual. To these ends, several recommendations from the study are discussed below.

WHAT IS NEW IN THIS STUDY?

This study makes several new contributions to the analysis of the rural investment climate:

1. A much more comprehensive analysis is made of data from the RIC pilots than was hitherto available.
2. More systematic use is made of community data than in previous studies.
3. Benchmark indicators are developed and employed in the analyses.
4. Responses about constraints in the investment climate are analyzed using the concepts Enterprise Investment Climate Outcomes (EICO) and Enterprise Investment Climate Interactions (EICI).
5. Novel Stata programs (documented in this report) were developed for random effect analysis with sampling weights.
6. Estimates are made of the contribution of groups of variables—enterprise, industry, and community characteristics—in explaining variation of outcome variables.
7. Household and enterprise weights are used in the analysis of data for Nicaragua and Sri Lanka. For Nicaragua, household and enterprise weights lost during data processing were reestimated. For Sri Lanka, the enterprise weights used by the Sri Lankan team were readily available, but household weights had to be reestimated.
8. By collecting community-based data, RIC introduces a spatial dimension not prominent in PICS. This expansion opens new options for analysis of differential enterprise development in a heterogeneous rural space and along the rural-urban axis, addressing the complex analytical and methodological questions raised when observations are clustered in communities.

WHAT HAS BEEN LEARNED ABOUT RURAL ECONOMIES?

RIC2 unearthed several important findings about the three economies that would not otherwise be available:

1. The household data show that multiactivity of self-employment in RNFEs, wage employment, and farming is a common feature of rural households. Available choices for individual households and household members differ with family structure, human capital, assets, and community characteristics. RNFE self-employment choice is part of a broader array of livelihood strategies, which includes pursuing opportunities and managing risks. The view among some analysts that RNFE self-employment results from push factors for the poor is too narrow. This study provides evidence that households engaged in RNFE tend to be better off than farming families and tend to pursue entrepreneurship for commercial reasons.
2. The large majority of rural enterprises are very small in size, with one or a few workers only. The large majority of enterprises (68 to 86 percent) employ only unpaid family workers. Enterprise population is not static, however. After comparing the three countries and differences within the countries, a general pattern emerges showing that, with increased per capita income, both enterprise density per 1,000 inhabitants and the average size of enterprises increases.
3. Large numbers of small rural enterprises buy and sell mainly locally, which is understandable given the nature of their businesses—services, retail trade, repair, and so on. Yet, evidence from Sri Lanka indicates that enterprises that sell mainly outside their communities have higher productivity.
4. It is often believed that access to credit is the most constraining factor in RNFE. Yet the large majority of enterprises in the three countries (58 to 79 percent) report having no formal or informal debt. Only 31 percent or less (varying by country) carry formal debt equivalent to 50 percent of equity or higher. Although entrepreneurs may have a genuine desire for inexpensive loans, the benchmark indicator and benchmark components for finance services do not strongly influence measures of enterprise productivity.
5. The share of agro-processing in rural enterprises ranges from only 2 percent in Sri Lanka to 14 percent in Nicaragua. This shows that a focus only on agribusiness is too narrow for private-sector development in rural areas.
6. In all three countries, enterprise productivity appears to differ with enterprise and community characteristics. Consequentially, major differences emerge within and between

countries. Generally, the age of enterprises and the experience of entrepreneurs do matter, indicating the importance of adaptation and innovation through learning by doing for increasing productivity.

7. As diverse as are the three countries in this analysis, the lists of most important investment climate concerns expressed by their entrepreneurs show remarkable consistency. Six items appear on each country's list of top-ten concerns and three more occur on two of the three lists. The common concerns are:

   - finance cost, availability, and procedures;
   - electricity and water (access, cost, or reliability); and
   - road quality and access.

   The remaining concerns vary more widely: market demand, economic policy uncertainty, and telecommunications. At the other end of the spectrum, probably because of their small scale, few rural entrepreneurs in any of these countries report as obstacles food safety regulations, regulations on land use, customs regulations, or work permits for expatriates, to mention a few.
8. Complaints by entrepreneurs about rural infrastructure do in fact reflect conditions that register in the community surveys. Where improvements in infrastructure are reported at the community level, fewer complaints about infrastructure appear in the surveys. In communities reporting a higher number of households with access to protected water sources, for example, entrepreneurs complain less about water. The same result was found with respect to electricity, telecommunication, and postal services.
9. Enterprise performance; household choices about entrepreneurship, farming, and wage employment; and perceptions about investment climate obstacles are all affected by the community's overall economic environment. Observed community characteristics (including benchmarks) matter, but apart from these, regression analysis shows that unmeasured community factors are highly influential in causing similarities in choices and behavior within a community.
10. Many rural enterprises are not registered: 70, 42, and 80 percent in Nicaragua, Sri Lanka, and Tanzania respectively. Registration status is associated with higher enterprise performance only in Tanzania, but registered enterprises employ more labor in Nicaragua and Sri Lanka and have more invested capital in Sri Lanka and Tanzania. Thus, the effect of registration in the three countries is mixed.

RECOMMENDATIONS ON RICS METHODOLOGY

The recommended method for RICS implementation is described in the Implementation Manual (World Bank 2007c), which is based on previous analytical work. Findings from the present study indicate benefits from updating and further sharpening the manual's recommended methodologies and approach. Main emphases include the following.

Sensitivity for Using Weights and Random Effects

The econometric techniques used to estimate the various models address two characteristics of RIC databases: (i) observations have unequal sampling weights, and (ii) observations are clustered in communities. The question arises how much difference it makes to account for weights and random effects in the actual empirical analysis of RIC data. Section 4 of Annex I considers representative examples of enterprise performance, entrepreneurship selection, and EICO models estimated in alternative ways, using data from Nicaragua and Sri Lanka. In particular, the effect of benchmark indicators and community characteristics is sensitive to the specification of the model with regard to application of sampling weights and use of random effects. Overall, the evidence suggests that accounting for weights should be mandatory and that accounting for clustering is strongly advisable.

Use of Community Variables

The community forms a core element in RICS. Selected as the lowest unit of public administration, with public services that perform public duties, it is also considered to be a basic level of economic organization, meaning it shows basic economic infrastructure, interaction between enterprises, and public functions that affect businesses. Hence, at this level, investment climate has

unique characteristics that distinguish one community from another in attractiveness for private investment and business operation. Because many community characteristics can be measured, the resulting descriptors can be used as exogenous variables in explaining enterprise performance and entrepreneurial choice, assessing investment climate outcomes, and explaining outcomes of community-level economic activity. The information can also be condensed in robust community-level benchmark indicators, a technique utilized in the previous chapters.

It sounds simple, but in reality there are some complicating factors. First, communities exhibit significant differences in size and, therefore, in available services and economic functions. Obviously, the incidence of certain services in a community increases with population size. If one community consists of five hamlets and another community of ten of similar size, the chance of finding public infrastructure, public services, and business services will be higher in the second community. This does not necessarily imply that the second community will offer a better investment climate, however.

Second, communities differ in character. Some communities are simply larger hamlets, others are a cluster of hamlets, and some can be rural market towns. This kind of variation can be helpful in explaining differences in enterprise activity. Some of these differences may be easily influenced by policies and public and private investment, but others may reflect geographic features or large past investments that are a given fact.

Third, even though communities are considered to be the basic units of administration, functions may be performed at lower levels of aggregation as well, for example, when some duties are charged to village headmen or councils. Many administrative functions, moreover, are performed not at the community level but at higher levels of local administration, such as the subdistrict, district, or province level.

Fourth, significant heterogeneity in access to infrastructure services can exist within a community. Even if a community has a concrete or asphalt road, many residents may still have major problems getting to that road because of impassable feeder roads, lack of a bridge, or other obstacles. Even in communities with electricity or fixed phone lines, many residents may still be deprived of such services due to incomplete coverage. In other cases, a community may lack basic economic infrastructure and public and business services readily available in neighboring communities.

On the basis of these observations it can be concluded that, by nature, measurement of community economic characteristics is difficult and the descriptors to some extent are imprecise and deficient. Some of the variables measured may have heterogeneous effects, and, as already noted, many variables correlate and interrelate in complex ways. Moreover, several important variables remain by necessity unobserved or insufficiently measured. When used in regressions, therefore, variables are likely to present problems of multicollinearity and effects from unobserved variables.

Does this represent a weakness in RICS as compared to PICS? No. PICS makes no systematic attempt to measure community-level variables, leaving most of them unobserved and their influence unassessed. Variables that go unmeasured don't create problems of multicollinearity. Some of the statistical problems encountered in RICS are, in other words, to an extent problems of "luxury" rather than "poverty." It is better to have data with some weaknesses than to have no data. More importantly, the present study clearly shows the statistical importance of community variables. In rural space, local conditions can differ widely, hence the need to measure their effect accurately.

But, having said that, the challenge faced by the RICS team and country team managers is to design and implement surveys in ways that circumvent avoidable statistical problems and optimize chances for reliable and useful results.

International Comparison

International comparison of benchmark indicators and their components is imprecise because of two main factors. First, sometimes major differences can exist between countries in their systems of public administration, the average size of communities and the functions performed at the community level. Second, the concept *rural* has no common international definition, and each country either employs its own definition or simply designates an inventory of jurisdictions and administrative zones as *rural* or *urban*. Therefore, the selection of communities for RIC assessments may likely extend beyond those the host country classifies as rural. Indeed, it will likely prove necessary for RICS task managers to sample both in

Box 7.1 Some Options for Exploring Improvement of Benchmark Indicators

The purpose of benchmark indicators is to condense information on the investment climate at the community and country level for comparison and analysis. Further optimization of definitions and empirical performance should focus on the following:

1. Overlap in definition between the benchmark indicators increases intercorrelation and reduces the effectiveness of both in the regressions. Especially the connectivity index and the infrastructure services index seem to have too much overlap.
2. For comparison and good statistical performance in regressions, benchmark indicators and their subcomponents should show sufficient variation. Options can be explored to increase variation by restructuring and rescaling subindicators. In particular, the governance and human capital indexes deserve attention. Some benchmark components relate closely to community size, which is not necessarily meaningful from an enterprise perspective. Analysis should guide the rescaling of these components.
3. Some determinants of economic outcomes can be changed by policy and investment efforts, others hardly or not at all. Options should be considered on how to distinguish such factors as determinants of economic structure on the one hand and investment climate obstacles on the other. The PICS provides no guidance since it has no focus on spatial diversity; but for rural areas, RICS should consider this, as infrastructural policies and investment programs are often region specific.
4. From a users' perspective, economic policy obviously involves a trade-off between robust generality of benchmark indicators and ability to identify more specific weaknesses in the business environment.
5. It is worthwhile to include a few more indicators while narrowing the scope of each of them. This should lead to a smaller within-community statistical variation among the subindicators while increasing the variation of indicators between communities.

Source: The Authors.

rural market towns and in peri-urban communities, regardless of how they are classified.

In addition, adjustments in definitions and scaling are likely to affect country rankings.[63]

Experience with Benchmark Indicators

Just as with other large surveys carried out by the World Bank Group, including Governance, Cost of Doing Business, and the many PICS, benchmark indicators from the RIC assessments are needed for cross-country comparison and analysis. This study's experience using benchmark indicators in the regressions is encouraging. In binary correlations they behaved as expected from a theoretical point of view, although some weaknesses remain to be addressed.

The correlation between benchmark indicators, in part stemming from overlap, could be reduced and the scaling of some components improved. A particular flaw is that several benchmark indicators increase with community size, because some underlying variables likely relate positively to population size and land area, for example, the incidence of business services and roads. Improved standardization could mitigate this. Moreover, scope remains for sharpening and optimizing BI conceptualization and measurement. Rather than making ad hoc adjustments, the work should be guided by statistical analysis. The present estimation models can be used to test the effect of changes in benchmark indicator definitions, and adding other countries to the sample—Bangladesh, Benin, Ethiopia, Indonesia, and Pakistan—will greatly enhance the effectiveness of this work. Suggestions are summarized in Box 7.1.

Questionnaires

1. Experience in the RIC pilots shows that more data were collected than were actually used in the analysis. This is understandable because of the three surveys' pilot character. Questionnaire length comes at a cost, however, not only in monetary terms, due to the amount of time needed to conduct the surveys and process the data, but also in loss of

quality, due to nonresponse and interview fatigue among interviewees. Some efforts have already been made to reduce the standard questionnaire included in the Implementation Manual (World Bank 2007c). Based on accumulated experience, however, a careful revision that leads to a substantially shortened questionnaire is recommended.

2. Further questionnaire standardization should be promoted. Country teams were generally inclined to include excessive numbers of country-specific variations in their survey designs. These additions, however, resulted mainly in data that were not even subsequently processed and analyzed, while unintentionally diluting the pilot's RIC focus and reducing its utility for inter-country comparisons.
3. Considerably more care should be given to measuring labor inputs. Labor is one of the most important inputs in the rural economy and a crucial variable in many analyses. The enumeration of labor input in the Sri Lankan and Nicaragua pilots, particularly, had some deficiencies that undoubtedly affected the analysis. Unpaid family labor, for example, could have been measured properly using information from the household database if the household and enterprise questionnaires had been better linked. Good labor enumeration deserves priority.

Survey Design

1. Stratified sampling is indispensable for drawing a cost-effective sample. This study has shown that weights are absolutely necessary in the analysis for obtaining unbiased results.[64] Therefore major effort should be made in the design and implementation of surveys to establish a proper sampling framework and to collect reliable weights.
2. Because rural space contains so many small and so few large enterprises, further stratification by enterprise size may make sense. If large enterprises appear in a community's sampling frame, they should be sampled with a higher probability than the small enterprises, with sampling weights adjusted for the differential sampling probabilities. This effectively increases the amount of variation in, or the heterogeneity among, enterprises and is likely to improve the quality of regression results.
3. Similar arguments apply in the selection of communities. Community variation is desirable. In countries with many more or less similar communities—typical rural communities—a stratified sampling of communities is recommended that undersamples the homogeneous group. This adds to the importance of a good sampling system and weights.
4. Given the targeted number of observations in a survey of, say, 1,000 enterprises, trade-offs arise between the number of communities and the number of observations per community. Increasing the number of communities is favorable because it lifts constraints on the number of community variables that can be used in regressions. Several benchmark components and other community-level variables are aggregates of household and enterprise data, however. If there are few observations per community, the quality of these aggregates declines and the justification for employing them as independent variables vanishes. Therefore, for each country, optimizing the number of communities and the number of observations per community deserves special attention. The optimal choice of the number of communities and the number of observations per community is a conceptual matter that should be resolved by a statistical specialist (who should also address the consequences of varying the number of observations per community). In particular, it is unclear whether the trade-off implemented in the pilot RICS or subsequent RIC implementations has been optimal.
5. The number of observations per community is small in many cases, which decreases the reliability of aggregates calculated from individual observations. To the degree that this is simply caused by the small number of observations in the sampling frame (that is, the community is just very small), this is an unavoidable situation, but if instead the sampling frame contains an adequate number of observations, the likely cause of the problem is nonresponse. As is well-known, nonresponse already causes other statistical complexities, such as reduced randomization, but the RICS methodology needs aggregates

calculated from a sufficient number of individual observations. High priority should therefore be given to reducing nonresponse.

Survey Implementation and Processing

Experience shows that the cost-effectiveness of RIC surveys can greatly be improved by the following:

1. Improved quality of data collection and processing. Poor quality data collection and processing has major effect on the outcomes and efficiency of RIC work. It limits potential output and the precision of findings. It also increases the cost of analysis, since the amount of time required is much greater than for good-quality data sets. If fewer records must be discarded because of data-quality problems, sample sizes can be reduced somewhat or coverage increased.
2. Related to this point and as recommended in the Implementation Manual, another objective should be consistent database design. The three RIC databases studied in this report follow completely different formats. Great effort had to be expended to replicate the analysis of one country with data from another. Inconsistency leads to substantial delays in research output and discourages cross-country analysis.

Data Analysis

Based on this study's experiences with estimation, country teams are recommended to include in their studies the following analyses, in order of priority:

1. overviews of household data, enterprise data, and community data;
2. calculation of the benchmark indicators;
3. analysis of enterprise performance;
4. analysis of household economic choices, including enterprise start-up;
5. analysis at the community level of rural nonfarm outcomes, such as enterprise density, nonfarm income, and nonfarm employment;
6. analysis of EICOs; and
7. analysis of EICIs.

The econometric models and Stata programs provided in Annex I facilitate items 3, 4, 6, and 7. At the same time, just as the RICS pilots represent a process of learning by doing, these models remain a work in progress: they do not yet address issues of self-selection and simultaneity, and interaction between EICO and EICI variables, on the one hand, and enterprise performance and household choices, on the other hand, remain to be explored (see Section 5.5).

Use of RICS for Operational Work

The purpose of conducting RICS is to support policy dialogue, to sharpen development policies, and to improve the effectiveness of interventions. In Sri Lanka, the RIC has clearly contributed to this.[65] (See Box 7.2.)

The RIC results can contribute in several ways to preparing policy interventions:

1. The RIC survey results provide much otherwise unavailable information about the economic activities of rural households; the population of local enterprises; and institutional, policy, and infrastructural constraints. This is important information for policy discussions between civil society, especially private-sector stakeholders, and Government. The quality of dialogue can be enhanced by good RIC analysis and further assessment of options.
2. The various regressions performed—enterprise performance and entrepreneurship choices—show that from micro-perspectives various factors constrain productivity, employment, investment, and income earning. Many of these constraining factors can be changed—some more easily and others—raising challenges for policy intervention. These regressions provide suggestions that will always need further work, however, rather than ready-to-use policy recommendations.
3. In general, the RIC analysis as outlined in this report can and should be followed by further exploration of the RIC databases with regard to specificity of constraints, locations, groups of enterprises affected, and other concerns. RIC surveys create multipurpose databases having a public goods character that can be used for several years to come as sources of information to feed policy and project design. Moreover, additional information can be obtained from other sources and through targeted interviews and small-scale follow-up surveys. In fact, information from RIC databases will be helpful when designing better tailor-made surveys of rural space.

Box 7.2 Operational Applications: The Sri Lanka RIC Pilot

An assessment of IC constraints faced by RNFEs was conducted in the Sri Lanka pilot using both subjective and objective criteria. The ranking of RIC constraints in Sri Lanka, whether determined by objective measures or from subjective responses, differs in several important aspects from constraints experienced by the larger urban-based enterprises. Access to road transport and markets, reliable public utilities (electric power), and access to finance proved to be the main factors influencing (or constraining) employment growth and total factor productivity in the rural areas. Urban firms did not experience the same level of difficulty in accessing infrastructure and finance; instead, they faced governance issues and were far more likely to be negatively affected by inconsistent macroeconomic policymaking.

While it is difficult to gauge the full policy effect of the Sri Lanka RIC, the findings have already proven useful for the WBG's country operations there. The findings were woven into the Development Policy Review (DPR) and the Development Forum Paper for the 2007 Development Forum. The findings were also cited by leading politicians and civil servants at various forums in Sri Lanka.

The South Asia Region has also applied the findings to the design of operations in the proposed public financial management and statistical institutional building project as well as the legal and judicial follow-up project. The RIC analysis has also been used by other international agencies and by Sri Lankan think tanks and universities. RIC findings were used by the Asia Foundation for their operations designed to improve provincial level investment climates. The Asian Development Bank, which coauthored the Sri Lanka RIC Assessment with the Bank, has relied on the analysis to inform its microfinance project.

Government is drawing on the data on missing infrastructure to develop a PPP capacity for large infrastructure projects. Bank staff also have used the materials to teach courses at the University of Colombo, the Moratuwa MBA in infrastructure, and the University of Sri Jayawardenpura.

It is generally agreed that the Sri Lanka RIC did not result in a breakthrough on the frontier of investment climate studies. Rather it has proved useful as a practitioner's tool. The assessment quantified many of the issues already under discussion and allowed a prioritization and focus on the most important critical areas. It managed to strengthen the evidence base for policy formulation and allowed stakeholders to speak with authority on particular issues covered by the study, such as the losses from electricity outages or the cost of labor regulations.

Source: Author communication with Ismail Radwan 2008.

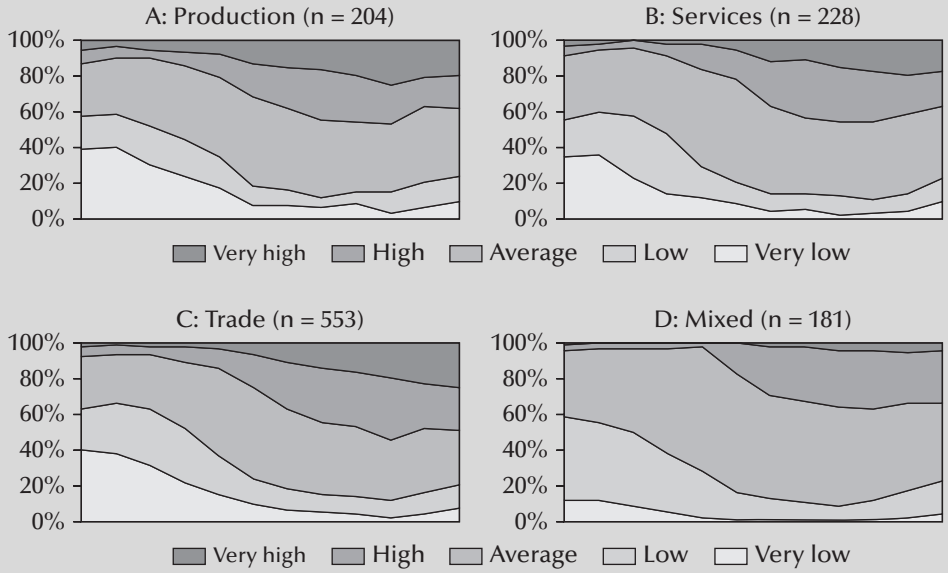
4. RICAs should only be conducted in countries where Government, relevant donors, and Bank operations strongly demand information on rural enterprises. RICAs should be part of regular country programs and should have strong potential for use in lending operations or analytical work for PRSPs, CASs, and sector work. Moreover, a team with the right qualifications and sufficient budget should be available to carry out the RICS.

FURTHER RESEARCH

Many potential future research topics present themselves.

1. The analytical tools applied in this report deserve to be applied with priority to other available data sets such as those for Bangladesh, Benin, Ethiopia, Indonesia, and Pakistan. This will add to the understanding of similarities and variance in rural economies and will help sharpen the analytical methods applied.
2. Relationships between the perceived investment climate, the actual conditions measured through the community questionnaire, enterprise performance, and entrepreneurship choice can be developed further. RIC2 relates actual conditions to perceptions, enterprise performance, and entrepreneurship choice; future studies should relate perceptions to enterprise performance and entrepreneurship choice, with careful attention to bidirectional causality and self-selection issues.
3. In a descriptive manner, perceived investment climate obstacles (termed EICOs in this report) are ranked in each country, allowing

Figure 7.1 Level of Seasonal Activity Among Enterprises in Tanzania



Source: RIC Surveys.

comparisons among countries in terms of their ranked EICOs. Descriptive differences are interesting, but a statistical test should be developed that indicates whether differences in ranking are indeed statistically significant.

4. The design of the questionnaires makes the analysis of EICOs very inviting, but aspects of interactions with government and society (termed EICIs) are more difficult to summarize and analyze. The RICS data offers far more than could be covered here. In particular, EICIs might be benchmarked as well.
5. Selective comparative studies of RICS and PICS will originate a better understanding of the rural-urban continuum and rural-urban divides, provided suitable databases can be identified.
6. Communities in the RICS databases are coded with geographic longitude-latitude information. This fixes them somewhere in rural space, but spatial implications have not yet been explored at all. Spatial regression methods could exploit correlations among nearby communities that are widely ignored in applied research to date.
7. Because questionnaires, and therefore data, were not comparable between the countries, the deeper details on labor input for Tanzania were not fully utilized. Additional analysis with adjusted labor variables is worth trying.
8. Entrepreneurs were asked about business seasonality. Figure 7.1 illustrates its importance for Tanzania, but this is only the very beginning of a description. How does the cycle compare with the seasonal cycle in agriculture? Does the degree of seasonality reported by entrepreneurs vary with the market served (that is, location of customers), the type of customers, or the location of of input sources?

Annex A. Description of Data

Description of variables and enterprises. The RIC survey procedures and methodology are described World Bank (2006). In total data on about 1,350 enterprises were collected for each of Nicaragua, Sri Lanka, and Tanzania (Table A.1). For various reasons, however, smaller numbers of enterprise samples were included in the analysis. For Nicaragua, the household-based enterprises may not belong to different households. Some households reported on as many as four enterprises, but since sales and expenditures were reported only at the household level, multiple household-based enterprises were aggregated to the household level, leading to a reduction of 308 in the enterprise count. For some communities in Sri Lanka, essential variables were incomplete, leading to a significant loss in the number of observations.[66] In Nicaragua, a data-entry failure to record the community identification number led to a loss of 247 enterprises. In a few cases in all three countries, explanatory enterprise-level variables were missing. A trade-off clearly occurs between the number of variables specified in the regression models and the number of enterprises available for the regression analysis. More elaborate specification implies an increased need to drop some records because of missing variables.

One type of sample reduction is deliberate. Some of the enterprises for which sufficient data are available reported zero sales, and in other cases the reported calculated net value added was very small or even negative. This may happen when nonperforming enterprises are temporarily closed, have just opened, or are about to exit. It may also represent self-employment on a very small scale.[67] This data problem may also occur when reference-period inputs yield future rather than reference-period output, and similarly, when reference-period output results from previously purchased inputs.[68] Because of the potentially disrupting role of this group, a threshold is adopted for inclusion in the regression analysis: when analyzing sales, enterprises must have sales at least equal to US$100 or a net value added at least US$60; when analyzing net value added, enterprises are included only if net value added is at least US$60.[69] As indicated in Table A.1, this restriction had the greatest effect on Tanzania. As a result, for all reasons together, the number of enterprises in the analyzed samples is about 40 percent less than that in the overall database.

Data availability. From the data collected in the RIC surveys, variables are taken and other variables are calculated for analysis of enterprise performance. Table A.2 provides an overview of data grouped as financial data, other enterprise characteristics, and community characteristics. Community-based benchmark indicators are calculated mainly from the community questionnaires and partly from aggregated household and enterprise data.

Reliability of verbally reported costs, sales, volumes, and prices can be weak; most enterprises have no bookkeeping, and reporting can be affected by the limitations of human memory, estimation errors, and unwillingness to fully disclose financial data. On the other hand, most small enterprises have simple, fairly simple to remember activities.

A weakness in the data concerns the measurement of labor, for which definitions differ between countries. Moreover, assessment of part-time labor and employment of household labor is imprecise, especially for Sri Lanka (see Box A.1 for details). Thus, despite its possible flaws, a stock rather than a flow concept is used. The value of paid labor input was recorded in the data as the actual paid wage sum, a flow concept, but the stock is not known because the number of part-time workers is not known exactly. Quality differences of paid labor are presumably reflected in the wage paid per person. For unpaid household members, no information is available about quality of labor; for the manager, the number of years of formal education

Table A.1 Enterprises Included in the RIC Surveys and the Enterprise Performance Regressions

	Nicaragua		Sri Lanka		Tanzania	
Regression	Sales	Net Value Added	Sales	Net Value Added	Sales	Net Value Added
Number of enterprises	1535*		1327		1239	
Enterprises dropped because of insufficient data	667*		278		134	358
Maximum used in regressions	868		1049		1105	881
Enterprises with reported sales < 100 and/or NVA < 60 US$	22	54	31	208	158	182
Remaining enterprises	846	814	1018	841	947	699

Source: RIC Surveys.

* includes multiple enterprises owned by one household that were aggregated.

and training is available for Sri Lanka and Tanzania but not for Nicaragua.

For capital, understandable difficulties arise in separating capital such as a house, equipment, and vehicles into use for consumption and use for productive purposes. Also, market prices for valuation of assets are not always known, and some liabilities may have been obtained from relatives and friends at concessionary prices.

Underemployment of household labor may be a serious problem, and hence the marginal productivity of household labor may be lower than that of paid labor. To some extent, that may also be the case with household-owned assets and borrowed cash and assets.

One way of dealing with uncertainties about the reliability of imputed labor and capital costs is to specify imputed and paid labor and capital costs as separate variables. This allows the estimation of different elasticities for paid and imputed inputs.

Parts of the data sets show deficiencies in the collection and processing of raw data. In the selection of the enterprise records for the regression analysis, interpretations must be made about how to handle missing data, possible data errors, and mistakes in coding.

In conclusion, it is fair to say that errors, in part the result of the surveys' pilot character, negatively affect the precision and efficiency of parameters estimated in the regressions.

Table A.2 Definition of Variables

Variable Name	STATA Name	Description
Variables Used in the Enterprise Performance Regressions		
Recorded and calculated financial data		
Sales	Sale	Total Sales (US $)
Net value added	NVA	NVA = GVA − Depreciation (US $)
Total labor input[1]	VLI	Value of labor input = paid wages + HLIV (US $)
Total capital input	VCI	Value of Capital input = ICAP + PCC (US $)
lnVLL (lnL*lnL)	lnVLL	lnVLI * lnVLI quadratic lnL term
lnVLC (lnL*lnC)	lnVLC	lnVLI * lnVLC interaction lnL lnC term
lnVCC (lnC*lnC)	lnVCC	lnVCI * lnVCI quadratic lnC term
LnIL*lnIL (imputed)	lL11	lnHLIV * lnHLIV quadratic imputed lnL term

Table A.2 Definition of Variables (*continued*)

Variable Name	STATA Name	Description
Variables Used in the Enterprise Performance Regressions		
Recorded and calculated financial data		
LnPL*lnPL (paid)	lL22	lnPLC * LnPLC quadratic paid lnL term
lnIC*lnIC (imputed)	lC11	lnICAP * lnICAP quadratic imputed lnC term
lnPC*lnPC (paid)	lC22	lnPCC * lnPCC quadratic paid lnC term
lnIL*lnIC (imputed)	lL1C1	lnHLIV * lnICAP interaction imputed lnL lnC term
lnPL*lnPC (paid)	lL2C2	lnPLC * lnPCC Interaction paid lnL lnC term
Depreciation	depr	(imputed) 8% percent of the value of Fixed Assets (excluding land)[2] (US $)
Nonfactor cost	NFC	Purchase of materials and items for resale; Purchase of electricity, water, gas, and fuels; Transport costs; Cost for telephone, mobile phone, fax, internet, postal and insurance service; Business services (US $)
Family labor input	HLIV	Household labor imputed value = family labor cost at imputed wage rate[3] (US $)
Paid labor input	PLC	Paid wages (US $)
Imputed capital cost	ICAP	Imputed capital cost = net value of assets * imputed interest rates[4, 5] (US $)
Paid capital costs	PCC	Paid Capital Costs = rental payments for land, equipment, buildings, etc. (US $)
Other enterprise characteristics		
Age of enterprise	firmage	firmage = Year of survey − year enterprise established
Experience of manager	exptmng	Years of experience reported for manager
Gender of manager	sextmng	Dummy variable for male manager (1 = male)
Education of manager	edutmng	Education of manager[6] (years)
Registration	regis	Registration status of enterprise (1 = registered)
Industry dummies/line of business		
Trade enterprise	B1	Dummy (1 = Retail or wholesale trading)
Services enterprise	B2	Dummy (1 = Services)
Manufacturing, nonagricultural, enterprise	B3	Dummy (1 = Manufacturing nonagricultural goods)
Agricultural processing enterprise	B4	Dummy (1 = Processing of agriculture, hunting, fishing products)
Mining enterprise	B5	Dummy (1 = Mining and quarrying)
Production	B345	Dummy (1 = Production [combined B3, B4, B5])
Other production	B567	Dummy (1 = Other production [combined mining, construction, gas water electricity])
Mixed enterprise	B8	Dummy (1 = Enterprise including at least two lines of business)
Community characteristics		
Agricultural seasonality	agseas	Average of the standard deviation in monthly agricultural labor input for male and female labor
Enterprise density	entdense_pop	Number of enterprises in community per 1,000 people
Community population	popnw	Number of inhabitants in the community
Paddy land per capita	apaddylpc	Area of paddy land per capita (acre or hectare)
Illiteracy	shilliterate	Share of illiterates (%)
Main market in		Market in which households buy/sell most of their goods
Own community	mainmkt1	Dummy (1 = Buying/selling in own community)
Neighboring communities	mainmkt2	Dummy (1 = Buying/selling in neighboring communities)
Commercial center	mainmkt3	Dummy (1 = Buying/selling in commercial center)
Nearest city	mainmkt4	Dummy (1 = Buying/selling in nearest city)
Main income from		
Agriculture	insoura	Dummy (1 = Main income source community is agriculture)
Wages	insourw	Dummy (1 = Main income source community is wages/salaries)
Self-employment	insourn	Dummy (1 = Main income source community is self-employment)

(*continued on next page*)

Table A.2 Definition of Variables (*continued*)

Variable Name	STATA Name	Description
Variables Used in the Enterprise Performance Regressions		
Benchmark indicators and components		
Connectivity	conn	Connectivity index (index %)
Infrastructure services	access	Access to infrastructure services (index %)
Business services	devserv	Business services index (index %)
Governance	govern	Governance and corruption index (index %)
Human capital	humcap	Human capital index (index %)
Finance services	finan	Finance services index (index %)
Inverse time to near city	c1	Inverse time taken by main means of transportation to nearest major city (index %)
Inverse cost of transportation	c2	Inverse cost of transportation to the nearest major city (index %)
Inverse time to main market	c3	Inverse time taken by main means of transportation to main market (index %)
Proximity post office	c5	Proximity to the main post office (index %)
Percent with electricity	a1	Percentage of households with electricity (index %)
Availability of electricity	a2	Dummy (1 = Electricity availability)
Percent with protected water	a3	Percentage of households with access to protected water
Percent with fixed phone	a4	Percentage of households with fixed-line telephone
Percent with cellular phone	a4B	Percentage of households with access to cellular phones
Sewage system	a5	Dummy (1 = Sewage system in the community)
Garbage collection	a6	Dummy (1 = garbage collection or disposal service in the community)
Concrete/asphalt road	a7	Dummy (1 = Most common road surface is concrete or asphalt)
Engineering service	d1	Dummy (1 = Engineering services available for businesses in the community)
Management consulting	d2	Dummy (1 = Management consulting services available for businesses in the community)
Marketing service	d3	Dummy (1 = Marketing services available for businesses in the community)
Accounting service	d4	Dummy (1 = Accounting services available for businesses in the community)
Insurance service	d6	Dummy (1 = Insurance services available for businesses in the community)
Information technology	d7	Dummy (1 = Information technology services available for businesses in the community)
Infrastr and services	go2	Infrastructure and services (index %)
Public services/institutions	go3	Dealing with government services and general policy and institutional constraints (combined indexes) (index %)
Public services	go3a	Dealing with government services (index %)
Public institutions	go3b	General policy and institutional constraints (index %)
Rule of law	go4	Rule of law (index %)
Number of banks	fi1	Number of formal financial sources weighed by mean of the distance (index %)
Number bank services	fi2	Number of formal finance services weighed by mean of the distance (index %)
Access to loans	fi3	Access to loans (index %)
Other Variables		
Gross value added	GVA	GVA = Sale − NFC (US $)
Total factor input	TFI	Total Factor Input = Value of capital input (VCI) + Value of labor input (VLI) (US $)
Total cost	TC	Total Cost = Total Factor Input + Non Factor Input + Depreciation (US $)
Net profit	netp	Net Profit = Sales − Total Cost (US $)
Profitability or total productivity	SATC	Profitability = Sales / Total Cost (ratio)
Net factor productivity	NVFC	Net Factor Productivity =NVA / TFI (ratio)
Net income	ninc	Net Income = sale − NFC − PCC − depr − PLC (US $)
Total liability	debt	Loans or credits from banks, private lenders and from suppliers; Money owed to friends and relatives, and others (leasing companies, etc.) (US $)
Total assets	totA	Total Assets[5] = Fixed assets + Inventories and stocks + Financial assets (US $)

Table A.2 Definition of Variables (*continued*)

Variable Name	STATA Name	Description
Variables Used in the Enterprise Performance Regressions		
Other Variables		
Net value of assets	Netass	Net value of assets = value of assets − liabilities (US $)
Employment	E	E = VLI/imputed daily wage rate (proxy for number of persons)
Imputed factor costs	IFC	IFC = HLIV + ICAP (US $)
Profitability of own resources	NIIFC	NIIFC = ninc/IFC (ratio)
Bank loan > 50 % of debt	formloan	Dummy that has a value of one if 50% or more of the total liability is from a Bank
Stand-alone enterprise	standalone	Dummy (1 = An enterprise not located at the house lot of the owner)
Enterprise with paid labor	PLC01	Dummy (1= Enterprise with hired labor)
Enterprise with debt	debt01	Dummy (1= Enterprise with some liabilities >0)
Ethnicity of manager	ethtmng	Dummy (1= Manager is Sinhalese − Sri Lanka only)

Source: The Authors.

Notes: 1. **Paid labor** is specified as a cost. For household labor, ambiguities about the coverage of household labor are specified in Box A.1.

2. **Depreciation** was imputed for estimation of net profit and net value added. In all countries (an arbitrary) 8 percent depreciation was adopted on Storage facilities (separate from the building); Machinery and equipment (excluding vehicles); Vehicles for transportation; Specialized vehicles; and Other fixed assets. Land was excluded from depreciation.

3. **Imputed wage rates** (opportunity cost for unpaid labor) chosen after comparing data from the enterprise and community surveys. Paid wages and numbers employed were calculated for groups of registered (formal) and unregistered (informal) enterprises with mainly full-time labor. Estimates that include enterprises with part-time labor will be biased. Registered enterprises in some cases pay high wages, which may not be accessible for most family labor. More relevant may be wages paid in unregistered enterprises. Wage rates reported from community surveys tend to reflect wage levels in registered enterprises. The following daily imputed wage rates were adopted: Nicaragua US$0.80, Sri Lanka US$1.30, Tanzania US$1.00.

4. **Imputed interest rates** were chosen after comparing interest paid on loans from formal enterprises and relevant rates reported in IMF financial tables and Central Banks. The imputed interest rates for land were corrected for inflation because in most cases, in the long term, land value will go up with inflation. The following imputed rates were chosen:

	Land	Other Assets
Nicaragua	10.58%	18.78%
Sri Lanka	7.5%	15%
Tanzania	12.9%	17%

5. **Total assets**

A. Fixed Assets

Land; Building and improvement in leasehold (excluding storage facilities); Storage facilities (separate from the building); Machinery and equipment (excluding vehicles); Vehicles for transportation; Specialized vehicles; Other fixed assets.

B. Inventories and stocks

Finished goods and commodities for sale; Work in progress; Raw materials excluding fuel; Fuel; Acquisition Cost: Machinery and equipment (including transport).

C. Financial Assets

Accounts receivable; Cash in hand/in bank.

6. Education

Nicaragua: Not applicable: education of manager was not recorded in the survey.

Sri Lanka: 0–15 years.

Tanzania: Codes were translated into years of schooling: *Primary education:* Under standard one = 0; Standard one = 1; Standard two = 2; Standard three = 3; Standard four = 4; Standard five = 5; Standard six = 6; Standard seven = 7; Standard eight = 8; Adult education = 8; Training after primary education = 9. *Secondary education:* Pre form one = 9; Form one = 10; Form two = 11; Form three = 12; Form four = 13; Form five = 14; Form six = 15. Training after secondary education = 16; University & other tertiary education = 17.

Box A.1 Specifications of Labor Input in Enterprise Questionnaires

The questionnaires differ widely in their descriptions of enterprise labor, and consequently possibilities for estimating labor input differ widely as well, as shown in these notes on the questionnaire definitions.

Tanzania has by far the best description of labor input, although it remains unclear whether various household membersare unpaid or receive a salary.

For **Nicaragua,** data about household labor and other employees in household enterprises are adequate. Data for stand-alone enterprises are ambiguous. A root of the problem is the overlooked possibility that stand-alone enterprises are household enterprises. Some questions skip possible relatives of the manager. Probably the manager is not included as an employee. It is possible that family members and the manager receive salaries. The categories listed ignore part-time temporary labor and assume that permanent labor is either full-time or part-time; this is an unrealistic assumption for seasonal enterprises.

The questions in the **Sri Lanka** survey are ambiguous and leave much room for diverging interpretations. It is not clear whether the manager is part-time or full-time. Only family members working full-time in the enterprise are listed. In some questions it is not clear whether household labor is included; in others household labor is included, but the manager may or may not be included. Also unclear is how to aggregate the different categories of labor.

Source: RIC enterprise questionnaires.

Annex B. Data Used for Enterprise Performance Analyses

Table B.1 Nicaragua: Selected Variables of Enterprises and Communities; Value and Dstribution

Variable name	mean	st dev	min	max	obs	pct20	pct40	pct60	pct80
Sales#	7.004	1.162	4.193	11.542	846	6.115	6.625	7.223	7.966
Net value added#	6.823	1.260	0.000	11.389	820	5.902	6.521	7.093	7.846
Total labor input#	5.627	1.071	0.000	11.201	846	4.908	5.355	5.820	6.421
Value of capital input#	4.893	1.642	0.000	10.598	846	3.567	4.479	5.277	6.192
lnVLL (ln*ln)	32.816	12.096	0.000	125.460	846	24.091	28.671	33.873	41.230
lnVLC (ln*ln)	28.081	12.461	0.000	101.309	846	18.422	24.275	29.834	36.668
lnVCC (ln*ln)	26.636	16.806	0.000	112.314	846	12.723	20.057	27.847	38.342
Depreciation#	2.424	1.872	0.000	9.291	846	0.695	1.613	2.795	4.029
Nonfactor cost#	3.350	2.520	0.000	11.100	846	0.000	3.095	4.464	5.578
Other enterprise characteristics									
Age of enterprise#	2.090	0.912	0.000	4.143	846	1.253	1.946	2.398	2.890
Registration	0.298	0.457	0.000	1.000	846	0.000	0.000	0.000	1.000
Industry dummies/line of business									
Trade enterprise	0.396	0.489	0.000	1.000	846	0.000	0.000	0.000	1.000
Services enterprise	0.188	0.390	0.000	1.000	846	0.000	0.000	0.000	0.000
Manufacturing, nonagricultural, enterprise	0.084	0.278	0.000	1.000	846	0.000	0.000	0.000	0.000
Agricultual processing enterprise	0.139	0.346	0.000	1.000	846	0.000	0.000	0.000	0.000
Other production enterprise	0.032	0.175	0.000	1.000	846	0.000	0.000	0.000	0.000
Mixed enterprise	0.161	0.368	0.000	1.000	846	0.000	0.000	0.000	0.000
Parent operating NFE	0.386	0.487	0.000	1.000	846	0.000	0.000	0.000	1.000
Parent was manager	0.002	0.040	0.000	1.000	846	0.000	0.000	0.000	0.000
Parent occupation. Missing	0.074	0.262	0.000	1.000	846	0.000	0.000	0.000	0.000
Community characteristics									
Agricultual seasonality	0.684	0.411	0.000	1.477	846	0.389	0.522	0.866	0.996
Enterprise density	49.915	93.972	0.538	445.000	846	2.749	4.067	11.917	43.695
Community population size#	8.785	1.862	5.303	11.775	846	6.553	8.216	9.893	10.528
Agricultural land per capita	0.422	0.707	0.000	4.762	640	0.050	0.094	0.275	0.667
Illiteracy	80.861	14.768	0.000	99.000	740	72.500	80.000	85.000	92.500
Benchmark indicators and components									
Connectivity	0.358	0.177	0.028	0.828	846	0.178	0.317	0.413	0.511
Infrastructure services	0.551	0.238	0.000	0.983	846	0.357	0.499	0.600	0.814
Business services	0.303	0.429	0.000	1.000	846	0.000	0.000	0.143	1.000
Governance	0.691	0.111	0.372	0.975	846	0.581	0.669	0.721	0.780
Human capital	0.216	0.074	0.067	0.402	846	0.144	0.187	0.233	0.289
Finance services	0.201	0.267	0.000	1.000	846	0.000	0.000	0.217	0.494

Source: RIC Surveys.

Note: See Annex A for definitions and description of data; # logarithmic value.

Table B.2 Nicaragua: Selected Variables of Enterprises and Communities; Value and Distribution

Variable name	mean	st dev	min	max	obs	pct20	pct40	pct60	pct80
Nonfactor cost	374	2388	0	66188	846	0	21	86	264
Sales	2438	5980	65	102954	846	452	753	1369	2881
Depreciation	89	449	0	10843	846	1	4	15	55
Gross value added	2064	4765	−10482	89209	846	348	678	1201	2483
Net value added	1975	4688	−11752	88378	846	324	646	1140	2275
Family labor input	324	280	0	1934	846	115	195	288	493
Paid labor cost	224	1692	0	72288	846	0	0	0	0
Total labor input	548	1747	0	73195	846	134	211	336	614
Net value of assets	4110	14862	−2667	221194	846	314	722	1613	3803
Imputed capital cost	581	2529	0	40046	846	34	81	173	461
Paid capital cost	39	345	0	9036	846	0	0	0	0
Value of capital input	619	2613	0	40046	846	34	87	195	488
Total factor input	1168	3451	0	81668	846	231	395	620	1112
Total cost	1630	5417	36	118003	846	290	476	792	1422
Net profit	808	4473	−43084	49815	846	−172	155	624	1506
Profitability	2.628	3.242	0.031	68.126	846	0.774	1.345	2.223	3.735
Net factor productivity	3.180	4.679	−2.052	76.724	845	0.719	1.421	2.519	4.241
Net income	1713	3819	−13619	51671	846	267	606	1123	2224
Total liability	135	987	0	19766	846	0	0	0	4
Total assets	4245	15276	0	221194	846	314	731	1663	3883
Employment	685	2184	0	91494	846	168	263	420	767
Imputed factor costs	905	2544	0	40046	846	218	379	556	927
Profitability of own resources	3.705	5.418	−16.352	76.853	844	0.817	1.699	2.788	4.726
Age of enterprise	10.715	10.019	0.000	62.000	846	2.500	6.000	10.000	17.000
Bank loan > 50% of debt	0.066	0.249	0.000	1.000	846	0.000	0.000	0.000	0.000
Stand-alone	0.074	0.262	0.000	1.000	846	0.000	0.000	0.000	0.000
Enterprise with paid labor	0.171	0.377	0.000	1.000	846	0.000	0.000	0.000	0.000
Enterprises with debt	0.212	0.409	0.000	1.000	846	0.000	0.000	0.000	1.000
Community population size	22199	30080	200	130000	846	700	3700	19800	37335

Source: RIC Surveys.

Notes: See Annex A for definitions and description of data.

Table B.3 Sri Lanka: Variables Used in Regressions of Enterprise Performance; Value and Distribution

Variable name	mean	st dev	min	max	obs	pct20	pct40	pct60	pct80
Sales#	7.355	1.445	4.429	14.439	1018	6.124	6.893	7.637	8.512
Net value added#	6.207	1.634	0.000	14.144	908	4.964	5.863	6.613	7.475
Total labor input#	6.029	0.790	4.190	13.117	1018	5.745	5.745	5.746	6.517
Value of capital input#	5.493	1.532	0.000	13.674	1018	4.254	5.146	5.854	6.752
lnVLL (ln*ln)	36.974	10.753	17.553	172.064	1018	33.009	33.009	33.019	42.476
lnVLC (ln*ln)	33.696	13.003	0.000	168.937	1018	24.675	29.311	34.351	41.326
lnVCC (ln*ln)	32.522	17.288	0.000	186.987	1018	18.097	26.485	34.269	45.590
Depreciation#	4.067	1.715	0.000	13.037	1018	2.598	3.748	4.496	5.520
Non Factor Cost#	6.479	1.907	0.000	13.064	1018	5.012	6.133	6.995	7.944
Other enterprise characteristics									
Age of enterprise#	1.720	1.080	0.000	4.431	1018	0.693	1.386	1.946	2.773
Experience of manager#	1.941	0.946	0.000	4.043	1018	1.099	1.792	2.197	2.773
Gender of manager	0.773	0.419	0.000	1.000	1018	0.000	1.000	1.000	1.000
Education of manager	9.523	3.082	0.000	15.000	1018	8.000	10.000	11.000	12.000
Registration	0.576	0.494	0.000	1.000	1018	0.000	0.000	1.000	1.000
Industry dummies/line of business									
Trade enterprise	0.392	0.488	0.000	1.000	1018	0.000	0.000	0.000	1.000
Services enterprise	0.197	0.398	0.000	1.000	1018	0.000	0.000	0.000	0.000
Manufacturing, nonagricultural enterprise	0.313	0.464	0.000	1.000	1018	0.000	0.000	0.000	1.000
Agricultural processing enterprise	0.023	0.150	0.000	1.000	1018	0.000	0.000	0.000	0.000
Mining enterprise	0.005	0.072	0.000	1.000	1018	0.000	0.000	0.000	0.000
Mixed enterprise	0.069	0.254	0.000	1.000	1018	0.000	0.000	0.000	0.000
Community characteristics									
Agricultural seasonality	0.739	0.316	0.000	1.379	1018	0.504	0.643	0.793	0.998
Enterprise density	52.298	34.753	4.458	172.308	1018	23.182	36.496	54.293	78.604
Community pop. size#	7.273	0.510	6.326	8.631	1018	6.801	7.091	7.321	7.689
Paddy land per capita	0.111	0.153	0.000	0.917	1018	0.016	0.044	0.098	0.157
Illiteracy	0.111	0.107	0.000	0.520	1018	0.020	0.045	0.100	0.200
Main market in:									
Own community	0.239	0.426	0.000	1.000	1018	0.000	0.000	0.000	1.000
Neighbor communities	0.189	0.392	0.000	1.000	1018	0.000	0.000	0.000	0.000
Commercial center	0.187	0.390	0.000	1.000	1018	0.000	0.000	0.000	0.000
Nearest city	0.385	0.487	0.000	1.000	1018	0.000	0.000	0.000	1.000
Main community income from:									
Agriculture	0.484	0.500	0.000	1.000	1018	0.000	0.000	1.000	1.000
Wages	0.390	0.488	0.000	1.000	1018	0.000	0.000	0.000	1.000
Self-employment	0.127	0.333	0.000	1.000	1018	0.000	0.000	0.000	0.000
Benchmark indicators and components									
Connectivity	0.457	0.151	0.051	0.770	1018	0.317	0.440	0.497	0.581
Access	0.363	0.142	0.055	0.741	1018	0.230	0.320	0.395	0.496
Business services	0.180	0.235	0.000	1.000	1018	0.000	0.000	0.143	0.286
Governance	0.664	0.042	0.570	0.756	1018	0.628	0.657	0.678	0.702
Human capital	0.352	0.079	0.133	0.577	1018	0.281	0.341	0.377	0.416
Finance services	0.532	0.138	0.126	0.842	1018	0.406	0.498	0.563	0.658

Source: RIC Surveys.

Notes: See Annex A for definitions and description of data; # logarithmic value.

Table B.4 Sri Lanka: Selected Variables of Enterprises and Communities; Value and Distribution

Variable name	mean	st dev	min	max	obs	pct20	pct40	pct60	pct80
Nonfactor cost	3768	16603	0	471421	1018	149	460	1090	2818
Sales	6588	41584	83	1864962	1018	456	984	2072	4973
Depreciation	361	6216	0	459195	1018	12	41	89	249
Gross value added	2819	33674	−20933	1847887	1018	154	368	746	1792
Net value added	2458	28636	−22094	1388692	1018	75	262	613	1579
Family labor input	298	141	39	1248	1018	176	312	312	312
Paid labor cost	548	6936	0	497323	1018	0	0	0	373
Total labor input	846	6932	65	497479	1018	312	312	312	676
Net value of assets	8192	100226	−10	5790863	1018	470	1126	2197	5714
Imputed capital cost	1073	13931	0	868319	1018	64	151	280	743
Paid capital costs	71	708	0	31083	1018	0	0	0	37
Value of capital input	1144	13972	0	868319	1018	69	171	348	855
Total factor input	1990	19008	110	889432	1018	383	520	791	1478
Total cost	6119	33095	150	1346158	1018	674	1302	2241	4840
Net profit	468	14556	−178701	960042	1018	−679	−300	−68	390
Profitability	1.052	1.314	0.022	17.550	1018	0.536	0.771	0.948	1.166
Net factor productivity	1.293	4.548	−13.199	97.699	1018	0.151	0.462	0.855	1.549
Net Income	1839	24013	−92878	1387325	1018	31	188	427	1113
Total liability	439	2103	0	56985	1018	0	0	21	259
Total assets	8631	100482	5	5790863	1018	567	1259	2487	6175
Employment	651	5332	50	382676	1018	240	240	240	520
Imputed factor costs	1371	13929	106	868520	1018	339	449	640	1061
Profitability of own resources	2.052	7.643	−21.485	357.622	1018	0.271	0.635	1.204	2.325
Education of manager	9.523	3.082	0.000	15.000	1018	8.000	10.000	11.000	12.000
Gender of manager	0.773	0.419	0.000	1.000	1018	0.000	1.000	1.000	1.000
Ethnicity of manager	0.928	0.259	0.000	1.000	1018	1.000	1.000	1.000	1.000
Age of enterprise	8.745	11.093	0.000	83.000	1018	1.000	3.000	6.000	15.000
Bank loan > 50% of debt	0.312	0.463	0.000	1.000	1018	0.000	0.000	0.000	1.000
Standalone	0.648	0.478	0.000	1.000	1018	0.000	1.000	1.000	1.000
Ent with paid labor	0.317	0.465	0.000	1.000	1018	0.000	0.000	0.000	1.000
Enterprises with debt	0.419	0.493	0.000	1.000	1018	0.000	0.000	1.000	1.000
Community pop size	1653	952	558	5600	1018	898	1200	1510	2183

Source: RIC Surveys.

Notes: See Annex A for definitions and description of data.

Table B.5 Tanzania: Variables Used in Regressions of Enterprise Performance; Value and Distribution

Variable name	mean	st dev	min	max	obs	pct20	pct40	pct60	pct80
Sales#	6.309	1.119	4.310	9.818	947	5.267	5.909	6.358	7.229
Net value added#	5.412	1.535	−2.506	9.664	807	4.400	5.078	5.741	6.491
Total labor input#	5.317	0.588	2.864	8.741	947	4.849	5.161	5.485	5.521
Value of capital input#	3.965	2.185	0.000	10.167	947	2.065	3.690	4.735	5.823
lnVLL (ln*ln)	28.620	6.636	8.202	76.408	947	23.517	26.636	30.083	30.486
lnVLC (ln*ln)	21.218	12.570	0.000	88.021	947	10.843	18.575	24.434	31.647
lnVCC (ln*ln)	20.494	17.160	0.000	103.375	947	4.266	13.614	22.424	33.902
Depreciation#	2.365	2.050	0.000	7.715	947	0.000	1.542	2.964	4.280
Non Factor Cost#	5.044	2.163	0.000	10.534	947	4.106	5.057	5.717	6.523
Other enterprise characteristics									
Age of enterprise#	1.841	0.955	0.000	4.007	947	1.099	1.609	2.079	2.708
Experience of manager#	1.606	0.620	0.693	3.970	947	1.099	1.386	1.838	1.946
Gender of manager	0.782	0.413	0.000	1.000	947	0.000	1.000	1.000	1.000
Education of manager	7.844	2.761	1.000	17.000	947	7.000	7.000	7.000	8.000
Registration	0.202	0.401	0.000	1.000	947	0.000	0.000	0.000	1.000
Industry dummies/line of business									
Trade enterprise	0.497	0.500	0.000	1.000	947	0.000	0.000	1.000	1.000
Services enterprise	0.184	0.387	0.000	1.000	947	0.000	0.000	0.000	0.000
Manufacturing, nonagricultural enterprise	0.051	0.219	0.000	1.000	947	0.000	0.000	0.000	0.000
Agricultural Processing enterprise	0.038	0.191	0.000	1.000	947	0.000	0.000	0.000	0.000
Mining, gas, etc., construction enterprise	0.024	0.154	0.000	1.000	947	0.000	0.000	0.000	0.000
Mixed enterprise	0.206	0.404	0.000	1.000	947	0.000	0.000	0.000	1.000
Community characteristics									
Community population size#	8.057	0.842	5.820	9.621	947	7.358	7.781	8.243	8.923
Agricultural land per capita	0.596	0.724	0.000	3.901	947	0.042	0.216	0.437	1.013
Illiteracy	27.600	22.644	1.000	100.000	947	8.000	15.000	27.500	49.500
Main community income from:									
Agriculture	0.844	0.363	0.000	1.000	947	1.000	1.000	1.000	1.000
Wages	0.136	0.343	0.000	1.000	947	0.000	0.000	0.000	0.000
Self-employment	0.020	0.140	0.000	1.000	947	0.000	0.000	0.000	0.000
Benchmark indicators and components									
Connectivity	0.214	0.143	0.008	0.560	947	0.082	0.148	0.231	0.342
Access	0.195	0.152	0.000	0.863	947	0.038	0.128	0.231	0.331
Business services	0.061	0.146	0.000	0.857	947	0.000	0.000	0.000	0.143
Governance	0.493	0.076	0.250	0.667	947	0.442	0.483	0.514	0.557
Human capital	0.215	0.056	0.092	0.379	947	0.170	0.195	0.220	0.257
Finance services	0.380	0.194	0.000	0.938	947	0.206	0.324	0.440	0.554

Source: RIC Surveys.

Notes: See Annex A for definitions and description of data; # logarithmic value.

Table B.6 Tanzania: Selected Variables of Enterprises and Communities; Value and Distribution

Variable name	mean	st dev	min	max	obs	pct20	pct40	pct60	pct80
Non Factor Cost	728	2060	0	37562	947	60	156	303	680
Sales	1186	2153	73	18368	947	193	367	576	1378
Depreciation	76	205	0	2241	947	0	4	18	71
Gross value added	456	2230	−37382	15769	947	55	138	276	620
Net value added	381	2230	−37382	15732	947	28	110	241	551
Family labor input	206	121	0	1200	947	125	155	240	240
Paid labor input	52	362	0	6254	947	0	0	0	0
Total labor input	258	374	17	6254	947	127	173	240	249
Net value of assets	1100	8720	−80175	115350	947	23	174	551	1745
Imputed capital cost	276	1042	0	19609	947	3	21	70	231
Paid capital costs	82	468	0	6612	947	0	0	0	33
Value of capital input	358	1269	0	26038	947	7	39	113	337
Total Factor Input	616	1453	33	31789	947	183	250	340	643
Total Cost	1420	2899	55	42717	947	319	492	787	1581
Net profit	−234	2867	−42597	15414	947	−383	−162	−47	184
Profitability	1.202	2.579	0.003	61.226	947	0.436	0.653	0.917	1.282
Net factor productivity	1.227	7.077	−155.76	61.226	947	0.074	0.386	0.783	1.693
Net income	247	2409	−37382	15732	947	−1	90	198	487
Total liability	948	5983	0	81737	947	0	0	0	12
Total assets	2020	6566	0	115350	947	46	211	643	1837
Employment	258	374	17	6254	947	127	173	240	249
Imputed factor costs	483	1050	0	19849	947	163	240	287	510
Profitability of own resources	1.303	8.134	−155.75	61.226	939	0.125	0.440	0.895	1.963
Education of manager	7.844	2.761	1.000	17.000	947	7.000	7.000	7.000	8.000
Gender of manager	0.782	0.413	0.000	1.000	947	0.000	1.000	1.000	1.000
Ethnicity of manager	0.218	0.413	0.000	1.000	947	0.000	0.000	0.000	1.000
Age of enterprise	8.522	8.851	0.000	54.000	947	2.000	4.000	7.000	14.000
Bank loan > 50% of debt	0.008	0.092	0.000	1.000	947	0.000	0.000	0.000	0.000
Standalone	0.365	0.482	0.000	1.000	947	0.000	0.000	0.000	1.000
Ent with paid labor	0.139	0.346	0.000	1.000	947	0.000	0.000	0.000	0.000
Enterprises with debt	0.233	0.423	0.000	1.000	947	0.000	0.000	0.000	1.000
Community pop size	4407	3639	336	15070	947	1568	2394	3800	7500

Source: RIC Surveys.

Notes: See Annex A for definitions and description of data.

Annex C. Enterprise Performance Regressions; Notes and Tables

Specification of enterprise performance regression models. Enterprise performance is measured by total productivity and net productivity. Total productivity is measured on gross output, which is equivalent to sales; net productivity is measured on net value added (NVA). The primary empirical specification uses the Cobb-Douglas production functions on sales and net value added (NVA) as a starting point; this is loglinear in output and inputs, augmented with variables hypothesized to enhance enterprise productivity. Throughout the report reference is made to *sales variants* and *NVA variants*. Due to inherent interest of the contribution of large enterprises to the rural economy, a particular question arises about describing potential size differences. Employing a dummy for larger enterprises is a poor strategy, as the size threshold to be used must be preselected and in any case represents something in the nature of the enterprises that should be described by variables that can be more meaningfully interpreted. Instead, quadratic and interactive specifications of the factor input variables *labor* and *capital* can better describe differences related to size than can log linear forms, and examination of preliminary regression results indicates that interactive specifications significantly improved the statistical fit.[70]

The aim in formulating the regression models is to choose variables that describe enterprise characteristics and the community environment and to avoid variables that can be seen as largely endogenous. Since the main objective of the analysis is to shed light on the contribution of the investment climate, three stepwise regressions are carried out. This reveals the structure of the data sets and illustrates the variance captured by enterprise variables, general community variables, and benchmark indicators. In addition, a fourth variant is estimated in which the benchmark indicators are replaced by some of their components with the highest individual explanatory power. Moreover, a fifth regression variant separates imputed and paid labor and capital, as mentioned above. Table C.1 provides an overview of the regression variants. Annex A, Table A.2 provides full specification of the variables used. Descriptive statistics of variables used are listed in Annex B, Tables B.1 to B.6. Regressions analyzing sales are run with a data set of enterprises with sales of at least US$100 or net value added of at least US$60; regressions exploring net value added use samples of enterprises with net value added of at least US$60.

The theoretical justification for the selection of variables is as follows. The factor inputs—labor and capital—correspond to the net value added output concept, and the factor and nonfactor inputs and depreciation correspond to sales (gross output concept). Depreciation was selected as a separate variable since it differs for the capital components of land and fixed assets.[71] The use of capital is perhaps the most endogenous factor on the right hand side. Own capital will depend on the owner's family history and level of income in the past, which is not easily captured in the enterprise data set. Borrowed capital, however, will depend on other endogenous variables such as solvability, profit level, and size of enterprise (for which we have data), and on supply of availability of sources approximated by the finance services index and perhaps other benchmark indicators. It might be worthwhile to pursue this in a subsequent study.

In addition to these production factors, a number of enterprise characteristics are considered that can add to gross and net productivity. Enterprise characteristics include enterprise age; entrepreneur experience, gender, and education; enterprise registration status; and line of business characterizing enterprise operations.[72] The contribution of registration status to productivity is not clear upfront

Table C.1 Specification of Regression Variants

Variants	(1)		(2)		(3)		(4)		(5)	
Variable to be explained	**Sales**	**NVA**	**Sales**	**NVA**	**Sales**	**NVA**	**Sales**	**NVA**	**Sales**	**NVA**
Specification of variables	Benchmark indicators (BI) with control of industry mix		BI with control of industry mix and enterprise variables		BI with control of industry mix, enterprise and community variables		BI and BI-components with control of industry mix, enterprise variables, and community variables		BI and BI-components with control of industry mix, enterprise variables with separation of imputed and paid labor, and capital and community variables	
Homogeneous labo and capital input			X		X		X			
Separation of imputed and paid labor and capital									X	
Other enterprise variables			X		X		X		X	
Industry dummies	X		X		X		X		X	
General community variables					X		X		X	
BI	X		X		X					
BI and BI components							X		X	

Source: This report.

since it may differ with policies and size. In some cases, registration may be a necessary condition for access to public and infrastructural services; in other cases, public officers may actively enforce registration even if there are only costs and no benefits. Most very small enterprises may not gain any benefit from registration and may remain under the radar of registration officers.[73] Bribery may play a role as well. Dummy variables for enterprise line of business or sector capture differences in productivity not explained by other variables.[74]

Community characteristics present direct incentives and disincentives for enterprises and, hence, may affect their productivity. For Nicaragua and Sri Lanka, data were collected about seasonality in agricultural labor use. This variable may indicate variation in the efficiency of labor and capital use between locations. On the one hand, in highly seasonal locations, enterprise activity may be secondary to the household's main pursuits; on the other hand, highly seasonal fluctuations in demand cause both highs and lows in production activity, increasing downtime and reducing efficiency. For Nicaragua and Sri Lanka, the number of enterprises per community is available and was used to compute enterprise density per 1,000 inhabitants. A higher density could well result in higher productivity due to a conglomeration effect and lower transaction costs. Community population size may be an indicator for the degree of urbanization in which the enterprise operates. Population size may also capture the quality of the economic and institutional infrastructure to the degree that these are not directly measured by other community variables in the regression model. The variable may not be a good indicator of urbanization in all

cases, however, as a community's geographic size can contribute to its population size; the pilot survey designs did not specifically define what could constitute a community.

Area of agricultural land cultivated per capita can affect a community's enterprise productivity. High areas per capita indicate relatively important lowland and upland agriculture and perhaps also low population density. A distinction between lowlands and uplands makes much sense in monsoon Asian countries, but less in many other developing countries. Paddy land (or irrigated lowland) may indicate high population densities and high productivity per unit of land. The quality of land data appears to be poor, allowing inclusion in the regression analysis of only paddy land per capita in Sri Lanka and total cultivated land per capita in Nicaragua and Tanzania.

The rate of illiteracy is included as a variable in the regression model, since it can affect enterprise performance. It may reflect the quality of labor enterprises can employ, but the indirect effect may be more important: illiteracy may be a proxy for the general level of development in a community or for a less favorable business environment.[75]

The Sri Lanka questionnaire includes a question about where households from the community buy and sell most of their goods, that is, in the community, a neighboring community, a commercial center, or the nearest city. The answers are used to specify dummies that to some extent at least indicate the community's degree of commercial openness. In Sri Lanka and Tanzania, questions were asked about the main source of income in the community, that is, agriculture, wages, or nonfarm enterprises. This says something about the types of production in which local enterprises engage. It may also capture some difference in income between communities. Dummies were created for these communities since they can explain differences in productivity.

All regressions models incorporate benchmark indicators, which are community-level composite indexes of characteristics of location, availability of infrastructure, utilities, public and private services, governance and corruption, human capital, and finance services, as described in Chapter 2 above. Benchmark indicators and their components describe both the enterprise environment at the community level and the investment climate. Many can be directly affected by policy interventions and public investments, but some (for example, distance to urban areas) are likely important for enterprises but not easily affected by economic policy. The benchmark indicators have certainly a degree of exogeneity, but questions of self-selectivity remain. The environment described by the benchmark indicators may determine whether enterprises are established and which investments are made. This study, however, will not pursue this issue.[76]

Some variables sometimes included in enterprise performance models are intentionally omitted here. Regional location, for example, is omitted since the regression model includes many variables that describe the community environment—precisely what regional dummy variables attempt to approximate. Similarly, price variables are omitted. The surveys measure cross-regional variation in the price of male daily labor and of commodities such as staple foods, Coca-Cola, and kerosene/diesel. The price of male labor could indicate labor scarcity/productivity. Community prices for staple foods, Coca-Cola,[47] and kerosene could reflect margin differences related to transport costs, value added, and price competition. These variables of course also reflect price variations inappropriate in a production function, however, and the regression model already includes measures of infrastructure services and connectivity that are proxies for determinants of margins. Chapter 6 explores the relation between benchmarks and prices.

Regression results. A dominant characteristic of the enterprise data sets is that by far most of the variance is explained by enterprise characteristics and that industry and community levels explain only a small part of it.[78] This is illustrated in Table C.2 using the results of the first four regression variants for the three countries; the fifth is similar to the fourth but has a different specification, with imputed and paid labor and capital separated. The contribution of community variables will be further analyzed later on. The regressions on benchmark indicators and enterprise dummies for the control of industry differences explain only between 3.6 and 9.3 percent of the variance. When enterprise variables are added, the explanation jumps by a factor of five to eight. General community variables explain 0.2 to 1.8 percent. Replacing the benchmark indicators with those of their components that add most increases explanation of the regression by another 0.5 to 2.0 percent.

Table C.2 Percentage of Variation Explained in Enterprise Performance Regressions

	Nicaragua		Sri Lanka		Tanzania	
	Sales	NVA	Sales	NVA	Sales	NVA
Numbers of observations	846	814	1018	841	947	699
Regression variants						
1. Benchmark indicators (BI) with control of industry mix	6.7	5.9	9.3	6.3	3.6	3.9
2. BI with control of industry mix and enterprise variables	42.6	33.3	74.1	50.4	33.9	19.4
3. BI with control of industry mix, other community variables, and enterprise variables	42.8	33.6	75.0	52.2	34.2	19.9
4. Mix of BI and BI components with control of other community variables, industry mix, and enterprise variables (preferred model)	44.1	35.3	75.7	54.2	36.0	20.4
5. Mix of BI and BI-components with control of industry mix, enterprise variables with separation of imputed and paid labor, and capital and community variables	46.5	38.2	76.2	53.6	35.5	21.4

Source: Table C.4 to Table C.9.

Note: The percentage of variation explained is measured by means of a pseudo-R^2, since the regression model is not estimated with Ordinary Least Squares but rather with a weighted random effects model. The pseudo-R^2 is computed as the squared correlation between the observed and predicted values of the dependent variables.

The NVA regressions have generally a one-third lower explanation in Sri Lanka and Tanzania; in Nicaragua the difference is smaller. When imputed and paid labor and capital are separated (regression variant 5), the explained variation increases by a few percent only, and for Sri Lanka's NVA and the Tanzania's sales model it even decreases.

Sales regressions: Sri Lanka. Since the Sri Lanka regressions have more variables than do the others, they will be discussed first. The estimated coefficients for enterprise characteristics are fairly stable in the three relevant regression variants (2) through (4), indicating that community variables do not have much effect on the parameter estimates. This isn't surprising, given the relatively small share of variance explained by community variables. Quadratic and interaction terms for logarithmic values of labor and capital dominate, with high significant levels in the regressions, and the coefficients for the log linear variables are not significant. In the variant with separated variables for imputed and paid labor and capital, variant (5), all factor input variables are significant except the loglinear and quadratic variables for paid capital. Interaction terms were not significant. The estimated coefficients for depreciation show negative signs but are not significant. Nonfactor cost has robust, highly significant coefficients of almost 0.50. For variant (4), the implied returns to scale at mean input levels equals 0.879 and rises with increasing levels of labor and capital.

The log of enterprise age, with a coefficient of about 0.12, contributes significantly to productivity. The log of manager experience has a coefficient of about 0.07 points in the same direction for the variants (2) through (4), but its significance is just at or below 10 percent. These findings suggest that learning by doing and innovation are important for survival and increased productivity. The dummy for male entrepreneurship shows a positive coefficient, but its significance is below the 10 percent level, except for variant (5). This indicates a tendency toward higher productivity among male-managed enterprises, although the statistical evidence is weak. The manager's years of education contributes significantly to productivity; each year of education adds about 2.5 to 3.0 percent to productivity. Enterprise registration status does not contribute to explanations of productivity. Some literature argues that registered enterprises have higher productivity because they receive more public services; others maintain that especially for smaller enterprises the cost of registration may outweigh the benefits; leading them to avoid registration. While simultaneously considering many other

characteristics, our findings don't support either of these views.

Industry dummies in the variants (2) through (4) reveal that services have about 20 percent higher productivity than does trade (which is included in the intercept), but for variant (5) no significant contribution was found. For the other sectors no significant differences are found that cannot be explained by enterprise and community characteristics.

Community characteristics are less stable in the regressions than are enterprise characteristics, in part because they are more interrelated among themselves and with the enterprise and industry variables, and also because they make a relatively weak contribution to overall explanations of variance. But, given the IC focus, they are in a sense the core of this study. As already seen, the contribution of benchmark indicators is weak. The business services index has in all specifications a positive and significant contribution. The connectivity and governance indexes are significant only in regression variant (1), where they apparently capture explanations contributed to other variables in the other regressions. As already noted, in variants (4) and (5) benchmark indicators are replaced by their best contributing subindexes. Among the connectivity subindexes, cost of transportation to the nearest city is positive and significant at about the 10 percent level. One infrastructure services component is significant: concrete or asphalt roads, which, surprisingly, has a negative sign. The governance subindicator of dealing with government services has a positive sign and a significance level around 10 percent. A second governance component, contract resolution and contract enforcement, has a negative effect, but a parameter value below 10 percent significance. The human capital index has no subindexes; its parameter estimates are sensitive to differences in model specifications but are in all cases without statistical significance: in essence, it has no effect. The finance services subindicator for loan access is significant. This subcomponent, however, is somewhat suspect. Aggregated from all enterprises in the community, it could well reflect a self-selectivity: more productive enterprises gain better access to loans.

Some of the other community characteristics show interesting results. Enterprises in communities with high agricultural seasonality in labor use have lower productivity; the effect is statistically not quite significant, but it is rather robust across model variants. The likely explanation for this is underuse of off-season capital and labor. Enterprise density, community population size, and illiteracy have unstable parameter estimates. In all regressions, the area of paddy land per capita is positively related to productivity, but does not reach the 10 percent significance level. Three dummies indicate where households buy and sell most of their goods, that is, in the community, a neighboring community, a commercial center, or the nearest city. These dummies indicate the community's degree of commercial openness. Enterprises in communities selling primarily in a neighboring community or commercial center have a 25 to 35 percent higher productivity. Surprisingly, for the 38 percent of communities selling in the nearest city, no significantly higher productivity was found. Also dummies for communities with a main source of income from wages (variant (3)) or from nonfarm enterprises (variants (4) and (5)) suggest a 15 and 20 percent higher productivity, respectively, although in most cases below the 10 percent significance level.

NVA regressions Sri Lanka. The results from the NVA and sales regressions show similarities and differences. In the NVA variants (2) through (4), all labor and capital variables are significant. For the variant (5) with separation of imputed and paid capital and labor, only the loglinear and quadratic variables for paid capital are significant. With R^2 lower than for (4), this is not the most attractive variant. Returns to scale for variant (4) are estimated at 0.574 when evaluated at mean values of log-labor and log-capital, rising with increasing levels of labor and capital.

Enterprise age is again a significant variable in explaining productivity, but not entrepreneur experience and gender. Manager education strongly contributes to higher productivity, suggesting that each additional year of education increases productivity by about 5 percent.

For NVA no industry dummy is significant in the regressions with enterprise variables. Among the community variables, only the dummies for marketing in the neighboring community and in the commercial centre are significant. Agricultural seasonality has a negative effect of a magnitude similar to the sales regression model. Income source does not matter to enterprise net value added.

Among the benchmark indicators, business services contributes significantly to higher productivity in all variants; the other benchmarks don't contribute significantly. The benchmark indicator components add more. The connectivity subindexes for transportation costs to the nearest city and time taken by the main means of transportation to the main market have significant coefficients, although with different signs. The access dummy for asphalt and concrete roads has a negative sign but is not quite significant at 10 percent. Governance and human capital don't contribute significantly. For the finance services benchmark, the loan access subindex makes a significant contribution to explaining higher productivity.

Sales regressions for Nicaragua. The parameter estimates for Nicaragua's enterprise variables are stable and consistent. The loglinear and quadratic variables for labor and capital in variants (2) through (4) are all significant; the interaction variable is not. For variant (4), the implied returns to scale at mean input levels is quite low at 0.624, but it increases in labor and capital. In variant (5) with separated variables for imputed and paid labor and capital, virtually all labor and capital variables have significant coefficients. Depreciation and nonfactor input have strongly significant contributions, although the parameter values for depreciation and nonfactor input of about 6 to 8 percent differ surprisingly from their 3.5 and 15 percent of the share in sales. Firm age has a strongly significant contribution to total productivity. Data on the top manager's experience, gender, and education were unfortunately not collected for Nicaragua. Enterprise registration status does not affect productivity in any of the regression variants.

Nicaragua shows significant differences in productivity between sectors, with enterprises in services exhibiting 11 percent lower productivity than trade enterprises, and manufacturing nonagricultural products and mixed enterprises having 10 and 18 percent higher productivity, respectively.

Nicaragua's community variables are limited compared to Sri Lanka's. It lacks the measure of commercial openness and income source, and their contribution differs much with model specification. Only for variant (3) are significant parameter estimates found: enterprise density and community population size.[79] The benchmark indicator for connectivity has a significant positive parameter value; the effect of human capital tends to be negative, becoming significant in variants (4) and (5). Infrastructure services, business services and governance don't contribute significantly. Even if an aggregate index with no effect, however, may have subindexes with an effect. This is the case for Nicaragua: except for governance (human capital has no subindexes) and finance services, many subindexes contribute significantly to productivity differences:

c1 Inverse time taken by main means of transportation to nearest major city (in minutes)

c2 Inverse cost of transportation to the nearest major city

c4 Inverse cost of transportation to the main market (public transportation)

c5 Proximity to the main post office

a1 Percentage of households with electricity

a2 Availability of electricity (such as lack of disruptions)

a6 Garbage collection or disposal service in the community

d7 Information technology services available for businesses in the community

In the cases of c2 and a2, the parameter signs are negative, contrary to the expectations for an index variable or dummy ranging between zero and one, with the value one assumed optimal for enterprise development.

NVA regressions for Nicaragua. Tanzania's NVA regressions show many similarities and some differences. Almost all labor and capital variables are significant in variants (2) through (5) and imply roughly the same returns to scale as the sales regressions. Enterprise age has a highly significant contribution of 9 percent to the explanation of productivity, virtually identical to the sales regression, but, contrary to the sales regressions, here registration status also contributes strongly, suggesting a 10 percent gain in productivity. The direction of causality is not clear since no significance was found in the sales regression. Perhaps more productive enterprises with high net income register for tax purposes.

The findings of differences between sectors are similar to the sales regression, with, relative to trading, lower productivity levels for services and higher levels for mixed enterprises. Also similarly, enterprise density yields a significant parameter

value only in variant (3). As for other community variables, the sign and magnitude of their estimated effect always corresponds with the sales regression results, but none is statistically significant, with the exception of illiteracy in variant (4), where a higher share of illiteracy contributes to lower productivity.

Connectivity contributes to higher productivity; human capital has a negative sign that becomes significant in variants (4) and (5), contrary to expectations. Again, several subindicators have significant signs. These are mostly the same as for the sales regressions, adding a3 (percent of households with access to protected water, with its negative sign, contrary to expectations) and switching d7 (information technology services) out for d4 (legal services). The contributions of c1, c4, c5, a1, a6, and d4 remain positive and are roughly similar to the sales regression; the contribution of a2 remains negative and of a similar magnitude.

Sales regressions for Tanzania. In Tanzania's sales regressions the variables for factor input are all significant except for the loglinear and quadratic variables of paid capital cost. Nonfactor cost has a highly significant parameter, but depreciation does not contribute to explanation of output. At mean input values, returns to scale equal only 0.464 for variant (4) and decrease in labor and capital. Enterprise age and entrepreneur education do not significantly contribute to productivity in any of the regression variants, although both have the expected positive sign; entrepreneur experience does help. Male led enterprises have about 15 percent higher productivity. Registration contributes about 33 percent higher productivity with a high level of significance.

No significant differences in productivity are found between industries.

None of the community variables contributes significantly to productivity differences. Communities with nonfarm enterprises as the main source of income show a positive tendency for higher productivity, but it is not significant at the 10 percent level.

The contribution of benchmark indicators and subindexes is unstable and levels of significance are in most cases rather low. Connectivity and its component c3 (time taken by the main means of transportation to the main market) are significant with the correct sign in all cases, but in variant (4) and (5) its component c2 (transportation cost to the nearest major city) has the opposite sign. Infrastructure services parameters have a negative sign in all variants, but only in variants (2) and (3) is statistical significance weak; component a5 (presence of a sewage system in the community) is more robustly negative. The business services index fails to exhibit any explanatory power when entered as an aggregate, but the estimated effects of two subcomponents do achieve almost 10 percent significance: legal services in an upward direction and insurance services downward. The governance indicator does not play a role. Human capital has a positive sign in all variants, but its significance never quite reaches the 10 percent level. The finance services indicator has no traction in the variants (1) through (3). In variants (4) and (5), its component fi1 (number of banks in the community) and fi3 (access to loans) exert a little more influence, but the estimates are statistically insignificant and of opposite sign. In all, community variables and benchmark indexes do not matter much to enterprise sales.

NVA regressions for Tanzania. In the NVA regressions the parameter values of most labor and capital input variables are not significant. Only the quadratic form of capital input is significant in variants (2) through (4); returns to scale are estimated at 0.292 and rise with increases in labor and capital. In variant (5) the logarithms of paid labor and imputed capital and their quadratic forms have significant parameters. Multicollinearity among these inputs obscures the relationship: in a specification that omits all quadratic terms, the parameter estimates of the linear terms are positive and significant but actually quite small.[80]

Manager experience has some positive contribution, but the estimated effect does not reach to the 10 percent significance level. Contrary to the sales regression, gender is not associated with differences in net factor productivity, although all variants suggest a 10 percent advantage for male managers. Each additional year of manager education significantly adds 3 to 4 percent to productivity. Registration makes a difference of about 40 percent.

Sectoral differences are not significant. All community variables have the expected sign, but none contributes significantly to explaining differences in productivity. Connectivity adds significantly to productivity in all variants, but none of the other benchmark indicators does. Of the benchmark components, c6 (rail stop within walking distance

of the community) has a positive effect, and d6 (insurance services available for businesses in the community) contributes with a negative sign.

Variation in productivity. Productivity varies in response to many often interrelated factors, some of which may not even be explicitly included in the data set but which may still be influential in the background as unobserved variables. Many explanatory variables may capture tiny parts of total variation, probably often of characteristics referring to limited locations or few enterprises only.

In principle, variation in the explained variable derives from (i) observed enterprise variables, (ii) unobserved enterprise variables, (iii) observed community variables, and (iv) unobserved community variables. By far, most of the variation in productivity in the Nicaragua and Sri Lanka data sets is explained by factors at the enterprise level, (combining (i) and (ii)). Box C.1 provides an

Box C.1 Variance Explained by Enterprise and Community Variables

"How much of the variation in the dependent variable can be explained by location as given by community?" Translated, this means: "How much do community dummy variables explain?" The answer is roughly 20 percent for Nicaragua, 28 percent for Sri Lanka,, and 35 percent for Tanzania.

Of course, some of the explanatory enterprise variables may differ systematically from community to community. In some communities, for example, enterprises may employ more labor and capital or more may be registered. Thus, some of the systematic variation in lnsale and lnNVA may be explained by systematic differences in the enterprise variables.

Therefore, as a follow-up analysis, simple regressions of lnsale and lnNVA on enterprise variables were run (this resulted for Nicaragua in $R^2 = 0.423$ for lnsale and $R^2 = 0.330$ for lnNVA), residuals were retrieved, and a dummy variable model was then run on these residuals. Again, for Nicaragua about 20 percent of the variation not explained by the observable enterprise variables is at the community level. This means that community variables can explain at best $0.2*(1 - 0.423) = 0.115$, which is 11.5 percent of the variation in lnsale, and $0.197*(1 - 0.33) = 0.132$ percent, which is 13.2 percent of the variation in lnNVA. Of course, they actually explain much less, because if community variables would explain 11.5 percent of the variation of lnsale, they would provide a perfect prediction of the community effect.

For Sri Lanka, the maximum for lnsale is much lower, mainly because such a large part of the variance (73.6 percent) is explained by the enterprise variables. By contrast, for Tanzania the variance explained by enterprise variables is relatively low, and the maximum total variance that can be can be explained by community variables is about 27 percent.

Percentage of Variance Explained

	Nicaragua		Sri Lanka		Tanzania	
Variables to be explained*	lnsale	lnNVA	lnsale	lnNVA	lnsale	lnNVA
1. ANOVA on explained variable (%)	20.0	19.7	27.5	29.5	39.6	33.3
2. Regression on enterprise and industry variables (%)	42.3	33.0	73.6	49.5	33.0	18.1
3. ANOVA on regression residuals (%)	19.3	20.4	18.3	22.7	34.9	29.1
4. Maximum possible explanation by community variables (%)	11.5	13.2	7.3	14.7	26.5	27.3
Number of enterprises	846	814	1018	841	947	699
Number of communities	93	93	118	118	136	127

Source: RIC Surveys.

*lnsale and ln NVA are the logarithmic values of sales and net value added.

overview of the break-down of variation in the data sets. Line 1 in the box table indicates the variation in productivity explained by community-level factors (combining (iii) and (iv)). It ranges from 20 percent in Nicaragua to 40 percent in Tanzania. Line 2 shows that observed enterprise variables explain between 18 percent of lnNVA in Tanzania to 74 percent of lnsale in Tanzania. The remainder, residuals of the enterprise regressions, covers variation sources (ii), (iii), and (iv). Line 3 shows that between 18 and 35 percent of this remainder can be explained by community-level factors (that is, (iii) and (iv)). Line 4 shows the maximum share that community variables (source (iii)) can explain when explanation is perfect and source (iv) vanishes. In Nicaragua it is only around 12 percent of total variation in productivity; in Sri Lanka between 7 and 14 percent; and in Tanzania about 27 percent. In practice, the explanation will be much lower.

An important reason for the low share of explanation of variation by community-level factors is that the communities are relatively homogeneous: if nothing else, they are all rural. Moreover, the number of communities is low: the RIC samples cover only 99, 151, and 149, respectively, in Nicaragua, Sri Lanka, and Tanzania, and because of missing data, the enterprise performance regression models are estimated with enterprise data from even fewer communities (93 in Nicaragua, 118 in Sri Lanka, and 136 or 127 in Tanzania). Therefore, while determinants of the variation in productivity across communities may be many, only a limited number can be explored. Additionally, few of these determinants are aggregated from households and enterprises: in some communities, the number of household and enterprise observations is rather small, which affects the reliability of components aggregated from these micro-data.

Employment and capital generation. What conditions generate enterprises with more employment and enterprises with more capital? Or, expressed differently, what explains differences in enterprise size, measured by labor and by capital input, respectively?

Regression models are run to explain employment and capital generation from industry mix enterprise characteristics, benchmark indicators and other community characteristics. The parameter estimates are presented in Table C.10 to Table C.15, a summary of the explanation found is presented in Table C.3. Industry mix and benchmark indicators together can explain 4 to 5 percent of the variation in employment generation. By adding enterprise characteristics the explanation doubles in Nicaragua, triples in Sri Lanka, and

Table C.3 Explanation of Employment and Capital Generation per Enterprise

	Employment generation			Capital generation		
	Nicaragua	Sri Lanka	Tanzania	Nicaragua	Sri Lanka	Tanzania
Number of observations	846	1018	947	846	1018	947
Industry mix and benchmark indicators	5.2	4.0	4.1	5.3	7.9	7.1
Industry mix, benchmark indicators and enterprise characteristics	10.8	13.8	6.7	21.6	29.4	15.4
Industry mix, benchmark indicators, enterprise characteristics and other community characteristics	12.3	17.6	8.6	21.3	30.6	17.8
Ibid. with benchmark indicators replaced by components	13.6	18.6	10.2	24.2	32.9	18.0

Source: Annex C, Table C.10 toTable C.15.

Note: See note to Table C.2.

increases by half in Tanzania. Other community variables add 1.5 percent in Nicaragua, almost 2 percent in Tanzania, and almost 4 percent in Sri Lanka. The higher percentage of explanation in Sri Lanka could well be the result of the larger number of enterprise and community variables than in the other countries. Differences in capital generation can be better explained than can differences in employment generation. Partly, this could be due to greater weaknesses in the data measuring labor input. Enterprise characteristics contribute most to differences; other community variables contribute much less. In Nicaragua adding community variables results in a lower percentage of explanation; for Sri Lanka and Tanzania the explanation increases by 1.2 and 2.4 percent.

Employment parameter estimates. The regression parameter estimates for employment in Nicaragua are nearly all significant at 10 percent. Employment increases significantly with enterprise. Registered enterprises employ 52 percent more labor. Compared to trading enterprises, enterprises in services, manufacturing nonagricultural products, and manufacturing agricultural products provide 17, 9, and 13 percent less employment, whereas mining and multiactivity enterprises have 40 and 52 percent more employees. Enterprises in communities with seasonality and more agricultural land per capita have 17 and 12 percent more employees, suggesting that the largest enterprises in terms of employment are in agricultural areas. Community size contributes significantly. Enterprise density and illiteracy don't contribute much. Connectivity, infrastructure services, and governance make significant contributions, but for infrastructure services and governance the sign is opposite what is expected. Replacing the benchmark indicators by the components providing the best explanation, shows that enterprises hire more workers in communities where more households carry cell phone service (a sign of development) but hire fewer where more households have electric service or a sewage system. Moreover, enterprises are smaller where dealing with government agencies is easier.

In Sri Lanka employment in enterprises relates significantly to male managers, manager education, and enterprise registration. Enterprise age does not play a role; entrepreneur experience does, but significance remains below 10 percent. Service enterprises are 16 percent smaller in employment than are trading enterprises. Agricultural processing, mining, and multiactivity enterprises on the other hand are 54, 95, and 15 percent larger. Enterprises in communities with higher enterprise density and marketing in neighboring communities, commercial centers, and cities have significantly more employees. A high level of income from enterprises or employment has no strong effect. Most surprisingly, in Sri Lanka none of the benchmark indicators contributes significantly to employment per enterprise in any of the regression variants. Only when selecting among the benchmark components do some investment climate elements show up significantly. In particular, secure provision of electricity and wider availability of telephone service are associated with larger enterprises, as are business services in this specification of the model. Among corruption and governance components, infrastructure and services have negative effect.

In Tanzania enterprises with a male top manager and a more educated manager have more employees. Differences between industries in numbers of employees are limited, with agricultural processing 22 percent smaller than trading enterprises and 27 percent bigger. Smaller communities tend to have enterprises with more employees. None of the benchmark indicators and other community variables contributes significantly to explaining differences in employment per enterprise. After replacing benchmark indicators by their components, only three contribute significantly with a positive sign to explaining employment generated per enterprise: proximity to post office, availability of marketing services, and governance components related to dealing with government services and general policy and institutional constraints.

Community unobservables are important in each country, as witnessed by the standard deviation of the community random effect. In Nicaragua, variation of community observables is nearly as large as the idiosyncratic variation at the enterprise level, and in Tanzania it is not far behind. Overall, community effect is largest in Tanzania and idiosyncratic variation most important in Sri Lanka.

Capital generated. In Nicaragua enterprise age and the registration status contribute strongly to the amount of capital invested per enterprise. Enterprises in services and multiactivity enterprises

have most capital invested, and enterprises in processing of agricultural products relatively less. Communities with seasonality have not only more employees but also a higher level of investment per enterprise. Community variables affecting investment are quite different from those affecting employment. Here we find robust significant explanation from the human capital and finance services benchmark indicators. After replacing the benchmark indicators by their components, several subcomponents of the connectivity, infrastructure services, and governance indicators become important. Positive contributors are communities with fixed phone lines, garbage collection, governance on infrastructure and services, and number of bank services. Negative parameter values are found for proximity to main market and to post office and stable electricity supply.

In Sri Lanka investment per enterprises relates significantly with entrepreneur experience but not enterprise age. A male top manager and a more educated manager also contribute to investment. Service enterprises have 40 percent lower levels of investment than trading enterprises, and enterprises in agricultural processing have 68 percent more. Community variables contribute to explaining differences only in the table's last column, listing selected benchmark components: enterprises in larger communities with greater enterprise density use more capital. Benchmark indicators are ineffectual except for business services, in which accounting services contribute. Three other benchmark components have significant parameter estimates. Positive are percentage of access to protected water and households with fixed phone line; negative is roads with asphalt or concrete surface.

In Tanzania registered enterprises have 280 percent[81] more investment. Enterprises with experienced managers and male managers have more investment. Among the industries, service enterprises and multiactivity enterprises have relatively high levels of investment, those in agricultural processing relatively less. Enterprises in communities with a high amount of agricultural land per capita have more investment. Of the benchmark indicators, human capital has a positive, significant parameter estimate. Three subindicators lead to higher investment: more widespread cell phone use, concrete or asphalt roads, and number of banks. Offsetting the latter is the negative effect of the number of bank services.

In conclusion, capital investment seems to be more dependent on enterprise characteristics than is employment generation. The community characteristics and benchmark indicators that explain investment and employment levels differ. Several benchmark subindicators contribute to explanation; however, in other cases the indicators' signs are contrary to expectations derived from theory, most likely because of multicollinearity and unobserved background variables.

Table C.4 Sales Regressions: Nicaragua

Variable name	Benchmark indicators (BI) with control of industry mix	BI with control of industry mix and enterprise variables		BI with control of industry mix and enterprise and community variables		BI and BI-components with control of industry mix, enterprise variables, and community variables		BI and BI-components with control of industry mix, enterprise variables with separation of imputed and paid labor, and capital and community variables	
Variant	(1)	(2)		(3)		(4)		(5)	
N	846	846		846		846		846	
r2	0.067*	0.426*		0.428*		0.441*		0.465*	
sv[a]	0.284	0.286		0.286		0.286		0.288	
smu[a]	0.317	0.246		0.247		0.251		0.242	
Factor and nonfactor inputs									
Total labor input#		−0.225	−4.964*	−0.231	−5.074*	−0.232	−5.110*		
Family labor input#								−0.041	−0.902
Paid labor input#								−0.281	−11.994*
Total capital input#		−0.167	−4.117*	−0.165	−4.072*	−0.164	−4.048*		
Imputed capital cost#								−0.210	−5.915*
Paid capital cost#								−0.105	−4.243*
lnVLL (ln*ln)		0.047	8.998*	0.048	9.128*	0.048	9.199*		
LnIL*lnIL (imputed)								0.024	5.856*
LnPL*lnPL (paid)								0.051	15.133*
lnVCC (ln*ln)		0.024	8.735*	0.024	8.785*	0.024	8.639*		
lnIC*lnIC (imputed)								0.028	9.936*
lnPC*lnPC (paid)								0.032	6.730*
lnVLC (lnL*lnC)		0.010	1.543	0.010	1.460	0.010	1.447		
lnIL*lnIC (imputed)								0.011	2.488*
lnPL*lnPC (paid)								−0.008	−3.384*
Depreciation#		0.060	6.327*	0.060	6.292*	0.061	6.483*	0.073	7.562*
Non Factor Cost#		0.077	12.902*	0.077	12.955*	0.079	13.117*	0.074	12.223*
Other enterprise characteristics									
Age of enterprise#		0.098	7.689*	0.098	7.632*	0.097	7.583*	0.101	7.836*
Registration		−0.044	−1.426	−0.045	−1.461	−0.046	−1.492	0.023	0.734

Industry dummies/line of business										
Services enterprise	−0.198	−5.990*	−0.113	−3.388*	−0.116	−3.482*	−0.111	−3.340*	−0.102	−3.027*
Manufacturing, nonagricultural enterprise	−0.021	−0.471	0.112	2.564*	0.114	2.610*	0.110	2.502*	0.096	2.182*
Agricultural processing enterprise	−0.041	−1.126	0.001	0.034	0.000	−0.002	−0.004	−0.097	−0.040	−1.090
Other production enterprise	0.006	0.092	−0.073	−1.117	−0.072	−1.100	−0.076	−1.164	−0.061	−0.930
Mixed enterprise	0.582	17.552*	0.184	5.313*	0.183	5.288*	0.172	4.967*	0.181	5.178*
Community characteristics										
Agric. seasonality					−0.042	−0.559	0.047	0.616	0.053	0.714
Enterprise density					0.001	2.362*	0.000	0.802	0.000	0.333
Community pop size#					0.068	2.161*	−0.011	−0.298	−0.032	−0.930
Agric land per capita					−0.048	−1.164	−0.035	−0.808	−0.052	−1.244
Illiteracy					0.000	0.155	−0.002	−0.965	−0.001	−0.458
Benchmark indicators and components										
Connectivity	0.873	2.706*	0.811	3.133*	0.673	2.485*				
Time to near city							0.402	3.025*	0.423	3.277*
Cost to near city							−2.000	−3.029*	−2.091	−3.262*
Cost to main market							1.800	2.956*	1.770	2.999*
Distance to post office							0.259	2.267*	0.253	2.280*
Infrastructure service	−0.035	−0.134	−0.060	−0.290	−0.112	−0.508				
Percent with electricity							0.493	3.531*	0.390	2.868*
Cost of transportation							−0.256	−2.525*	−0.213	−2.164*
Garbage collection							0.267	2.662*	0.267	2.738*
Business services	0.079	0.682	0.133	1.439	0.079	0.846				
Technology services							0.160	1.859*	0.186	2.237*
Governance	0.054	0.156	0.419	1.495	0.212	0.747	0.627	1.999*	0.446	1.462
Human capital	0.442	0.792	−0.634	−1.413	−0.590	−1.257	−1.471	−2.829*	−1.170	−2.313*
Finance services	0.071	0.406	−0.083	−0.596	−0.163	−1.041	−0.154	−0.878	−0.193	−1.133
Intercept	6.450	26.831*	5.573	21.956*	5.212	15.198*	5.564	13.962*	5.571	14.313*

Source: RIC Surveys.

Notes: See Annex A for definitions and description of data; # logarithmic value

a. "sv" is the standard deviation of the enterprise-specific disturbance. "smu" is the standard deviation of the community random effect.

"*": significant at least at 10 percent.

Table C.5 Net Value Added Regressions: Nicaragua

Variable name	Benchmark indicators (BI) with control of industry mix		BI with control of industry mix and enterprise variables		BI with control of industry mix and enterprise and community variables		BI and BI-components with control of industry mix, enterprise variables, and community variables		BI and BI-components with control of industry mix, enterprise variables with separation of imputed and paid labor, and capital and community variables	
Variant	(1)		(2)		(3)		(4)		(5)	
N	814		814		814		814		814	
r2	0.059*		0.333*		0.336*		0.353*		0.382*	
sv[a]	0.291		0.292		0.292		0.292		0.294	
smu[a]	0.311		0.300		0.295		0.295		0.273	
Total labor input#			−0.238	−5.076*	−0.242	−5.162*	−0.247	−5.257*		
Family labor input#									0.039	0.815
Paid labor input#									−0.392	−16.031*
Total capital input#			−0.024	−0.560	−0.024	−0.543	−0.023	−0.538		
Imputed capital cost#									−0.094	−2.466*
Paid capital cost#									−0.100	−3.838*
lnVLL (ln*ln)			0.065	11.926*	0.065	12.025*	0.066	12.125*		
LnIL*lnIL (imputed)									0.024	5.364*
LnPL*lnPL (paid)									0.068	19.363*
lnVCC (ln*ln)			0.034	11.619*	0.035	11.695*	0.034	11.559*		
lnIC*lnIC (imputed)									0.029	9.769*
lnPC*lnPC (paid)									0.030	6.047*
lnVLC (lnL*lnC)			−0.022	−3.028*	−0.022	−3.092*	−0.022	−3.056*		
lnIL*lnIC (imputed)									−0.002	−0.416
lnPL*lnPC (paid)									−0.011	−4.270*
Other enterprise characteristics										
Age of enterprise#			0.091	6.788*	0.090	6.721*	0.090	6.695*	0.095	7.074*
Registration			0.108	3.578*	0.108	3.572*	0.110	3.668*	0.209	6.725*
Industry dummies/line of business										
Services enterprise	−0.225	−6.513*	−0.139	−3.970*	−0.142	−4.062*	−0.139	−3.981*	−0.130	−3.687*
Manufacturing, nonagricultural enterprise	−0.055	−1.224	0.034	0.750	0.036	0.801	0.034	0.762	0.023	0.502
Agricultural processing enterprise	−0.088	−2.339*	−0.005	−0.140	−0.006	−0.149	−0.008	−0.211	−0.052	−1.364
Other production enterprise	−0.102	−1.540	−0.232	−3.454*	−0.231	−3.449*	−0.236	−3.523*	−0.227	−3.370*
Mixed enterprise	0.525	15.158*	0.239	6.695*	0.238	6.679*	0.232	6.505*	0.250	6.919*

Community characteristics										
Agricultural seasonality					−0.018	−0.201	0.046	0.516	0.058	0.695
Enterprise density					0.001	2.476*	0.001	0.894	0.000	0.330
Community population size#					0.067	1.840	−0.024	−0.583	−0.056	−1.458
Agricultural land per capita					−0.046	−0.956	−0.044	−0.846	−0.056	−1.166
Illiteracy					−0.002	−0.831	−0.004	−1.669*	−0.003	−1.358
Benchmark indicators and components										
Connectivity	1.033	3.245*	1.133	3.670*	0.930	2.934*				
Time to near city							0.371	2.415*	0.391	2.715*
Cost to near city							−1.551	−1.998*	−1.625	−2.238*
Cost to main market							1.878	2.688*	1.797	2.756*
Distance post office							0.340	2.558*	0.349	2.810*
Infrastructure service	−0.128	−0.499	−0.264	−1.064	−0.252	−0.981				
Percent with electricity							0.653	3.975*	0.501	3.244*
Availability electricity							−0.356	−3.050*	−0.328	−3.005*
Percent with protected water							−0.299	−2.126*	−0.224	−1.697*
Garbage collection							0.294	2.538*	0.320	2.948*
Business services	0.060	0.525	0.073	0.662	0.021	0.188				
Accounting service							0.154	1.532	0.193	2.050*
Governance	0.183	0.530	0.549	1.641	0.357	1.072	0.824	2.268*	0.588	1.727*
Human capital	0.313	0.568	−0.798	−1.492	−0.725	−1.322	−1.798	−2.902*	−1.481	−2.550*
Finance services	0.000	−0.001	−0.143	−0.855	−0.191	−1.037	−0.189	−0.945	−0.236	−1.266
Intercept	6.297	26.457*	5.210	18.148*	4.970	12.672*	5.519	12.203*	5.554	12.968*

Source: RIC Surveys.

Notes: See Annex A for definitions and description of data; # logarithmic value

a. "sv" is the standard deviation of the enterprise-specific disturbance. "smu" is the standard deviation of the community random effect.

"*": significant at least at 10 percent.

Table C.6 Sales Regressions: Sri Lanka

Variable name	Benchmark indicators (BI) with control of industry mix		BI with control of industry mix and enterprise variables		BI with control of industry mix and enterprise and community variables		BI and BI-components with control of industry mix, enterprise variables, and community variables		BI and BI-components with control of industry mix, enterprise variables with separation of imputed and paid labor, and capital and community variables	
Variant	(1)		(2)		(3)		(4)		(5)	
N	1018		1018		1018		1018		1018	
r2	0.093		0.741		0.750		0.757		0.762	
sv[a]	1.311		0.737		0.737		0.737		0.729	
smu[a]	0.636		0.262		0.210		0.179		0.179	
Factor and nonfactor inputs										
Total labor input#			−0.309	−1.033	−0.325	−1.094	−0.281	−0.948		
Family labor input#									1.893	1.750*
Paid labor input#									−0.080	−1.861*
Total capital input#			0.095	0.554	0.091	0.532	0.076	0.441		
Imputed capital cost#									−0.251	−2.474*
Paid capital cost#									0.063	1.071
lnVLL (ln*ln)			0.081	2.476*	0.081	2.495*	0.077	2.364*		
LnIL*lnIL (imputed)									−0.167	−1.772*
LnPL*lnPL (paid)									0.024	4.007*
lnVCC (ln*ln)			0.046	3.976*	0.047	4.125*	0.047	4.133*		
lnIC*lnIC (imputed)									0.034	3.841*
lnPC*lnPC (paid)									−0.007	−0.712
lnVLC (lnL*lnC)			−0.073	−2.200*	−0.074	−2.239*	−0.071	−2.165*		
Depreciation#			−0.036	−1.066	−0.039	−1.156	−0.043	−1.279	−0.024	−0.718
Non Factor Cost#			0.474	20.513*	0.473	20.607*	0.479	20.950*	0.471	20.806*
Other enterprise characteristics										
Age of enterprise#			0.111	3.027*	0.116	3.178*	0.118	3.234*	0.127	3.464*
Experience of manager#			0.075	1.808*	0.075	1.791*	0.066	1.585	0.045	1.085
Gender of manager			0.124	1.523	0.117	1.448	0.110	1.376	0.136	1.722*
Education of manager			0.025	2.071*	0.026	2.204*	0.027	2.323*	0.024	2.027*
Registration			0.094	1.158	0.094	1.167	0.089	1.124	0.089	1.134
Industry dummies/line of business										
Services enterprise	−0.605	−3.553*	0.211	2.044*	0.215	2.082*	0.208	2.023*	0.101	0.920
Manufacturing, nonagricultural enterprise	−0.560	−3.903*	0.031	0.343	0.020	0.219	0.021	0.234	−0.053	−0.554
Agricultural processing enterprise	0.316	0.891	0.068	0.332	0.085	0.420	0.086	0.424	0.028	0.141
Other production enterprise	0.483	1.214	−0.068	−0.298	−0.097	−0.427	−0.027	−0.118	−0.085	−0.373
Mixed enterprise	−0.015	−0.062	−0.154	−1.150	−0.122	−0.912	−0.109	−0.820	−0.141	−1.048

Community characteristics										
Agricultural seasonality					−0.131	−0.825	−0.193	−1.349	−0.184	−1.294
Enterprise density					0.000	0.276	0.002	0.992	0.001	0.725
Community population size#					−0.006	−0.069	0.017	0.201	0.019	0.233
Agricultural land per capita					0.412	1.628	0.121	0.489	0.157	0.641
Illiteracy					−0.064	−0.165	0.033	0.087	0.084	0.228
Main market in:										
Neighboring communities					0.369	2.613*	0.353	2.556*	0.326	2.374*
Commercial center					0.108	0.758	0.256	1.809*	0.173	1.232
Nearest city					0.036	0.311	0.112	0.964	0.074	0.642
Main community income from:										
Wages					0.166	1.664*	0.132	1.425	0.144	1.575
Self-employment					0.156	1.028	0.199	1.364	0.220	1.523
Benchmark indicators and components										
Connectivity	1.167	1.669*	0.194	0.573	−0.094	−0.267				
Cost of transportation							0.276	1.497	0.320	1.750*
Infrastructure service	−0.085	−0.112	−0.271	−0.743	−0.198	−0.534				
Concrete/asphalt road							−0.241	−2.601*	−0.217	−2.372*
Business services	1.018	2.426*	0.394	1.968*	0.438	2.192*				
Management consult							0.416	2.521*	0.361	2.190*
Governance	1.055	2.548*	−0.117	−0.115	0.190	0.178				
Conflict resolution and contract enforcement							−0.359	−1.211	−0.258	−0.879
Public services/ institutions							2.420	1.714*	2.187	1.561
Human capital	−1.266	−0.606	−0.083	−0.139	0.126	0.208	−0.152	−0.280	−0.118	−0.220
Finance services	0.139	0.113	0.262	0.732	0.349	1.027				
Access to loans							0.428	2.105*	0.471	2.327*
Intercept	7.669	5.155*	2.890	2.564*	2.634	2.052*	−2.175	−0.685	−6.279	−1.447

Source: RIC Surveys.

Notes: See Annex A for definitions and description of data; # logarithmic value.
a. "sv" is the standard deviation of the enterprise-specific disturbance. "smu" is the standard deviation of the community random effect.
"*": significant at least at 10 percent.

Table C.7 Net Value Added Regressions: Sri Lanka

Variable name	Benchmark indicators (BI) with control of industry mix		BI with control of industry mix and enterprise variables		BI with control of industry mix and enterprise and community variables		BI and BI-components with control of industry mix, enterprise variables, and community variables		BI and BI-components with control of industry mix, enterprise variables with separation of imputed and paid labor, and capital and community variables	
Variant	(1)		(2)		(3)		(4)		(5)	
N	841		841		841		841		841	
r2	0.063		0.504		0.522		0.542		0.536	
sv[a]	1.258		0.950		0.950		0.950		0.965	
smu[a]	0.652		0.378		0.323		0.243		0.235	
Factor and nonfactor inputs										
Total labor input#			−0.840	−1.956*	−0.946	−2.218*	−0.912	−2.158*		
Family labor input#									1.523	0.910
Paid labor input#									−0.212	−3.431*
Total capital input#			0.951	3.960*	0.953	3.958*	0.927	3.879*		
Imputed capital cost#									0.368	0.977
Paid capital cost#									−0.031	−0.370
lnVLL (ln*ln)			0.211	4.406*	0.218	4.551*	0.213	4.498*		
LnIL*lnIL (imputed)									−0.103	−0.718
LnPL*lnPL (paid)									0.054	6.034*
lnVCC (ln*ln)			0.057	3.138*	0.059	3.213*	0.059	3.293*		
lnIC*lnIC (imputed)									0.015	1.110
lnPC*lnPC (paid)									0.020	1.229
lnVLC (lnL*lnC)			−0.215	−4.289*	−0.217	−4.327*	−0.214	−4.313*		
lnIL*lnIC (imputed)									−0.059	−0.981
lnPL*lnPC (paid)									−0.012	−1.559
Other enterprise characteristics										
Age of enterprise#			0.096	1.836*	0.095	1.825*	0.102	1.962*	0.119	2.219*
Experience of manager#			0.055	0.918	0.060	1.006	0.040	0.679	0.018	0.299
Gender of manager			0.026	0.223	0.025	0.213	0.032	0.278	0.076	0.644
Education of manager			0.051	2.968*	0.050	2.903*	0.053	3.087*	0.049	2.813*
Registration			0.062	0.546	0.064	0.567	0.073	0.659	0.105	0.928
Industry dummies/line of business										
Services enterprise	0.082	0.454	0.159	1.157	0.148	1.072	0.128	0.926	−0.028	−0.182
Manufacturing, nonagricultural enterprise	−0.133	−0.882	−0.070	−0.582	−0.085	−0.705	−0.081	−0.671	−0.173	−1.307
Agricultural processing enterprise	1.003	2.766*	0.082	0.294	0.106	0.377	0.094	0.334	0.037	0.128
Other production enterprise	1.578	3.821*	0.093	0.291	0.086	0.270	0.155	0.488	0.206	0.626
Mixed enterprise	0.045	0.177	−0.231	−1.201	−0.207	−1.075	−0.191	−1.007	−0.250	−1.261

Community characteristics										
Agricultural seasonality					−0.166	−0.709	−0.235	−1.107	−0.211	−0.983
Enterprise density					0.002	0.992	0.003	1.131	0.002	0.850
Community population size#					0.036	0.280	0.011	0.099	0.027	0.233
Agricultural land per capita					0.491	1.259	0.224	0.622	0.280	0.775
Illiteracy					0.113	0.197	0.178	0.333	0.167	0.310
Main market in:										
Neighboring communities					0.542	2.642*	0.462	2.501*	0.457	2.459*
Commercial center					0.294	1.407	0.417	2.225*	0.335	1.784*
Nearest city					0.097	0.563	0.078	0.497	0.042	0.267
Main community income from:										
Wages					0.043	0.295	−0.042	−0.322	−0.017	−0.133
Self-employment					0.166	0.737	0.287	1.360	0.341	1.602
Benchmark indicators and components										
Connectivity	0.831	1.132	0.298	0.611	0.017	0.033				
Cost of transportation							0.700	2.586*	0.723	2.664*
Time to main market							−0.346	−1.815*	−0.337	−1.758*
Infrastructure service	−0.176	−0.220	−0.533	−1.005	−0.337	−0.620				
Concrete/asphalt road							−0.187	−1.434	−0.171	−1.305
Business services	0.862	1.989*	0.526	1.830*	0.587	1.987*	0.484	1.857*	0.440	1.665*
Governance	0.033	0.015	0.207	0.141	0.108	0.069				
Public services/ institutions							2.195	1.109	2.001	1.007
Human capital	0.551	0.429	0.325	0.382	0.319	0.365	−0.340	−0.433	−0.100	−0.128
Finance services	0.214	0.268	0.151	0.289	0.464	0.926				
Access to loans							0.605	2.127*	0.681	2.384*
Intercept	5.657	3.637*	2.709	1.687*	2.506	1.357	−1.699	−0.388	−5.559	−0.840

Source: RIC Surveys.

Notes: See Annex A for definitions and description of data; # logarithmic value.
a. "sv" is the standard deviation of the enterprise-specific disturbance. "smu" is the standard deviation of the community random effect.
"*" — significant at least at 10 percent.

Table C.8 Sales Regressions: Tanzania

Variable name	Benchmark indicators (BI) with control of industry mix	BI with control of industry mix and enterprise variables		BI with control of industry mix and enterprise and community variables		BI and BI-components with control of industry mix, enterprise variables, and community variables		BI and BI-components with control of industry mix, enterprise variables with distinction of imputed and paid labor, and capital and community variables	
Variant	(1)	(2)		(3)		(4)		(5)	
N	947	947		947		947		947	
r2	0.036*	0.339*		0.342*		0.360		0.355	
sv[a]	0.940	0.800		0.800		0.800		0.797	
smu[a]	0.625	0.477		0.476		0.458		0.469	
Factor and nonfactor inputs									
Total labor input#		0.740	1.840*	0.746	1.853*	0.705	1.751*		
Family labor input#								−0.387	−3.284*
Paid labor input#								0.125	1.861*
Total capital input#		0.318	2.923*	0.326	2.992*	0.317	2.914*		
Imputed capital cost#								−0.209	−3.050*
Paid capital cost#								−0.043	−0.932
lnVLL (ln*ln)		−0.024	−0.694	−0.024	−0.683	−0.021	−0.597		
LnIL*lnIL (imputed)								0.043	2.738*
LnPL*lnPL (paid)								−0.011	−0.970
lnVCC (ln*ln)		0.019	3.344*	0.019	3.295*	0.020	3.422*		
lnIC*lnIC (imputed)								0.013	2.164*
lnPC*lnPC (paid)								0.015	1.798*
lnVLC (lnL*lnC)		−0.076	−3.899*	−0.077	−3.960*	−0.077	−3.932*		
lnIL*lnIC (imputed)								0.022	2.086*
lnPL*lnPC (paid)								−0.015	−2.364*
Depreciation#		0.009	0.412	0.009	0.418	0.004	0.167	0.038	1.574
Nonfactor Cost#		0.219	13.680*	0.219	13.613*	0.217	13.517*	0.217	13.484*
Other enterprise characteristics									
Age of enterprise#		0.041	1.292	0.039	1.220	0.040	1.271	0.032	0.997
Experience of manager#		0.087	1.748*	0.091	1.839*	0.097	1.962*	0.092	1.855*
Gender of manager		0.155	2.088*	0.150	2.009*	0.146	1.959*	0.152	2.031*
Education of manager		0.014	1.242	0.014	1.243	0.013	1.145	0.014	1.259
Registration		0.330	4.067*	0.336	4.128*	0.329	4.059*	0.323	3.930*

Industry dummies/line of business										
Services enterprise	−0.011	−0.114	0.079	0.937	0.079	0.933	0.069	0.819	0.042	0.492
Manufactruing, nonagricultural enterprise	−0.162	−0.927	−0.047	−0.318	−0.047	−0.318	−0.027	−0.180	−0.051	−0.342
Agricultural processing enterprise	−0.125	−0.653	0.136	0.825	0.119	0.721	0.133	0.808	0.096	0.582
Other production enterprise	−0.155	−0.706	−0.039	−0.204	−0.036	−0.189	−0.036	−0.186	−0.054	−0.277
Mixed enterprise	0.118	1.182	0.020	0.231	0.022	0.256	0.018	0.212	0.018	0.205
Community characteristics										
Community population size#					0.047	0.680	−0.011	−0.162	−0.003	−0.038
Agricultural land per capita					0.015	0.181	−0.034	−0.426	−0.018	−0.221
Illiteracy					0.001	0.353	0.002	0.884	0.002	0.862
Main community income from:										
Wages					−0.045	−0.124	0.189	0.502	0.127	0.334
Self-employment					0.228	1.406	0.092	0.595	0.070	0.441
Benchmark indicators and components										
Connectivity	0.773	1.410	1.051	2.420*	1.042	2.348*				
Cost of transportation							−0.486	−1.659	−0.473	−1.582
Time to main market							0.244	1.810*	0.251	1.825*
Infrastructure services	−0.341	−0.618	−0.759	−1.730*	−0.878	−1.890*				
Sewage system							−0.496	−2.319*	−0.552	−2.535*
Business services	0.541	1.183	0.032	0.090	−0.021	−0.057				
Legal services							0.347	1.479	0.383	1.596
Insurance services							−0.674	−1.526	−0.686	−1.528
Governance	−0.612	−0.737	−0.270	−0.411	−0.099	−0.148	−0.031	−0.047	0.108	0.158
Human capital	1.955	1.501	1.181	1.136	0.906	0.836	0.736	0.725	0.838	0.812
Finance services	0.409	0.905	0.048	0.132	0.077	0.214				
Number of banks							−0.232	−1.081	−0.239	−1.095
Access to loans							0.268	1.111	0.169	0.688
Intercept	5.939	11.902*	1.134	0.896	0.656	0.473	1.286	0.926	5.148	6.741*

Source: RIC Surveys.

Notes: See Annex A for definitions and description of data; # logarithmic value.
a. "sv" is the standard deviation of the enterprise-specific disturbance. "smu" is the standard deviation of the community random effect.
"*": significant at least at 10 percent.

Table C.9 Net Value Added Regressions: Tanzania

Variable name	Benchmark indicators (BI) with control of industry mix	BI with control of industry mix and enterprise variables		BI with control of industry mix and enterprise and community variables		BI and BI-components with control of industry mix, enterprise variables, and community variables		BI and BI-components with control of industry mix, enterprise variables with distinction of imputed and paid labor, and capital and community variables	
Variant	(1)	(2)		(3)		(4)		(5)	
N	699	699		699		699		699	
r2	0.039*	0.194		0.199		0.204		0.214	
sv[a]	1.046	0.980		0.980		0.980		0.968	
smu[a]	0.559	0.484		0.482		0.489		0.490	
Factor and nonfactor inputs									
Total labor input#		−0.357	−0.530	−0.383	−0.567	−0.291	−0.432		
Family labor input#								−0.148	−0.925
Paid labor input#								−0.190	−1.759*
Total capital input#		0.072	0.420	0.071	0.411	0.091	0.527		
Imputed capital cost#								−0.198	−1.917*
Paid capital cost#								−0.046	−0.672
lnVLL (ln*ln)		0.065	1.121	0.068	1.172	0.061	1.060		
LnIL*lnIL (imputed)								0.018	0.834
LnPL*lnPL (paid)								0.046	2.427*
lnVCC (ln*ln)		0.039	4.446*	0.039	4.451*	0.040	4.617*		
lnIC*lnIC (imputed)								0.032	3.511*
lnPC*lnPC (paid)								0.021	1.673*
lnVLC (lnL*lnC)		−0.045	−1.465	−0.046	−1.472	−0.051	−1.649*		
lnIL*lnIC (imputed)								0.006	0.367
lnPL*lnPC (paid)								0.014	1.335
Other enterprise characteristics									
Age of enterprise#		0.018	0.400	0.012	0.268	0.006	0.129	0.008	0.175
Experience of manager#		0.105	1.482	0.112	1.577	0.110	1.554	0.104	1.485
Gender of manager		0.109	1.037	0.101	0.955	0.094	0.888	0.108	1.022
Education of manager		0.036	2.176*	0.036	2.170*	0.035	2.121*	0.027	1.623*
Registration		0.429	3.750*	0.439	3.827*	0.447	3.895*	0.408	3.549*

Industry dummies/line of business										
Services enterprise	0.083	0.685	−0.085	−0.736	−0.082	−0.716	−0.087	−0.752	−0.139	−1.208
Manufacturing, nonagricultural enterprise	−0.080	−0.364	0.059	0.285	0.068	0.328	0.105	0.505	0.035	0.171
Agricultural processing enterprise	0.037	0.166	0.068	0.324	0.052	0.245	0.080	0.375	0.043	0.204
Other production enterprise	0.191	0.638	0.189	0.662	0.217	0.753	0.222	0.771	0.168	0.588
Mixed enterprise	0.150	1.157	0.047	0.380	0.081	0.644	0.090	0.720	0.075	0.602
Community characteristics										
Community population size#					0.044	0.541	0.051	0.627	0.038	0.459
Agricultural land per capita					0.138	1.382	0.093	0.975	0.088	0.922
Illiteracy					0.001	0.366	0.000	−0.063	0.000	0.097
Main community income from:										
Wages					0.084	0.161	0.696	1.185	0.636	1.089
Self-employment					0.224	1.186	0.226	1.211	0.156	0.827
Benchmark indicators and components										
Connectivity	0.980	1.740*	1.019	2.006*	1.169	2.238*				
Rail stop							0.285	1.921*	0.256	1.721*
Infrastructure services	0.436	0.663	−0.104	−0.175	−0.078	−0.126				
Concrete/asphalt road							−0.140	−0.837	−0.129	−0.773
Business services	−0.038	−0.076	−0.287	−0.641	−0.310	−0.677				
Legal services							0.266	0.903	0.285	0.965
Insurance services							−1.338	−2.323*	−1.316	−2.296*
Governance	−0.385	−0.446	−0.437	−0.565	−0.319	−0.411				
Public service/ institution							−0.312	−0.969	−0.307	−0.956
Human capital	1.114	0.771	−0.039	−0.029	−0.173	−0.126	0.330	0.248	0.142	0.107
Finance services	−0.044	−0.091	−0.158	−0.361	−0.116	−0.265	−0.167	−0.381	−0.294	−0.668
Intercept	5.375	10.193*	5.122	2.489*	4.612	2.149*	4.510	2.117	5.256	6.098*

Source: RIC Surveys.

Notes: See Annex A for definitions and description of data; # logarithmic value.
a. "sv" is the standard deviation of the enterprise-specific disturbance. "smu" is the standard deviation of the community random effect.
"*": significant at least at 10 percent.

Table C.10 Labor Input Regressions: Nicaragua

Variable name	Industry mix and benchmark indicators		Industry mix, benchmark indicators, and enterprise characteristics		Industry mix, benchmark indicators, enterprise characteristics, and other community characteristics		As in column 3, but with benchmark indicators replaced by components	
Variant	(1)		(2)		(3)		(4)	
N	846		846		846		846	
r2	0.052		0.108		0.123		0.136	
sv[a]	0.284		0.285		0.285		0.285	
smu[a]	0.280		0.277		0.238		0.231	
Enterprise characteristics								
Age of enterprise#			0.079	6.24*	0.077	6.09*	0.079	6.22*
Registration			0.510	19.64*	0.516	19.93*	0.516	19.97*
Industry dummies								
Services enterprise	−0.167	−5.06*	−0.167	−5.05*	−0.169	−5.12*	−0.174	−5.27*
Manufacturing nonagricultural enterprise	−0.117	−2.68*	−0.096	−2.20*	−0.093	−2.14*	−0.083	−1.91*
Agricultural processing enterprise	−0.131	−3.64*	−0.126	−3.49*	−0.128	−3.55*	−0.124	−3.42*
Other production enterprise	0.373	5.77*	0.402	6.19*	0.397	6.14*	0.402	6.22*
Mixed enterprise	0.549	16.57*	0.523	15.77*	0.518	15.62*	0.522	15.78*
Community characteristics								
Agricultural seasonality					0.173	2.36*		
Enterprise density					0.001	1.51	0.001	2.17*
Community population size#					0.051	1.67*	0.057	2.55*
Agricultural land per capita					0.122	3.03*	0.097	2.46*
Illiteracy					0.002	0.99		
Benchmark indicators and components								
Connectivity	0.629	2.18*	0.742	2.59*	0.667	2.54*		
Infrastructure services	−0.509	−2.19*	−0.603	−2.62*	−0.630	−2.95*		
Percent with electricity							−0.503	−4.83*
Percent with cellar phone							0.190	2.80*
Sewage system							−0.231	−2.12*
Business services	−0.008	−0.08	0.010	0.09	−0.050	−0.56		
Governance	−0.573	−1.84*	−0.324	−1.05	−0.455	−1.65*		
Public services							−0.410	−2.05*
Human capital	0.610	1.22	0.053	0.11	0.139	0.30		
Finance services	0.145	0.93	0.084	0.54	0.003	0.02		
Intercept	5.842	27.09*	5.494	25.55*	4.855	16.40*	5.313	21.69*

Source: RIC Surveys.

Notes: See Annex A for definitions and description of data; # logarithmic value.
a. "sv" is the standard deviation of the enterprise-specific disturbance. "smu" is the standard deviation of the community random effect.
"*": significant at least at 10 percent.

Table C.11 Capital Input Regressions: Nicaragua

Variable name	Industry mix and benchmark indicators		Industry mix, benchmark indicators and enterprise characteristics		Industry mix, benchmark indicators, enterprise characteristics and other community characteristics		As in column 3, but with benchmark indicators replaced by components	
Variant	(1)		(2)		(3)		(4)	
N	846		846		846		846	
r2	0.053		0.216		0.213		0.242	
sv[a]	0.284		0.285		0.285		0.285	
smu[a]	0.479		0.368		0.357		0.339	
Enterprise characteristics								
Age of enterprise#			0.134	10.56*	0.134	10.56*	0.136	10.69*
Registration			1.329	50.91*	1.330	50.95*	1.324	50.77*
Industry dummies								
Services enterprise	0.065	1.97*	0.058	1.74*	0.059	1.77*	0.053	1.61
Manufacturing, nonagricultural enterprise	−0.075	−1.71*	−0.044	−1.01	−0.043	−0.99	−0.033	−0.76
Agricultural processing enterprise	−0.222	−6.12*	−0.204	−5.63*	−0.204	−5.61*	−0.201	−5.55*
Other production enterprise	0.063	0.98	0.098	1.50	0.096	1.47	0.101	1.56
Mixed enterprise	0.312	9.40*	0.246	7.39*	0.246	7.40*	0.246	7.39*
Community characteristics								
Agricultural seasonality					0.181	1.75*		
Enterprise density					0.000	−0.33		
Community population size#					0.003	0.07	−0.067	−1.65*
Agricultural land per capita					−0.084	−1.47	−0.118	−2.03*
Illiteracy					0.001	0.24		
Benchmark indicators and components								
Connectivity	−0.662	−1.40	−0.409	−1.11	−0.409	−1.09		
Time to main market							−0.348	−2.27*
Distance to post office							−0.486	−3.32*
Infrastructure services	0.407	1.07	0.211	0.71	0.205	0.68		
Cost of transportation							−0.475	−4.30*
Percent with fixed phone							0.295	2.71*
Garbage collection							0.572	4.35*
Business services	0.090	0.53	0.115	0.87	0.086	0.66		
Governance	−0.363	−0.71	0.284	0.71	0.241	0.61		
Public services							−0.915	−3.24*
Public institutions							0.696	2.48*
Human capital	2.766	3.40*	1.435	2.25*	1.734	2.68*	1.916	3.11*
Finance services	0.687	2.68*	0.500	2.50*	0.404	1.85*		
Number of bank services							0.594	2.50*
Intercept	4.346	12.42*	3.568	12.96*	3.411	8.16*	4.762	10.69*

Source: RIC Surveys.

Notes: See Annex A for definitions and description of data; # logarithmic value;
a. "sv" is the standard deviation of the enterprise-specific disturbance. "smu" is the standard deviation of the community random effect.
"*" — significant at least at 10 percent.

Table C.12 Labor Input Regressions: Sri Lanka

Variable name	Industry mix and benchmark indicators		Industry mix, benchmark indicators, and enterprise characteristics		Industry mix, benchmark indicators, enterprise characteristics, and other community characteristics		As in column 3, but with benchmark indicators replaced by components	
Variant	(1)		(2)		(3)		(4)	
N	1018		1018		1018		1018	
r2	0.040		0.138		0.176		0.186	
sv[a]	0.759		0.736		0.736		0.736	
smu[a]	0.340		0.276		0.220		0.198	
Enterprise characteristics								
Age of enterprise#			0.006	0.16	0.007	0.18	0.003	0.08
Experience manager#			0.058	1.38	0.063	1.52	0.064	1.55
Gender of manager			0.206	2.58*	0.189	2.38*	0.186	2.37*
Education of manager			0.042	3.67*	0.043	3.80*	0.041	3.69*
Registration			0.246	3.23*	0.249	3.30*	0.228	2.96*
Industry dummies								
Services enterprise	−0.088	−0.89	−0.144	−1.51	−0.160	−1.68*	−0.160	−1.68*
Manufacturing, nonagricultural enterprise	0.056	0.68	0.128	1.52	0.110	1.30	0.110	1.31
Agricultural processing enterprise	0.594	2.90*	0.532	2.68*	0.543	2.75*	0.515	2.62*
Other production enterprise	1.042	4.54*	0.943	4.26*	0.951	4.31*	0.965	4.46*
Mixed enterprise	0.238	1.73*	0.154	1.16	0.154	1.16	0.157	1.19
Community characteristics								
Agricultural seasonality					−0.227	−1.41		
Enterprise density					0.004	2.44*	0.003	2.14*
Community population size#					−0.021	−0.24		
Agricultural land per capita					0.258	1.01	0.226	1.03
Illiteracy					0.039	0.10		
Main market in:								
Neighboring communities					0.368	2.55*	0.346	2.58*
Commercial center					0.361	2.50*	0.251	1.79*
Nearest city					0.202	1.70	0.218	1.93*
Main community income from:								
Wages					0.145	1.43		
Self-employment					0.089	0.58		
Benchmark indicators and components								
Connectivity	0.245	0.63	0.211	0.61	−0.077	−0.22		
Infrastructure services	0.277	0.66	0.216	0.58	0.313	0.84		
Availability electricity							0.180	1.72*
Percent with fixed phone							1.036	2.63*
Business services	0.277	1.21	0.284	1.40	0.304	1.50	0.353	1.89*
Governance	−0.432	−0.37	−1.016	−0.98	−0.910	−0.84		
Infrastructure and services							−0.503	−1.66*
Rule of law							0.709	1.01
Human capital	−0.182	−0.27	−0.322	−0.53	−0.404	−0.67		
Finance services	0.022	0.05	0.072	0.20	0.210	0.61		
Intercept	6.063	7.36*	5.689	7.65*	5.464	5.70*	4.715	17.71*

Source: RIC Surveys.

Notes: See Annex A for definitions and description of data; # logarithmic value.

a. "sv" is the standard deviation of the enterprise-specific disturbance. "smu" is the standard deviation of the community random effect.
"*" — significant at least at 10 percent.

Table C.13 Capital Input Regressions: Sri Lanka

Variable name	Industry mix and benchmark indicators		Industry mix, benchmark indicators, and enterprise characteristics		Industry mix, benchmark indicators, enterprise characteristics, and other community characteristics		As in column 3, but with benchmark indicators replaced by components	
Variant	(1)		(2)		(3)		(4)	
N	1018		1018		1018		1018	
r2	0.079		0.294		0.306		0.329	
sv[a]	1.403		1.269		1.269		1.269	
smu[a]	0.657		0.467		0.410		0.395	
Enterprise characteristics								
Age of enterprise#			−0.028	−0.445	−0.025	−0.404	−0.010	−0.15
Experience manager#			0.152	2.119*	0.155	2.154*	0.156	2.20*
Gender of manager			0.654	4.753*	0.641	4.668*	0.603	4.44*
Education of manager			0.112	5.697*	0.108	5.532*	0.111	5.65*
Registration			0.822	6.277*	0.828	6.345*	0.818	6.30*
Industry dummies								
Services enterprise	−0.232	−1.274	−0.389	−2.363*	−0.403	−2.442*	−0.403	−2.45*
Manufacturing, nonagricultural enterprise	−0.438	−2.859*	−0.193	−1.323	−0.188	−1.286	−0.174	−1.20
Agricultural processing enterprise	0.851	2.247*	0.690	2.020*	0.682	1.997*	0.635	1.87*
Other production enterprise	0.820	1.929*	0.531	1.393	0.581	1.520	0.621	1.63
Mixed enterprise	0.196	0.771	−0.022	−0.094	−0.012	−0.052	0.025	0.11
Community characteristics								
Agricultural seasonality					0.154	0.542		
Enterprise density					0.002	0.698	0.005	2.00*
Community population size#					0.230	1.457	0.281	2.02*
Agricultural land per capita					0.126	0.276		
Illiteracy					−0.527	−0.759		
Main market in:								
Neighboring communities					−0.050	−0.195		
Commercial center					0.047	0.184		
Nearest city					−0.033	−0.155		
Main community income from:								
Wages					−0.028	−0.159		
Self-employment					−0.294	−1.073	−0.385	−1.53
Benchmark indicators and components								
Connectivity	1.050	1.431	0.938	1.597	0.860	1.365		
Infrastructure services	0.856	1.079	0.672	1.056	0.426	0.642		
Percent with protected water							0.856	2.78*
Percent with fixed phone lines							1.640	2.20*
Concrete/asphalt road							−0.333	−2.09*
Business services	0.585	1.347	0.609	1.750*	0.770	2.142*		
Accounting service							0.716	3.42*
Governance	−0.092	−0.042	−2.117	−1.192	−2.448	−1.285		
Human capital	0.708	0.547	0.305	0.295	0.370	0.343		
Finance services	−0.348	−0.443	−0.191	−0.304	−0.089	−0.147		
Intercept	4.704	3.015*	3.919	3.078*	2.415	1.422	0.221	0.21

Source: RIC Surveys.

Notes: See Annex A for definitions and description of data; # logarithmic value.

a. "sv" is the standard deviation of the enterprise-specific disturbance. "smu" is the standard deviation of the community random effect.
"*": significant at least at 10 percent.

Table C.14 Labor Input Regressions: Tanzania

Variable name	Industry mix and benchmark indicators		Industry mix, benchmark indicators, and enterprise characteristics		Industry mix, benchmark indicators, enterprise characteristics, and other community characteristics		As in column 3, but with benchmark indicators replaced by components	
Variant	(1)		(2)		(3)		(4)	
N	947		947		947		947	
r2	0.041		0.067		0.086		0.102	
sv[a]	0.492		0.486		0.486		0.486	
smu[a]	0.324		0.320		0.308		0.305	
Enterprise characteristics								
Age of enterprise#			−0.008	−0.40	−0.007	−0.38	−0.006	−0.33
Experience manager#			−0.030	−0.99	−0.033	−1.09	−0.036	−1.22
Gender of manager			0.158	3.51*	0.159	3.54*	0.155	3.49*
Education of manager			0.021	3.18*	0.021	3.12*	0.022	3.30*
Registration			0.051	1.10	0.050	1.08	0.047	1.02
Industry dummies								
Services enterprise	0.068	1.34	0.041	0.81	0.041	0.81	0.043	0.84
Manufacturing, nonagricultural enterprise	−0.107	−1.16	−0.096	−1.06	−0.086	−0.95	−0.085	−0.94
Agricultural processing enterprise	−0.206	−2.05*	−0.223	−2.24*	−0.220	−2.22*	−0.229	−2.32*
Other production enterprise	0.313	2.73*	0.272	2.37*	0.272	2.37*	0.260	2.27*
Mixed enterprise	0.085	1.63	0.064	1.22	0.064	1.21	0.071	1.37
Community characteristics								
Community population size#					−0.109	−2.53*	−0.077	−2.03*
Agricultural land per capita					−0.052	−0.99		
Illiteracy					0.001	0.59		
Main community income from:								
Wages					0.174	0.77		
Self-employment					0.017	0.16		
Benchmark indicators and components								
Connectivity	0.348	1.22	0.407	1.45	0.377	1.34		
Distance post office							0.250	2.66*
Infrastructure services	−0.121	−0.42	−0.146	−0.52	−0.189	−0.64		
Business services	0.389	1.64	0.318	1.35	0.361	1.56		
Marketing service							0.216	2.07*
Governance	0.048	0.11	0.068	0.16	0.004	0.01		
Public services/ institutions							0.284	1.69*
Human capital	0.410	0.61	0.319	0.47	0.701	1.02		
Finance services	0.103	0.44	0.067	0.29	0.065	0.28		
Intercept	5.083	19.59*	4.867	18.29*	5.719	13.08*	5.346	15.52*

Source: RIC Surveys.

Notes: See Annex A for definitions and description of data; # logarithmic value.

a. "sv" is the standard deviation of the enterprise-specific disturbance. "smu" is the standard deviation of the community random effect.

"*": significant at least at 10 percent.

Table C.15 Capital Input Regressions: Tanzania

Variable name	Industry mix and benchmark indicators		Industry mix, benchmark indicators, and enterprise characteristics		Industry mix, benchmark indicators, enterprise characteristics, and other community characteristics		As in column 3, but with benchmark indicators replaced by components	
Variant	(1)		(2)		(3)		(4)	
N	947		947		947		947	
r2	0.071		0.154		0.168		0.180	
sv[a]	1.835		1.743		1.743		1.743	
smu[a]	1.145		1.103		1.077		1.047	
Enterprise characteristics								
Age of enterprise#			−0.016	−0.23	−0.023	−0.34	−0.031	−0.46
Experience manager#			0.356	3.32*	0.367	3.42*	0.369	3.46*
Gender of manager			0.507	3.16*	0.489	3.04*	0.501	3.14*
Education of manager			−0.024	−1.00	−0.024	−0.98	−0.025	−1.05
Registration			1.332	7.99*	1.337	8.03*	1.311	7.89*
Industry dummies								
Services enterprise	0.502	2.66*	0.298	1.64*	0.304	1.68*	0.306	1.69*
Manufacturing, nonagricultural enterprise	−0.179	−0.53	−0.078	−0.24	−0.068	−0.21	−0.093	−0.29
Agricultural processing enterprise	−0.851	−2.28*	−0.900	−2.53*	−0.916	−2.57*	−0.861	−2.43*
Other production enterprise	−0.033	−0.08	−0.185	−0.45	-0.141	−0.34	−0.147	−0.36
Mixed enterprise	0.565	2.92*	0.459	2.47*	0.505	2.68*	0.497	2.66*
Community characteristics								
Community population size#					0.203	1.33		
Agricultural land per capita					0.334	1.79*	0.307	1.88*
Illiteracy					0.002	0.43		
Main community income from:								
Wages					−0.394	−0.49		
Self-employment					0.381	1.05		
Benchmark indicators and components								
Connectivity	−1.512	−1.48	−1.238	−1.26	−0.921	−0.93		
Infrastructure services	1.466	1.42	1.337	1.35	1.513	1.46		
Percent with cell phones							1.367	2.23*
Concrete/asphalt road							0.576	2.01*
Business services	0.353	0.41	0.119	0.15	−0.016	−0.02		
Legal services							0.595	1.37
Governance	−0.979	−0.63	−1.268	−0.85	−0.897	−0.60		
Human capital	5.409	2.23*	5.022	2.14*	4.561	1.89*	4.282	1.92*
Finance services	0.180	0.21	0.340	0.42	0.470	0.58		
Number of banks							0.926	1.64
Number of bank services							−1.591	−1.89*
Intercept	3.106	3.34*	2.314	2.49*	0.168	0.11	1.355	2.37*

Source: RIC Surveys.

Notes: See Annex A for definitions and description of data; # logarithmic value.

a. "sv" is the standard deviation of the enterprise-specific disturbance. "smu" is the standard deviation of the community random effect.

"*": significant at least at 10 percent.

Annex D. Household Characteristics in Sri Lanka, Tanzania, and Nicaragua: Findings from the RIC Pilot Surveys

INTRODUCTION

This annex discusses the household demographics and characteristics among the survey households with brief notes on estimation methods used for key variables such as income, assets, investments, and a human capital index. It also reviews distribution of households across these key variables as well as averages reported. Assets, investment, and income are recorded in U.S. dollars using the 2004 exchange rate of each country.[82]

The estimation of statistics about rural households and the rural population employed sampling weights for Sri Lanka and Nicaragua, but unfortunately not for Tanzania, where weights are not available. Sampling weights are necessary because household sampling was stratified by the presence of a household-based enterprise. This implies that the statistics for Tanzania must be interpreted with caution: statistics of variables expected to be correlated with the presence of an enterprise might be significantly biased.

Sampling weights for Sri Lanka were constructed ex-post from a comparison of sample information with a community listing. Unfortunately, the listings of households were not available (or incomplete) for 18 communities, leading to 106 households (10 percent) for which weights are not available. As Table D.1 shows, these households are primarily concentrated in the war-torn areas in the east and northeast. Statistics computed with weights can thus represent only those communities for which weights are available, therefore excluding the war-torn areas of Sri Lanka.

HOUSEHOLD DEMOGRAPHICS AND CHARACTERISTICS

Age Distribution

The sample households' age distributions in all three countries skewed to the left, as shown in Table D.2, meaning the countries' households have a larger number of younger people. Tanzania has the highest percentage share (42.5 percent) of population below 16 years of age followed by Nicaragua (38.4 percent) and Sri Lanka (26.4 percent). In contrast, Sri Lanka has the highest share (66.8 percent) of working age population (16 to 60 years) followed by Nicaragua (56.1percent) and Tanzania (53.5 percent). Percentage shares of the male population of the sample households are consistently lower than their female counterparts; however the share is lowest in Tanzania (48.3 percent) and highest in Sri Lanka (49.6 percent), with Nicaragua in the

Table D.1 Sri Lanka: Availability of Sampling Weights by Province

Province	Number of household	Percentage with available weight
Western	219	100.0
Central	123	91.9
Southern	172	100.0
North Western	116	100.0
North Central	71	100.0
Uva	84	100.0
Sabaragamuwa	116	100.0
Eastern	79	3.8
North & East	83	75.9
Overall	1,063	90.0

Source: RIC Survey, Tanzania.

Table D.2 Age Distribution (%) of Household Members: Sri Lanka, Tanzania, and Nicaragua

	Sri Lanka			Tanzania			Nicaragua		
Age group	Male	Female	All	Male	Female	All	Male	Female	All
Household members (including household heads):									
<16	28.0	24.8	26.4	42.6	42.4	42.5	40.6	36.4	38.4
16–30	28.5	28.8	28.7	26.1	29.4	27.8	27.9	28.8	28.4
31–45	21.1	23.1	22.1	18.0	17.5	17.8	17.0	18.1	17.6
46–60	16.1	15.9	16.0	9.0	6.6	7.8	7.5	9.2	8.4
>60	6.2	7.5	6.9	4.2	4.0	4.1	7.0	7.6	7.3
Total	100	100	100	100	100	100	100	100	100
N	2082	2113	4195	3909	4106	8015	3799	4071	7870
Household heads:									
16–30	9.9	2.8	8.7	21.9	19.2	21.5	19.3	11.4	16.6
31–45	39.0	18.7	35.4	42.8	35.8	41.7	41.5	30.5	37.7
46–60	36.8	38.2	37.0	24.1	24.2	24.2	20.2	29.9	23.5
>60	14.4	40.3	18.9	11.2	20.8	12.7	19.1	28.2	22.2
Total	100	100	100	100	100	100	100	100	100
N	814	143	957	1351	260	1611	963	572	1535

Source: RIC Surveys.

middle (48.7 percent). Despite the lower shares of males, the distribution by age group does not vary significantly when it is partitioned by the male-female divide, except in Nicaragua.

A fair number of the household heads in the three countries are female, ranging between 15 and 16 percent in Sri Lanka and Tanzania. Surprisingly, almost 37 percent of households are headed by females in Nicaragua. Although most of the household heads are within the working-age cohort, notable numbers of them are over 60 years of age. A higher proportion of female household heads does not belong to the working-age category in these countries. All together, the percentage share of non-working-age household heads is the highest in Nicaragua (19.5 percent) and lowest in Tanzania (12.7 percent), which is consistent with the higher percentage of younger population in the country.

Household Size

The average household size is highest in Tanzania (4.98) and lowest in Sri Lanka (4.26). It is slightly less in Nicaragua (4.85) than in Tanzania (Table D.3). On average, numbers of female members were higher than male members in all of the pilot surveys. In Sri Lanka most of the households (92.3 percent) have 1 to 6 members, while in Tanzania and Nicaragua the percentage share of household members within this range are 75.5 percent and 81.6 percent, respectively.

Education

The raw data on education are not comparable across the selected countries owing partly to their different education systems and partly to the ways the questionnaires were framed. Sri Lanka followed the standard format: 1 to 5 being primary, 6 to 8 pre-secondary, 9 to12 secondary, and above 12 tertiary or college. Although Nicaragua has a system of grades and levels similar to Sri Lanka's, data were collected at two levels. The first level collected by categories, such as preschool, primary, secondary, university, and so on, and the second level ascertained the number of years in each category. These two levels were converted to grade level by adding the number of years in the previous category to the number of years completed in the assigned category. If an individual completed three years of secondary, for example, the number of years required to complete the primary level were added to the three secondary-level years. In addition to these categories, the Sri Lanka and Nicaragua surveys also utilized other categories, such as pre-primary, adult education, and basic, medium, and superior level technician. The same method was followed to obtain a corresponding grade level for each respondent. In Tanzania, standard 1 through standard 8 is considered primary and pre-form 1 through form 6 as secondary, followed by college or university. For comparability, standards 1 through 5 were converted to *primary*,

Table D.3 Distribution (%) of Households by Household Size and Gender

	Sri Lanka (n = 957)			Tanzania (n = 1611)			Nicaragua (n = 1535)		
Household size	**Male**	**Female**	**All**	**Male**	**Female**	**All**	**Male**	**Female**	**All**
0	3.1	0.8	0.0	5.0	4.8	0.0	5.6	3.54	0.0
1	28.4	29.9	2.7	27.6	22.6	5.9	25.6	24.61	3.3
2	33.9	40.7	9.4	23.9	27.3	10.4	27.0	29.89	10.1
3	23.2	18.6	19.2	21.1	21.9	12.7	22.5	21.1	15.3
4	9.2	7.0	26.9	13.1	11.7	16.5	11.5	11.73	18.8
5	1.4	2.6	23.4	5.4	6.9	16.6	5.6	4.97	21.0
6	0.3	0.5	10.7	3.0	3.2	13.4	1.5	1.91	13.1
7	0.6		4.8	0.6	1.1	9.0	0.5	0.98	7.0
8			1.6	0.2	0.4	5.8	0.0	0.25	5.2
9			1.0	0.1	0.1	4.8	0.2	0.84	2.6
>10			0.3	0.1		4.9	0.0	0.19	3.6
All	100	100	100	100	100	100	100	100	100
Average	2.15	2.11	4.26	2.4	2.5	5.0	2.36	2.49	4.85

Source: RIC Surveys.

standards 6 through nine9 to *junior high,* and preform 1 through form 6 to *secondary*.

Table D.4 provides information on the education of the household members and household heads. It is clear from the table that the percentage of the population that has received at least some schooling is impressive in all three pilot countries, ranging from 70 to 90 percent, with Sri Lanka highest and Nicaragua lowest. In Tanzania most (72 percent) of the eligible members have completed primary level education or more, while in other two countries most of members have completed post-primary education, including the junior high and secondary levels. More than 10 percent of household members in Sri Lanka, 4 percent in Nicaragua, and 5 percent in Tanzania also have a tertiary level education. Both literacy rates and educational achievement among female household members is lower than for males in both Sri Lanka and Tanzania but higher in Nicaragua.

The education of household heads shows a similar pattern. Household heads rates of schooling are highest in Sri Lanka (93 percent) and lowest in Nicaragua (73 percent). Among household heads in Tanzania it is 86 percent. In Sri Lanka and Tanzania, however, female household heads received almost two years of schooling less than male heads.

Human Capital

A human capital stock value and index is computed on the basis of the methodology developed by Bils and Klenow (2000), which is based on the classic Mincerian returns to schooling and experience. Define h as the human capital stock, and let a denote age, and let s refer to years of schooling. Then, abstracting from an irrelevant scaling constant,[83] the log of the human capital stock is measured with the following formula.

$$\ln(h(s, a)) = \theta s + \gamma_1 (a - s - 6) + \gamma_2 (a - s - 6)^2$$

Bils and Klenow (2000) obtain parameter estimates for θ, γ_1 and γ_2 from averages of 52 countries: $\theta = 0.099$, $\gamma_1 = 0.512$, and $\gamma_2 = -0.00071$.[84] Thus, the natural logarithm of the human capital stock is measured for each individual as

$$\ln(h) = 0.099s + 0.0512\,(a-s-6) - 0.00071(a-s-6)^2$$

and the human capital stock is found simply by taking the antilog:

$$h = e^{\ln(h)}$$

The human capital stock is computed for each individual between 16 and 65 years of age. The lowest possible value of h over all values of s and a equals approximately 1.55, achieved at $a = 16$ and $s = 0$; the highest value depends on the maximum attainable level of education: for $s = 14$, the maximum equals about 10.06 for someone aged $a = 56$ years old; for $s = 15$, it equals about 11.11 for someone aged $a = 57$ years old. Since the longest length of schooling effort depends on the school system in a country (and possibly the precision of the coding system used in the survey), h is scaled to a number between 0 and 1, and we shall refer to this as the

Table D.4 Distribution (%) of Household Members and Household Heads by Level of Education in Selected Countries

	Sri Lanka			Tanzania			Nicaragua		
	Male	Female	All	Male	Female	All	Male	Female	Total
Household members:									
0	5.5	9.7	7.6	26.7	30.6	28.7	31.8	28.5	30.1
1–5	27.3	27.0	27.1	26.0	25.6	25.8	32.7	30.8	31.7
6–8	20.4	15.4	17.9	36.9	37.4	37.1	20.4	21.9	21.1
9–12	35.9	38.6	37.2	3.9	2.9	3.4	11.6	13.9	12.8
>=13	11.0	9.4	10.2	6.5	3.5	5.0	3.5	4.9	4.3
Total	100	100	100	100	100	100	100	100	100
Mean years	7.6	7.2	7.4	5.3	4.7	5.0	4.0	4.6	4.3
N	1984	2011	3995	3909	4106	8015	3790	4064	7856
Household Heads:									
0	4.8	16.4	6.8	11.3	27.3	13.8	27.8	27.4	27.6
1–5	30.1	35.0	31.0	16.7	15.8	16.5	29.9	37.5	32.5
6–8	20.9	19.5	20.7	57.5	47.7	55.9	24.9	18.5	22.7
9–12	34.1	25.8	32.6	3.6	2.7	3.5	13.0	9.8	11.9
>=13	10.1	3.3	8.9	11.0	6.5	10.2	4.5	6.8	5.3
Total	100	100	100	100	100	100	100	100	100
Mean years	7.4	5.7	7.1	6.5	5.0	6.2	4.5	4.4	4.5
N	810	142	952	1351	260	1611	959	570	1529

Source: RIC Surveys.

"human capital index":

$$h_{index} = \frac{h - \min(h)}{\max(h) - \min(h)}$$

This ensures comparability between countries, while obviously retaining comparability between individuals within a country or between communities of a country.

For the variable *s*, years of schooling, the coding in the surveys does not measure years of schooling as much as levels of attained education. Thus, the survey information must be transformed to derive the estimate of years of schooling *s* used in the computations of human capital stock. The coding in the household survey was scaled as indicated in Table D.5. Nicaragua is not represented in this table: this questionnaire used a different coding scheme, allowing the respondent to state how many years were completed at a particular schooling level. To turn this into years of schooling, assumptions are made about the number of years required to reach the indicated level of schooling: 6 for junior secondary, 9 for senior secondary or mid-level technical training, and 12 for university and advanced technical training, with a peak of 15 years on the total years of schooling taken.[85]

At the household level, both the human capital stock and the index can be presented as average, sum, or maximum individual stock or index. In Table D.6 and Table D.7 provide household level averages among adults of human capital stocks and indexes, respectively. The average human capital stock is the highest (4.66) in Sri Lanka with no significant difference between males and females, lowest in Nicaragua (3.61) with no gender difference, and in between in Tanzania (3.77) with higher average stock among males than females. When the distributions of households by average human capital stock are compared, it shows an almost normal distribution for Sri Lanka and for Nicaragua and Tanzania distributions highly skewed to the left. The average human capital stock is less than four for 66 percent of households in Tanzania and 69 percent in Nicaragua, as opposed to 34 percent in Sri Lanka. In Sri Lanka 15 percent of households have an average human capital stock of more than 6; the same is only 2 percent in Tanzania and 6 percent in Nicaragua.

Table D.7 provides distribution of households by their average human capital index. Like human capital stock, the average human capital index is also

Table D.5 Coding of Schooling Level: Original and Converted

	Sri Lanka		Tanzania	
Measured schooling level	Original coding	Assigned years of schooling	Original coding	New coding
Year 1	1	1	Standard 1	1
Year 2	2	2	Standard 2	2
Year 3	3	3	Standard 3	3
Year 4	4	4	Standard 4	4
Year 5	5	5	Standard 5	5
Year 6	6	6	Standard 6	6
Year 7	7	7	Standard 7	7
Year 8	8	8	Standard 8	8
Year 9	9	9	Training	9
Year 10	10	10	Pre form 1	9
Year 11	11	11	Form 1	10
Year 12	12	12	Form 2	11
Year 13	13	13	Form 3	12
University	14	15	Form 4	13
Professional	15	14	Form 5	14
Technical College	16	14	Form 6	15
Pre-school	17	0	Training	16
No schooling	18	0	University	16
Other	19	Missing	Adult education	1
20 (?)	20 (?)	Missing	Under std. 1	0
Don't know	99	Missing		

Source: RIC Surveys.

Note: At the university level, years of schooling are not differentiated any further.

Table D.6 Distribution (%) of Households by Average Human Capital Stock

	Sri Lanka			Tanzania			Nicaragua		
Average human capital stock	Male	Female	All	Male	Female	All	Male	Female	All
<2.5	4.9	9.0	4.7	10.3	16.8	10.1	22.8	21.1	19.6
2.5–3.0	8.5	8.8	9.9	10.5	15.5	12.7	17.1	19.6	16.7
3.0–3.5	8.7	10.2	6.4	13.8	19.9	19.4	16.4	16.3	22.0
3.5–4.0	14.3	10.3	12.9	20.5	21.0	23.7	12.5	13.0	11.1
4.0–4.5	12.0	9.6	17.0	13.1	12.4	14.4	11.7	10.7	10.3
4.5–5.0	8.5	12.9	12.6	15.0	8.1	9.7	4.5	6.2	7.3
5.0–5.5	10.5	12.2	9.9	6.3	2.7	3.8	4.0	4.8	4.3
5.5–6.0	7.5	7.5	8.7	2.5	1.3	1.9	1.5	1.8	2.6
6.0–6.5	8.8	5.8	6.7	1.7	0.5	2.0	3.4	2.3	2.0
>6.5	16.4	13.9	11.1	6.5	1.8	2.4	6.3	4.2	4.1
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Mean[a]	4.83	4.63	4.66	4.11	3.54	3.77	3.68	3.61	3.61
Standard Dev	1.63	1.59	1.45	1.48	1.07	1.10	1.56	1.40	1.30
N	903	919	947	1425	1487	1585	1309	1448	1506

Source: RIC Surveys.

Note: The overall average is not a simple average of the averages of the gender-specific columns.

Table D.7 Distribution (%) of Households by Average Human Capital Index

Average human capital stock index	Sri Lanka			Tanzania			Nicaragua		
	Male	Female	All	Male	Female	All	Male	Female	All
0.0–0.1	5.2	9.3	4.7	13.3	20.3	12.5	24.1	23.5	20.1
0.1–0.2	16.4	17.9	16.0	29.5	40.4	39.6	32.0	32.6	37.5
0.2–0.3	24.6	19.5	27.8	33.4	30.1	33.9	22.9	23.3	21.2
0.3–0.4	17.5	22.8	21.8	14.5	6.2	9.2	8.9	10.2	11.4
0.4–0.5	17.1	13.8	16.0	4.0	1.8	3.0	4.2	5.8	4.8
0.5–0.6	10.3	11.3	8.0	2.1	0.6	1.1	4.0	2.0	3.2
0.6–0.7	4.6	3.3	3.9	1.7	0.5	0.4	2.3	1.3	1.3
> 0.7	4.3	2.1	1.7	1.5	0.2	0.3	1.7	1.4	0.5
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Mean	0.35	0.32	0.33	0.24	0.19	0.21	0.22	0.21	0.21
Standard Dev	0.17	0.17	0.15	0.14	0.10	0.10	0.16	0.15	0.14
N	903	919	947	1425	1487	1585	1309	1448	1506

Source: RIC Surveys.

higher in Sri Lanka (0.33), compared with a value of 0.21 in both Nicaragua and Tanzania. The distribution of the households by average human capital index is almost perfectly normal in Sri Lanka and relatively skewed to the left in Nicaragua and Tanzania. The distribution in Tanzania is concentrated below 0.30. In Sri Lanka, 48 percent of households have an index less than 0.30, and in Tanzania and Nicaragua 86 percent and 79 percent of households, respectively, fall below the 0.30 threshold.

LABOR FORCE

Potential Labor Force: Number of Adults in the Rural Population

As a first step in examining the rural labor force, Table D.8 considers the number of working adults (within the age range of 16 to 65 years) in the sample households in the pilot countries. It is clear from the table that the average number of working-age adults is the highest (2.99) in Sri Lanka, followed by Nicaragua (2.80). Tanzania has the lowest (2.76) number. Average numbers of working-age female members are, however, higher than male members in Nicaragua, with no significant difference between male and female adults in both Sri Lanka and Tanzania. In Sri Lanka, 84 percent of the households have 2 to 4 adult male and female members eligible to work, while in Tanzania and Nicaragua the figures are 77 percent and 74 percent, respectively. Distribution of adult members by gender reveals that 83 percent and 88 percent of the households in Sri Lanka have 1 to 2 male and female members, respectively, while these figures are 80 percent and 83 percent in Tanzania and 76 percent and 82 percent in Nicaragua.

Whereas the average household in Sri Lanka has 1.5 males and 1.5 females, 1.2 males and 0.6 females are engaged in an economic activity of some kind, whether working on the farm, being self-employed in a household enterprise, or being employed for a wage or salary. This implies a labor force participation rate of 0.77 for males and 0.39 for females.

Occupation at the Time of the Survey

The present surveys offer two types of information about occupation. First is the activity occupying respondents at the time of the survey. Actually, only the Sri Lanka questionnaire asks for "Activities involved at present" and does not offer respondent a chance to list several activities. The response should therefore indicate the primary activity. Other countries did not collect this current occupational profile. Occupational profiles of the household members and household heads are presented in Table D.9. It shows that in Sri Lanka about 36 percent of eligible household members are either salaried or wage employed and 16 percent are self-employed. Another 7.5 percent and 8.1 percent of eligible household members are unemployed or students, respectively. Almost 25 percent of

Table D.8 Potential and Actual Workforce: Number of Adults Present and Working

Number of adult members present	Sri Lanka (n = 957)			Tanzania (n = 1586)			Nicaragua (n = 1490)		
	Male	Female	All	Male	Female	All	Male	Female	All
0	4.7	2.3	0.0	10.0	6.0	0.0	13.1	5.1	0.0
1	59.1	57.9	4.3	60.5	62.9	11.3	58.9	59.8	12.3
2	24.0	30.0	41.4	19.5	19.8	45.2	17.3	22.2	44.4
3	9.0	6.4	24.1	6.0	8.2	21.4	6.5	9.0	18.4
4	2.2	3.3	18.0	3.1	2.1	10.8	2.9	2.5	10.9
5	0.4	0.2	7.8	0.6	0.6	6.3	1.1	0.9	7.8
6	0.6		2.8	0.1	0.3	2.5	0.2	0.5	4.3
7			0.6	0.0	0.1	1.0		0.1	0.8
>=8			1.0	0.1	0.1	1.5			1.1
Total	100	100	100	100	100	100	100	100	100
Average	1.48	1.51	2.99	1.35	1.42	2.76	1.31	1.49	2.80

Number of working adult members	Sri Lanka (n = 957)			Tanzania (n = 1586)			Nicaragua (n = 1490)		
	Male	Female	All	Male	Female	All	Male	Female	All
0	8.8	48.9	0.0	12.4	10.1	1.3	13.1	39.9	0.0
1	72.5	43.3	47.0	64.2	65.5	15.5	64.0	46.8	42.9
2	13.8	7.2	36.8	16.0	16.8	48.9	16.5	10.0	33.1
3	4.7	0.4	11.3	4.3	5.4	17.7	4.3	2.4	13.6
4	0.3	0.2	4.3	2.4	1.5	8.6	1.5	0.3	6.3
5			0.6	0.4	0.3	4.2	0.5	0.2	2.7
6			0.1	0.1	0.3	2.0	0.2	0.5	0.5
7			0.0	0.0	0.1	0.7		0.1	0.6
>=8				0.1	0.1	1.0			0.2
Total	100	100	100	100	100	100	100	100	100
Average	1.15	0.60	1.75	1.22	1.26	2.48	1.19	0.79	1.98
Labor force participation rate	0.76	0.39	0.57	0.91	0.89	0.90	0.83	0.49	0.65

Source: RIC Surveys.

Table D.9 Distribution (%) of Household Members (age ≥ 16 and ≤ 65) and Household Heads by Occupation in Sri Lanka

Sri Lanka	All household members			Household heads		
	Total	male	female	Total	male	female
Salaried/wage employed	36.11	51.45	21.02	52.72	58.81	19.53
Self-employed	15.74	20.93	10.62	28.24	28.87	24.80
Employer	0.04	0.08	0.00	0.02	0.03	0.00
Unpaid family member	1.20	1.34	1.06	2.15	2.36	1.01
Unemployed	7.48	8.22	6.76	0.15	0.02	0.84
Student	8.13	9.43	6.86	0.11	0.13	0.00
Housekeeping	24.82	2.34	46.91	5.14	1.04	27.43
Retired	1.47	2.02	0.94	3.87	3.53	5.71
Various inactive	5.00	4.18	5.81	7.34	5.21	20.67
Total	100.00	100.00	100.00	100.00	100.00	100.00

Source: RIC Surveys.

Table D.10 Adult Labor Force Participation Rate and Labor Force Composition

Type of household	Labor force participation rate			Proportion of labor force in:			Number of adults
	Overall	Male	Female	Wage employment	Nonfarm self-employment	Farming	
Sri Lanka							
All	0.574	0.763	0.388	0.643	0.233	0.310	1816
With NFHE(w)	0.652	0.835	0.474	0.351	0.688	0.233	1161
Without NFHE(w)	0.541	0.732	0.352	0.794	0.000	0.350	655
With NFHE(i)	0.687	0.860	0.514	0.571	0.538	0.195	683
Without NHFE(i)	0.565	0.757	0.375	0.724	0.083	0.359	932
Tanzania							
All	0.898	0.910	0.887	0.153	0.349	0.900	3937
With NFHE(i)	0.891	0.904	0.879	0.159	0.453	0.887	2454
Without NHFE(i)	0.911	0.920	0.902	0.144	0.176	0.921	1483
Nicaragua							
All	0.647	0.832	0.485	0.631	0.219	0.375	3286
With NFHE(i)	0.756	0.823	0.698	0.360	0.644	0.219	2280
Without NHFE(i)	0.597	0.828	0.399	0.742	0.038	0.423	718

Source: RIC Surveys.

Notes: NFHE denotes "nonfarm household enterprise".
(w) indicates that the presence of a household enterprise is judged by employment in nonfarm self-employment activities;
(i) indicates that the presence of a household enterprise is judged by the receipt of income from a nonfarm business or from nonfarm self-employment activities.

adult household members report housekeeping as their main occupation, of which the majority are women: more than 47 percent of eligible women household members engage in housekeeping activities, against only 2.3 percent or their male counterparts. Although in smaller numbers than their male counterparts, a fair percentage of the women are either self-employed (10 percent) or salaried/wage employed (21 percent). Also of interest is that women members are relatively less unemployed than are male members. Somewhat more male than female household members are enrolled in school: 9.4 percent against 6.9 percent. Only 1.5 percent of eligible members took early retirement, with only a small difference between males and females.

Unlike the household members, most of the household heads (82 percent) are employed either as self-employed or salaried/wage workers. Participation of male household heads in these two categories of employment together is higher (88 percent) than those of their female (44 percent) counterparts. A significant proportion of female household heads, however, either engages in household activities (27 percent) or is retired/inactive (26 percent). Together, only 5 percent of household heads list their current (primary) occupation as housekeeper, and 11.2 percent are retired or inactive.

Economic Activity During the Previous 12 Months

The second measure of economic activity makes specific reference to the previous 12 months. Table D.10 provides an overview. The labor force participation rate is 57.4 percent in Sri Lanka, with a large gender gap (76.3 percent for men, 38.8 percent for women), lower than both Tanzania at 89.8 percent with no gender gap and Nicaragua at 64.7 percent also with a large gender gap. The proportion of the labor force holding a wage job is similar in Sri Lanka and Nicaragua, as is the proportion self-employed in a nonfarm business. Nicaragua has a slightly higher percentage of working adults in farming. The labor force structure in Tanzania is sharply different: 90 percent of the rural labor force works on farms and only 15 percent hold a wage job.

As with any statistical comparison, it must be kept in mind that statistics for Sri Lanka and Nicaragua are weighted with sampling weights,

whereas Tanzania's statistics are unweighted. This caution is especially appropriate here. Since the survey design oversampled households with a nonfarm enterprise, these labor force composition statistics for Tanzania overstate the percentage of the labor force active in nonfarm self-employment.

Table D.10 also distinguishes labor force statistics according to the entrepreneurship status of the household. The presence of a nonfarm enterprise may be established by income receipt within the household or activity status among its members. Ideally, these criteria should make no difference, but in Sri Lanka's case a questionnaire flaw makes a large difference here.[86] Because of this, the income criterion provides the evidence of entrepreneurship in Tanzania and Nicaragua, and work status does so in Sri Lanka. In Sri Lanka, then, the labor force participation rate is higher, for both men and women, if the household operates an enterprise; wage employment is more prevalent among households that do not. Nicaragua resembles Sri Lanka in this regard, but Tanzania does not show such a difference.

Table D.11 provides information on household members' involvement in at least one major economic activity, such as self-employment, wage employment, and farming. The central message of the table is that few households specialize in only one activity, whether a nonfarm enterprise, wage employment, or farming. This is true however households are categorized: by income source (either including or omitting marginal income sources yielding less than US$60 per year) or by work activity. In Sri Lanka, for example, 29.4 percent of the rural households have at least one member working in an enterprise, but only three of every ten of these (8.6 percent of the total) do so as a specialty; this ratio is the same in Nicaragua, where 22.7 percent of the households operate an enterprise. The lack of specialization is even more pronounced in Tanzania: 7 of 100 among the 55.5 percent operating an enterprise. (Note again that the income criterion biases the statistics in Sri Lanka.)

Proportions of households relying absolutely wage employment are 37 percent in Sri Lanka and 45 percent in Nicaragua. Complete dependence on wage employment is very negligible (1.5 percent) in Tanzania. Almost 27 percent of households in Tanzania rely on farming as their only source of income. In Sri Lanka and Nicaragua 8 percent and 10 percent of households, respectively, rely only on farming. About 43 percent of households in Sri Lanka, 56 percent in Tanzania, and 35 percent in Nicaragua have two or three economic activities simultaneously.

The table shows that 9.5 percent of households in Nicaragua, 1.4 percent in Sri Lanka, and 7.3 percent in Tanzania do not have any economic activity. These percentages rise to 11.8 percent, 6.7 percent, and 27.2 percent, respectively, if marginally relevant income sources (producing less than $60 per year) are excluded.

Much of the diversification across activities occurs because different household members do different types of work. Still, 10.5 percent of working adults in Sri Lanka, 29.1 percent in Tanzania, and 14.6 percent in Nicaragua engage in more than one activity, and these percentages are higher for men than for women. Moreover, these statistics hide occurrences of people holding two of the same type of activities (for example, two wage jobs). Household heads specialize more than do other household members.

Table D.12 reports the share of the active labor force engaged in a particular activity. This sums to more than 100 percent because people may be (and often are) active in more than one type of employment, as was already illustrated in Table D.11. Especially in Sri Lanka, the variation in the structure of the labor force is large.

As the last table in this section on the labor force, Table D.13 highlights the educational achievement, gender composition, age structure, and human capital indexes among labor force participants by their activity status. In Sri Lanka, those who are self-employed in a nonfarm enterprise tend to be slightly more educated that those working on a farm; they are comparable to wage employees, except for the more highly educated, who often work for someone else. Wage employees are younger and less likely to be female than are the self- or farm-employed. As a result, wage employees' human capital lags behind that of the self- and farm-employed, but not by much.

In Nicaragua and Tanzania, the differentiation between the three labor force groups according to education and gender is similar to that of Sri Lanka. Age patterns vary, however. In Tanzania, wage employees are the oldest group, farmers the youngest, and the nonfarm self-employed are in between; in Sri Lanka, the pattern is the exact opposite; and in Nicaragua, it is the nonfarm self-employed who are clearly the oldest.

Table D.11 Accounting for Economic Activities of Households and Household Members

	Overall household			Employment[c] of all adult household members[d]			Employment of heads of household		
	Any income[a] from stated source	Income from stated source > $60[b]	Any activity	Total	Male	Female	Total	Male	Female
Sri Lanka									
Enterprise only	0.0	1.5	8.6	8.8	10.2	7.4	12.9	13.5	9.5
Enterprise and wage	9.9	10.8	9.1	1.5	2.7	0.2	3.6	4.2	0.2
Enterprise and farming	1.5	0.3	5.3	2.8	3.2	2.4	6.8	5.9	11.7
Enterprise, wage, and farming	6.6	4.3	6.4	0.4	0.7	0.0	1.0	1.1	0.1
Wage only	36.0	43.1	36.7	29.3	39.4	19.5	35.1	38.5	17.1
Farming only	18.3	13.3	8.4	8.9	9.9	7.9	13.5	12.9	16.9
Wage and farming	26.4	20.1	22.5	5.8	10.2	1.5	14.5	16.8	2.2
Not employed	1.4	6.7	3.0	42.6	23.7	61.2	12.6	7.2	42.4
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Tanzania									
Enterprise	14.6	25.0	3.8	4.9	5.3	4.5	5.1	4.3	9.9
Enterprise and wage	2.2	3.4	3.4	3.3	4.5	2.2	4.7	4.4	6.7
Enterprise and farming	38.9	18.2	28.2	16.2	20.3	12.3	28.5	30.0	20.2
Enterprise, wage, and farming	6.2	2.4	20.1	6.9	10.7	3.4	14.9	16.0	9.0
Wage	1.5	2.7	0.4	0.8	0.9	0.7	1.1	1.0	1.4
Farming	26.6	19.2	36.7	55.0	46.1	63.6	38.8	37.5	45.7
Wage and farming	2.7	2.0	4.6	2.7	3.4	2.1	5.1	5.1	4.9
Not employed	7.3	27.2	2.9	10.2	9.0	11.3	1.8	1.8	2.2
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Nicaragua									
Enterprise	7.7	8.7	7.2	10.1	8.5	11.5	13.8	10.7	20.4
Enterprise and wage	8.3	8.9	9.4	2.5	2.7	2.3	2.6	2.5	3.1
Enterprise and farming	3.5	2.7	2.6	1.6	2.3	0.9	2.6	3.2	1.3
Enterprise, wage, and farming	2.7	1.7	3.5	0.5	0.7	0.3	0.9	0.8	1.2
Wage	45.5	50.5	37.3	29.2	36.4	23.1	33.6	36.1	28.5
Farming	10.0	8.5	9.3	13.0	18.1	8.6	13.4	18.3	3.2
Wage and farming	12.8	7.4	19.7	10.0	18.3	2.9	16.0	22.9	1.4
Not employed	9.5	11.8	11.1	33.1	13.0	50.4	17.0	5.6	40.9
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Source: RIC Surveys.

Notes: a. For Sri Lanka, adults living in 149 households with zero or negative total incomes are excluded.
b. Column 2 counts involvement in an activity only if it generates at least $60 over the previous 12 months.
c. Employment refers to the previous 12 months.
d. Including heads of households.

LAND AND PHYSICAL ASSETS

Land Holdings

Land is still the most prominent asset in the developing countries' portfolio, as it determines household economic position. The pilots' data collection on land ownership and value did not follow common questions and responses, and measurement units of land differ within and across countries. Sri Lanka, for example, has two categories of land using different measuring units (acres and perches), and Nicaragua measures land in manzanas (mzs).

In Sri Lanka, perches were converted into acres by the international standard of 65 perches

Table D.12 Variation in the Structure of the Labor Force by Region

	Wage employment	Self-employed in nonfarm enterprise	Self-employed in farm activity
A: Sri Lanka			
Western	72.7	19.0	17.4
Central	73.5	13.5	26.4
Southern	56.6	23.3	39.1
North Western	53.1	27.3	33.5
North Central	58.7	20.9	51.5
Uva	53.2	22.6	54.8
Sabaragamuwa	73.0	17.1	23.4
Eastern	33.3	0.0	83.3
North & East	58.3	53.6	17.9
Overall	64.3	23.3	31.0
Variation across regions[a]	12.9	14.2	21.5
B: Tanzania			
Kilimanjaro	16.7	39.4	85.5
Morogoro	14.8	35.1	92.4
Mtwara	8.8	39.5	91.8
Mbeya	19.8	42.7	87.2
Tabora	16.5	25.6	91.4
Kigoma	11.1	32.5	92.8
Kagera	16.5	30.8	89.9
Overall	15.3	34.9	90.0
Variation across regions[a]	3.7	5.9	2.8
C: Nicaragua			
Cabezera departamental	56.8	13.0	56.5
Municipios del pacífico	73.4	25.2	17.5
Municipios del resto del pais	58.6	29.7	37.7
Bluefields, Puerto Cabezas, San Carlos	81.7	19.6	4.7
Overall	63.1	21.9	37.5
Variation across regions[a]	12.0	7.2	22.8

Source: RIC Surveys.

Note: a. Variation across regions is measured by the simple standard deviation of the percentages listed.

per acre. For a small number of plots, measurements were reported in both acres and perches. This could indicate that the plot size was the sum of them or that the plot size was reported in both measures. A comparison with plot values as well as an inspection of the ratio of perches over acres showed no reliable pattern,[87] but it seemed most likely that households offered two alternative measures. Accordingly, land size was computed as the average of the two reported values.

The land unit measure *manzana* is in use in Latin American countries, but between countries its exact meaning may vary. In Nicaragua, 1 manzana equals 7000 square meters or 1.729 acres.[88]

Table D.14 summarizes land holdings per household. In Sri Lanka, the average household owns 0.92 acres of land. Not all households actually own land: 89.8 percent of households reported owning land, and among those that do, average holdings are 1.03 acres. In Nicaragua, only 22.7 percent of the sample households own land, which on average equals 84 acres. In Tanzania, almost 90 percent of the sample households own on average 2.7 acres of land.

For comparison across countries, physical land units are converted into dollar values. As information on land prices was sometimes unavailable, this problem was resolved by taking the average price of land per unit and imputing the average price where it was missing. In Sri Lanka, when the value of the plot was not reported, the sample average of 479,529 rupees (US $4955.35) per acre was used for imputation. Table D.15 provides

Table D.13 Characteristics of Groups of Labor Force Participants

	All adult rural population	Working for wage or salary	Self-employed in nonfarm enterprise	Self-employed on the farm
A: Sri Lanka				
Schooling*				
No schooling	5.5	3.6	3.2	5.2
Elementary school	22.2	22.0	22.1	27.3
Junior high school	15.5	17.3	18.7	21.4
High school	43.5	38.9	46.6	35.9
Tertiary education/training	13.3	18.1	9.4	10.3
Percent female	50.7	28.9	37.6	33.3
Average age	35.8	36.3	39.9	42.5
Average human capital stock	4.73	4.88	4.95	4.90
Average human capital stock index	0.334	0.350	0.358	0.353
B: Tanzania				
Schooling*				
No schooling	13.8	6.6	6.7	15.6
Elementary school	12.5	8.3	10.1	13.3
Junior high school	59.5	53.8	63.3	61.0
High school	5.2	5.5	4.3	3.4
Tertiary education	9.1	25.8	15.5	6.6
Percent female	51.2	31.2	36.6	51.5
Average age	32.9	37.7	35.9	33.9
Average human capital stock	3.76	5.01	4.45	3.73
Average Human capital stock index	0.21	0.32	0.27	0.20
C: Nicaragua				
Schooling*				
Able to read, write	83.7	84.8	87.2	73.1
No schooling	18.1	16.2	14.7	30.4
Elementary school	26.2	26.5	22.8	36.0
Junior high school	26.4	24.6	29.5	21.2
High school	21.9	22.8	23.5	9.5
Tertiary education/training	7.5	9.9	9.5	2.9
Percent female	53.2	36.3	55.1	27.3
Average age	32.9	31.8	36.5	33.8
Average human capital stock	3.56	3.80	3.95	3.00
Average human capital stock index	0.21	0.24	0.25	0.15

Source: RIC Surveys.

Note: * At each level, this includes people who completed only a part of the indicated level.

Table D.14 Average per Household Landownership in the Selected Countries

	Sri Lanka	Tanzania	Nicaragua
Unit	Acre	Acre	Acre
Mean	0.92	2.40	16.60
Standard deviation	1.15	3.22	131.13
Percent owning land	89.8	89.9	22.7
If owning land:			
Mean	1.03	2.67	84.04
Standard deviation	1.17	3.28	319.21

Source: RIC Surveys.

distribution of households by land value. Although most of the households own land in Tanzania, the value of land is less than US$500 for 79 percent of households. Distribution of value of land is more even in Sri Lanka than in Tanzania and Nicaragua. Although 85 percent of households are landless, the average per household land value is highest in Nicaragua. This means the top 5 percent of households own most of the land.

A brief note on land leasing: In Sri Lanka, land leasing is relatively uncommon. During a period of three years before the survey, about 7 percent of

Table D.15 Distribution (%) of Households by Value of Land

Land asset values (US $ equivalent)	Sri Lanka	Tanzania	Nicaragua
<500	45.2	79.2	85.2
500–1000	13.8	10.8	3.0
1000–1500	6.5	2.8	2.0
1500–2000	7.1	1.9	1.2
2000–2500	3.6	0.6	1.0
2500–3000	4.0	0.9	0.5
3000–3500	2.4	0.4	0.5
3500–4000	4.1	0.6	0.4
4000–4500	1.1	0.1	0.3
4500–5000	2.9	0.6	5.9
>5000	9.3	2.2	
Total	100	100	100
Average	1505	667.3	2156.5
SD	3101	2522.1	18833.3
N of obs 998			

Source: RIC Surveys.

the households leased land for purposes of crop production, with the majority paying by means of sharecropping arrangements. Another 2 percent of households lease land out, again usually under a sharecropping arrangement. In Tanzania, of the 1,610 households in the sample, 202 (one-eighth) lease land in, and 57 (3 percent) household lease land out. Sharecropping is exceedingly rare; payment is often in the form of straight rent payments, though land is also sometimes shared between households without expectation of any payment. In Nicaragua, the situation is similar: 3 percent of households provided land to someone else; half of the plots provided in this way were explicitly rented out, and the other half were shared or lent or something else of such nature. Leasing in is slightly more common: 9 percent received land for agricultural use.

Agricultural Assets Other Than Land

As farming is the most important economic activity for most of the household respondents, estimates were made of the value of farm-related assets held by the sample households. The households in all three country samples have multiple agricultural activities, such as crop production, livestock rearing, aquaculture, and plantation. Production-related tools and equipments are treated as agricultural assets. Trees whose products are not sold, however, are considered household assets. Table D.16 provides estimates of agricultural assets. The average household agricultural asset is highest in Nicaragua, although it includes only aquaculture and livestock, followed by Sri Lanka, which includes crops, trees, livestock,

Table D.16 Distribution (%) of Households by Total Value of Agricultural Assets

Value of Agricultural assets (US $)	Sri Lanka	Nicaragua**	Tanzania*	
	% of households	% of households	Agricultural asset	% of households
<50	83.8	80.1	<10	60.8
50–00	4.1	5.8	10–20	15.5
100–150	2.0	2.0	20–30	4.8
150–200	1.6	1.5	30–40	2.1
200–250	0.7	1.0	40–50	2.6
250–300	1.3	0.7	50–60	1.5
300–350	0.7	0.7	60–70	1.5
350–400	0.3	0.7	70–80	0.7
400–450	0.5	0.6	80–90	10.6
>450	5.1	7.0	>90	14.2
Total	100.0	100.0		100.0
Average	274	339.1		96.0
St. Dev.	1042	3106.8		1123.8
N	998	1535		1611

Source: RIC Surveys.

* Does not include tree assets, as data were collected in counts only.

** Includes only aquaculture and livestock assets. Data does not allow estimating crop assets and tree assets.

and aquaculture. Tanzania has the lowest level of agricultural assets. The distribution of households by agricultural asset group show that almost 84 percent of households in Sri Lanka and 80 percent of households in Nicaragua own less than US$50, and only 61 percent of households in Tanzania own less than US$50 worth of agricultural assets.

Durable Assets

Durable assets include durable consumables like electronics, furniture, transport equipments, and so on. Table D.17 provides distributions of households by durable asset group. The durable asset base is the highest in Sri Lanka. The average durable asset is more than 13 times higher in Sri Lanka than in Nicaragua and more than 5 times higher in Nicaragua than Tanzania. The distribution of households by durable assets is skewed to the left in all the countries.

Value of Residential Houses

Table D.18 provides the distribution of value of houses owned by the sample households. The average value of houses is higher in Sri Lanka

Table D.17 Distribution (%) of Households by Value of Durable Assets

Sri Lanka		Tanzania		Nicaragua	
Durable asset	**% of households**	**Durable asset**	**% of households**	**Durable asset**	**% of households**
<500	47.60	<5	49.78	<10	12.70
500–1000	18.72	5–10	20.36	10–20	14.92
1000–1500	11.57	10–15	9.93	20–30	10.68
1500–2000	6.21	15–20	4.90	30–40	9.38
2000–2500	4.42	20–25	3.10	40–50	6.32
2500–3000	2.82	25–30	1.92	50–60	6.06
3000–3500	1.41	30–35	1.37	60–70	6.84
3500–4000	0.94	35–40	0.87	70–80	4.89
>4000	6.30	>40	7.76	>80	28.21
Total	100.00		100.00		100.00
Average	1072		16.57		104.89
SD	1836		56.63		445.42
N	998		1611		1535

Source: RIC Surveys.

Table D.18 Distribution of Households by House Value

Sri Lanka		Tanzania		Nicaragua	
House value	**% of households**	**House value**	**% of households**	**House value**	**% of households**
<1000	32.64	<100	37.93	<500	14.72
1000–2000	17.12	100–200	9.06	500–1000	18.31
2000–3000	11.19	200–300	7.45	1000–1500	9.32
3000–4000	7.71	300–400	4.22	1500–2000	17.65
4000–5000	9.13	400–500	5.34	2000–2500	3.52
5000–6000	4.42	500–600	2.79	2500–3000	7.62
6000–7000	1.41	600–700	1.86	3000–3500	5.21
8000–8000	3.67	700–800	2.61	3500–4000	4.17
>8000	12.70	>800	28.74	>4000	19.48
Total	100		100		100
Average	4481		1683		3851
SD	5821		6364		13782
	998		1611		1535

Source: RIC Surveys.

(US$4,481) than in Nicaragua (US$3,851); it is lowest (US$1,683) in Tanzania. In Sri Lanka the majority (61 percent) of households live in houses worth less than US$3,000, and 13 percent live in highly valued houses worth more than US$8,000. In Nicaragua, the majority of households (60 percent) live in low-quality houses worth less than US$2,000, and almost 20 percent of households live in high-quality houses worth US$4,000. In Tanzania, the value of houses in general is much lower than in the other two countries. Almost 60 percent of households have houses valued below US$400; only 29 percent of Tanzania's households have houses worth more than US$800.

Total Assets

As shown in Table D.19, the average total asset holding is highest in Sri Lanka, followed by Nicaragua. The average asset holding is the lowest in Tanzania: 43 percent of households there own less than US$500 worth of assets. Contrary to this, 27 percent of households in Nicaragua own less than US$1,000 worth of assets, and 47 percent own between US$1,000 to US$4,000 worth of assets. In Sri Lanka 11 percent of the households own US$1,000, and 30 percent own between US$1,000 to US$4,000 worth of assets.

Table D.19 illustrates substantial inequality in the households' asset holdings. It is not clear how countries rank in the level of inequality. Table D.20 and Figure D.1 shed light on this. As shown in Table D.20, Sri Lanka exhibits a much lower coefficient of variation and Gini coefficientc; Nicaragua and Tanzania are similar. Figure D.1 makes the comparison visual by standardizing the distributions at their mean value. Tanzania and Nicaragua's distributions look very similar; the lowest value is more pronounced in Tanzania, but the categories right above it show a higher frequency in Nicaragua. Sri Lanka has a much more even distribution in the diagram's left half.

Table D.19 Distribution (%) of Households by Total Assets

Sri Lanka		Tanzania		Nicaragua	
Total asset	**% of households**	**Total asset**	**% of households**	**Total asset**	**% of households**
<1000	11.38	<500	43.08	<1000	26.84
1000–2000	9.88	500–1000	20.67	1000–2000	24.95
2000–3000	11.29	1000–1500	7.88	2000–3000	12.51
3000–4000	9.22	1500–2000	5.28	3000–4000	9.12
4000–5000	8.75	2000–2500	4.16	4000–5000	4.50
5000–6000	6.68	2500–3000	2.23	5000–6000	3.78
6000–7000	6.02	3000–3500	1.55	6000–7000	2.41
7000–8000	4.42	3500–4000	1.61	7000–8000	1.76
>8000	32.36	>4000	13.53	>8000	14.14
Total	100		100		100
Average	7332		2462.62		6453.94
St.Dev.	8331		7471.38		24883.92
N	998		1611		1535

Source: RIC Surveys.

Table D.20 Measures of Inequality in Total Assets

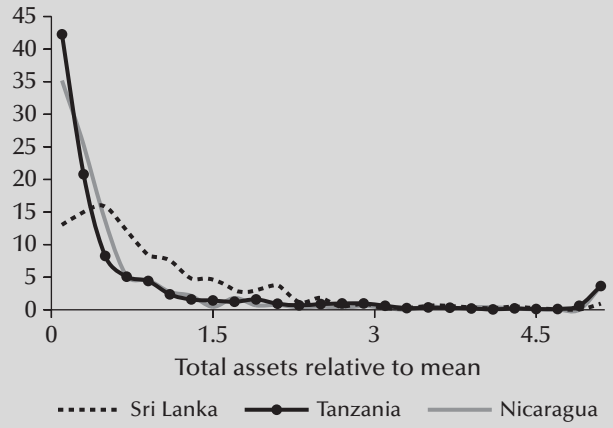
	Sri Lanka	Tanzania	Nicaragua
Coefficient of variation	1.103	3.034	3.610
Gini Coefficient	0.590	0.838	0.821

Source: RIC Surveys.

Household Investment

Household investment has been considered as net addition to total assets over a one-year period in categories of agricultural production, including crop production, aquaculture, and livestock; of net additions to house buildings; and of durable assets. Table D.21 provides distribution of households by investment. An average household

Figure D.1 Comparing Distribution of Total Assets Across Countries



Source: RIC Surveys.

Note: [a] Total assets are standardized by the country's mean in order to compare the distributions. The right tail is accumulated at a value of 5.0

Table D.21 Household Investment by Category

Investment in		Sri Lanka	Tanzania	Nicaragua
Agricultural assets	Mean	56	27	n.a.
	St.dev.	699	950	
	% positive	24.9	41.6	
	% negative	4.2	14.22	
Durable assets	Mean	80	1	n.a
	St.dev.	347	16	
	% positive	35.7	57.2	
	% negative	24.9	32.6	
Housing	Mean	n.a.	16	n.a.
	St.dev.		136	
	% positive		10.6	
	% negative		0.0	
Total assets	Mean	136	44	n.a.
	St.dev.	783	959	
	% positive	51.7	68.4	
	% negative	3.3	26.3	

Source: RIC Surveys.

increased its assets by US$136 in Sri Lanka and by US$44 in Tanzania; the large standard deviations indicate how widely household investment varies across households. Measurement error may of course be to blame: it is always difficult to obtain accurate financial statistics of assets, let alone of investment. Unfortunately, the Nicaragua survey data do not permit an estimate of household investment.

Household Income

Before describing household incomes, it is important to note that the Sri Lanka questionnaire had a flaw that most likely causes income components other than wages and salaries and farming to be understated. In particular, this flaw caused information on enterprise income and other income to be gathered only if at least one member of the household was employed for a wage. As a result, incomes are likely understated for a significant number of households. Indeed, 134 households show a total income reported of 0, of which 119 report nonfarm enterprise employment. In the statistics reported below, households with 0 or negative incomes are removed from the analysis. Secondly, for Nicaragua, six enterprises with negative total household incomes are removed from the calculation of income-related statistics. Third, the Tanzania statistics do not use sampling weights, and the stratification is explicitly based on a feature (entrepreneurship) related to income. Therefore, income statistics of these two countries and comparisons with the third should be considered with appropriate caution. Fourth, it should be noted that income statistics were reported by household respondents, which, in the case of Sri Lanka and Tanzania, is separate from any information recorded in the enterprise questionnaire. The Nicaragua questionnaire merged enterprise income and expenditure information into the household questionnaire for small, household-based enterprises.

Further, employment and earnings do not always go hand in hand in the survey data. It is fairly typical of household surveys to find discrepancies, which may be caused by response error, enumerator error, data-entry problems, unclear questionnaire design, phrasing of questions, and so on. The degree of inconsistency is demonstrated in Table D.22. The questionnaire design error in Sri Lanka is obvious: members of 315 household are involved in nonfarm self-employment but did not report income—half of all households with employment in an enterprise. Inconsistencies in Tanzania also run high: 165 households report nonfarm self-employment but no enterprise income, and 269 record enterprise income but no employment. The Nicaragua data are more consistent, perhaps because the enterprise and household questionnaires were blended.

Table D.22 Consistency in Reports on Household Enterprise Activity

Household reports income from nonfarm self-employment	Household reports activity in nonfarm self-employment								
	Sri Lanka			Tanzania			Nicaragua		
	No	Yes	Total	No	Yes	Total	No	Yes	Total
No	447	315	762	448	165	613	440	36	476
Yes	14	287	301	269	729	998	59	1000	1,059
Total	461	602	1,064	717	894	1,611	499	1036	1535

Source: RIC Surveys.

Note: These statistics are raw counts of households and therefore unweighted.

In a normal situation, the presence of a household enterprise is judged by the reported receipt of income from a nonfarm business or from nonfarm self-employment activities. In Sri Lanka, the design of the questionnaire led to nonresponses on self-employment and other income because of an erroneous skip pattern. As a result, the criterion of assuming the presence of a nonfarm enterprise by income is biased.

Table D.23 describes household income and its component for each country. The distribution is divided into different intervals for each country, which are selected to highlight the contribution of each component effectively. Separate categories of *0* and *Negative* describe households that received no income from a given source or for which expenditures from the given activity (for example, in Tanzania, farming as well as wage earnings) exceeded the revenues it generated.

In Table D.23 income is divided into five components. Households earn income from wages and salary; self-employment; farming operations, which include crop production, aquaculture and livestock, and plantation; remittance income from friends and relatives; and "other income," which includes interest income from savings and dividends, lotteries, gifts from friends and relatives, receipts from government, and retirement benefits. Households may also have many indigenous expenses that cannot be categorized as productive expenses and thus were excluded from *gross other income* to estimate *other net income.*

Mean income among rural households equals US$2,696 in Sri Lanka, US$670 in Tanzania, and US$2,016 in Nicaragua. In every country, standard deviations exceed the mean, and the median is substantially less than the mean. These are indications that the income distribution has a long right tail. Median household income in Tanzania is only US$247.

In Nicaragua, nearly half of household income derives from wage earnings. The next largest contributor is nonfarm enterprise income at nearly one-fourth. Farm income and remittances each contribute about 11.5 percent.

In Tanzania, nonfarm enterprises contribute nearly 40 percent of average household income, farming 30 percent, and wage earnings 12 percent. It may seem surprising that nonfarm enterprise income is so dominant, but it should be remembered that the Tanzanian RICS survey oversampled households with enterprises but did not provide sampling weights; this prevents the sample from being representative of the Tanzanian rural population and may bias any statistics computed from it, especially if the statistic is related to the entrepreneurial status of the household, as is clearly the case here.

In Sri Lanka, remittances are the largest average contributor to household income, even though nearly 88 percent of the households do not receive any remittance income. Wage income contributes about one-third; nonfarm enterprises contribute less than one-tenth. But, as mentioned, these income statistics for Sri Lanka should be treated with great caution since they are seriously biased.

Partly in response to the finding that remittance income appears to be so important in Sri Lanka, appearing at the top of the income distribution, Figure D.2 examines the functional distribution of household income: it considers the contribution of each component to total household income within quintiles. Clearly, Sri Lankan households in the upper quintile benefit greatly from remittances. Enterprise income expands slightly from quintiles 1 to 4; wage income is most important for

Table D.23 Total Household Income and Its Components

A. Sri Lanka

	Wage	Enterprise	Farm	Remittance	Other	Total
Negative	0.00	0.00	1.93	0.00	0.00	n.a.
0	21.18	82.09	45.32	87.73	59.53	n.a.
0–500	22.68	8.71	35.51	1.07	33.95	20.85
500–1000	27.24	3.78	6.93	2.01	1.51	22.64
1000–1500	14.11	3.09	3.58	0.38	1.56	17.42
1500–2000	5.93	1.24	1.63	0.16	1.23	9.84
2000–2500	4.88	0.05	2.68	1.67	0.04	6.44
2500–3000	1.89	0.42	0.24	0.54	0.61	6.05
3000–3500	1.06	0.00	0.57	0.41	0.08	3.84
3500–4000	0.17	0.01	0.13	0.75	0.30	1.24
4000–4500	0.03	0.13	0.19	0.15	0.14	1.64
4500–5000	0.42	0.13	0.39	0.00	0.23	1.05
>5000	0.41	0.34	0.91	5.11	0.83	8.98
Mean	835	162	343	1088	268	2696
Standard dev	1237	565	904	6386	1544	7883
Median	620	0	10	0	0	1216

B. Tanzania

	Wage	Enterprise	Farm	Other	Total
Negative	0.93	0.00	19.12	0.00	4.84
0	86.41	38.05	6.08	58.66	0.37
0–200	4.97	36.00	56.61	28.06	40.29
200–400	1.49	10.80	9.37	4.78	16.82
400–600	1.12	5.77	3.41	2.98	11.17
600–800	1.18	2.79	1.99	1.37	6.83
800–1000	0.93	1.55	0.62	1.30	3.48
1000–1200	0.81	1.06	0.25	0.87	2.86
1200–1400	0.56	0.68	0.31	0.50	2.17
1400–1600	0.37	0.19	0.37	0.25	1.80
1600–1800	0.37	0.50	0.31	0.31	1.99
1800–2000	0.19	0.43	0.19	0.25	1.06
>2000	0.68	2.17	1.37	0.68	6.33
Mean	81	261	203	126	670
Standard deviation	392	739	1469	500	1873
Median	0	55	42	0	247

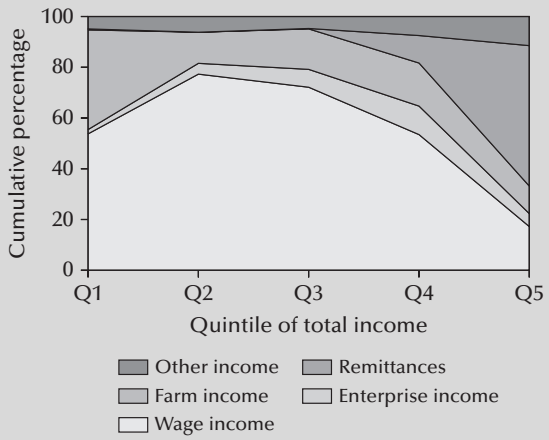
C. Nicaragua

	Wage	Enterprise	Farm	Remittance	Other	Total
Negative	0.00	0.4	4.47	0.00	0.00	0.23
0	30.69	77.43	66.49	68.94	77.43	0.16
0–600	19.41	7.03	22.1	22.07	15.36	20.69
600–1200	18.41	6.21	3.97	4.72	5.11	26.42
1200–1800	13.05	3.14	0.54	1.72	0.76	18.47
1800–2400	8.13	1.62	0.67	0.78	0.75	10.14
2400–3000	3.46	0.69	0.35	0.75	0.01	7.34
3000–3600	2.18	0.62	0.28	0.16	0.2	3.82
3600–4200	1.98	0.54	0.2	0.02	0.05	4.13
4200–4800	0.66	0.51	0.05	0.39	0.04	1.98
4800–5400	0.41	0.23	0.05	0.02	0	0.92
5400–6000	0.42	0.39	0.14	0.06	0.02	1.14
>6000	1.21	1.18	0.7	0.38	0.28	4.58
Mean	988	434	228	230	136	2016
Standard dev.	1363	2247	1527	1089	648	3444
Median	586	0	0	0	0	1237

Source: RIC Surveys.

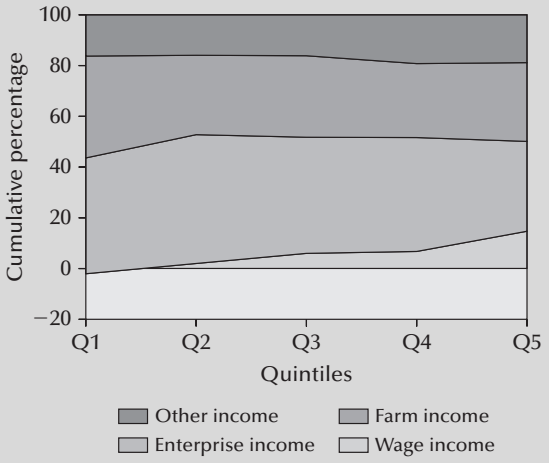
Figure D.2 Income Shares as a Percentage of Total Household Income, by Quintile

A. Sri Lanka[a]



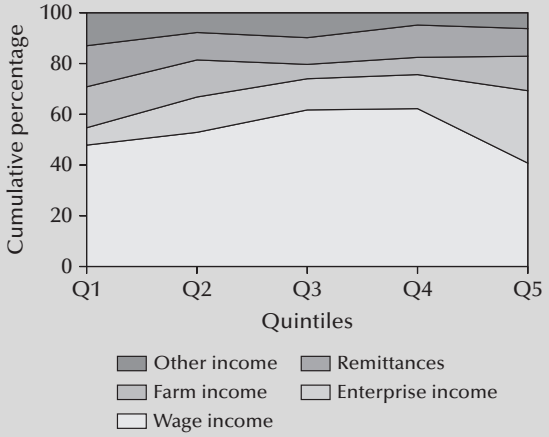
Note: [a] Households with negative or zero total incomes have been omitted.

B. Tanzania[a]



Note: [a] Households with zero or negative total income are omitted, but a few households reported expenditures for earning a wage income that exceed the wage income itself. This causes a curious negative wage share at the lowest quintile.

C. Nicaragua



households in quintiles 2 and 3. Farm income is most important in the lowest quintile and diminishes from there on.

In Tanzania, enterprise income dominates, which is to be expected as the survey oversampled enterprise households. The share of farm income is roughly constant across the five quintiles; wage income expands. In Nicaragua, the share of remittance income is surprisingly constant across households. The enterprise share expands; the wage share rises until sharply dropping in the fifth quintile, and the share of farm income drops steadily.

The concept of household income is useful when sources are brought together and aggregated. After all, households often gather income from all sources together and then allocate it to the various needs among household members. But to express the standard of living of these household members better, it is of course common to compute household income per capita. Table D.24 reports those statistics. Averages range from US$164 in Tanzania to US$661 in Sri Lanka, with Nicaragua in between at US$483. Median per capita incomes are actually about the same in Sri Lanka and Nicaragua, at slightly over US$280; in Tanzania, the median is only US$56.

In Tanzania, where the median per capita income is US$56, 5 percent of households have a per capita income of over US$600. A similar contrast exists in Sri Lanka, where more than 4 percent of households exceed US$2,000 in per capita income when the median is US$282. The contrast seems less pronounced in Nicaragua. Table D.25 formalizes this comparison by means of the coefficient of variation and the Gini coefficient. Indeed, Tanzania and Sri Lanka have more unequally distributed per capita incomes. Figure D.3 makes this more visual, describing the distribution of per capita income in standardized form by dividing each household's income by the country's mean.

Table D.24 Per Capita Income and Its Components

A. Sri Lanka

	Wage	Enterprise	Farm	Remittance	Other	Total
Negative	0.00	0.00	1.93	0.00	0.00	n.a.
0	21.18	82.09	45.32	87.73	59.53	n.a.
0–200	41.12	12.50	41.47	3.05	34.97	37.69
200–400	23.69	2.88	4.46	0.63	2.69	26.51
400–600	8.83	1.21	2.60	0.64	0.72	11.96
600–800	2.18	0.52	2.63	0.68	1.12	7.07
800–1000	1.37	0.54	0.38	0.55	0.14	2.75
1000–1200	0.39	0.21	0.35	2.05	0.18	2.58
1200–1400	0.59	0.02	0.06	0.32	0.00	3.64
1400–1600	0.19	0.00	0.60	0.48	0.00	1.10
1600–1800	0.00	0.02	0.20	0.94	0.00	1.47
1800–2000	0.00	0.00	0.00	0.33	0.00	1.13
>2000	0.46	0.00	0.00	2.60	0.66	4.11
Mean	212	38	85	265	61	661
Standard dev	337	128	209	1576	375	1943
Median	143	0	2	0	0	282

B. Tanzania

	Wage	Enterprise	Farm	Other	Total
Negative	0.93	0.00	19.12	0.00	4.84
0	86.41	38.05	6.08	58.66	0.37
0–60	5.77	40.35	60.40	29.55	46.74
60–120	2.11	9.62	8.01	5.71	17.13
120–180	1.61	4.22	2.48	1.92	10.24
180–240	0.81	2.67	0.93	1.37	5.46
240–300	0.68	1.06	0.50	1.06	3.85
300–360	0.50	0.56	0.56	0.43	1.99
360–420	0.43	0.68	0.25	0.19	1.49
420–480	0.25	0.43	0.43	0.19	1.43
480–540	0.12	0.43	0.12	0.12	0.62
540–600	0.06	0.25	0.06	0.00	0.87
>600	0.31	1.68	1.06	0.81	4.97
Mean	16	64	52	32	164
Standard dev	75	222	464	194	570
Median	0	11	9	0	56

C. Nicaragua

	Wage	Enterprise	Farm	Remittance	Other	Total
Negative	0.00	0.4	4.47	0.00	0.00	0.23
0	30.69	77.43	66.49	68.94	77.43	0.16
0–250	35.45	13.16	25.56	24.86	18.77	41.41
250–500	21.32	4.18	1.71	3.59	2.37	29.89
500–750	6.99	1.82	0.42	0.89	0.75	13.47
750–1000	2.53	0.68	0.18	1.1	0.04	5.96
1000–1250	1.36	0.66	0.22	0	0.36	2.7
1250–1500	1.05	0.75	0.27	0.02	0.02	2.29
1500–1750	0.08	0.21	0.16	0.28	0.05	0.81
1750–2000	0.03	0.2	0.01	0.06	0.00	0.63
2000–2250	0.08	0.03	0.16	0	0.00	0.51
2250–2500	0	0.09	0	0	0.00	0.22
>2500	0.42	0.39	0.36	0.25	0.22	1.71
Mean	226	111	49	57	40	483
Standard dev.	325	984	361	221	208	1133
Median	126	0	0	0	0	287

Source: RIC Surveys.

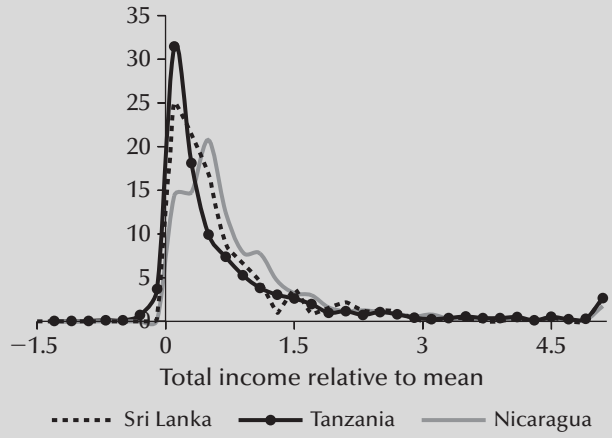
Table D.25 Measures of Inequality in per Capita Total Income

	Sri Lanka	Tanzania	Nicaragua
Coefficient of variation	2.942	3.468	2.347
Gini coefficient[a]	0.769	0.757	0.568

Source: RIC Surveys.

Note: a. Households with negative incomes are omitted from the computation.

Figure D.3 Comparing Distributions of Income per Capita, Normalized at the Country-Specific Mean



Source: RIC Surveys.

Sri Lanka and Tanzania show a more pronounced clustering of households at very low income levels. It should be noted that the communities in Nicaragua are somewhat larger than those in Tanzania and Sri Lanka.

Another perspective on household income is gained by expressing income per worker and by distinguishing households with and without a nonfarm enterprise. Table D.26 makes clear that total income in households with enterprises is higher than in those without them, both in total and per capita or per worker. The exception is Sri Lanka's (w) definition, but this should be discounted, since enterprise income is understated as a result of the questionnaire design flaw.

Table D.26 Income Total, per Capita and per Worker, by Category and Presence of Nonfarm Enterprise

	Sri Lanka					Tanzania			Nicaragua		
	All hh	hh with NFHE(w)	hh without NFHE(w)	hh with NFHE(i)	hh without NFHE(i)	All hh	hh with NFHE(i)	hh without NFHE(i)	All hh	with NFHE(i)	hh without NFHE(i)
Household income											
Total	2697	2443	2779	3228	2581	670	749	542	2025	3120	1705
Wage income	836	805	846	1132	771	81	71	97	992	677	1083
Enterprise income	163	637	8	906	0	261	421	0	436	1927	0
Farm income	344	358	339	308	351	202	163	265	231	259	222
Other	1355	643	1586	881	1459	126	93	179	367	257	399
Income per capita											
Total	661	601	680	788	633	164	182	135	485	743	410
Wage income	212	194	217	272	199	16	15	19	227	128	256
Enterprise income	38	148	2	211	0	64	103	0	111	493	0
Farm income	85	86	85	69	89	52	39	71	49	67	44
Other	326	173	376	237	345	32	25	45	98	56	110
Income per worker											
Total	1572	1095	1731	1424	1605	310	351	241	1260	1541	1166
Wage income	727	805	706	801	706	237	196	314	970	838	998
Enterprise income	545	545	n.a.	770	0	257	315	0	1151	1297	0
Farm income	580	524	602	587	579	117	82	172	384	569	338

Source: RIC Surveys.

Note: (w) denotes that the presence of a household enterprise is judged by reported employment in nonfarm self-employment activities; (i) indicates that the presence of a household enterprise is judged by the reported receipt of income from a nonfarm business or from nonfarm self-employment activities.

Relation to Benchmark Indicators and Population Size

Table D.27 provides mean values of income and employment variables by quintiles of the benchmark indicators and population size. This table contains a wealth of information that will not be discussed here. Suffice it to say that the trend is rarely uniformly upward or downward, but that many suggestions of trends are manifest. These trends are equivalent to simple correlations, but they should be explored in a multivariate context, as, for example, household income or entrepreneurship rates are influenced by a conglomerate of factors. Therefore, to quantify these trends more carefully is a topic for a more advanced regression analysis.

Table D.27 Household Income and Employment by Benchmark Quintiles and Population Size

A. Sri Lanka

	1	2	3	4	5	Total
Connectivity						
Total household income	1726	2048	1553	1652	4621	2457
Enterprise income	141	148	86	106	240	148
Wage income	738	660	810	775	758	756
Farm income	483	459	166	434	112	308
Other income	46	222	240	96	530	246
Remittance income	317	558	252	241	2981	999
Other + remittance income	363	781	492	337	3511	1246
Activity: nfhe	0.187	0.311	0.224	0.286	0.378	0.284
Activity: wage	0.715	0.743	0.782	0.709	0.794	0.752
Activity: farm	0.590	0.398	0.390	0.587	0.232	0.429
Infrastructure services						
Total household income	2219	1401	1598	1935	3937	2421
Enterprise income	263	94	105	104	172	146
Wage income	823	679	678	760	803	752
Farm income	573	351	299	445	53	302
Other income	92	46	156	184	516	241
Remittance income	468	231	361	441	2393	979
Other + remittance income	560	278	516	625	2909	1221
Activity: nfhe	0.273	0.211	0.349	0.264	0.331	0.293
Activity: wage	0.653	0.704	0.674	0.845	0.802	0.747
Activity: farm	0.575	0.587	0.396	0.543	0.209	0.426
Business services						
Total household income	1621		5840	3070	1184	2453
Enterprise income	140		232	118	115	147
Wage income	708		593	1186	572	744
Farm income	270		520	241	287	308
Other income	129		916	107	107	247
Remittance income	375		3580	1418	103	1007
Other + remittance income	504		4495	1524	210	1255
Activity: nfhe	0.326		0.315	0.220	0.265	0.299
Activity: wage	0.777		0.687	0.724	0.683	0.740
Activity: farm	0.427		0.408	0.366	0.551	0.433
Corruption/governance						
Total household income	1643	2930	1768	1505	4979	2557
Enterprise income	97	145	100	206	65	125
Wage income	807	953	724	579	745	759
Farm income	337	503	210	420	116	320
Other income	261	73	197	143	682	264
Remittance income	141	1256	536	157	3371	1090
Other + remittance income	402	1329	734	300	4053	1354
Activity: nfhe	0.191	0.264	0.307	0.314	0.186	0.257
Activity: wage	0.804	0.764	0.712	0.643	0.818	0.744
Activity: farm	0.473	0.627	0.294	0.530	0.340	0.453

Table D.27 Household Income and Employment by Benchmark Quintiles and Population Size (*continued*)

A. Sri Lanka

	1	2	3	4	5	Total
Human capital						
Total household income	1550	1335	2756	1373	5143	2421
Enterprise income	79	268	193	91	128	146
Wage income	635	670	1082	672	724	752
Farm income	323	231	492	156	291	302
Other income	36	64	229	102	800	241
Remittance income	478	102	761	352	3199	979
Other + remittance income	514	166	990	454	4000	1221
Activity: nfhe	0.219	0.400	0.331	0.278	0.267	0.293
Activity: wage	0.797	0.753	0.747	0.697	0.727	0.747
Activity: farm	0.517	0.324	0.435	0.388	0.432	0.426
Finance						
Total household income	1635	1971	2479	1699	4889	2557
Enterprise income	143	77	167	137	97	125
Wage income	685	751	968	685	671	759
Farm income	296	200	327	479	317	320
Other income	42	129	179	285	700	264
Remittance income	469	814	838	114	3103	1090
Other + remittance income	510	943	1017	399	3803	1354
Activity: nfhe	0.262	0.214	0.317	0.219	0.260	0.257
Activity: wage	0.792	0.780	0.762	0.714	0.666	0.744
Activity: farm	0.450	0.393	0.504	0.515	0.406	0.453
Population size						
Total household income	1498	1860	2589	1561	3560	2422
Enterprise income	64	93	249	128	159	146
Wage income	683	652	1053	764	660	753
Farm income	368	386	387	247	225	302
Other income	210	65	124	93	529	241
Remittance income	172	665	776	330	1986	980
Other + remittance income	382	730	900	422	2516	1221
Activity: nfhe	0.200	0.274	0.347	0.256	0.329	0.293
Activity: wage	0.581	0.741	0.694	0.766	0.813	0.747
Activity: farm	0.600	0.463	0.540	0.419	0.295	0.426

B. Tanzania

	1	2	3	4	5	Total
Connectivity						
Total household income	501	619	723	839	632	658
Enterprise income	113	255	219	418	256	249
Wage income	74	69	62	85	130	83
Farm income	175	144	367	211	89	196
Other income	138	149	74	124	156	129
Activity: nfhe	0.410	0.492	0.587	0.655	0.673	0.555
Activity: wage	0.242	0.210	0.323	0.297	0.382	0.286
Activity: farm	0.938	0.964	0.910	0.826	0.831	0.897
Infrastructure services						
Total household income	630	441	756	646	830	658
Enterprise income	119	182	307	295	375	249
Wage income	87	32	114	39	142	83
Farm income	296	127	181	171	190	196
Other income	128	97	153	140	122	129
Activity: nfhe	0.469	0.498	0.534	0.631	0.681	0.555
Activity: wage	0.216	0.243	0.309	0.292	0.392	0.286
Activity: farm	0.963	0.951	0.877	0.854	0.821	0.897

Table D.27 Household Income and Employment by Benchmark Quintiles and Population Size (*continued*)

B. Tanzania

	1	2	3	4	5	Total
Business services						
Total household income	628			450	1105	666
Enterprise income	246			198	374	256
Wage income	95			78	26	84
Farm income	150			82	603	198
Other income	136			92	100	126
Activity: nfhe	0.555			0.565	0.529	0.553
Activity: wage	0.267			0.364	0.314	0.284
Activity: farm	0.899			0.913	0.880	0.899
Corruption/governance						
Total household income	512	545	768	630	859	664
Enterprise income	240	287	282	280	205	259
Wage income	67	43	147	51	107	84
Farm income	126	132	171	170	410	202
Other income	75	81	167	128	137	118
Activity: nfhe	0.599	0.593	0.567	0.548	0.571	0.575
Activity: wage	0.310	0.248	0.317	0.314	0.287	0.295
Activity: farm	0.881	0.894	0.917	0.860	0.924	0.895
Human capital						
Total household income	445	541	578	625	1233	670
Enterprise income	158	252	223	307	383	261
Wage income	20	31	92	74	205	81
Farm income	170	170	107	125	460	202
Other income	97	87	154	117	184	126
Activity: nfhe	0.397	0.504	0.545	0.646	0.715	0.555
Activity: wage	0.203	0.172	0.266	0.360	0.451	0.284
Activity: farm	0.957	0.939	0.872	0.868	0.831	0.896
Finance						
Total household income	589	624	519	786	832	664
Enterprise income	191	275	269	316	255	260
Wage income	78	76	45	91	145	85
Farm income	160	136	106	280	329	197
Other income	159	137	99	98	102	120
Activity: nfhe	0.550	0.558	0.605	0.580	0.585	0.575
Activity: wage	0.287	0.293	0.319	0.269	0.307	0.295
Activity: farm	0.886	0.890	0.911	0.926	0.859	0.895
Population size						
Total household income	581	399	928	673	726	663
Enterprise income	228	206	245	248	349	252
Wage income	72	45	105	144	53	84
Farm income	121	79	425	115	228	196
Other income	160	65	153	165	95	130
Activity: nfhe	0.492	0.540	0.568	0.607	0.608	0.559
Activity: wage	0.293	0.211	0.330	0.311	0.285	0.287
Activity: farm	0.867	0.872	0.940	0.929	0.878	0.897

Table D.27 (*continued*)

C. Nicaragua

	1	2	3	4	5	Total
Connectivity						
Total household income	1053	1487	2589	2592	2422	2068
Enterprise income	133	245	667	620	573	458
Wage income	586	880	840	1172	1245	974
Farm income	213	225	422	341	75	252
Other income	68	60	181	247	98	136
Remittance income	53	78	480	212	430	247
Other + remittance income	121	137	661	459	528	383
Activity: nfhe	0.143	0.166	0.301	0.266	0.264	0.227
Activity: wage	0.746	0.659	0.611	0.693	0.706	0.685
Activity: farm	0.774	0.531	0.279	0.252	0.071	0.377
Infrastructure services						
Total household income	1562	1730	2040	2197	2547	2068
Enterprise income	175	318	443	398	791	458
Wage income	732	931	852	1011	1242	974
Farm income	356	261	333	274	95	252
Other income	58	144	251	86	135	136
Remittance income	241	75	162	426	284	247
Other + remittance income	299	219	413	512	420	383
Activity: nfhe	0.143	0.197	0.223	0.277	0.280	0.227
Activity: wage	0.838	0.699	0.500	0.632	0.744	0.685
Activity: farm	0.673	0.551	0.359	0.251	0.141	0.377
Business services						
Total household income	1766			2879	2383	2068
Enterprise income	299			860	638	458
Wage income	858			1107	1185	974
Farm income	221			678	113	252
Other income	142			126	126	136
Remittance income	246			108	320	247
Other + remittance income	388			234	446	383
Activity: nfhe	0.202			0.306	0.245	0.227
Activity: wage	0.696			0.625	0.692	0.685
Activity: farm	0.449			0.348	0.213	0.377
Corruption/governance						
Total household income	1661	1774	2424	2166	2153	2070
Enterprise income	397	412	508	458	489	458
Wage income	868	672	1003	1120	1168	975
Farm income	169	190	457	239	142	253
Other income	100	212	163	118	83	136
Remittance income	127	289	293	232	270	248
Other + remittance income	227	500	456	350	354	384
Activity: nfhe	0.195	0.199	0.234	0.249	0.243	0.226
Activity: wage	0.737	0.494	0.744	0.697	0.717	0.685
Activity: farm	0.374	0.490	0.361	0.363	0.317	0.377

Table D.27 Household Income and Employment by Benchmark Quintiles and Population Size (*continued*)

C. Nicaragua

	1	2	3	4	5	Total
Human capital						
Total household income	800	1674	2139	2280	3246	2068
Enterprise income	92	238	593	586	707	458
Wage income	439	977	995	1052	1322	974
Farm income	205	147	218	158	522	252
Other income	25	100	146	138	251	136
Remittance income	39	213	188	346	443	247
Other + remittance income	64	313	333	484	694	383
Activity: nfhe	0.125	0.172	0.254	0.276	0.303	0.227
Activity: wage	0.675	0.724	0.658	0.667	0.703	0.685
Activity: farm	0.747	0.500	0.232	0.236	0.210	0.377
Finance						
Total household income	1513	1896	2019	2390	2682	2068
Enterprise income	193	269	492	517	834	458
Wage income	691	1307	1106	994	1257	974
Farm income	263	103	204	380	145	252
Other income	88	180	62	210	185	136
Remittance income	277	37	155	289	262	247
Other + remittance income	366	216	217	499	446	383
Activity: nfhe	0.168	0.166	0.275	0.241	0.275	0.227
Activity: wage	0.686	0.750	0.732	0.625	0.707	0.685
Activity: farm	0.530	0.408	0.280	0.403	0.172	0.377
Population size						
Total household income	1436	948	2143	2230	2562	1950
Enterprise income	276	91	506	563	686	457
Wage income	786	567	967	816	1375	967
Farm income	138	141	279	320	87	177
Other income	72	61	204	244	129	138
Remittance income	165	88	187	287	285	212
Other + remittance income	236	148	390	531	414	349
Activity: nfhe	0.133	0.140	0.295	0.283	0.263	0.224
Activity: wage	0.711	0.709	0.610	0.607	0.743	0.687
Activity: farm	0.561	0.668	0.383	0.224	0.181	0.382

Source: RIC Surveys.

Annex E. Specimen Indexes Derived from RICS Data

The six benchmark indicators identified in Chapter 2 are defined below, with illustrations from the Sri Lanka pilot. The six indicators are:

1. Connectivity
2. Infrastructure services
3. Business services
4. Governance
5. Human capital
6. Finance services

Each benchmark indicator is based on a number of subindicator indexes, with values calculated at the community level. Most subindicator values are directly obtained from the community survey; some also use data from the enterprise and household surveys. The calculation is carried out in two steps. First, subindicators are calculated at the community level; and second, the subindicators are aggregated at the community level by taking their sum and dividing by the number of observations per community (missing observations are ignored).

1. Connectivity. The index of *rural-urban connectivity/ regional economic integration* is computed as the average of the following eight subindexes, transforming the average value such that the maximum equals 1 and scaling the value of the index for each community relative to this maximum. Note that each subindex is constructed such that a more remote community is characterized by a smaller value of the subindex. A scaling of the indexes has been selected in such a way that international comparison is possible.

1) Time taken by main means of transportation to the nearest major city.

The longer the time, the less integrated the community. Let X be the distance in minutes from a given community in the sample to a bigger town, and let Xmin be the minimum of X, and set Xmin = 10. Then we establish an index of connectivity for each observation as Xmin/X. If X < Xmin, the index is set at 1.

2) Relative cost of transportation to the nearest major city (by public transportation).

The higher the cost, the less integrated the community. Let X be the cost in US$ of a trip to the nearest city, and let Xmin be the minimum of X, and set Xmin = 0.05. Then an index of connectivity for each observation is defined as Xmin/X. If X < Xmin, the index is set at 1. If there is no public transportation, this indicator is coded as 0.

3) Time taken by main means of transportation to the main market.

The longer the time, the less integrated the community. Let X be the time from a given community in the sample to the main market, and let Xmin be the minimum of X, and set Xmin = 10. Then an index of connectivity for each observation is defined as Xmin/X. If X < Xmin, the index is set at 1.

4) Cost of transportation to the main market (by bus).

The lower the cost of transportation, the more integrated the community. Find the cost of transportation in the price questionnaire. Let X be the cost of transportation by bus in US$ from a given community in the sample to the main market, and let Xmin be the minimum of X, and set Xmin = 0.05. Then an index of connectivity for each observation is defined as Xmin/X. If X < Xmin, the index is set at 1.

5) Distance to the post office.

The lower the distance to the post office, the more integrated the community. Define D as the distance to the post office in kilometers. The index is of connectivity calculated as 1/D. If D < 1, the index is set at 1.

6) Rail stop in walking distance.

The connectivity subindex in this case is a dummy variable: 1 if there is a rail stop within walking distance, 0 otherwise.

Table E.1 Distribution of Connectivity Index Over Communities

		0 to .1	.1 to .2	.2 to .3	.3 to .4	.4 to .5	.5 to .6	.6 to .7	.7 to .8	.8 to .9	.9 to 1	N	average	standard deviation
	%	2.7	10.3	17.1	19.2	22.6	16.4	6.2	5.5	0.0	0.0	**146**	**0.40**	**0.17**
Connectivity Index	**mean**	0.057	0.160	0.258	0.341	0.453	0.539	0.644	0.733	.	.			
Time taken by main means of transportation to the nearest major city	%	9.6	15.1	7.5	16.4	4.1	15.1	13.7	0.0	0.0	18.5	146	0.47	0.31
	mean	0.069	0.143	0.236	0.333	0.409	0.500	0.667	.	.	1.000			
Cost of transportation to the nearest major city (by public transportation)	%	1.5	5.9	8.1	7.4	10.3	5.1	14.7	0.0	15.4	31.6	136	0.66	0.29
	mean	0.032	0.165	0.242	0.351	0.435	0.538	0.637	.	0.806	0.993			
Time taken by main means of transportation to the main market	%	9.7	12.4	14.5	12.4	1.4	11.7	17.2	0.0	0.0	20.7	145	0.49	0.32
	mean	0.070	0.146	0.243	0.333	0.427	0.500	0.667	.	.	1.000			
Cost of transportation to the main market (by bus)	%	15.8	2.7	3.4	5.5	11.6	4.8	10.3	0.0	13.7	32.2	146	0.61	0.36
	mean	0.007	0.146	0.249	0.353	0.433	0.538	0.647	.	0.806	0.995			
Distance to the post office	%	2.1	6.2	12.4	6.2	0.0	20.7	0.0	0.0	0.0	52.4	145	0.69	0.35
	mean	0.081	0.140	0.225	0.333	.	0.500	.	.	.	1.000			
Rail stop in walking distance	%	87.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	12.4	145	0.12	0.33
	mean	0.000	.	.	.	.	.	.	.	.	1.000			
Percentage of households with fixed telephone lines	%	67.6	16.6	9.0	2.8	2.1	0.7	0.7	0.7	0.0	0.0	145	0.08	0.12
	mean	0.022	0.122	0.215	0.333	0.417	0.500	0.600	0.750	.	.			
Percentage of households with cell phone	%	69.7	13.1	6.2	3.4	2.8	4.1	0.0	0.7	0.0	0.0	145	0.10	0.14
	mean	0.027	0.126	0.217	0.330	0.400	0.508	.	0.750	.	.			

Source: The Authors.

7) Percentage of households with fixed-line telephones.
This connectivity subindex is simply the share of households in the community with fixed-line telephones.

8) Percentage of households with a cell phone.
The connectivity subindex is the share of households with a cell phone.

2. *Infrastructure services.* The index that measures access to infrastructure services and service delivery is constructed as the average of the following eight subindexes. Note that each subindex is constructed such that a community with poorer access is characterized by a smaller value of the subindex.

1) Percentage of households that use electricity.
This subindex measures the share of households with electricity.

2) Availability of electricity.
This subindex measures the reliability of the electricity supply. Index = (4/number of blackouts per month) if the number of blackouts per month is greater than or equal to 4. The index equals 1 if the number of blackouts is less than 4, and it equals 0 if electricity is unavailable.

3) Percentage of households with access to drinking water from a protected source.
This subindex measures the share of households with access to drinking water from a protected pipe or well.

4) Percentage of households with a fixed-line telephone.
This subindex measures the share of households with fixed-line telephones.

5) Percentage of households with a cell phone.
This subindex measures the share of households with a cell phone.

*6) Sewage channels in the community.**
The access subindex in this case is a dummy variable: 1 if there are sewage channels, 0 otherwise.

*7) Garbage collection or disposal service in the community.**
This subindex is a dummy variable: 1 if there is garbage collector/disposal service, 0 otherwise.

8) *Most common road surface (internal road) is concrete or asphalt.*
The access subindex in this case is a dummy variable: 1 if the main road surface is concrete/asphalt, 0 otherwise.

*Access to sewage channels and garbage collection could potentially be a cost and not a benefit for businesses, depending on the cost of alternative ways of disposal. In the Sri Lanka report, access to sewage and garbage collection was considered a benefit.

3. *Business services.* The index that measures the availability of business services is derived from seven dummy variables, measured in each community.

1) Engineering services available for businesses in the community.
2) Management consulting services available for businesses in the community.
3) Marketing services available for businesses in the community.
4) Accounting services available for businesses in the community.
5) Legal services available for businesses in the community.
6) Insurance services available for businesses in the community.
7) Information services available for businesses in the community.

The community index is the sum of the indexes divided by the number of indexes for which data are available.

4. *Governance.* This index measures the existence of governance and corruption. A high index indicates a low level of administrative burden and corruption.

Five subindexes have been defined:

1) General policy and institutional constraints.
2) Infrastructure and services.
3) Dealing with government agencies.
4) Rule of law.
5) Conflict resolution and contract enforcement.

All subindexes are constructed from the enterprise survey data by taking the average of scores. Each community subindex is the average of observations from all businesses in that community. An aggregated index of corruption and governance is constructed at the community level from these subindexes. Missing scores are ignored in the calculations.

1) General policy and institutional constraints.
This subindex measures general policy and institutional constraints. The subindex is high if there are no constraints. Variables included are:

a. Corruption as a constraint for rural investment climate;

Table E.2 Distribution of Infrastructure Service Index Over Communities

Infrastructure services Index		0 to .1	.1 to .2	.2 to .3	.3 to .4	.4 to .5	.5 to .6	.6 to .7	.7 to .8	.8 to .9	.9 to 1	N	average	standard deviation
	%	6.2	13.0	19.2	28.8	14.4	13.0	2.7	2.7	0.0	0.0	**146**	**0.35**	**0.16**
	mean	0.027	0.154	0.253	0.356	0.446	0.545	0.632	0.750	.	.			
Percentage of households that use electricity	%	7.6	0.7	4.8	5.5	2.8	4.1	14.5	9.7	29.7	20.7	145	0.69	0.29
	mean	0.001	0.150	0.240	0.328	0.443	0.513	0.647	0.755	0.864	0.965			
Availability of electricity index	%	13.1	9.0	13.8	8.3	9.0	2.8	1.4	0.0	4.1	38.6	145	0.55	0.39
	mean	0.026	0.127	0.235	0.331	0.403	0.500	0.667	.	0.800	1.000			
Percentage of households with access to potable water	%	4.8	3.4	0.0	1.4	2.8	4.8	7.6	4.8	15.9	54.5	145	0.80	0.28
	mean	0.003	0.132	.	0.350	0.413	0.510	0.631	0.750	0.839	0.990			
Percentage of households with fixed-line telephone	%	67.6	16.6	9.0	2.8	2.1	0.7	0.7	0.7	0.0	0.0	145	0.08	0.12
	mean	0.022	0.122	0.215	0.333	0.417	0.500	0.600	0.750	.	.			
Percentage of households with access to cellular phones	%	69.7	13.1	6.2	3.4	2.8	4.1	0.0	0.7	0.0	0.0	145	0.10	0.14
	mean	0.027	0.126	0.217	0.330	0.400	0.508	.	0.750	.	.			
Sewage channels in the community	%	89.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11.0	145	0.11	0.31
	mean	0.000	.	.	.	.	.	.	.	.	1.000			
Garbage collection or disposal service in the community	%	88.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11.7	145	0.12	0.32
	mean	0.000	.	.	.	.	.	.	.	.	1.000			
Most common road surface (internal road) is concrete or asphalt	%	67.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	32.9	146	0.33	0.47
	mean	0.000	.	.	.	.	.	.	.	.	1.000			

Source: The Authors.

Table E.3 Distribution of Business Service Index Over Communities

Availability Index		0 to .1	.1 to .2	.2 to .3	.3 to .4	.4 to .5	.5 to .6	.6 to .7	.7 to .8	.8 to .9	.9 to 1	N	average	standard deviation
	%	54.2	14.8	16.9	0.0	8.5	0.7	0.0	1.4	1.4	2.1	**142**	**0.15**	**0.22**
	mean	0.000	0.143	0.286	.	0.429	0.571	.	0.714	0.857	1.000			
Engineering services available for businesses in the community	%	94.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.6	142	0.06	0.23
	mean	0.000	.	.	.	.	.	.	.	.	1.000			
Management consulting services available for businesses in the community	%	93.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.0	142	0.07	0.26
	mean	0.000	.	.	.	.	.	.	.	.	1.000			
Marketing services available for businesses in the community	%	91.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.5	142	0.08	0.28
	mean	0.000	.	.	.	.	.	.	.	.	1.000			
Accounting services available for businesses in the community	%	89.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	10.6	142	0.11	0.31
	mean	0.000	.	.	.	.	.	.	.	.	1.000			
Legal services available for businesses in the community	%	73.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	26.1	142	0.26	0.44
	mean	0.000	.	.	.	.	.	.	.	.	1.000			
Insurance services available for businesses in the community	%	60.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	39.4	142	0.39	0.49
	mean	0.000	.	.	.	.	.	.	.	.	1.000			
Information technology services available for businesses in the community	%	90.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.9	142	0.10	0.30
	mean	0.000	.	.	.	.	.	.	.	.	1.000			

Source: The Authors.

Table E.4 Distribution of Governance Indexes Over Communities

Governance index		0 to .1	.1 to .2	.2 to .3	.3 to .4	.4 to .5	.5 to .6	.6 to .7	.7 to .8	.8 to .9	.9 to 1	N	average	standard deviation
	%	0.0	0.0	0.0	0.0	0.0	8.0	65.6	26.4	0.0	0.0	**125**	**0.67**	**0.04**
	mean	.	.	.	.	.	0.583	0.661	0.717	.	.			
General policy and institutional constraints	%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.4	97.6	125	0.98	0.03
	mean	.	.	.	.	.	.	.	.	0.878	0.987			
Infrastructure and services	%	34.4	51.2	13.6	0.8	0.0	0.0	0.0	0.0	0.0	0.0	125	0.13	0.06
	mean	0.064	0.143	0.228	0.306	.	.	.	.	.	.			
Dealing with government agencies	%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	125	1.00	0
	mean	.	.	.	.	.	.	.	.	.	0.999			
Rule of law	%	1.6	1.6	2.4	17.6	36.8	20.8	9.6	8.8	0.8	0.0	125	0.49	0.14
	mean	0.052	0.167	0.273	0.351	0.455	0.556	0.648	0.733	0.813	.			
Conflict Resolution and contract enforcement	%	0.0	0.0	0.0	3.2	3.2	4.0	28.0	27.2	20.8	13.6	125	0.75	0.14
	mean	.	.	.	0.370	0.459	0.561	0.665	0.761	0.845	0.952			

Source: The Authors.

b. Economic policy as a constraint for rural investment climate;
c. Crime, theft, and disorder as constraints for rural investment climate;
d. Legal system as a constraint for rural investment climate; and
e. War or other social frictions as a constraint for rural investment climate.

The variables are classified in the database on a scale from 1 to 5 (from no obstacle to *very severe*). For this subindex the variable is reclassified as follows:

0.75 if item is a minor obstacle
0.50 if item is a moderate obstacle
0.25 if item is a major obstacle
0.00 if item is very severe obstacle

2) Infrastructure and services.
This subindex measures reports of unofficial extra fees for service delivery. The index is 1 when there are no extra payments.

Variables included are:

a. Extra fee to register/renew;
b. Extra fee to apply for a basic activity license/permit/renew;
c. Extra fee to apply for a construction permit;
d. Extra fee to apply for an electricity connection for industry use;
e. Extra fee to apply for an electricity connection for domestic use; and
f. Extra fee to apply for other services.

Note: These variables are *not* included in the core.

Each index at the enterprise level is defined as $1 - (X/X_{95})$, where X is the value of the unofficial, extra fee and X_{95} is the value at the ninety-fifth percentile of the distribution of available observations on a given item. The index is only calculated if businesses applied for the service; other cases will be missing values. The index is set at 0 when $X > X_{95.}$

3) Dealing with government agencies.
This subindex measures cost related to dealing with government agencies. A high subindex indicates low costs.

Variables included are:

a. Tax-related issues;
b. Labor-related issues;
c. Issues related to fire and building safety;
d. Issues related to sanitation and epidemiology;
e. Issues relating to environmental regulation; and
f. Others.

Note: These variables are *not* included in the core.

For each type of deal or issue, the business owner or manager is asked if an unofficial payment or gift was ever expected or requested. When the answer to such questions was *yes,* the value of the index is recorded as 0; when the answer was *no,* the value is 1. For businesses that did not have a relevant deal or issue in the 12 months prior to the survey, the observation was treated as a missing value.

4) Rule of law.
This subindex indicates the rule of law and the predictability of its application. A high value indicates strong rule of law. Variables included are:

a. Predictability of laws and regulations that affect the operation and growth of businesses; and
b. Rules and regulations that can be manipulated or misinterpreted by officials.

Predictability of laws and regulations that affect the operation and growth of businesses is scaled as follows:

1.00 fully predictable (originally coded as 1)
0.75 highly predictable (originally coded as 2)
0.50 somehow predictable (originally coded as 3)
0.25 unpredictable (originally coded as 4)
0.00 highly unpredictable (originally coded as 5)

In cases where the question was answered *don't know,* the observation was treated as a missing value.

The possibility that rules and regulations can be manipulated or misinterpreted by officials is scaled as follows:

0	if	*strongly agree* (originally 1)
0.3333	if	*agree* (originally 2)
0.6666	if	*disagree* (originally 3)
1	if	*strongly disagree* (originally 4)

The observation is treated as a missing value if the answer is *don't know, can't say.*

5) Conflict resolution and contract enforcement.
This subindex registers conflict resolution and contract enforcement. A high value indicates favorable conditions. The following variables are included:

a. Must rely on the reputation of others with whom you enter into agreements;

b. A contract will protect you from being cheated by others; and
c. The legal system will uphold your contract and property rights in disputes.

We reclassify the variable so that the index will be as follows:

0	if	*strongly agree* (originally 1)
0.3333	if	*agree* (originally 2)
0.6666	if	*disagree* (originally 3)
1	if	*strongly disagree* (originally 4)

An observation is treated as a missing value if the answer is *don't know, can't say.*

5. Human capital. This indicator provides an index of human capital. A high value indicates a high level of education and working experience. The index is constructed on the basis of the household survey.

Bils and Klenow (2000) describe human-capital stock according to the classic Mincerian returns to schooling and experience:

$$\ln[h(a, t)] = \frac{\theta s^{1-\psi}}{1-\psi} + \gamma_1(a - s - 6) + \gamma_2(a - s - 6)^2$$

where a is age and s is years of schooling, and the quadratic term is standard in the empirical literature on wages.

The returns to schooling and experience (γ_1 and γ_2) are based on estimates of the sources of wage differences, that is, the following Mincer equation:

$$\ln[h(w)] = \lambda_0 + \lambda_1 s + \lambda_2(a - s - 6) + \lambda_3(a - s - 6)^2 + \varepsilon$$

Bils and Klenow obtain estimates for γ_1 and γ_2 from averages of 52 countries of λ_2 and λ_3. The estimated parameter values are $\gamma_1 = 0.0512$, $\gamma_2 = -0.00071$, $\theta = 0.099$, and $\psi = 0$ (which corresponds to the classic case in the labor literature showing no diminishing returns to education).

The individual human stock index is obtained as follows:

$$\ln(stock) = (0.099 * s) + (0.0512 * (a - s - 6)) - (0.00071(a - s - 6)^2)$$

Thus an individual's stock is computed as the antilog of the result of this expression:

$$stock = e^{\ln(stock)}$$

For the variable s, years of schooling, we rescaled the coding in the household survey as follows:

Variable value	Original coding	New coding for index	Percentage of the sample
Year 1	1	1	2.08
Year 2	2	2	6.31
Year 3	3	3	3.95
Year 4	4	4	4.48
Year 5	5	5	6.63
Year 6	6	6	4.96
Year 7	7	7	4.66
Year 8	8	8	8.97
Year 9	9	9	4.23
Year 10	10	10	13.49
Year 11	11	11	15.52
Year 12	12	12	4.33
Year 13	13	13	8.14
University (post-graduate)	14	15	0.98
Professional	15	14	0.26
Technical College	16	14	0.06
Preschool	17	0	2.54
No Schooling	18	0	3.83
Other	19	Missing	0.57
20 (?)	20 (?)	Missing	3.80
Don't know	99	Missing	0.20

Note: The schooling variable is censored at 14 years of formal education.

In case of missing values of education and/or age, the observation is treated as a missing value.

The human-capital stock is computed for each individual between 16 and 65 years of age. The lowest possible value of *stock* over all values of s and a equals approximately 1.5, achieved at using $a = 16$ and $s = 0$; the highest equals about 10.0 for someone with $s = 14$ and $a = 57$. These two values become meaningful when the community human capital index is computed.

The human capital index of the community is aggregated from individuals between 16 and 65 years old, without any restrictions in terms of occupational status.

Thus, the index is obtained through the following steps:

1. Using data from the household survey, impute a human capital stock for each individual on the basis of the model estimated by Bils and Klenow.
2. By community, compute the average human capital stock, that is, h_ave = (sum of human capital/number of individuals).
3. Finally, compute the community-level human capital index as $(h\_ave - 1.55)/(10 - 1.5)$.

6. Finance services. This indicator describes the level of development of financial intermediation.

Table E.5 Distribution of Human Capital Indexes Over Communities

		0 to .1	.1 to .2	.2 to .3	.3 to .4	.4 to .5	.5 to .6	.6 to .7	.7 to .8	.8 to .9	.9 to 1	N	average	standard deviation
Human resources index	%	0.0	6.7	31.3	41.0	17.9	3.0	0.0	0.0	0.0	0.0	**134**	**0.33**	**0.09**
	mean	.	0.165	0.259	0.350	0.427	0.551	.	.	.	.			

Source: The Authors.

Table E.6 Distribution of Finance Service Indexes Over Communities*

		0 to .1	.1 to .2	.2 to .3	.3 to .4	.4 to .5	.5 to .6	.6 to .7	.7 to .8	.8 to .9	.9 to 1	N	average	standard deviation
Finance Index	%	1.4	5.4	3.4	15.6	19.7	25.9	19.7	7.5	1.4	0.0	**147**	**0.50**	**0.16**
	mean	0.039	0.172	0.251	0.359	0.450	0.549	0.655	0.735	0.827	.			
Number of formal financial sources weighed by mean of the distance	%	2.7	4.1	6.8	11.6	10.2	12.9	21.8	18.4	8.8	2.7	147	0.56	0.22
	mean	0.023	0.141	0.255	0.361	0.447	0.560	0.655	0.748	0.855	0.938			
Number of formal finance services weighed by mean of the distance	%	7.5	8.2	8.2	19.0	10.2	16.3	15.6	10.2	2.0	2.7	147	0.47	0.23
	mean	0.048	0.162	0.260	0.363	0.444	0.560	0.659	0.739	0.843	0.963			
Access to loans	%	4.1	6.1	11.6	12.2	15.6	21.8	15.0	3.4	7.5	2.7	147	0.47	0.23
	mean	0.023	0.135	0.230	0.337	0.426	0.526	0.637	0.761	0.872	1.000			

Source: The Authors.

* This table differs from the standard definitions used in the text due to differences in the definitions of financial services used across the questionnaires. (This should be avoided in the next phase.)

The financial intermediation will be more mature if the number of formal institutions is larger, the number of services available is larger, and the share of enterprises that want to borrow money is large.

It consists of four variables:

1) *Number of formal financial and insurance institutions that offer services in the community, weighted by distance from the community;*
2) *Number of financial and insurance services provided in the community by formal institutions, weighted by distance from the community;*
3) *Community access to a commodity exchange for futures or options contracts; and*
4) *Access to loans.*

1) *Number of formal financial and insurance institutions that offer services in the community, weighted by distance from the community.*

This subindex describes the number of formal institutions that serve customers in the community as reported, weighted by distance to the institution.

Step 1: Construct the distance index 1 – {(D)/(30)}, where D is the distance, in kilometers, to each type of financial institution in the community. If D >30, set D = 30.

Step 2: Obtain the mean of the indexes calculated in step1 for each community; call this "M."

Let X be the number of institutions. Calculate a variable: A = X * (1/M), then calculate the index as follows: A/max(A).

2) *Number of financial and insurance services provided in the community by formal institutions, weighted by distance from the community.*

The questionnaire asks whether the following 13 services or financial tools are available:

1. Loan up to 6 months
2. Loan 6 months to 2 years
3. Loan longer than 2 years
4. Credit line for business linked to checking account
5. Warehouse receipts accepted as collateral
6. Leasing contracts available
7. Land accepted as collateral
8. Savings account offering withdrawal on short notice
9. Term deposits
10. Current account
11. Money transfer (sending and receiving)
12. Life and funeral insurance
13. Insurance for health, accident, disability, fire, and livestock

Let X be the total number of formal services available in the community; calculate: I = X * (1/M), then calculate the index as follows: I/max I.

3) *Community access to a commodity exchange for futures or options contracts.*

This variable is a binary index.

4) *Access to loans.*

From the enterprise questionnaire, information is obtained for each community concerning the proportion of businesses that obtained formal loans for day-to-day operations and investments in their businesses during the last 5 years. The index is measured as the percentage of such enterprises.

The aggregate finance services index is the average of the four subindexes.

Box E.1. Definition of Formal Financial Institutions (Including Insurance)

Institutions with a permanent office, proper marquee, and regular business hours, including:

- Commercial banks;
- Savings banks and post offices that offer financial services (mostly savings);
- Microfinance institutions with full-time professional staff;
- Government agencies engaged in retailing credit;
- Cooperatives, credit unions, and business associations providing financial services; and
- Companies offering life, health, accident, disability, fire, or livestock insurance.

Source: RICS Enterprise and Community Questionnaires, Appendix C.

Annex F. Entrepreneurship: Notes and Tables

Specification Issues

Dependent Variables: Activity Status. In Nicaragua, household entrepreneurship is defined as deriving a positive income from an enterprise. Because of an error in the questionnaire design in Sri Lanka, this definition would understate the incidence of entrepreneurship. Instead, in Sri Lanka entrepreneurship is defined as "household members working on own account."[89] Involvement in farming or wage employment is defined similarly.

The income- and work-based definitions of entrepreneurship should largely overlap, leaving a gap only if work "on own account" earned no income. In practice, some households indeed report work but no income; however, some others report income but no work. In Nicaragua, a discrepancy arises between the two definitions in 7.3 percent of the sample households, split evenly between those reporting work but no income and those reporting income but no work. In Sri Lanka, 1.3 percent reported income but no work, and, to a large degree because of the questionnaire design flaw, 29.6 percent reported work but no income.[90] The income-based definition is preferable from a perspective of productivity and poverty analysis, but for Sri Lanka the work-based definition is an acceptable alternative.

Dependent Variables: Enterprise Start-Up. Technically, information on whether a household recently started up an enterprise is not directly available in the RIC data, but for the purpose of this analysis, households are considered to have started up an enterprise if the enterprise they operate is less than two years old. This understates the true two-year start-up rate, since these enterprises must have survived for up to two years to be included; enterprises that recently started up but had already died before the survey was held cannot be counted.

Moreover, because of the design of the survey, enterprise start-up must be based on the premises of the household's residence at the time of the survey; it cannot be a stand-alone business. In Sri Lanka, where stand-alone businesses are common (64.8 percent of the sample), the estimated start-up rate of 1.2 percent is therefore significantly understated. The true start-up rate might be as high as 6.05 percent (Box F.1). In Nicaragua, the situation is quite different: the enterprise start-up rate is measured to be 3.0 percent, probably fairly accurate because the number of stand-alone enterprises is rather small (7.5 percent of the sample).

As mentioned above, an enterprise is considered a start-up if it was established during the two years prior to the survey. More accurately, this means that for Sri Lanka, the enterprise began operations during 2002 or 2003, and for Nicaragua, the enterprise was established during 2003 or 2004. The RICS survey was held in Sri Lanka between September and November, 2003, but the actual date was not recorded in the database. Therefore, the measured start-up rate covers a little less than two years. Fortunately, in any given year, most start-up enterprises appear to start in the months between January and August, as 88 percent of the existing enterprises did so. Thus, the start-up rate for Sri Lanka is still close to the true two-year start-up rate (after allowing for the caveats mentioned above). The Nicaragua survey was held between February and May, 2004, and the date of the survey is missing in many of the records. The month in which an enterprise began operations is unknown in Nicaragua, but it is safe to assume that the Nicaragua statistic may capture only one-third to one-half of start-ups in 2004. Allowing for this, a better estimate of the two-year start-up rate would range from 4.0 percent to 4.5 percent.

Box F.1 RIC Survey Design and the Undercount of Enterprise Start-Up

To gain insight into the effect of survey design on the measurement of enterprise start-up, consider that the rate of enterprise start-up is measured to be 1.20 percent in Sri Lanka.[a] That is, 1.20 percent of the Sri Lankan households started a household-based enterprise during the two years prior to the RICS survey. The weighted percentage of households with a (household-based) enterprise in the survey is 5.47 percent. Thus, 21.93 percent of the enterprises in the survey are rated as start-ups. But another 20.30 percent of the households report work on own account but did not provide data about the enterprise. Now in the RICS enterprise database, 35.24 percent of the enterprises are household-based and 64.76 percent are stand-alone enterprises; 24.96 percent and 23.90 percent of them are start-ups, respectively.[b] If one were to infer that the 20.30 percent of the households that report work on own account but did not provide data about the enterprise are operating a stand-alone enterprise and that the start-up rate of 23.90 percent applies to them uniformly, then another 4.85 percent ($= 0.239 \times 20.30$) of the households would be operating a start-up enterprise, and the enterprise start-up rate among Sri Lankan rural households would be 6.05 percent.[c]

As for Nicaragua, there are 1,372 household-based enterprises in 1,060 households. Of these, 179 began operations during the previous two years (13.28 percent).[d] These are aggregated to the household level, and a household is assumed to have started an enterprise when one of its several operations began during the previous two years; this means that 18.15 percent of the households with one of more enterprises began at least one.[e] As mentioned in the discussion about the definition of the entrepreneurship variable, there were only a few households with inconsistent enterprise income and work activity in the database. Thus, only a few of the stand-alone enterprises might be under household ownership of one of the households in the sample, and thus any correction would be minor.

Source: The Authors.

[a] This derives from the dataset with which the determinants of enterprise start-up are examined. Households with missing information or in communities without benchmark indicators or community variables are omitted.

[b] The difference between 24.96 percent and 21.93 percent arises from the fact that not all household-based enterprises are actually linked with a household and also that the second statistic is computed in a subsample that is subject to data availability.

[c] This statistic further assumes that stand-alone start-ups did not first begin as a household-based entity that was turned into a stand-alone within the first two years of its existence. If this happens, double counting occurs among the start-up percentage.

[d] The start-up rate among stand-alone enterprises is 26 out of 110 or 21.3 percent as a weighted percentage.

[e] While Nicaragua surveyed every enterprise activity in the household, Sri Lanka selected only one activity. As shown, the start-up rate among activities in Nicaragua is 13.28 percent but among households it is 18.15 percent. By virtue of this survey design feature, the start-up rate in Sri Lanka should be relatively lower than in Nicaragua. The difference is further accentuated by the way this one activity in Sri Lanka is selected: the main activity in the household is likely the more established one.

Explanatory Variables. The explanatory variables are divided into three sets. The first pertains to household characteristics and includes demographics such as the number of male and female adult household members, their ages and human capital, and the gender and ethnicity of the head. These variables represent determinants of the available supply of household labor that may be drawn upon in a household enterprise, the quality of this labor, and barriers both to the operation of an enterprise and to other opportunities of employment; that is, these variables capture any barriers on the basis of gender, ethnicity, or age, for example, whether encountered when seeking wage employment in the labor market or in procuring inputs or financial resources for the enterprise. Similarly, human capital may signal skills used either in the enterprise (directly for production or indirectly for procuring inputs) or elsewhere in wage employment.

Another variable among household characteristics describes whether the household head's parents were entrepreneurs. The household head would have been able to experience entrepreneurship close up, receiving an in-house apprenticeship, as it were. Parents may have introduced the head to their business network.

The last group of household variables describes financial resources: remittance income, income from other sources, and household assets. These clearly facilitate enterprise operation. Household assets, in particular, however, might be endogenously related

to the decision to operate an enterprise. To be sure, the assets do not include enterprise assets available in the enterprise database, but it is possible that entrepreneurs use household assets for their businesses, especially if the enterprises are based at the residence. Moreover, previous profitability may well have generated savings or investments in consumer durables that are now observed as household assets. The econometric analysis recognizes household assets as a potential endogenous household-level variable.

The second set of variables describes the investment climate in the community where the household resides. As described in Chapter 2, the investment climate is measured by a multitude of variables that are condensed into six benchmark indicators. Both these indicators and their components are used in the econometric analysis to measure the investment climate's effect on entrepreneurship. The benchmark indicators cover characteristics of location, availability of infrastructure, utilities, public and private services, governance, human capital, and finance services. These variables are constructed such that a higher value is expected to favor the enterprise performance. Connectivity may raise the value of entrepreneurship if it opens up distant markets; it may lower it if distant competitors seek out clients in the community. Infrastructure services are necessary for many business exploits: greater access should increase the value of entrepreneurship. The availability of business services may facilitate entrepreneurship. The governance index measures rule of law, security, contract enforcement, and so forth, without which enterprise operation is more risky. The community human-capital index is a proxy for the quality of labor, which can raise enterprise productivity and thus the value of entrepreneurship. At the same time, human capital relates positively to income and thus to potential market size. The finance services index summarizes finance and insurance services that facilitate enterprise start-up and operation and thus should encourage household entrepreneurship.

As mentioned above, however, all of these arguments do not necessarily imply that the benchmark indicators must favor the entrepreneurship choice. A robust investment climate may create more lucrative opportunities outside the household-owned-and-operated enterprise and thus may lead to enterprise closure while nonetheless fostering an expanded nonfarm economy. The next section returns to this issue, but for now it is sufficient to temper expectations regarding effects of investment climate indicators on entrepreneurship choice.

The third set of variables describes the community apart from immediate associations with the policy-related investment climate. These variables are motivated as follows. Community size and the level of per capita income measure market opportunities an enterprise could exploit inside a community. Since a successful enterprise in turn generates income, per capita income may be an endogenous community-level variable. The male wage rate measures the cost of labor, the opportunity cost of the entrepreneur and household members, and the level of well-being among potential customers: the effect on household entrepreneurship is therefore ambiguous. Enterprise openness indicates connection with markets outside the community and thus, as it broadens the potential market, it is expected to encourage entrepreneurship. Finally, in a community with pronounced agricultural seasonality, households may seek ways to make their slack season more profitable (Haggblade, Hazell, and Brown, 1989; Lanjouw and Lanjouw, 2001); alternatively, the cyclical unreliable availability of labor may be a stumbling block for many enterprises.

Endogeneity among Explanatory Variables. It was mentioned above that household assets and community income levels are plausibly endogenous. Because of the random effects structure of the entrepreneurship probit equation, this endogeneity may express itself at two levels. Household assets is a household-level variable for which a random effect model is appropriate. Thus, endogeneity may express itself through a correlation between the (community-level) random effects of the assets and entrepreneurship equations as well as through a correlation of the household-level disturbances. Meanwhile, community income may be modeled with a community-level simple regression model, where its disturbance may be correlated with the community-level random effect in the assets and entrepreneurship equations. The most appropriate way to estimate the entrepreneurship model while allowing for endogeneity is to build the assets equation into a likelihood function that also includes the residual of the community income equation. Estimation of this model, however, which relies on numerical optimization methods, runs into frequent convergence problems.

As an alternative, the assets equation may be estimated with a fixed effects regression model; the community income equation may be estimated with the ordinary least squares method; and the random effect and household-level residual of the assets equation and the residual of the community-income equation may be inserted into the entrepreneurship model: if these added variables have a statistically significant effect on entrepreneurship choice, assets and/or community income are endogenous. As it turns out, the three added variables are never jointly significant, and in all the models only one of them is statistically significant once at a 10 percent level and not at the 5 percent level. Thus, endogeneity does not appear to be an issue.

Explanatory Power of the Regression Models. An *analysis of variance* assists in the assessment of the degree to which benchmark indicators and community variables help explain the variation in entrepreneurship among rural households. Since the explained variable (entrepreneurship) is dichotomous, the criterion value is not the sum of squared residuals but rather the value of the log-likelihood function, which, moreover, is negative and maximized. The evidence is summarized in Table F.3. If no explanatory variables enter the model, the criterion value equals –611.8 for Nicaragua and –489.8 for Sri Lanka. In the ideal case where entrepreneurship is perfectly explained, the criterion value equals 0. As reported in Panel A of the table, household characteristics explain (that is, reduce the criterion function by) 8.09 percent in Nicaragua and 4.58 percent in Sri Lanka, where it should be noted, the Sri Lankan model does not include remittance and other income, which are influential variables in Nicaragua. The community random effect contributes 1.09 percent in Sri Lanka but is absent in Nicaragua. The benchmark indicators help to explain 2.98 percent in Nicaragua and 1.46 percent in Sri Lanka, and community characteristics further add about half as much. In all, benchmark indicators and community characteristics contribute 35 percent (= (2.98 + 1.35)/12.42) of the explanation of entrepreneurship choice in Nicaragua and 28 percent (= (1.46 + 0.72)/7.85) in Sri Lanka. The overall explanatory power of the model is 12.42 percent in Nicaragua and 7.85 percent in Sri Lanka. Panel B indicates the fit when the best components are used: relative to the model with only household characteristics, a few well-selected variables provide significant improvements in both countries, and in Sri Lanka this even exceeds the explanatory power of panel A. Panels C and D describe the explanatory power of benchmark indicators and community characteristics in isolation: the benchmark indicators carry relatively less information in Sri Lanka.

Tables

Table F.1 Definitions and Descriptive Statistics of Variables Used in the Econometric Models

A: Variable definitions

Variable	Definition
Household Characteristics	
Entrepreneurship[a]	Dummy variable, 1 = household operates an enterprise, 0 = otherwise
Wage employment[a]	Dummy variable, 1 = members of the household hold a wage job, 0 = otherwise
Nonfarm economic activity[a,b]	Dummy variable, 1 = household operates an enterprise and/or household members hold a wage job, 0 = otherwise
Farming[a]	Dummy variable, 1 = household operates a farm, 0 = otherwise
Enterprise start-up[c]	Dummy variable, 1 = household operates an enterprise that started up at most two years prior to the time of the survey, 0 = otherwise
Number of male adults	Number of male adults in the household
Number of female adults	Number of female adults in the household
Average age	Average age among adults in the household
Human-capital index	Average human capital (defined as in Annex E) among adults in the household
Head female	Dummy variable, = 1 if the head of household is female, = 0 otherwise
Not Sinhalese	Dummy variable, = 1 if the head of household is not of Sinhalese ethnicity, 0 = otherwise
Head's parents were entrepreneur	Dummy variable, = 1 if the parents of the head of household, 0 = otherwise
ln(Other income)	Natural log of the sum of all types of income except wage, enterprise, farm, and remittance income (US$)
ln(Remittances)	Natural log of remittance income (US$)
ln(Assets)	Natural log of household assets (US$)
Benchmarks and components	
Connectivity index	Connectivity index, defined in Annex E
Infrastructure services index	Infrastructure services index, defined in Annex E
Business services index	Business services index, defined in Annex E
Governance index	Governance index, defined in Annex E
Human-capital index	Community-level human-capital index, defined in Annex E
Finance services index	Finance services index, defined in Annex E
Community variables	
ln(Community size)	Natural log of number of residents in the community
ln(Income per capita)	Natural log of the average income per capita (US$) in the community
Agricultural seasonality	Average of the standard deviation in monthly agricultural labor input for male and female labor
Enterprise openness	Index of dealings of the enterprise outside the community with respect to clients and input providers
ln(Male wage rate)	Natural log of the average male wage rate in agriculture, service, and manufacturing

Table F.1 Definitions and Descriptive Statistics of Variables Used in the Econometric Models (*continued*)

B: Descriptive statistics

	Nicaragua		Sri Lanka	
Variable	**Mean**	**StDev**	**Mean**	**StDev**
Household Characteristics				
Entrepreneurship[a]	0.219	0.414	0.253	0.435
Wage employment[a]	0.704	0.456	0.760	0.427
Nonfarm economic activity[a,b]	0.817	0.387	0.894	0.308
Farming[a]	0.272	0.445	0.479	0.500
Enterprise start-up[c]	0.030	0.171	0.013	0.112
Number of male adults	1.316	0.992	1.496	0.914
Number of female adults	1.459	0.901	1.494	0.783
Average age	33.975	8.726	37.323	7.701
Human capital index	0.209	0.136	0.333	0.158
Head female	0.329	0.470	0.191	0.393
Not Sinhalese	n.a	n.a	0.140	0.347
Head's parents were entrepreneur	0.362	0.480	0.272	0.445
ln(Other income)	1.190	2.345	n.a	n.a
ln(Remittances)	1.574	2.580	n.a	n.a
ln(Assets)	7.258	1.717	8.595	1.147
Benchmarks and components				
Connectivity index	0.350	0.186	0.436	0.166
Infrastructure services index	0.535	0.256	0.440	0.162
Business services index	0.296	0.428	0.179	0.241
Governance index	0.686	0.111	0.677	0.037
Human capital index	0.205	0.074	0.324	0.092
Finance services index	0.288	0.244	0.357	0.098
Community variables				
ln(Community size)	8.702	1.929	7.539	0.594
ln(Income per capita)	5.824	0.702	8.144	0.554
Agricultural seasonality	0.721	0.413	0.731	0.307
Enterprise openness	1.438	0.349	2.227	0.785
ln(Male wage rate)	0.002	0.199	5.542	0.200
Number of households	1163		849	

Source: RIC Surveys.

Notes: a. Dependent variable. Column percentages do not add up to 100 as households may participate in several activities simultaneously. Nicaragua follows income-based definitions; Sri Lanka follows work-based definitions.

b. Combining nonfarm entrepreneurship and wage employment.

c. Dependent variable. In Sri Lanka, enterprise start-up is measured for 790 households, rather than 849.

Table F.2 Determinants of Entrepreneurship Choices

A: Nicaragua: Weighted probit[a]

	(1)		(2)		(3)		(4)	
	dP	t[b]	dP	t	dP	t	dP	t
Household characteristics								
Number of male adults			3.08	2.23	3.00	2.24	3.42	2.55
Number of female adults			2.55	2.01	2.82	2.30	2.71	2.20
Average age			4.07	2.95	3.89	2.88	3.83	2.89
Human capital index			0.83	0.58	0.84	0.59	1.06	0.75
Head female			6.14	2.06	5.31	1.81	5.74	1.98
Head parents entrepreneur			2.27	0.86	2.06	0.80	2.19	0.86
ln(Other income)			−2.66	2.08	−2.82	2.21	−2.42	1.91
ln(Remittances)			−7.53	5.83	−7.26	5.71	−7.27	5.71
ln(Assets)			5.14	3.59	4.99	3.56	4.71	3.40
Benchmarks and components								
Connectivity	−0.80	0.38	−0.42	0.19	−0.53	0.23		
Proximity to post office							2.78	2.07
Infrastructure services	5.45	2.32	5.78	2.44	2.78	1.10		
Access to water							3.33	2.50
Business services	−1.91	1.25	−2.28	1.48	−1.80	1.16		
Governance	−0.74	0.59	−0.45	0.35	0.09	0.07		
Human capital	7.91	4.90	6.29	3.79	4.89	2.45	4.49	2.70
Finance services	−0.37	0.25	−1.00	0.69	−1.29	0.85		
Community variables								
ln(Community size)					0.78	0.36		
ln(Income per capita)					2.76	1.36		
Agricultural seasonality					1.94	1.40		
Enterprise openness					3.67	2.48	5.30	3.91
ln(Male wage)					−0.48	0.35		
Regression statistics								
Log-likelihood	−580.19		−544.10		−535.83		−535.90	
Number of observations	1163		1163		1163		1163	

(*continued on the next page*)

Table F.2 Determinants of Entrepreneurship Choices (*continued*)

B: Sri Lanka: Weighted random effect probit[a]

	(1)		(2)		(3)		(4)	
	dP	t	dP	t	dP	t	dP	t
Household characteristics								
Number of male adults			2.73	1.26	2.92	1.38	3.25	1.54
Number of female adults			−2.17	1.00	−2.03	0.94	−1.88	0.88
Average age			0.91	0.33	0.84	0.32	0.58	0.23
Human capital index			0.93	0.34	1.04	0.38	1.57	0.60
Head female			−8.61	1.03	−8.83	1.08	−10.40	1.42
Not Sinhalese			−4.99	0.51	−4.87	0.51	−14.24	1.80
Head parents entrepreneur			1.56	0.31	2.11	0.42	2.09	0.43
ln(Assets)			9.74	3.13	9.82	3.18	10.55	3.59
Benchmarks and components								
Connectivity	5.70	1.29	5.34	1.28	4.29	1.04		
Proximity to post office							4.84	1.84
Infrastructure services	−2.19	0.64	−2.08	0.58	−4.14	1.13		
Sewage system							−7.33	2.54
Business services	3.49	1.13	3.66	1.27	3.36	1.08		
Engineering services							−4.49	1.69
Information technology services							5.93	1.58
Governance	2.89	1.07	3.91	1.43	3.75	1.40		
Human capital	4.83	1.35	2.41	0.60	3.36	0.81		
Finance services	−2.95	0.67	−3.68	0.83	−4.26	1.02		
Community variables								
ln(Community size)					4.44	1.25	3.98	1.35
ln(Income per capita)					0.90	0.31		
Agricultural seasonality					−0.96	0.27		
Enterprise openness					−0.72	0.20		
ln (Male wage rate)					2.59	0.78	4.36	1.67
Regression statistics								
Standard error of random effect	10.38	3.80	9.92	3.65	9.80	3.49	6.99	2.81
Log-likelihood	−474.70		−454.90		−451.39		−438.06	
Number of observations	849		849		849		849	

Source: RIC Surveys.

Notes: a. Columns headed with "dP" report the percentage point increase in the probability that an average household operates an enterprise in response to a one-standard-deviation increase in the explanatory variable. For the variables "Head female," "Head parents entrepreneur," and "Not Sinhalese," which are dummy variables, the change in the explanatory variable is one unit.

b. Significance levels implied by the t-statistics are: 1% if $t \geq 2.58$, 5% if $1.96 \leq t < 2.58$, 10% if $1.645 \leq t < 1.96$.

Table F.3 Contributions to Explanation of Entrepreneurship Choice

Explanatory variables	Nicaragua		Sri Lanka	
	Criterion value[a]	Increment as % of base	Criterion value[a]	Increment as % of base
None (base)	–611.797		–489.841	
A. Household characteristics only	–562.312	8.09%	–467.398	4.58%
+ Community random effect	–562.312	0.00%	–462.053	1.09%
+ Benchmark indicators	–544.097	2.98%	–454.898	1.46%
+ Community characteristics	–535.835	1.35%	–451.391	0.72%
B. Household characteristics only	–562.312		–467.398	
+ Community random effect and best components	–535.900	4.32%	–438.063	5.88%
C. Benchmark indicators and community random effect only	–580.186	5.17%	–474.698	3.09%
D. Community characteristics and community random effect only	–579.148	5.34%	–479.544	2.10%

Source: RIC Surveys.

Note: a. The criterion is the maximized log-likelihood function, which in the case of probit models is always negative and rises as explanation improves.

Table F.4 Determinants of Household Activity Choices

A: Nicaragua: Weighted probit and weighted random effect probit[a]

	Nonfarm household enterprise		Wage employment		Nonfarm economic activity		Farming	
	dP	t[b]	dP	t	dP	t	dP	t
Household characteristics								
Number of male adults	3.42	2.55	7.91	3.03	3.41	1.80	4.01	1.40
Number of female adults	2.71	2.20	1.47	0.69	1.03	0.71	0.16	0.05
Average age	3.83	2.89	−5.46	2.47	−1.61	1.17	10.15	3.64
Human capital index	1.06	0.75	9.11	4.43	7.07	4.26	−10.62	3.36
Head female	5.74	1.98	2.60	0.52	0.53	0.15	−6.16	0.90
Head parents entrepreneur	2.19	0.86	1.69	0.42	2.03	0.67	−8.89	1.37
ln(Other income)	−2.42	1.91	−4.86	2.32	−4.80	3.99	−4.59	1.61
ln(Remittances)	−7.27	5.71	−6.42	3.13	−7.89	5.50	−4.32	1.36
ln(Assets)	4.71	3.40	−18.77	6.87	−9.59	4.92	19.33	5.47
Benchmarks and components								
Proximity to post office[c]	2.78	2.07						
Percent of households with electricity[d]			−8.18	2.15	−4.84	1.42		
Infrastructure services, access to water[d]	3.33	2.50						
Sewage system[d]							−8.97	1.47
Dealing with government agencies[f]			−7.19	2.06	−4.09	1.42		
Rule of law[f]							−7.13	1.31
Human capital	4.49	2.70	−6.72	1.22				
Community variables								
ln(Income per capita)			11.01	1.99	5.47	1.31	−15.14	2.56
Agricultural seasonality			−7.31	1.72			8.37	1.62
Enterprise openness	5.30	3.91			3.87	1.15		
Regression statistics								
Standard error of random effect	0.00		15.46	5.41	12.00	4.41	19.07	5.12
Log-likelihood	−535.90		−544.73		−330.56		−498.38	
Number of observations	1163		1163		1163		1163	

(*continued on the next page*)

Table F.4 Determinants of Household Activity Choices (*continued*)

B: Sri Lanka: Weighted random effect probit[a]

	Nonfarm household enterprise		Wage employment		Nonfarm economic activity		Farming	
	dP	t	dP	t	dP	t	dP	t
Household characteristics								
Number of male adults	3.25	1.54	5.32	2.27	0.64	0.75	12.26	4.08
Number of female adults	−1.88	0.88	10.10	4.72	2.66	3.49	−4.40	1.45
Average age	0.58	0.23	−5.41	2.45	−2.44	3.05	4.84	1.69
Human capital index	1.57	0.60	7.68	3.20	3.29	3.57	−3.09	0.91
Head female	−10.40	1.42	12.17	2.34	−1.68	0.95	−7.50	0.87
Not Sinhalese	−14.24	1.80	4.33	0.49	−5.85	2.23	17.76	1.39
Head's parents entrepreneur	2.09	0.43	−0.21	0.04	2.19	1.21	−1.89	0.27
ln(Assets)	10.55	3.59	−6.04	2.13	−0.42	0.47	20.64	4.87
Benchmarks and components								
Inverse cost of transport to major city[c]			−6.55	1.65				
Inverse cost of transport to main market[c]			5.73	1.63				
Proximity to post office[c]	4.84	1.84						
Infrastructure services			5.61	1.78	2.28	1.76		
Percent of households with electricity[d]							−8.54	1.32
Percent of households with fixed phone line[d]							−13.38	1.93
Sewage system[d]	−7.33	2.54						
Engineering services[e]	−4.49	1.69						
Management consulting services[e]							−15.62	2.17
Marketing services[e]			4.94	2.18			18.92	2.75
Information technology services[e]	5.93	1.58						
Conflict resolution, contract enforcement[f]					1.67	1.37		
Financial penetration[g]			−6.38	2.38			9.32	1.56
Community variables								
ln(Community size)	3.98	1.35	4.90	1.66	3.35	2.34	−11.52	1.88
Agricultural seasonality							9.20	1.36
ln(Male wage rate)	4.36	1.67						
Regression statistics								
Standard error of random effect	6.99	2.81	8.79	3.85	5.38	2.86	24.81	5.18
Log-likelihood	−438.06		−390.69		−188.94		−401.02	
Number of observations	849		849		849		849	

Source: RIC Surveys.

Notes: a. Columns headed with "dP" report the percentage point increase in the probability that an average household operates an enterprise in response to a one-standard-deviation increase in the explanatory variable. For the variables "Head female," "Head parents entrepreneur," and "Not Sinhalese," which are dummy variables, the change in the explanatory variable is one unit.
b. Significance levels implied by the t-statistics are: 1% if $t \geq 2.58$, 5% if $1.96 \leq t < 2.58$, 10% if $1.645 \leq t < 1.96$.
c. A component of the connectivity index.
d. A component of the infrastructure services index.
e. A component of the business services index.
f. A component of the governance index.
g. A component of the finance services index.

Table F.5 Determinants of Enterprise Start-Up[a]

A: Nicaragua

	(1)		(2)		(3)		(4)	
	dP	t	dP	t	dP	t	dP	t
Household characteristics								
Number of male adults			−0.04	0.13	−0.10	0.37	−0.08	0.29
Number of female adults			0.08	0.21	0.12	0.31	0.14	0.37
Average age			−0.39	1.05	−0.50	1.31	−0.41	1.13
Human capital index			0.14	0.42	0.23	0.74	0.21	0.75
Head female			1.15	1.40	1.09	1.32	1.05	1.33
Head's parents entrepreneur			−0.10	0.14	−0.18	0.25	−0.06	0.09
ln(Other income)			−0.61	1.81	−0.68	2.03	−0.67	1.96
ln(Remittance income)			−1.08	2.85	−0.99	2.74	−0.94	2.65
ln(Assets)			0.31	1.05	0.33	1.09	0.36	1.18
Benchmarks and components								
Connectivity	−0.96	1.60	−0.89	1.55				
Inverse cost of transport to major city							−0.57	1.62
Infrastructure services	0.72	1.29	0.79	1.45				
Percent of households with fixed phone line							−0.52	1.20
Business services	0.26	0.51	0.13	0.31				
Governance	−0.35	1.04	−0.35	1.14				
Human capital	0.54	1.41	0.38	0.99				
Finance	0.26	0.65	0.27	0.74				
Community variables								
ln(Community size)					0.45	1.16	0.72	1.83
ln(Income per capita)					−0.17	0.39		
Agricultural seasonality					0.50	1.52	0.40	1.24
Enterprise openness					0.64	1.96	0.63	2.11
ln(Male wage rate)					0.44	1.51		
Regression statistics								
lnL	−156.51		−152.46		−151.34		−150.91	
n	1170		1170		1170		1170	

(*continued on the next page*)

Table F.5 Determinants of Enterprise Start-Up[a] (*continued*)

B: Sri Lanka

	(1)		(2)		(3)		(4)	
	dP	**t**	**dP**	**t**	**dP**	**t**	**dP**	**t**
Household characteristics								
Number of male adults			−0.11	0.66	−0.19	1.13	−0.16	1.06
Number of female adults			0.20	1.45	0.15	1.19	0.17	1.38
Average age			−0.27	1.60	−0.26	1.48	−0.25	1.52
Human capital index			−0.17	0.93	−0.11	0.71	−0.21	1.18
Head female			0.28	0.76	0.33	0.88	0.26	0.77
Not Sinhalese			−1.05	1.68	−0.77	1.37	−1.04	1.89
Head parents entrepreneur			0.24	0.66	0.51	1.41	0.37	1.06
ln(Assets)			0.43	2.11	0.46	2.28	0.43	2.38
Benchmarks and components								
Connectivity	−0.02	0.10	0.03	0.13				
Inverse cost of transport to major city							0.33	1.81
Inverse cost of transport to main market							−0.26	1.65
Infrastructure services	−0.37	1.46	−0.47	1.96			−0.40	2.27
Business services	0.19	0.83	0.21	1.06				
Governance	−0.17	1.20	−0.17	1.28				
Human capital	0.38	2.02	0.44	1.93			0.28	1.47
Finance	0.10	0.53	0.08	0.42				
Community variables								
ln(Community size)					−0.16	0.96		
ln(Income per capita)					0.00	0.02		
Agricultural seasonality					0.23	1.26		
Enterprise openness					−0.29	1.71	−0.30	1.88
ln(Male wage rate)					0.36	2.05	0.24	1.38
Regression statistics								
lnL	−50.54		−49.29		−49.00		−48.48	
n	790		790		790		790	

Source: RIC Surveys.

Notes: a. Estimates of a weighted probit model. Columns headed with "dP" report the percentage point increase in the probability that an average household operates an enterprise in response to a one-standard-deviation increase in the explanatory variable. For the variables "Head female," "Head parents entrepreneur," and "Not Sinhalese," which are dummy variables, the change in the explanatory variable is one unit.

b. Significance levels implied by the t-statistics are: 1% if $t \geq 2.58$, 5% if $1.96 \leq t < 2.58$, 10% if $1.645 \leq t < 1.96$.

Annex G.
Benchmark Indicators

Table G.1 Missing Data for Communities and Components

	Nicaragua						Sri Lanka						Tanzania					
	obs	obs	average	st dev	missing communities	missing components	obs	obs	average	st dev	missing communities	missing components	obs	obs	average	st dev	missing communities	missing components
Number of communities	**98**						**147**						**149**					
E. Benchmark Indicators																		
Indicator 1: Connectivity	**98**		**0.34**	**0.18**	**0**		**146**		**0.40**	**0.17**	1		**149**		**0.20**	**0.14**	**0**	
Subindex 1: Time taken by main means of transportation to the nearest major city		98	0.32	0.30		0		146	0.47	0.31		1		149	0.26	0.31		0
Subindex 2: Cost of transportation to the nearest major city (by public transportation)		98	0.09	0.08		0		136	0.66	0.29		11		149	0.08	0.17		0
Subindex 3: Time taken by main means of transportation to the main market		85	0.33	0.31		13		145	0.49	0.32		2		149	0.52	0.39		0
Subindex 4: Cost of transportation to the main market (by bus)		98	0.07	0.09		0		146	0.61	0.36		1		149	0.08	0.17		0
Subindex 5: Distance to the post office		98	0.47	0.46		0		145	0.69	0.35		2		136	0.26	0.33		13
Subindex 6: Rail stop in walking distance		n.a.						145	0.12	0.33		2		149	0.23	0.42		0
Subindex 7: Percentage of households with fixed telephone lines		98	0.59	0.49		0		145	0.08	0.12		2		146	0.03	0.07		3
Subindex 8: Percentage of households with cell phones		98	0.51	0.50		0		145	0.10	0.14		2		147	0.13	0.20		2

Indicator 2: Infrastructure Services	**98**		**0.51**	**0.24**	**0**		**146**		**0.35**	**0.16**	1		**149**		**0.19**	**0.16**	**0**	
Subindex 1: Percentage of households that use electricity		97	0.67	0.31		1		145	0.69	0.29		2		148	0.12	0.22		1
Subindex 2: Availability of electricity index		94	0.62	0.39		4		145	0.55	0.39		2		145	0.31	0.43		4
Subindex 3: Percentage of households with access to protected water		98	0.62	0.32		0		145	0.80	0.28		2		149	0.47	0.38		0
Subindex 4: Percentage of households with fixed-line telephone		98	0.59	0.49		0		145	0.08	0.12		2		146	0.03	0.07		3
Subindex 5: Percentage of households with cellular phones		98	0.51	0.50		0		145	0.10	0.14		2		147	0.13	0.20		2
Subindex 6: Sewage channels in the community		98	0.08	0.28		0		145	0.11	0.31		2		148	0.07	0.25		1
Subindex 7: Garbage collection or disposal service in the community		98	0.49	0.50		0		145	0.12	0.32		2		149	0.13	0.33		0
Subindex 8: Most common road surface (internal road) is concrete or asphalt		n.a.						146	0.33	0.47		1		149	0.22	0.42		0
Indicator 3: Business Services	**98**		**0.24**	**0.40**	**0**		**142**		**0.15**	**0.22**	5		**146**		**0.07**	**0.15**	**3**	
Subindex 1: Engineering services available		98	0.23	0.43		0		142	0.06	0.23		5		146	0.04	0.20		3
Subindex 2: Management consulting services		98	0.22	0.42		0		142	0.07	0.26		5		146	0.16	0.37		3
Subindex 3: Marketing services		98	0.19	0.40		0		142	0.08	0.28		5		146	0.13	0.34		3
Subindex 4: Accounting services		98	0.27	0.44		0		142	0.11	0.31		5		146	0.02	0.14		3
Subindex 5: Legal services		98	0.31	0.46		0		142	0.26	0.44		5		146	0.07	0.25		3
Subindex 6: Insurance services		98	0.21	0.41		0		142	0.39	0.49		5		146	0.03	0.16		3
Subindex 7: Information technology		98	0.24	0.43		0		142	0.10	0.30		5		146	0.03	0.16		3

(*continued on the following page*)

Table G.1 Missing Data for Communities and Components (*continued*)

	Nicaragua						Sri Lanka						Tanzania					
	obs	obs	average	st dev	missing communities	missing components	obs	obs	average	st dev	missing communities	missing components	obs	obs	average	st dev	missing communities	missing components
Indicator 4: Governance	**96**		**0.67**	**0.13**	**2**		**125**		**0.67**	**0.04**	22		**140**		**0.50**	**0.08**	**9**	
Subindex 1: General policy and institutional constraints		96	0.79	0.16		2		125	0.98	0.03		22		140	0.85	0.19		9
Subindex 2: Infrastructure and services		60	0.00	0.00		38		125	0.13	0.06		22		140	0.01	0.03		9
Subindex 3: Dealing with government agencies		53	0.94	0.18		45		125	1.00	0.00		22		n.a.				
Subindex 4: Rule of law		94	0.50	0.30		4		125	0.49	0.14		22		140	0.49	0.16		9
Subindex 5: Conflict resolution and contract enforcement		96	0.59	0.22		2		125	0.75	0.14		22		140	0.66	0.18		9
Indicator 5: Human Capital	**97**		**0.21**	**0.08**	**1**		**134**		**0.33**	**0.09**	13		**146**		**0.21**	**0.06**	**3**	
Indicator 6: Finance Services*	**98**		**0.17**	**0.27**	**0**		**147**		**0.50**	**0.16**	0		**149**		**0.18**	**0.15**	**0**	
Subindex 1: Number of formal financial sources that serve a community		98	0.20	0.29		0		147	0.56	0.22		0		149	0.24	0.25		0
Subindex 2: Number of finance services that are provided by formal institutions in a community		98	0.15	0.25		0		147	0.47	0.23		0		149	0.12	0.16		0
Subindex 3: Access to loans		n.a.						147	0.47	0.23		0		140	0.19	0.22		9
Total missing number of observations on communities and components respectively					3	109					42	183					15	95
Total missing observations on communities and components respectively (%)					0.3	2.6					4.6	3.9					1.7	2.1

Source: the Authors.

* See note to Table E.6

Table G.2 Nicaragua: Correlation Coefficients Between Benchmark Indicators and Prices at Community Level

	Connectivity	Infra structure services	Business services	Governance	Human capital	Finance services
Index, price of food	0.21*	0.11	0.08	−0.04	0.13	0.04
Diesel price	−0.34*	−0.37*	−0.18	0.01	0.21	0.04
Fertilizer price	−0.17	−0.11	0.07	0.02	−0.19	0.08
Index, price of inputs	0.08	0.12	−0.08	−0.07	0.21	0.13
Average price of phone call	−0.21*	−0.14	0.17	−0.28*	−0.10	0.05
Index, price of transportation	0.17	0.01	0.07	0.30*	0.30*	0.05
Average price of salt	−0.36*	−0.16	0.06	−0.04	0.13	0.07
Index, price of Coca-Cola	0.02	−0.01	−0.05	−0.13	0.23*	−0.04
Price of soap	−0.05	0.03	0.04	0.01	0.34*	0.06
Index, wage (male)	−0.15	0.12	0.11	−0.29*	0.07	0.04
Index, wage (female)	−0.16	−0.04	−0.03	−0.22*	0.10	−0.01

Source: RIC Surveys.

Note: * indicates at least 10 percent significance level.

Table G.3 Sri Lanka: Correlation Coefficients Between Benchmark Indicators and Prices at Community Level[a]

	Connectivity	Infrastructure services	Business services	Governance	Human capital	Finance services
Average fertilizer (Rs/kg)	−0.17*	0.04	0.02	0.03	−0.17*	0.01
Average cement (Rs/50 kg bag)	−0.27*	−0.10	−0.14	0.25*	−0.04	−0.25*
Asbestos sheet (one sheet 96*53)	−0.27*	−0.20*	−0.21*	−0.08	0.05	−0.37*
Lawyer fee for a civil case	0.29*	0.18*	0.16*	−0.04	0.11	0.15*
Average price rice per kg.	0.16*	0.19*	0.12	0.27*	−0.06	0.30*
Price of wheat flour per kg.	0.14*	0.15*	0.26*	−0.24*	−0.07	0.15*
Price of Coca-Cola	−0.25*	−0.21*	−0.13	−0.04	−0.13	−0.35*
Price of kerosene per liter	−0.17*	−0.20*	−0.14*	−0.14	−0.11	−0.24*
Price of toilet soap	−0.04	−0.21*	−0.04	−0.12	−0.01	−0.17*
Price of a 40w bulb	−0.07	−0.26*	−0.06	−0.13	0.01	−0.21*
Price of a bike tire	−0.14*	−0.03	0.02	−0.15	−0.13	−0.07
Male casual wage	0.15*	0.02	0.13	−0.05	0.23*	0.02
Female casual wage	0.17*	−0.11	0.23*	−0.08	0.17*	0.04
Casual wage	0.15	−0.06	0.20*	−0.05	0.22*	−0.01
Time to clear a check in this area (days)	−0.30*	−0.18*	−0.11	0.11	−0.19*	−0.21*
Time to clear a clearance check in this area	−0.35*	−0.28*	−0.12	−0.03	−0.27*	−0.44*
Cost of bus fee to commercial center	−0.50*	−0.36*	−0.10	0.15	−0.10	−0.52*
Minutes to commercial center by bus	−0.48*	−0.17*	−0.09	0.01	−0.19*	−0.20*
Cost of 3-wheeler fee to commercial center	−0.60*	−0.23*	−0.19*	0.12	−0.19*	−0.38*
Minutes to commercial center by 3-wheeler	−0.45*	−0.16*	−0.06	0.03	−0.25*	−0.22*

Source: RIC Surveys.

Note: * indicates at least 10 percent significance level.
a. Several cost indicators are expressed in time required for service.

Table G.4 Tanzania: Correlation Coefficients Between Benchmark Indicators and Prices at Community Level

	Connectivity	Infrastructure services	Business services	Governance	Human capital	Finance services
Diesel per liter	−0.17	−0.02	−0.27	−0.59*	−0.22	−0.37*
Cement	−0.07	−0.05	−0.02	−0.05	−0.05	−0.19*
Galvanized steel sheet for roofing	0.05	0.08	0.00	0.07	0.21*	0.02
Ordinary rice per kg.	0.04	−0.08	−0.17*	−0.03	−0.09	−0.03
Coca-Cola per can	−0.06	−0.02	−0.09	0.06	−0.11	−0.07
Kerosene per liter	0.12	0.11	0.10	0.07	0.06	0.11
Bar of toilet soap	−0.12	−0.12	−0.19*	0.18*	−0.08	−0.06
Light bulb	−0.13	−0.18	−0.09	−0.08	−0.25*	0.07
Agriculture wage per day	0.09	0.10	0.05	−0.06	0.07	−0.03
Construction wage per day	0.05	0.07	0.12	−0.13	0.00	−0.05
Rural public works per month	0.02	−0.05	−0.07	−0.21*	−0.01	0.08
Agriculture wage per day (female)	−0.06	−0.04	−0.06	0.02	0.04	−0.14*
Construction wage per day (female)	−0.03	0.12	0.05	0.03	0.06	−0.11
Rural public works per month (female)	0.01	−0.03	−0.08	−0.10	0.06	0.05
Telephone call to the capital town of the nearby region	−0.22*	−0.10	−0.04	−0.09	−0.26*	−0.03
Cell phone/mobile phone call to the capital town of the nearby region	−0.26*	−0.22*	−0.22*	0.04	−0.15	−0.17
Average transport cost	−0.08	0.08	−0.01	0.20*	0.03	−0.19*

Source: RIC Surveys.

Note: * indicates at least 10 percent significance level.

Table G.5 Nicaragua: Correlation Coefficients Between Benchmark Indicators and Community Characteristics

	Connectivity	Infrastructure services	Business services	Governance	Human capital	Finance services
Population	0.53*	0.51*	0.46*	0.08	0.27*	0.51*
Road surface	0.37*	0.42*	0.27*	−0.08	0.32*	0.06
Percent households with gas	0.44*	0.57*	0.32*	0.09	0.59*	0.17
Number of public services	0.69*	0.83*	0.46*	−0.04	0.45*	0.31*
Highest level of school type	0.32*	0.51*	0.34*	0.13	0.43*	0.53*
Highest level of health center type	0.28*	0.44*	0.30*	0.08	0.29*	0.39*
Index: Highest level of health center type, weighted by distance	0.46*	0.56*	0.38*	0.15	0.38*	0.47*
Number of types of banks available	0.42*	0.50*	0.43*	0.20*	0.40*	0.72*
Percentage of households that buy outside						
Immigrants (as % of population)	−0.29*	−0.38*	−0.31*	−0.04	−0.23*	−0.39*
Number of business services provided by chamber of commerce	0.27*	0.24*	0.27*	0.03	0.03	0.20*
Number of business services provided by overall business association	0.15	0.13	0.15	0.16	0.11	0.17*
Number of business services provided by sector business association	0.35*	0.27*	0.20*	0.06	0.14	0.25*
Number of services available to businesses	0.36*	0.30*	0.28*	0.09	0.12	0.28*
Number of taxes	0.50*	0.44*	0.27*	0.20*	0.41*	0.45*
Percent households with electricity	0.29*	0.62*	0.22*	−0.10	0.37*	0.08
Quality of electricity	0.22*	−0.16	−0.10	−0.02	0.15	0.04
Distance to market	−0.23*	0.08	0.12	−0.18*	0.08	0.14
Distance to city	−0.06	0.16	0.28*	0.01	0.22*	0.38*
Seasonality of male labor in agriculture	0.06	0.01	0.09	−0.03	−0.12	−0.08
Seasonality of female labor in agriculture	0.06	−0.03	−0.02	0.08	−0.11	−0.05
Percent of households that benefit from a series of programs (cumulative, might be over 100%)	0.14	0.11	0.25*	0.20*	0.22*	0.25*
Percent of households that benefit from infrastructure programs (cumulative, might be over 100%)	0.12	0.09	0.21*	0.19*	0.22*	0.22*
Percent of households that benefit from social programs (cumulative, might be over 100%)	0.14	0.12	0.34*	0.14	0.19*	0.29*
Index: Formality of conflict-solving institution	0.10	0.05	0.06	0.01	0.10	0.13
Number of crimes	0.36*	0.25*	0.11	0.04	0.17*	0.25*
Percent of reported incidents solved	−0.20*	−0.39*	−0.27*	−0.11	−0.26*	−0.45*
Percent of conflicts solved by neighbors/friends	0.19*	0.14	0.17*	0.06	0.16	0.03
Percent of conflicts solved by community leaders	−0.15	−0.08	0.01	−0.04	−0.15	0.17*
Percent of conflicts solved by local judge	0.18*	0.06	0.02	0.04	0.24*	0.12
Percent of conflicts solved by district judge	0.01	0.02	0.05	−0.01	−0.07	0.06
Percent of conflicts solved by magistrate court	0.14	0.12	0.01	−0.04	0.14	0.13
Percent of conflicts solved by labor court	−0.10	−0.16	−0.06	−0.02	−0.05	−0.11
Total cultivated land per capita	−0.14	−0.12	−0.12	−0.08	0.00	0.04
Amount of lowland per capita	−0.03	−0.02	−0.07	−0.05	0.01	−0.08
Total upland per capita	−0.23*	−0.20*	−0.13	−0.11	−0.06	0.18

Source: RIC Surveys.

Note: * indicates at least 10 percent significance level.

Table G.6 Sri Lanka: Correlation Coefficients Between Benchmark Indicators and Community Characteristics

	Connectivity	Infrastructure services	Business services	Governance	Human capital	Finance services
Population	0.19*	0.37*	0.27*	0.04	0.04	0.17*
Road surface	−0.19*	−0.16*	0.09	0.20*	−0.19*	−0.14
Percent households with gas	0.41*	0.52*	0.26*	−0.02	0.40*	0.35*
Number of public services	0.61*	0.57*	0.21*	0.00	0.34*	0.43*
Highest level of school type	0.01	0.06	−0.12	−0.11	0.04	−0.02
Number of hospital services	0.08	0.14	0.08	0.09	0.14	0.04
Highest level of health center type	0.05	0.06	0.02	0.16*	0.12	0.02
Index: Highest level of health center type, weighted by distance	0.15*	0.17*	0.08	0.05	0.18*	0.16*
Number of banks	0.32*	0.19*	0.28*	−0.05	0.16*	0.32*
Percent of households series of programs benefit (cumulative, might be over 100%)	−0.04	−0.04	0.02	0.11	−0.15*	−0.02
Percent of households infrastructure programs benefit (cumulative, might be over 100%)	0.05	0.01	−0.05	0.15	0.03	0.24*
Percent of households social programs benefit (cumulative, might be over 100%)	0.02	0.02	−0.05	−0.03	0.03	0.16*
Index: Formality of conflict-solving institution	−0.03	0.03	0.19*	−0.13	0.04	0.10
Percent of reported incidents solved	−0.13	0.01	0.09	−0.10	−0.14	0.10
Number of business services provided by chamber of commerce	0.00	0.02	−0.17*	0.02	0.00	0.02
Number of business services provided by overall business association	0.02	0.14	−0.12	−0.15	0.05	−0.12
Number of business services provided by sector business association	0.01	−0.06	−0.10	0.08	0.10	0.06
Immigrants, as % of population	0.04	−0.03	0.03	0.03	0.08	0.12
Number of business services (overall)	0.01	0.13	0.21*	0.26*	−0.07	−0.01
Number of taxes	0.40*	0.19*	0.21*	0.06	0.20*	0.26*
Percent households with electricity	0.44*	0.55*	0.16*	−0.02	0.38*	0.51*
Quality of electricity	0.00	−0.26*	−0.10	0.14	−0.06	0.07
Distance to market	−0.38*	−0.23*	0.09	0.02	0.07	0.05
Distance to city	−0.56*	−0.33*	0.08	0.03	0.06	0.03
Standard deviation of agricultural male labor	−0.22*	−0.27*	−0.21*	0.23*	−0.21*	−0.14
Standard deviation of agricultural female labor	−0.29*	−0.28*	−0.17*	0.20*	−0.17*	−0.16*
Percent conflicts solved by neighbors/friends	−0.11	−0.11	0.02	−0.08	−0.12	0.00
Percent conflicts solved by community leaders	0.03	0.06	0.08	−0.18*	0.05	0.18*
Percent conflicts solved by local judge	0.05	0.03	0.05	−0.21*	0.08	0.11
Percent conflicts solved by court of appeal	0.01	0.13	0.06	−0.11	0.02	0.02
Percent conflicts solved by magistrate court	0.01	−0.02	0.03	0.05	−0.06	0.02
Percent conflicts solved by labor court	−0.16*	−0.05	0.30*	0.03	−0.08	0.03
Number of crimes	−0.06	0.02	0.03	0.03	−0.03	0.08
Total cultivated land per capita	−0.35*	−0.34*	−0.13	−0.06	−0.14	−0.30*
Total lowland per capita	−0.29*	−0.34*	−0.15*	0.09	−0.11	−0.29*
Total upland per capita	−0.31*	−0.28*	−0.10	−0.11	−0.13	−0.25*

Source: RIC Surveys.

Note: * indicates at least 10 percent significance level.

Table G.7 Tanzania: Correlation Coefficients Between Benchmark Indicators and Community Characteristics

	Connectivity	Infrastructure services	Business services	Governance	Human capital	Finance services
Population	0.00	0.01	−0.01	0.14*	−0.02	−0.02
Road surface	0.10	0.27*	0.06	−0.29*	0.01	0.20*
Percent households with cell phone	0.47*	0.70*	0.17*	0.10	0.36*	0.15*
Number of public services	0.67*	0.72*	0.26*	0.01	0.38*	0.26*
Highest level of school type	0.11	0.24*	0.17*	−0.02	0.17*	0.02
Highest level of health center type	0.19*	0.25*	0.13	−0.07	0.26*	−0.05
Index: Highest level of health center type, pondered by distance	−0.26*	−0.35*	−0.27*	0.05	−0.21*	−0.04
Index: Highest level of health center type, weighted by distance	0.25*	0.42*	0.27*	−0.01	0.22*	0.12
Number of banks	0.18*	0.27*	0.43*	−0.08	0.14*	0.38*
Number of banks available, weighted by distance	0.05	−0.07	0.12	−0.04	−0.10	−0.24*
Index: Mean distance to banking centers	0.06	−0.06	0.12	−0.03	−0.08	−0.29*
Index: Formality of conflict-solving institution	0.02	0.05	0.08	0.10	0.02	0.01
Percent of conflicts solved by neighbors/friends	0.13	0.20*	0.01	0.22*	0.13	0.00
Percent of conflicts solved by community leaders	0.09	−0.01	0.08	0.16*	−0.05	0.00
Percent of conflicts solved by ward tribunal	−0.18*	−0.13	0.00	0.14*	−0.13	−0.14*
Percent of conflicts solved by magistrate court	−0.04	−0.02	0.04	−0.06	−0.02	−0.05
Percent of conflicts solved by resident court	0.05	0.10	0.02	−0.01	0.10	−0.02
Emigrants, as % of population	0.06	0.15*	0.14*	0.06	−0.04	−0.01
Number of business services provided by chamber of commerce	0.20*	0.07	0.24*	−0.09	0.06	0.18*
Number of business services provided by overall business association	0.08	0.04	0.40*	−0.10	0.02	0.06
Number of business services provided by sector business association	0.06	0.01	0.26*	−0.09	0.04	0.12
Number of business services (overall)	0.14*	0.05	0.35*	−0.11	0.05	0.15*
number of taxes	0.18*	0.16*	0.18*	−0.06	0.17*	0.22*
Number of disasters/events	−0.09	−0.04	0.09	−0.02	−0.17*	0.11
Number of crimes	0.21*	0.35*	0.15*	0.08	0.14*	0.10
Percent of reported incidents solved	−0.10	−0.14	−0.12	0.14	−0.13	0.01
Percent households with electricity	0.46*	0.68*	0.12	0.06	0.35*	0.14*
Quality of electricity	−0.16	−0.19	−0.10	0.18	−0.09	−0.04
Distance to market	−0.40*	−0.27*	−0.02	0.07	−0.17*	−0.05
Distance to city	−0.31*	−0.08	0.02	0.14	−0.04	−0.14
Percent households that benefit from investment programs	−0.07	0.03	−0.01	−0.11	0.01	0.04
Total cultivated land per capita	−0.27*	−0.28*	−0.08	0.04	−0.11	−0.07
Amount of lowland per capita	−0.14	−0.13	−0.10	−0.05	−0.13	−0.20*
Amount of upland per capita	−0.18*	−0.23*	−0.07	0.19*	−0.11	−0.15

Source: RIC Surveys.

Note: * indicates at least 10 percent significance level.

Table G.8 Descriptive Statistics of Variables Used in Regression Analysis

A: Dependent variables

	Variable name	Obs	Mean	St.Dev.	Min	Max	Definition
entdense_hh	Enterprise density	147	0.162	0.132	0.018	0.935	number of enterprises per household
lnworkincpc[a]	Income from work#	130	5.510	0.662	1.562	7.376	log of per capita income from work (farm, nfhe, wage) [a]
share_rnfinc[a]	Ratio RNF income	130	0.787	0.586	0.000	6.706	ratio of rural nonfarm income over total income from work[a]
share_rnfwork[a]	Ratio RNF workers	130	0.808	0.213	0.263	1.000	ratio of rural nonfarm workers over total number of workers[a]
lnrnfincpc[a]	RNF income per capita#	128	5.166	0.748	1.044	7.324	log of rural nonfarm income per capita[a]
lnrnfincpw[a]	RNF income per worker#	128	6.320	0.646	3.285	8.474	log of rural nonfarm income per worker[a]
lfpr[a]	Labor force participation rate	130	0.592	0.127	0.337	0.995	labor force participation rate[a]
lnwmale[a]	Male wage#	136	5.456	0.242	4.828	5.991	log of reported male wage[a]
ratio_salecost[b]	Ratio sales / total cost	147	2.643	7.156	0.577	86.06	ratio of sales over total cost[b]
ratio_nvatfc[b]	Ratio NVA / factor cost	142	21.12	67.41	0.455	658.0	ratio of net value added over total factor cost[b]
lnGVA[b]	Gross Value Added (GVA)#	147	7.166	1.386	4.284	12.51	log of gross value added per enterprise[b]
lnSD_GVA[b]	St dev GVA#	144	7.307	1.908	2.083	13.39	log of standard deviation of gross value added[b]
seasadj[b]	Seasonal adjustment	147	10.11	1.387	6.167	12.00	seasonal adjustment factor in enterprises (higher is less seasonal) [b]
agseas	Agricultural seasonality	141	0.744	0.293	0.000	1.379	agricultural seasonality (higher is more seasonal)

B: Explanatory variables

lnpopnw	Community population size#	142	7.312	0.596	6.310	8.990	log of population size
shsinh	Share Sinhalese population	142	77.148	36.950	0.000	100.000	share of Sinhalese among the population
shilliterate	Illiteracy	142	13.173	13.032	0.000	70.000	share of illiterate
dschoolindex	Proximity to school	142	0.360	0.208	0.000	1.000	proximity to school index
dhospindex	Proximity to hospital	142	0.227	0.136	0.000	0.571	proximity to hospital index
distcityindex	Proximity to city	142	0.594	0.372	0.000	1.000	proximity to city index
openness[b]	Openness	124	2.135	0.689	1.048	4.834	openness (inputs, customers) among enterprises[b]
lowlandpc	Lowland per capita	139	0.123	0.192	0.000	1.019	acres of lowland per capita
uplandpc	Upland per capita	139	0.247	0.441	0.000	4.587	acres of upland per capita
conn_index	Connectivity	146	0.398	0.167	0.026	0.781	connectivity benchmark
infra_index	Infrastructure services	146	0.345	0.165	0.000	0.795	infrastructure services benchmark
devs_index	Business services	142	0.153	0.224	0.000	1.000	business services benchmark
gov_index[b]	Governance	125	0.670	0.041	0.569	0.756	governance benchmark[b]
hc_index[a]	Human capital	134	0.329	0.086	0.133	0.577	human capital benchmark[a]
finance_index[c]	Finance services	147	0.501	0.162	0.002	0.842	finance services benchmark[c]
pricerice	Price rice	131	30.319	2.640	20.000	38.000	consumer price of rice
pcoke	Price Coca-Cola	147	20.755	3.035	12.870	32.540	consumer price of can of Coca-Cola
pkerosine	Price kerosene	146	27.248	2.452	21.500	36.000	consumer price of kerosene

Source: RIC Surveys.

Note: a. Aggregated from household data.
b. Aggregate from enterprise data.
c. In part, an aggregate from enterprise data.
Logarithmic value.

Table G.9 Regressions at the Community Level

A: Full set of explanatory variables

	Enterprise density		Income from work#		Ratio RNF income		Ratio RNF workers		RNF income per capita#	
	b	t	b	t	b	t	b	t	b	t
Community population size#	−0.0211	−0.98	0.2108	2.19	0.0356	0.62	0.0972	2.93	0.1261	0.95
Share Sinhalese population	0.0008	1.92	0.0003	0.14	−0.0035	−2.61	0	0	−0.0044	−1.39
Illiteracy	0.0015	1.61	0.0033	0.73	−0.0011	−0.46	−0.0022	−1.25	0.0036	0.63
Proximity to school	0.1431	2	0.1754	0.6	−0.1033	−0.64	−0.0725	−0.73	0.0264	0.07
Proximity to hospital	0.3049	3.15	−0.4815	−1.18	0.2021	0.92	0.1955	1.45	0.3559	0.8
Proximity to city	0.01	0.23	0.0857	0.46	−0.2781	−2.62	−0.061	−0.89	−0.2076	−0.94
Openness	−0.0318	−2.61	−0.1983	−2.53	−0.0413	−1.29	0.0066	0.23	−0.2262	−1.99
Lowland per capita	0.0139	0.3	0.1571	0.39	−0.2954	−1.38	−0.4269	−2.64	−0.6903	−1.35
Upland per capita	0.0202	1.12	−0.1302	−0.84	−0.0827	−1.57	−0.0108	−0.23	−0.1847	−0.86
Connectivity	0.1865	2.14	−0.8791	−1.75	0.2499	1.08	−0.0448	−0.3	−0.2344	−0.39
Infrastructure services	−0.025	−0.29	−1.0754	−1.83	−0.1675	−0.63	0.0524	0.34	−0.7345	−0.98
Business services	0.1486	3.45	0.8895	4.03	−0.0755	−0.53	0.08	1.24	0.5357	1.63
Governance	−0.2631	−0.8	−4.09	−2.41	−1.0194	−1.52	0.1503	0.34	−3.7865	−1.94
Human capital	0.3354	2.26	2.1749	3.64	−0.1016	−0.28	−0.1634	−0.67	2.28	2.92
Finance services	−0.1443	−1.71	0.1243	0.22	0.8128	2.3	0.0104	0.06	0.6406	0.83
Price rice	0.0104	2.3	−0.0561	−2.27	−0.0154	−1.37	−0.0031	−0.35	−0.0793	−2.56
Price Coca-Cola	0.0064	0.93	−0.0889	−1.6	0.0088	0.37	0.0052	0.4	−0.0394	−0.56
Price kerosene	0.0049	1.28	−0.023	−0.92	−0.0134	−1.22	−0.0125	−1.41	−0.0381	−1.1
Intercept	−0.3773	−1.04	11.0703	4.78	2.0168	2.5	0.4588	0.7	11.2461	3.8
N of communities	103		103		103		103		101	
R2	0.487		0.41		0.334		0.421		0.258	
F	4.43		3.23		3.32		6.23		2.1	
p-value	0		0		0		0		0.013	
Standard error	0.079		0.48		0.24		0.155		0.626	

(*continued on the following page*)

Table G.9 Regressions at the Community Level (*continued*)

	Labor force participation rate		Labor force participation rate		Male wage#		Ratio sales / total cost		Ratio NVA / factor cost	
	b	t	b	t	b	t	b	t	b	t
Community population size#	−0.0236	−0.17	0.0482	1.69	−0.0163	−0.4	−2.945	−1.46	−34.43	−1.89
Share Sinhalese population	−0.0034	−0.97	−0.001	−1.79	0.0012	1.37	−0.119	−1.69	−0.88	−1.47
Illiteracy	0.0076	1.22	−0.0013	−1.02	0.0024	1.06	0.061	0.64	1.04	1.32
Proximity to school	0.191	0.47	−0.022	−0.28	0.0393	0.29	−8.987	−1.23	−70.93	−1.13
Proximity to hospital	0.2756	0.58	0.0238	0.21	−0.2257	−1.28	−2.899	−0.39	−44.93	−0.72
Proximity to city	−0.0766	−0.33	−0.008	−0.14	0.1102	1.38	−1.128	−0.36	2.69	0.1
Openness	−0.1548	−1.33	−0.0119	−0.69	−0.0002	−0.01	2.913	1.48	20.25	1.28
Lowland per capita	0.3597	0.66	−0.1367	−1.72	−0.1603	−1.09	0.536	0.08	−23.16	−0.34
Upland per capita	−0.344	−1.4	0.0607	2.71	0.1726	3.53	−1.124	−0.4	13.81	0.72
Connectivity	−0.2898	−0.51	0.01	0.08	0.2423	1.24	−23.614	−1.87	−188.38	−1.64
Infrastructure services	−0.5713	−0.69	−0.1792	−1.59	0.0122	0.07	21.505	1.65	173.33	1.51
Business services	0.096	0.31	0.1182	2.55	0.1319	1.67	21.117	1.94	148.14	1.75
Governance	−4.5566	−2.32	0.6151	1.86	0.043	0.07	21.576	0.78	390.36	1.4
Human capital	1.8947	2.57	0.448	2.66	0.58	1.78	−6.291	−0.6	−56.25	−0.57
Finance services	0.7399	0.9	−0.1622	−1.7	0.0752	0.46	20.334	1.57	167.82	1.53
Price rice	−0.0589	−1.68	−0.0073	−1.27	0.0301	2.62	−1.633	−1.81	−12.54	−1.47
Price Coca-Cola	−0.0167	−0.21	0.0041	0.39	0.0045	0.31	0.16	0.23	−5.81	−0.74
Price kerosene	−0.0267	−0.84	0.0049	0.88	−0.0267	−2.58	0.647	1.24	6.85	1.63
Intercept	12.2117	3.94	−0.0583	−0.13	4.7297	5.86	36.941	1.11	302.27	0.9
N of communities	101		103		96		103		101	
R2	0.225		0.293		0.45		0.554		0.495	
F	1.51		3.01		8.12		0.26		0.47	
p-value	0.108		0		0		0.999		0.964	
Standard error	0.626		0.107		0.178		9.842		85.452	

	Gross Value Added (GVA)#		St dev GVA#		Seasonal adjustment		Agricultural seasonality	
	b	t	b	t	b	t	b	t
Community population size#	0.1179	0.44	0.4663	1.18	−0.4713	−1.51	−0.0425	−0.71
Share Sinhalese population	−0.0133	−2.09	−0.0182	−1.95	−0.0146	−2.41	0.0025	2.09
Illiteracy	−0.0027	−0.23	0.0024	0.16	0.0101	0.98	0.0064	2.54
Proximity to school	0.8479	1.05	1.1159	1	1.2544	1.43	−0.1231	−0.85
Proximity to hospital	1.4538	1.44	2.8082	2.13	−0.4248	−0.41	0.3718	1.7
Proximity to city	−0.071	−0.15	−0.4235	−0.63	−0.1298	−0.24	0.3039	3.14
Openness	0.9783	5.53	1.3591	6.14	0.2846	1.73	0.0198	0.64
Lowland per capita	2.0124	1.79	2.336	1.83	−0.3681	−0.53	0.1994	1.59
Upland per capita	−0.5757	−1.27	−0.6686	−1.24	−0.0252	−0.06	−0.045	−0.92
Connectivity	−1.868	−1.64	−1.4115	−0.88	−0.3394	−0.26	−0.3934	−1.79
Infrastructure services	−0.8706	−0.58	−1.7739	−0.85	0.7207	0.58	0.0476	0.2
Business services	1.1365	1.92	0.9195	1.23	−1.4604	−2.29	−0.1501	−1.16
Governance	−11.3944	−2.63	−16.0493	−2.62	−3.3118	−0.76	3.301	4.33
Human capital	−2.1677	−1.43	−1.7799	−0.76	−1.4056	−0.66	−0.9383	−2.69
Finance services	3.4432	2.4	4.8838	2.4	3.8692	2.83	−0.4954	−2.08
Price rice	−0.0613	−0.97	−0.0638	−0.74	−0.069	−1.01	−0.0253	−2.02
Price Coca-Cola	0.0415	0.3	0.1421	0.7	−0.0343	−0.27	−0.016	−0.7
Price kerosene	−0.0899	−1.38	−0.1186	−1.42	0.0215	0.39	0.0461	3.75
Intercept	15.758	2.5	13.9641	1.65	17.0432	3.9	−1.1687	−1.58
N of communities	103		103		103		101	
R2	0.585		0.564		0.289		0.596	
F	7.53		9.61		1.8		9.24	
p-value	0		0		0.039		0	
Standard error	1.156		1.603		1.197		0.225	

Source: RIC Surveys.

Note: # Logarithmic value.

(*continued on the following page*)

Table G.9 Regressions at the Community Level (*continued*)

B: Reduced set of explanatory variables (removing variables with |t| < 0.50)

	Enterprise density		Income from work#		Ratio RNF income		Ratio RNF workers		RNF income per capita#	
	b	t	b	t	b	t	b	t	b	t
Community population size#	−0.0232	−1.14	0.1976	2.14			0.0654	1.96	0.1417	1.03
Share Sinhalese population	0.0008	2.46			−0.0036	−3.23			−0.0042	−1.37
Illiteracy	0.0016	1.69	0.003	0.69			−0.0021	−1.71	0.0034	0.59
Proximity to school	0.1434	2.03	0.1963	0.69			0.1556	1.66		
Proximity to hospital	0.2986	3.06	-0.4819	-1.2	0.2173	1.01			0.375	0.94
Proximity to city			0.1039	0.6	−0.2844	−2.76	−0.1199	−2.03	−0.2442	−1.24
Openness	−0.0312	−2.62	−0.1979	−2.63	−0.0489	−1.73			−0.215	−2.02
Lowland per capita					−0.3202	−1.59	−0.281	−1.68	−0.7197	−1.49
Upland per capita	0.0228	1.54	−0.1043	−0.88	−0.0616	−1.21			−0.194	−0.92
Connectivity	0.1955	2.95	−0.8361	−1.81	0.2949	1.33				
Infrastructure services			−1.0952	−2.67	−0.1849	−0.74	0.2464	1.55	−0.7354	−1.05
Business services	0.146	3.69	0.8849	4.15			0.0349	0.61	0.5003	1.57
Governance	−0.243	−0.86	−3.7898	−2.68	−0.9992	−1.46			−3.484	−1.95
Human capital	0.3237	2.62	2.2523	4.23					2.1859	2.82
Finance services	−0.1458	−2.22			0.8022	2.22			0.5812	0.77
Price rice	0.0103	2.35	−0.0585	−2.49	−0.0115	−1.09			−0.0834	−2.8
Price Coca-Cola	0.0065	1	−0.0884	−1.66						
Price kerosene	0.005	1.29	−0.02	−0.81	−0.0128	−1.15	−0.0146	−1.9	−0.0367	−1.02
Intercept	−0.3851	−1.17	10.9974	5.4	2.2242	3.14	0.7248	2.36	10.1925	4.7
N of communities	103		103		103		123		101	
R2	0.485		0.408		0.325		0.328		0.252	
F	5.29		3.76		3.4		7.32		2.22	
p-value	0		0		0		0		0.011	
Standard error	0.078		0.472		0.234		0.173		0.617	

	Labor force participation rate		Labor force participation rate		Male wage#		Ratio sales / total cost		Ratio NVA / factor cost	
	b	t	b	t	b	t	b	t	b	t
Community population size#			0.0438	1.78			−3.643	−1.72	−33.05	−1.88
Share Sinhalese population	−0.0033	−0.97	−0.001	−2	0.0012	2.02	−0.124	−1.78	−0.9	−1.5
Illiteracy	0.0067	1.15	−0.0013	−1.05			0.063	0.71	0.95	1.27
Proximity to school							−8.583	−1.24	−79.34	−1.34
Proximity to hospital	0.2744	0.6			−0.177	−1.16			−49.61	−0.84
Proximity to city					0.1355	1.91				
Openness	−0.1375	−1.29	−0.013	−0.79			2.791	1.63	20.45	1.42
Lowland per capita	0.3791	0.73	−0.1354	−1.79	−0.1756	−2.34				
Upland per capita	−0.3661	−1.57	0.0642	3.08	0.163	4.36				
Connectivity	−0.4177	−0.7			0.1985	1.28	−23.884	−2.06	−194.61	−1.86
Infrastructure services	−0.4966	−0.64	−0.1776	−1.6			20.784	1.68	172.09	1.55
Business services			0.1199	2.74	0.1539	2.31	20.466	2.02	145.48	1.83
Governance	−4.4186	−2.72	0.6117	2.28			25.002	0.97	344.91	1.39
Human capital	1.8344	2.28	0.4541	2.78	0.6901	2.65	−8.856	−0.91	−63.17	−0.7
Finance services	0.6635	0.87	−0.1686	−1.93			20.084	1.65	171.17	1.63
Price rice	−0.0626	−1.95	−0.0063	−1.11	0.0192	1.99	−1.616	−1.91	−13.04	−1.75
Price Coca-Cola									−5.41	−0.7
Price kerosene	−0.0261	−0.83	0.005	0.88	−0.0227	−2.74	0.734	1.34	6.57	1.6
Intercept	11.8047	5.61	0.0264	0.09	4.9896	12.37	40.511	1.43	350.93	1.12
N of communities	101		103		99		105		103	
R2	0.221		0.289		0.447		0.55		0.494	
F	1.99		4.01		16.25		0.36		0.51	
p-value	0.03		0		0		0.98		0.928	
Standard error	0.61		0.105		0.173		9.38		82.007	

(*continued on the following page*)

Table G.9 Regressions at the Community Level (*continued*)

	Gross Value Added (GVA)#		St dev GVA#		Seasonal adjustment		Agricultural seasonality	
	b	t	b	t	b	t	b	t
Community population size#			0.4759	1.28	−0.4662	−1.47		
Share Sinhalese population	−0.0135	−2.2	−0.0184	−1.94	−0.0142	−2.38	0.0025	2.24
Illiteracy					0.0111	1.12	0.006	2.28
Proximity to school	0.9567	1.14	1.1068	1.01	1.3255	1.58	−0.1545	−1.14
Proximity to hospital	1.4444	1.55	2.809	2.14			0.3848	1.7
Proximity to city			−0.4091	−0.62			0.3033	3.06
Openness	0.9733	6.18	1.3618	6.1	0.2751	1.76	0.0164	0.55
Lowland per capita	1.976	1.77	2.3503	1.85	−0.356	−0.52	0.2197	1.95
Upland per capita	−0.5336	−1.26	−0.6808	−1.27			−0.0519	−1.04
Connectivity	−1.91	−1.71	−1.4622	−0.92			−0.418	−2.04
Infrastructure services	−0.7847	−0.55	−1.8033	−0.87	0.6258	0.55		
Business services	1.1788	2.04	0.9036	1.2	−1.4742	−2.42	−0.1536	−1.19
Governance	−11.4427	−2.86	−15.9317	−2.66	−2.9963	−0.7	3.2262	4.2
Human capital	−2.2286	−1.39	−1.7947	−0.76	−1.6403	−0.78	−0.8808	−2.64
Finance services	3.3834	2.58	4.8582	2.44	3.5932	2.94	−0.468	−1.97
Price rice	−0.0509	−0.79	−0.067	−0.78	−0.0689	−1.05	−0.0274	−2.02
Price Coca-Cola			0.1451	0.72			−0.013	−0.58
Price kerosene	−0.0907	−1.47	−0.116	−1.46			0.0451	3.76
Intercept	17.1409	4.28	13.873	1.68	16.5634	5.01	−1.3894	−1.87
N of communities	103		103		103		101	
R2	0.583		0.564		0.283		0.596	
F	8.74		9.17		2.13		9.24	
p-value	0		0		0.022		0	
Standard error	1.132		1.593		1.161		0.225	

Source: RIC Surveys.

Note: # Logarithmic value.

Annex H. Enterprise Investment Climate Outcomes and Interactions (EICOs and EICIs): Notes and Tables

MEASUREMENT AND SPECIFICATION ISSUES

Corruption

Corruption is an investment climate constraint since it increases the cost of doing business for producers and traders and may distort incentives. It causes business risk through the ensuing uncertainty about the rule of law and the integrity of the legal system. Consumers may also be affected through higher prices and inferior quality services.

The RIC enterprise questionnaire aimed at registering corruption through questions about informal payments for public services (registration, licenses, permits, inspections) and hooking up to infrastructural services (phone, water, electricity, sewage, gas). In addition, entrepreneur perceptions were registered about the integrity of the rule of law.

In the three RICS pilot countries, respondents were asked whether corruption affected their businesses. In Nicaragua all interviewees were directly asked whether corruption was an investment climate constraint. In contrast, entrepreneurs in Tanzania and Sri Lanka were asked whether "governance [poses] a problem [for the operation and growth of your establishment],"[91] which functioned as a screening question for follow-up questions on corruption, economic policy uncertainty, crime and theft, the legal system, and war. The most likely effect of this questionnaire structure was that corruption was relatively underreported. Table H.10 provides insights on this issue. In Nicaragua 412 enterprises, or 36 percent of the total, reported that corruption as a problem. In Sri Lanka only 5 percent and in Tanzania 27 percent did so.

Other than this EICO question, most enterprises have to deal to some extent with government agencies, for example, to obtain various permits and services; those who did were asked about payment of informal fees. Information about such unofficial fees is incorporated in the infrastructure and services component of the governance benchmark indicator (see Annex E). Furthermore, responses on the predictability of laws, kickbacks, and the manipulability of rent seeking by government officials are examined under the heading of enterprise investment climate interactions (EICIs) in Chapter 5.

The registration of questions about corruption is sensitive. In general, respondents may be reluctant to give answers because they may be unsure about confidentiality. In Nicaragua, for example, the survey was conducted before the election, and the political fears of the time contributed to low responses. But respondents are not the only ones who might worry about the implications of responses to questions about corruption: public and private survey agencies may have disincentives to pursue sensitive questions as well. They may fear that their association with a critical report may affect future government contracts or involve political risks.

The analysis of corruption and governance, while inconclusive owing to the reluctance of some respondents, is nonetheless instructive. The descriptive findings discussed here and in Section 5.1 of Chapter 5 are examined in greater detail by means of regression analytical models in Sections 5.3 and 5.4.

Computation of Standard Errors

The formula that econometric theory advocates for the computation of standard errors (labeled "V3"

for convenience) is based on asymptotic considerations, but the number of communities only ranges between 100 and 150. In the absence of sampling weights, it is easily shown that the asymptotic formula should simplify to the standard formula ("V1") that statistical software always uses, but actually the asymptotic formula V3 yields estimates of the standard errors that vastly overstates those computed with the standard formula V1. From a limited amount of experimentation, it was determined that the use of sampling weights typically increases the V1 standard errors in Nicaragua by 10.7 percent for enterprise characteristics and 9.7 percent for community characteristics and in Sri Lanka by 31.8 percent for enterprise characteristics and 47.4 percent for community characteristics. The standard error of the estimated standard deviation of the random error did not need adjustment in the case of Nicaragua but needed adjustment upward by 59 percent in Sri Lanka.

The effect of ignoring weights and random effects and the difference in using V1 or V3 is illustrated in Table H.11, showing a typical EICO model for cost of finance in Sri Lanka. The use of weights propels some explanatory variables to greater prominence (for example, Sinhalese manager, household-based enterprise, infrastructure services, business services, and finance services) and diminishes the measured effect of others (female manager). Incorporating a random effect again leads to changes in parameter estimates (industry dummies, income per capita, infrastructure services, and finance services).

The s(V1) standard errors incorporate the adjustment mentioned above. Among the enterprise characteristics, the s(V1) and s(V3) standard errors are not dramatically different, with three exceptions (age of enterprise, Sinhalese manager, and other production enterprise). But major differences are found among all community characteristics: s(V3) standard errors are about two and a half times as large as s(V1) standard errors, even after the adjustment of s(V1). T-statistics based on s(V3) collapse: if they are to be believed, nothing is statistically significant anymore. The adjustment of s(V1) has a degree of arbitrariness, but t-statistics are not implausible. This methodological econometric issue must be examined in greater detail at some point.

Effect of Size and Productivity on EICO Responses

The estimated EICO models omit enterprise size and productivity as determinants of the entrepreneur's response. But it is easy to argue that size matters. A large enterprise may be severely disrupted by poor quality of electricity delivery, for example; it may have an inside track to bank loans and other financial instruments; it may face more harassment from government agencies. But if EICO obstacles hinder the operation and growth of the enterprise, as is indeed the specific question posed to entrepreneurs, one would expect to find that a greater obstacle reduces the size and productivity of the enterprise. In other words, the direction of causality goes both ways, and the estimated effect of size and productivity on EICO responses would be negatively biased. This resulting endogeneity bias could be solved through instrumental variables. Alternatively, the size and productivity measures may be omitted and a reduced-form equation may be estimated-which indeed was the strategy followed here. The drawback of this approach is that it remains unclear whether the estimated effect of, say, female entrepreneurship on the EICO response is direct or manifested indirectly via through size and productivity.

To satisfy curiosity, however, Table H.5 reports the estimates of the effect of size, as measured by the log of sales, and productivity, as measured by the log of the ratio of net value added over total factor cost (V/C). The reported values are not scaled estimates as in the other tables; rather, they are unscaled estimates of the parameters of the weighted random effect ordered probit model. In Nicaragua, many of the two effects are statistically significant; in Sri Lanka, the effect is statistically relevant for only one EICO, telecommunication; and in Tanzania, five of the twenty parameter estimates are statistically significant. It is notable that the size effect is positive for 23 of the 30 estimates, that the productivity effect is negative for 24 of the 30 estimates, and that the effects have the same sign only five times. Taken at face value, this implies that operators of larger enterprises tend to complain more often about investment climate conditions and that entrepreneurs overseeing more productive businesses tend to view the environment in less problematic terms. But as mentioned, these effects are probably biased downward: the size effect is probably more strongly positive, and the productivity effect is less negative.

Table H.1 Detailed Responses to EICO Questions, by Country and Category

A: Nicaragua (0 = No obstacle, . . . , 2 = Severe obstacle)

Description	0	1	2	Total	N	Mean
Public services						
electricity	59.97	20.17	19.86	100	1135	0.599
telecommunication	94.93	3.27	1.80	100	1133	0.069
water	78.17	11.12	10.72	100	1135	0.325
postal service	98.83	0.72	0.44	100	1132	0.016
Transport						
road quality	77.81	11.46	10.72	100	1136	0.329
road block	83.54	8.12	8.34	100	1136	0.248
forms of transport	88.98	7.62	3.40	100	1135	0.144
access to transport facility	87.60	6.89	5.51	100	1135	0.179
Financing						
interest rate of loan	54.13	15.58	30.28	100	1135	0.762
loan process	68.58	9.46	21.96	100	1119	0.534
access to loans	56.83	14.17	29.01	100	1118	0.722
Commercialization						
market information	92.39	5.78	1.83	100	1132	0.094
market demand	69.86	15.00	15.14	100	1135	0.453
Permits and licenses						
obtaining licenses	96.25	2.37	1.38	100	1116	0.051
problems with registry	97.30	1.63	1.07	100	1115	0.038
business permit inefficiency	97.93	1.38	0.69	100	1114	0.028
Taxes						
tax rate	89.10	5.39	5.52	100	1113	0.164
bureaucracy in tax collection	95.59	2.69	1.71	100	1105	0.061
Work						
labor hiring	94.60	4.15	1.24	100	98	0.066
complementary labor cost	97.78	1.11	1.11	100	97	0.033
labor firing	98.89	1.11	0.00	100	97	0.011
work permit	100.00	0.00	0.00	100	90	0.000
availability of skillful labor	98.07	1.93	0.00	100	96	0.019
Land policy						
regulation on land use	98.64	0.83	0.54	100	897	0.019
legalizing delay	99.48	0.09	0.43	100	895	0.010
construction permit	99.79	0.21	0.00	100	907	0.002
land collateral	99.06	0.22	0.72	100	898	0.017
Agricultural policy						
agricultural product movement	98.66	0.86	0.48	100	909	0.018
agricultural cooperative	100.00	0.00	0.00	100	898	0.000
agricultural trade	97.46	2.25	0.29	100	910	0.028
food safety regulation	99.34	0.66	0.00	100	907	0.007
Nonagricultural policies						
custom rules	99.40	0.11	0.49	100	891	0.011
environmental regulation	99.09	0.91	0.00	100	896	0.009
Governance						
corruption	58.04	17.47	24.49	100	1127	0.664
economic policy uncertainty	46.47	17.02	36.52	100	1136	0.901
crime, theft, etc.	74.46	14.20	11.34	100	1136	0.369
legal system	94.13	2.67	3.20	100	1134	0.091

(*continued on next page*)

Table H.1 Detailed Responses to EICO Questions, by Country and Category (*continued*)

B: Sri Lanka (0 = No obstacle, . . . , 4 = Very severe obstacle)

Description	0	1	2	3	4	Total	N	Mean
Public utilities								
electricity	50.19	11.83	13.22	14.53	10.23	100	1321	1.228
telecommunication	73.80	8.82	9.23	5.88	2.28	100	1280	0.540
water	70.71	9.63	7.45	7.45	4.75	100	1302	0.659
postal service	91.13	5.13	2.41	0.77	0.57	100	1292	0.145
Transportation								
road access	65.61	9.73	9.24	10.7	4.72	100	1323	0.792
road quality	64.10	10.04	9.08	9.65	7.13	100	1322	0.857
road block	85.69	7.77	3.22	2.88	0.44	100	1296	0.246
access to transport facility	69.82	8.14	6.97	11.94	3.12	100	1305	0.704
Financing								
interest rate of loan	53.21	5.20	11.56	23.84	6.19	100	1313	1.246
loan procedures	55.48	6.26	11.09	21.91	5.25	100	1314	1.152
availability of loan sources	79.77	7.58	6.23	4.96	1.47	100	1295	0.408
Marketing								
market information	63.05	13.36	11.78	10.31	1.49	100	1316	0.738
market demand	52.91	6.15	13.11	20.05	7.78	100	1323	1.236
Registration, license and permits								
cost of registration	95.57	1.18	0.71	1.75	0.80	100	1326	0.110
cost of license/permit	95.17	0.99	1.23	1.62	1.00	100	1324	0.123
registration/license process	96.21	0.94	0.92	1.19	0.74	100	1323	0.093
Taxation								
tax rate	94.70	0.98	1.54	2.17	0.61	100	1325	0.130
bureaucracy in tax collection	96.61	1.26	1.00	0.75	0.38	100	1323	0.070
Labor								
flexibility in hiring and firing	98.05	0.58	0.71	0.57	0.10	100	1323	0.041
government labor policy	98.08	0.42	0.68	0.64	0.18	100	1319	0.044
obtaining work permit for expatriate	99.34	0.32	0.19	0.14	0.00	100	1314	0.011
availability of skillful labor	94.36	0.48	0.75	2.72	1.70	100	1323	0.169
Land								
regulation on land use	97.85	0.62	0.57	0.76	0.20	100	1324	0.048
obtaining land construction permit	97.37	0.99	0.51	0.50	0.63	100	1321	0.060
land ownership	97.14	0.81	0.26	0.89	0.90	100	1323	0.076
Agricultural policy								
agricultural subsidy	98.13	0.47	0.25	0.96	0.19	100	1326	0.046
agricultural import/export tariff	99.03	0.09	0.21	0.56	0.11	100	1319	0.026
protection of rice production	99.02	0.19	0.50	0.29	0.00	100	1325	0.020
protection of agricultural price	98.57	0.19	0.23	0.80	0.22	100	1326	0.039
Nonagricultural trade policy								
import/export regulation	99.15	0.45	0.30	0.07	0.03	100	1326	0.014
custom rules	99.60	0.05	0.26	0.03	0.06	100	1326	0.009
Environmental policy								
food safety regulation	98.11	0.64	0.35	0.60	0.31	100	1319	0.044
environmental rules	96.36	1.23	1.19	0.48	0.74	100	1325	0.080
Governance								
corruption	96.98	1.26	1.29	0.33	0.14	100	1318	0.054
economic policy uncertainty	93.57	0.88	1.96	2.10	1.49	100	1326	0.171
crime, theft, etc.	96.61	0.86	1.69	0.64	0.19	100	1323	0.069
legal system	96.71	1.11	0.99	0.59	0.61	100	1318	0.073
war	96.90	1.00	1.01	0.18	0.92	100	1317	0.072

Table H.1 (*continued*)

C: Tanzania (0 = No obstacle, . . . , 4 = Very severe obstacle)

Description	0	1	2	3	4	Total	N	Mean
Public utilities								
electricity access	50.42	6.14	6.40	21.46	15.57	100	1188	1.456
electricity quality	55.75	7.98	8.91	18.00	9.37	100	1078	1.173
telecommunication	58.49	8.49	9.34	17.83	5.86	100	1178	1.041
water	56.98	8.70	7.47	19.29	7.55	100	1218	1.117
postal service	63.87	8.91	7.21	12.55	7.46	100	1179	0.908
Transportation								
road access	64.58	7.65	7.08	15.07	5.62	100	1228	0.895
road quality	63.46	6.77	7.01	15.82	6.93	100	1226	0.960
road block	85.07	5.68	3.82	3.90	1.53	100	1179	0.311
access to transport facility	69.07	7.30	6.97	11.03	5.64	100	1206	0.769
Financing								
interest rate of loan	38.46	5.69	6.88	36.76	12.22	100	1178	1.786
loan procedures	36.85	6.20	7.20	35.85	13.90	100	1194	1.838
availability of loan sources	37.75	5.86	7.81	35.96	12.61	100	1229	1.798
Marketing								
market access	64.94	6.57	6.66	17.86	3.98	100	1232	0.894
market information	66.20	7.47	7.47	15.50	3.36	100	1219	0.824
market demand	66.14	8.81	7.58	15.16	2.31	100	1214	0.787
Registration, license and permits								
cost of registration	78.18	4.40	3.99	11.56	1.87	100	1228	0.546
cost of license/permit	77.70	3.24	4.70	11.68	2.68	100	1233	0.584
registration/license process	78.37	3.01	4.23	12.03	2.36	100	1230	0.570
Taxation								
tax rate	78.22	3.00	2.59	12.79	3.40	100	1235	0.602
bureaucracy in tax collection	81.11	4.56	3.40	9.11	1.82	100	1207	0.460
Labor								
flexibility in hiring and firing	94.23	1.95	0.81	2.44	0.57	100	1231	0.132
government labor policy	94.47	1.30	1.46	2.44	0.33	100	1230	0.128
obtaining work permit for expatriate	94.63	1.06	1.14	2.52	0.65	100	1228	0.135
availability of skillful labor	93.83	1.30	0.97	3.17	0.73	100	1231	0.157
Land								
regulation on land use	84.72	4.53	2.51	7.60	0.65	100	1237	0.349
obtaining land construction permit	85.29	2.83	3.40	7.11	1.37	100	1237	0.365
land ownership	84.75	3.49	3.08	7.30	1.38	100	1233	0.371
Agricultural policy								
agricultural subsidy	83.14	0.89	2.03	10.70	3.24	100	1234	0.500
agricultural import/export tariff	86.95	1.75	3.33	6.32	1.66	100	1203	0.340
protection of rice production	88.89	2.22	3.29	4.94	0.66	100	1215	0.263
protection of agricultural price	84.43	2.85	2.69	7.66	2.36	100	1227	0.407
Nonagricultural trade policy								
import/export regulation	92.18	0.98	1.55	3.59	1.71	100	1227	0.217
custom rules	92.01	0.90	1.87	3.67	1.55	100	1227	0.218
Environmental policy								
food safety regulation	93.17	1.73	1.15	3.45	0.49	100	1216	0.164
environmental rules	92.59	1.89	1.89	3.37	0.25	100	1215	0.168
Governance								
corruption	72.31	2.85	2.52	12.05	10.26	100	1228	0.851
economic policy uncertainty	74.14	6.22	4.83	13.09	1.72	100	1222	0.620
crime, theft, etc.	73.27	4.97	4.65	11.25	5.87	100	1227	0.715
legal system	77.17	7.53	5.54	8.35	1.41	100	1209	0.493
war	86.15	4.65	2.45	4.98	1.77	100	1184	0.316

Source: RIC Surveys.

Table H.2 Descriptive Statistics of Explanatory Variables

	Nicaragua		Sri Lanka		Tanzania	
	Mean[a]	St. dev.[a]	Mean[a]	St. dev.[a]	Mean[b]	St. dev.[b]
Enterprise characteristics						
Enterprise age	10.739	10.140	9.026	10.426	8.570	9.279
in logarithm	2.085	0.921	1.872	0.931	1.823	0.978
Education of manager			9.549	3.072	7.748	2.704
Female manager			0.250	0.433	0.223	0.417
Sinhalese manager			0.786	0.410		
Enterprise sales (US$)	2371	5911	5914	38495	1030	2032
in logarithm	6.912	1.267	7.031	1.630	5.966	1.366
Net value added/total factor cost	2.672	2.564	1.015	1.010	1.211	1.451
in logarithm	0.279	1.657	−0.590	1.328	−0.529	1.353
Household-based enterprise	0.917	0.275	0.424	0.494	0.610	0.488
Industry dummies						
Services enterprise	0.193	0.395	0.170	0.376	0.171	0.376
Manufacturing, nonagricultural enterprise	0.085	0.278	0.311	0.463	0.061	0.240
Agricultural processing enterprise	0.138	0.345	0.025	0.155	0.035	0.183
Other production enterprise	0.033	0.180	0.004	0.065	0.024	0.152
Mixed enterprise	0.158	0.365	0.138	0.345	0.200	0.400
Community variables						
Community population size	21987	30108	1752	1221	4387	3673
in logarithm	8.756	1.870	7.297	0.553	8.050	0.841
Income per capita (US$)	438	236	3666	2367	141	170
in logarithm	5.909	0.649	8.047	0.550	4.518	0.924
Agricultural seasonality	0.685	0.413	0.734	0.303		
Share income from agriculture	0.092	0.104	0.153	0.712	0.332	0.302
Benchmark indexes						
Connectivity	0.356	0.178	0.455	0.165	0.211	0.144
Infrastructure services	0.550	0.239	0.377	0.161	0.195	0.162
Business services	0.305	0.430	0.173	0.241	0.066	0.151
Human capital	0.215	0.075	0.338	0.083	0.213	0.056
Finance services	0.200	0.266	0.535	0.133	0.187	0.143
Number of enterprises[c]	1113		869		1178	

Source: RIC Surveys.
Notes: a. Weighted statistics.
b. Unweighted statistics.
c. For a few variables the number of observations is less because of missing values.

Table H.3 Determinants of Top Ten EICO Responses: Models with Benchmark Indicators[a]

A: Nicaragua, Part 1

	Interest rate of loan (2)	Loan procedures (6)	Electricity (5)	Availability of loan sources (3)	Road quality (9)
Enterprise characteristics					
Age of enterprise[b]	−0.022	−0.007	−0.028	−0.056*	−0.003
Household-based enterprise	−0.381**	−0.255	−0.403***	−0.155	0.045
Industry dummies					
Services enterprise	0.084	0.044	0.080	0.000	−0.102
Manufacturing, nonagricultural enterprise	−0.022	−0.111	0.211	−0.025	0.026
Agricultural processing enterprise	−0.168	−0.232	0.028	−0.132	−0.099
Other production enterprise	0.221	0.142	−0.063	0.181	0.053
Mixed enterprise	0.061	0.016	0.380***	−0.025	0.072
Community variables					
Community population size[b]	−0.033	−0.057	0.135***	−0.042	0.091
Income per capita[b]	0.014	0.066	0.101	−0.011	−0.012
Agricultural seasonality	0.009	0.115	−0.059	−0.002	−0.074
Share income from agriculture	0.068	0.025	0.022	−0.002	−0.039
Benchmark indexes					
Connectivity	−0.082	−0.004	−0.194*	−0.037	−0.389***
Infrastructure services	0.102	0.139	−0.314***	0.035	0.013
Business services	0.103	0.220**	0.014	0.198**	0.063
Human capital	−0.059	−0.154	0.095	−0.070	0.036
Finance services	−0.034	−0.099	−0.001	−0.028	−0.034
Regression statistics					
St.dev of comm. RE	0.515***	0.576***	0.421***	0.495***	0.439***
lnLikelihood	−800.20	−612.00	−760.99	−765.02	−540.89
Number of obs	865	853	864	854	865
p (community variables)	0.849	0.491	0.016**	0.936	0.306
p (benchmark indexes)	0.405	0.095*	0.000***	0.163	0.005***

(*continued on next page*)

Table H.3 Determinants of Top Ten EICO Responses: Models with Benchmark Indicators[a] (*continued*)

A: Nicaragua, Part 2

	Low market demand (7)	Water (10)	Economic policy uncertainty (1)	Corruption (4)	Crime (8)
Enterprise characteristics					
Age of enterprise[b]	0.054	−0.024	0.049	0.041	0.042
Household-based enterprise	0.267	−0.324*	−0.343**	−0.241	−0.293**
Industry dummies					
Services enterprise	0.028	0.163	−0.151	−0.137	0.074
Manufacturing, nonagricultural enterprise	0.252	0.004	−0.153	−0.179	−0.093
Agricultural processing enterprise	−0.110	0.074	−0.137	−0.050	0.032
Other production enterprise	−0.124	0.001	0.271	0.311	0.488**
Mixed enterprise	−0.089	0.240*	0.131	0.112	0.021
Community variables					
Community population size[b]	0.019	0.035	−0.102*	−0.065	−0.086*
Income per capitab	0.026	0.081	0.135	0.039	−0.043
Agricultural seasonality	−0.011	0.006	−0.131	−0.054	−0.048
Share income from agriculture	−0.054	0.016	0.016	0.049	0.117
Benchmark indexes					
Connectivity	0.130*	−0.211**	−0.316***	−0.208*	−0.302***
Infrastructure services	−0.015	−0.152	0.279**	0.165	0.286***
Business services	−0.005	−0.011	0.209*	0.166*	0.192**
Human capital	−0.194***	0.019	0.122	−0.069	−0.038
Finance services	0.028	0.089	−0.057	−0.041	0.126
Regression statistics					
St.dev of comm. RE	0.147*	0.314***	0.575***	0.576***	0.489***
lnLikelihood	−722.08	−568.40	−805.06	−718.01	−492.70
Number of obs	864	864	860	865	865
p(community variables)	0.785	0.735	0.068*	0.762	0.090*
p(benchmark indexes)	0.046**	0.005***	0.005***	0.012**	0.000***

Table H.3 (*continued*)

B: Sri Lanka, Part 1

	Interest rate of loan (1)	Loan procedures (4)	Electricity (3)	Road quality (5)	Low market demand (2)
Enterprise characteristics					
Age of enterprise[b]	0.000	−0.054	−0.004	−0.126**	−0.015
Education of manager	−0.090	0.004	0.104	0.146**	−0.032
Female manager	−0.217	−0.215	−0.361**	−0.056	0.053
Sinhalese manager	0.757***	0.958***	−0.013	0.574**	−0.088
Household-based enterprise	−0.205	−0.168	−0.516***	0.390**	−0.205
Industry dummies/line of business					
Services enterprise	−0.462***	−0.519***	0.615***	−0.134	−0.541***
Manufacturing, nonagricultural enterprise	0.321**	0.231	0.989***	0.323*	−0.109
Agricultural processing enterprise	−0.494	−0.653	0.722*	0.385	−0.626
Other production enterprise	−0.437	0.406	−0.810	0.422	0.779
Mixed enterprise	0.167	−0.059	0.412*	−0.290	0.290
Community variables					
Community population size[b]	0.163	0.157	0.082	−0.121	−0.035
Income per capita[b]	−0.048	−0.063	−0.159	−0.032	0.170
Agricultural seasonality	0.009	−0.052	0.069	−0.358**	−0.271
Share income from agriculture	0.009	0.010	−0.093	−0.265**	−0.045
Benchmark indexes					
Connectivity	0.064	0.104	0.019	−0.299*	−0.078
Infrastructure services	−0.023	−0.053	−0.030	−0.025	0.063
Business services	0.149	0.035	0.000	−0.297**	−0.154
Human capital	0.235	0.189	−0.276**	0.077	−0.017
Finance services	−0.137	−0.161	−0.082	−0.342**	0.081
Regression statistics					
St. dev of comm. RE	0.840***	0.646***	0.793***	0.908***	0.762***
lnLikelihood	−1240.19	−1262.36	−1426.91	−1126.29	−1394.49
Number of obs	1104	1108	1110	1110	1111
p(community variables)	0.652	0.369	0.116	0.001***	0.007***
p(benchmark indexes)	0.051**	0.119	0.030***	0.000***	0.600

(*continued on next page*)

Table H.3 Determinants of Top Ten EICO Responses: Models with Benchmark Indicators[a] (*continued*)

B: Sri Lanka, Part 2

	Easy access to roads (6)	Water (9)	Telecommu-nication (10)	Lack of market information (7)	Availability of transport facilities (8)
Enterprise characteristics					
Age of enterprise[b]	−0.090	−0.023	0.007	−0.046	−0.013
Education of manager	0.002	−0.022	0.267***	0.017	0.078
Female manager	0.092	−0.443**	−0.014	0.050	−0.108
Sinhalese manager	0.199	−0.074	0.197	−0.238	0.290
Household-based enterprise	0.391**	−0.212	−0.114	0.001	0.264
Industry dummies/line of business					
Services enterprise	−0.293	0.382*	−0.050	−0.619***	−0.905***
Manufacturing, nonagricultural enterprise	0.170	0.211	−0.207	−0.008	0.034
Agricultural processing enterprise	0.202	0.904**	0.231	−0.393	−0.648
Other production enterprise	0.223	0.770	−0.565	−0.935	−0.260
Mixed enterprise	−0.291	0.632***	0.154	0.176	−0.074
Community variables					
Community population size[b]	0.130	−0.098	0.152	−0.047	−0.045
Income per capita[b]	−0.029	−0.202*	−0.033	0.095	−0.169
Agricultural seasonality	0.005	0.208	0.053	−0.211**	−0.168
Share income from agriculture	−0.163	−0.072	0.226	−0.112	−0.092
Benchmark indexes					
Connectivity	−0.192	−0.259	0.069	−0.029	−0.387**
Infrastructure services	−0.224	0.133	−0.320*	0.140	0.021
Business services	−0.177	0.162	−0.295**	−0.246**	−0.002
Human capital	0.254	−0.115	0.116	−0.132	−0.002
Finance services	−0.392**	0.088	−0.008	0.020	−0.157
Regression statistics					
St. dev of comm. RE	0.896***	0.838***	0.806***	0.458***	0.846***
lnLikelihood	−1090.54	−1024.65	−905.31	−1172.45	−975.65
Number of obs	1111	1099	1089	1105	1096
p(community variables)	0.296	0.033**	0.389	0.004***	0.145
p(benchmark indexes)	0.000***	0.087*	0.004***	0.007***	0.003***

Table H.3 (*continued*)

C: Tanzania, Part 1

	Interest rate of loan (3)	Loan procedures (1)	Quality of electricity (5)	Availability of loan sources (2)	Road quality (8)
Enterprise characteristics					
Age of enterprise[b]	0.015	0.034	−0.005	−0.002	0.028
Education of manager	−0.052	−0.032	−0.032	−0.043	0.011
Female manager	0.074	0.185*	−0.162	0.072	−0.143
Household-based enterprise	−0.116	0.031	−0.062	−0.076	−0.123
Industry dummies					
Services enterprise	0.040	−0.153	−0.191	−0.041	−0.189
Manufacturing, nonagricultural enterprise	0.155	0.008	0.076	−0.176	−0.190
Agricultural processing enterprise	0.326	0.220	−0.066	0.519**	0.228
Other production enterprise	0.003	−0.081	0.143	−0.008	−0.554
Mixed enterprise	0.263**	0.086	−0.192	0.137	0.197
Community variables					
Community population size[b]	0.071	0.015	0.255***	0.046	−0.049
Income per capita[b]	0.147	0.257**	0.499***	0.237**	0.099
Share income from agriculture	−0.381***	−0.390***	−0.394***	−0.272***	−0.064
Benchmark indexes					
Connectivity	−0.220**	−0.197**	−0.103	−0.358***	−0.150
Infrastructure services	−0.041	−0.211	−0.410***	0.046	−0.175
Business services	0.049	−0.035	−0.148	−0.063	−0.303***
Human capital	−0.115	−0.082	−0.126	−0.155	−0.319***
Finance services	−0.262***	−0.158	−0.033	−0.254***	0.004
Regression statistics					
St. dev of comm. RE	0.988***	0.963***	1.329***	0.974***	1.008***
lnLikelihood	−1283.60	−1377.61	−1063.97	−1361.76	−1120.82
Number of obs	1117	1133	1023	1168	1165
p(community variables)	0.000***	0.000***	0.000***	0.001***	0.697
p(benchmark indexes)	0.000***	0.000***	0.004***	0.000***	0.000***

(*continued on next page*)

Table H.3 Determinants of Top Ten EICO Responses: Models with Benchmark Indicators[a] (*continued*)

C: Tanzania, Part 2

	Easy access to roads (10)	Water (6)	Telecomm-unications (7)	Access to Electricity (4)	Postal service (9)
Enterprise characteristics					
Age of enterprise[b]	0.041	−0.013	−0.040	−0.010	−0.047
Education of manager	−0.011	0.015	−0.053	0.002	0.012
Female manager	0.012	0.050	−0.009	−0.284***	−0.046
Household-based enterprise	−0.129	−0.075	0.023	0.035	−0.001
Industry dummies					
Services enterprise	−0.163	−0.005	−0.142	0.019	−0.254*
Manufacturing, nonagricultural enterprise	−0.187	0.137	0.266	0.184	0.088
Agricultural processing enterprise	−0.142	−0.310	0.107	0.065	0.177
Other production enterprise	−0.701**	0.031	0.081	0.205	−0.037
Mixed enterprise	0.077	−0.128	0.120	0.060	−0.008
Community variables					
Community population size[b]	−0.075	−0.056	−0.033	−0.014	0.030
Income per capita[b]	0.237*	0.334**	0.612***	0.407***	0.327*
Share income from agriculture	−0.061	0.005	−0.083	0.085	0.057
Benchmark indexes					
Connectivity	−0.017	−0.240*	−0.027	−0.036	−0.089
Infrastructure services	−0.241**	−0.477***	−0.514***	−0.414***	−0.635***
Business services	−0.369***	0.001	−0.198**	−0.264***	−0.195*
Human capital	−0.248**	0.034	0.019	−0.049	−0.002
Finance services	0.028	−0.033	−0.237**	−0.149*	−0.094
Regression statistics					
St. dev of comm. RE	1.056***	1.117***	1.177***	1.505***	1.216***
lnLikelihood	−1101.38	−1237.26	−1133.80	−1181.88	−1041.92
Number of obs	1168	1160	1120	1127	1122
p(community variables)	0.265	0.073*	0.001***	0.016**	0.063*
p(benchmark indexes)	0.001***	0.000***	0.000***	0.000***	0.000***

Source: RIC Surveys.

Notes: a. EICO items are sorted by the average rankings across the three countries; the number in parentheses in the column headings is the country-specific rank.

b. Variable enters the model in logarithmic form.

*** Significant at 1 percent; ** Significant at 5 percent; * Significant at 10 percent.

Table H.4 Determinants of Top Ten EICO Responses: Models with Specific Community Characteristics[a]

A: Interest rate of loan

	Nicaragua	Sri Lanka	Tanzania
Enterprise characteristics			
Age of enterprise[b]	−0.007	0.009	0.014
Education of manager		−0.080	−0.055
Female manager		−0.260*	0.068
Sinhalese manager		0.740***	
Household-based enterprise	−0.452***	−0.183	−0.102
Services enterprise	0.074	−0.449***	0.042
Manufacturing, nonagricultural enterprise	−0.017	0.284*	0.043
Agricultural processing enterprise	−0.126	−0.419	0.419
Other production enterprise	0.220	−0.413	−0.031
Mixed enterprise	0.084	0.153	0.260**
Community characteristics[c]			
Income per capita[b]			0.158
Share income from agriculture			−0.435***
Time to near city	−0.056	−0.144	−0.247**
Cost of transportation to city	0.111	−0.027	.068
Insurance service	0.082	0.141	−0.042
Human capital benchmark	−0.068	0.221	−0.146
Number of banks	−0.195	0.551**	0.215**
Number bank services	0.201	−0.346*	−0.452***
Regression statistics			
St.dev of comm. RE	0.503***	0.804***	1.046***
lnLikelihood	−829.21	−1291.59	−1276.31
Number of obs	893	1146	1117

B: Loan procedures

	Nicaragua	Sri Lanka	Tanzania
Enterprise characteristics			
Age of enterprise[b]	0.005	−0.042	0.040
Education of manager		0.013	−0.036
Female manager		−0.246	0.145
Sinhalese manager		0.883***	
Household-based enterprise	−0.302*	−0.134	0.023
Services enterprise	0.071	−0.496***	−0.122
Manufacturing, nonagricultural enterprise	−0.075	0.195	−0.002
Agricultural processing enterprise	−0.177	−0.530	0.312
Other production enterprise	0.318	0.447	−0.073
Mixed enterprise	0.080	−0.035	0.102
Community characteristics[a]			
Income per capita[b]			0.290***
Share income from agriculture			−0.389***
Time to near city	−0.049	−0.216	−0.152
Cost of transportation to city	0.095	0.104	−0.047
Insurance service	0.269*	0.110	−0.052
Human capital benchmark	−0.130	0.159	−0.225**
Number of banks	−0.126	0.527***	0.214**
Number bank services	0.045	−0.353**	−0.461***
Regression statistics			
St. dev of comm. RE	0.623***	0.651***	0.972***
lnLikelihood	−641.89	−1311.99	−1372.78
Number of obs	881	1150	1133

(*continued on next page*)

Table H.4 Determinants of Top Ten EICO Responses: Models with Specific Community Characteristics[a] (*continued*)

C: Electricity

	Nicaragua	Sri Lanka	Tanzania[d]
Enterprise characteristics			
Age of enterprise[b]	−0.019	0.023	−0.005
Education of manager		0.132*	−0.035
Female manager		−0.286*	−0.136
Sinhalese manager		−0.138	
Household-based enterprise	−0.390***	−0.501***	−0.052
Services enterprise	0.075	0.570***	−0.159
Manufacturing, nonagricultural enterprise	0.214	1.000***	0.052
Agricultural processing enterprise	0.069	0.839**	0.043
Other production enterprise	−0.125	−0.745	0.193
Mixed enterprise	0.375***	0.469**	−0.169
Community characteristics[c]			
Income per capita[b]			0.415**
Share income from agriculture			−0.333***
Percent households with electricity	−0.268***	−0.300**	−0.371***
Availability electricity			−0.478***
Information technology	0.066	0.087	−0.023
Human capital benchmark	0.164***	−0.199	0.170
Regression statistics			
St.dev of comm. RE	0.479***	1.119***	0.661***
lnLikelihood	−792.34	−1483.24	−1062.10
Number of obs	892	1152	1023

D: Availability of loan sources

	Nicaragua	Sri Lanka[e]	Tanzania
Enterprise characteristics			
Age of enterprise[b]	0.052		0.003
Education of manager			−0.052
Female manager			0.043
Sinhalese manager			
Household-based enterprise	−0.150		−0.091
Services enterprise	−0.001		−0.023
Manufacturing, nonagricultural enterprise	−0.027		−0.149
Agricultural processing enterprise	−0.078		0.577**
Other production enterprise	0.168		−0.005
Mixed enterprise	−0.002		0.175
Community characteristics[c]			
Income per capita[b]			0.286**
Share income from agriculture			−0.322***
Time to near city	−0.028		−0.305***
Cost of transportation to city	0.069		−0.076
Insurance service	0.132		−0.140
Human capital benchmark	−0.106		−0.159*
Number of banks	−0.264		0.135
Number bank services	0.235		−0.380***
Regression statistics			
St.dev of comm. RE	0.481***		0.935***
lnLikelihood	−795.01		−1358.78
Number of obs	882		1168

Table H.4 (*continued*)

E: Road quality

	Nicaragua	Sri Lanka	Tanzania
Enterprise characteristics			
Age of enterprise[b]	−0.012	−0.109*	0.032
Education of manager		0.174**	−0.012
Female manager		−0.018	−0.166
Sinhalese manager		0.151	
Household-based enterprise	−0.021	0.415**	−0.123
Services enterprise	−0.088	−0.161	−0.089
Manufacturing, nonagricultural enterprise	−0.005	0.353**	−0.194
Agricultural processing enterprise	−0.106	0.354	0.241
Other production enterprise	0.088	0.469	−0.560
Mixed enterprise	0.037	−0.137	0.191
Community characteristics[c]			
Income per capita[b]			0.153
Share income from agriculture			−0.051
Time to near city	−0.331***	−0.362**	−0.193**
Time to main market	0.061	−0.086	0.076
Concrete/asphalt road		−0.039	−0.252***
Human capital benchmark	−0.054	0.007	−0.420***
Regression statistics			
St.dev of comm. RE	0.472***	0.888***	1.105***
lnLikelihood	−557.59	−1186.04	−1157.15
Number of obs	893	1152	1191

F: Low market demand

	Nicaragua	Sri Lanka	Tanzania[e]
Enterprise characteristics			
Age of enterprise[b]	0.056*	−0.017	
Education of manager		−0.045	
Female manager		0.059	
Sinhalese manager		−0.015	
Household-based enterprise	0.248	−0.166	
Services enterprise	0.016	−0.558***	
Manufacturing, nonagricultural enterprise	0.241	−0.108	
Agricultural processing enterprise	−0.118	−0.589	
Other production enterprise	−0.159	0.770	
Mixed enterprise	−0.088	0.304	
Community characteristics[c]			
Time to near city	0.034	−0.385**	
Cost of transportation to city	0.067	0.130	
Time to main market	0.022	0.206	
Cost of transportation to market	−0.037	0.097	
Marketing service	0.058	−0.007	
Human capital benchmark	−0.086*	0.053	
Regression statistics			
St.dev of comm. RE	0.133	0.801***	
lnLikelihood	−749.58	−1437.62	
Number of obs	892	1153	

(*continued on next page*)

Table H.4 Determinants of Top Ten EICO Responses: Models with Specific Community Characteristics[a] (*continued*)

G: Road access

	Nicaragua[f]	Sri Lanka	Tanzania
Enterprise characteristics			
Age of enterprise[b]		−0.091*	0.030
Education of manager		0.008	−0.015
Female manager		0.075	−0.022
Sinhalese manager		0.155	
Household-based enterprise		0.376**	−0.115
Services enterprise		−0.340*	−0.161
Manufacturing, nonagricultural enterprise		0.156	−0.168
Agricultural processing enterprise		0.144	−0.060
Other production enterprise		0.283	−0.604*
Mixed enterprise		−0.310	0.093
Community characteristics[c]			
Income per capita[b]			0.210*
Share income from agriculture			−0.081
Time to near city		−0.147	−0.231***
Time to main market		−0.193	−0.003
Concrete/asphalt road		−0.056	−0.194**
Human capital benchmark		−0.001	−0.453***
Regression statistics			
St. dev of comm. RE		0.918***	1.137***
lnLikelihood		−1138.09	−1124.01
Number of obs		1152	1194

H: Water

	Nicaragua	Sri Lanka	Tanzania
Enterprise characteristics			
Age of enterprise[b]	−0.040	−0.025	−0.008
Education of manager		−0.039	0.009
Female manager		−0.504***	0.023
Sinhalese manager		0.142	
Household-based enterprise	−0.256	−0.237	−0.091
Services enterprise	0.167	0.366*	0.045
Manufacturing, nonagricultural enterprise	0.068	0.235	0.031
Agricultural processing enterprise	0.125	0.908**	−0.156
Other production enterprise	−0.076	0.565	0.048
Mixed enterprise	0.254*	0.610***	−0.079
Community characteristics[c]			
Income per capita[b]			0.375***
Share income from agriculture			−0.070
Time to main market	−0.105	−0.142	−0.231**
Percent households with protected water	−0.094*	0.021	−0.231***
Sewage system		0.210*	−0.353
Human capital benchmark	0.011	−0.251*	−0.253***
Regression statistics			
St. dev of comm. RE	0.347***	0.967***	1.262***
lnLikelihood	−593.01	−1047.70	−1259.48
Number of obs	892	1141	1185

Table H.4 (*continued*)

I: Telecommunication

	Nicaragua[e]	Sri Lanka	Tanzania
Enterprise characteristics			
Age of enterprise[b]		0.005	−0.033
Education of manager		0.250 ***	−0.074
Female manager		−0.080	−0.029
Sinhalese manager		0.442	
Household-based enterprise		−0.122	−0.002
Services enterprise		−0.063	−0.128
Manufacturing, nonagricultural enterprise		−0.251	0.227
Agricultural processing enterprise		0.188	0.023
Other production enterprise		−0.539	0.074
Mixed enterprise		0.095	0.083
Community characteristics[c]			
Income per capita[b]			0.569***
Share income from agricultural			−0.215**
Percent households with fixed phone line		−0.569***	−0.545***
Percent households with mobile phone		0.216	−0.189
Availability electricity		−0.074	−0.572***
Information technology		0.168	0.098
Human-capital benchmark		0.008	−0.007
Regression statistics			
St. dev of comm. RE		0.907***	1.143***
lnLikelihood		−935.32	−1126.64
Number of obs		1130	1120

J: Other EICO variables specific to Nicaragua

	Economic policy uncertainty	Corruption	Crime, theft, etc.
Enterprise characteristics			
Age of enterprise[b]	0.052*	0.049	0.047
Household-based enterprise	−0.364**	−0.239	−0.267*
Services enterprise	−0.122	−0.104	0.073
Manufacturing, nonagricultural enterprise	−0.126	−0.195	−0.101
Agricultural processing enterprise	−0.138	−0.046	0.013
Other production enterprise	0.220	0.322	0.485**
Mixed enterprise	0.185	0.157	0.033
Community characteristics[c]			
Population size[b]			−0.029
Income per capita[b]			−0.096
Time to near city	−0.089	−0.104	−0.222***
Time to main market			0.006
Percent households with fixed phone line	−0.143	0.045	
Percent households with mobile phone	0.112	−0.064	
Business services benchmark	0.142	0.163*	
Human capital benchmark	0.211**	−0.047	−0.002
Number of banks			0.115
Number bank services	−0.171**	−0.183**	
Regression statistics			
St. dev of comm. RE	0.556***	0.591***	0.556***
lnLikelihood	−832.30	−741.45	−501.23
Number of obs	893	888	865

(*continued on next page*)

Table H.4 Determinants of Top Ten EICO Responses: Models with Specific Community Characteristics[a] (*continued*)

K: Other EICO variables specific to Sri Lanka

	Market information	Access to transport facility
Enterprise characteristics		
Age of enterprise[b]	−0.060	−0.007
Education of manager	−0.009	0.104
Female manager	0.041	−0.121
Sinhalese manager	−0.216	0.210
Household-based enterprise	−0.004	0.269*
Services enterprise	−0.602***	−0.906***
Manufacturing, nonagricultural enterprise	0.005	0.082
Agricultural processing enterprise	−0.370	−0.620
Other production enterprise	−0.974	−0.246
Mixed enterprise	0.179	−0.060
Community characteristics[c]		
Connectivity benchmark		−0.341**
Time to main market	0.117	
Cost of transportation to market	−0.061	
Percent households with fixed phone line	0.043	
Percent households with mobile phone	0.021	
Concrete/asphalt road		−0.057
Marketing service	−0.091	
Information technology	0.034	
Human capital benchmark	−0.073	−0.044
Regression statistics		
St. dev of comm. RE	0.580***	0.853***
lnLikelihood	−1225.11	−1023.88
Number of obs	1147	1137

L: Other EICO variables specific to Tanzania

	Electricity access	Postal service
Enterprise characteristics		
Age of enterprise[b]	−0.001	−0.033
Education of manager	−0.017	0.021
Female manager	-0.318***	−0.175
Household-based enterprise	0.032	−0.074
Services enterprise	0.068	−0.145
Manuacturing,f nonagricultural enterprise	0.246	−0.054
Agricultural processing enterprise	0.092	0.253
Other production enterprise	0.273	0.122
Mixed enterprise	0.106	0.001
Community characteristics[c]		
Income per capita[b]	0.561***	0.334***
Share income from agricultural	−0.166*	−0.006
Distance to post office		−0.412***
Percent households with electricity	−0.167	
Availability electricity	−0.875***	
Concrete/asphalt road		−0.026
Information technology	−0.001	
Human capital benchmark	0.217*	−0.290***

Table H.4 (*continued*)

L: Other EICO variables specific to Tanzania

	Electricity access	Postal service
Regression statistics		
St.dev of comm. RE	1.388***	1.200***
lnLikelihood	−1165.04	−1068.57
Number of obs	1127	1146

Source: RIC Surveys.

Notes: a. Estimates represent the effect of a one-standard-deviation change in a continuous variable and a 0/1 change in a dummy variable on the tendency to report obstacles. For an average enterprise, this tendency is scaled on a range of 2 for Nicaragua and a range of 4 for Sri Lanka and Tanzania.

b. Variable is entered in logarithmic form.

c. With the exception of "Income per capita" and "Share of income from agriculture," all community characteristics are coded such that a higher value implies a presumed better investment climate.

d. For Tanzania, this measure refers to quality of electricity, whereas access to electricity is covered elsewhere. For Nicaragua and Sri Lanka, the questionnaire was not specific about the obstacles that electricity provision might pose to the entrepreneur.

e. This variable was not in the top-ten list of most important obstacles.

f. This variable was not included in the questionnaire used in Nicaragua.

*** Significant at 1 percent; ** Significant at 5 percent; * Significant at 10 percent.

Table H.5 Effect of Enterprise Size and Productivity on EICO Responses[ab]

	Nicaragua		Sri Lanka		Tanzania	
EICO Variable	**ln (Sales)**	**ln (V/C)**	**ln (Sales)**	**ln (V/C)**	**ln (Sales)**	**ln (V/C)**
Interest rate of loan	0.187***	−0.075**	−0.048	0.060	0.081	−0.030
Loan procedures	0.149***	−0.006	−0.001	−0.010	0.070	−0.024
Electricity	0.104**	−0.098***	0.073	−0.021	0.164*	−0.165**
Availability of loan sources	0.124***	−0.015	n.e.	n.e.	0.066	0.007
Road quality	0.203***	−0.079*	0.008	−0.006	0.059	−0.027
Low market demand	−0.222***	0.083**	−0.065	−0.041	n.e.	n.e.
Road access	n.a.	n.a.	−0.085	0.034	0.116*	−0.084
Water	0.000	−0.077*	0.095	−0.060	0.013	0.016
Economic policy uncertainty	0.109**	−0.020	n.e.	n.e.	n.e.	n.e.
Telecommunication	n.e.	n.e.	0.169***	−0.102**	0.008	−0.054
Corruption	0.212***	−0.040	n.e.	n.e.	n.e.	n.e.
Crime	0.175***	−0.066	n.e.	n.e.	n.e.	n.e.
Market information	n.e.	n.e.	−0.052	−0.089	n.e.	n.e.
Access to trans facility	n.e.	n.e.	−0.063	0.052	n.e.	n.e.
Electricity access	n.e.	n.e.	n.e.	n.e.	0.123*	−0.068
Postal service	n.e.	n.e.	n.e.	n.e.	0.035	−0.121**

Source: RIC Surveys.

Notes: a. Parameters reported here are estimates of the weighted random effect ordered probit model, not transformed according to the scale of measurement of the EICO (0/2 for Nicaragua and 0/4 for Sri Lanka and Tanzania). Thus, the magnitudes represent a 1-unit change in the log-sales and log-value added on the unscaled tendency to report obstacles.

b. ln (V/C) denotes the logarithm of net value added over total factor cost and measures enterprise productivity.

*** Significant at 1 percent; ** Significant at 5 percent; * Significant at 10 percent.

Table H.6 Effect of Enterprise and Community Characteristics on Total Investment Climate Burden, by Country[a]

	Nicaragua				Sri Lanka				Tanzania			
	EICO Burden		EICO Obstacle		EICO Burden		EICO Obstacle		EICO Burden		EICO Obstacle	
	b	t	b	t	b	t	b	t	b	t	b	t
Enterprise characteristics												
Age of enterprise[b]	0.102		0.119		−0.158		0.054		0.049		0.060	
Education of manager					0.161		0.024		0.308*		0.114**	
Female manager					−1.960***		−0.608***		−0.929		−0.212	
Sinhalese manager					2.665**		0.680**					
Household-based enterprise	−2.083***		−0.481		−1.225*		−0.205		0.249		0.172	
Industry dummies												
Services enterprise	0.148		0.016		−2.369***		−0.552**		−3.780***		−1.077***	
Manufacturing, nonagricultural enterprise	−0.502		−0.277		2.667***		0.589***		−1.376		−0.505	
Agricultural processing enterprise	−0.887		−0.418		0.325		−0.337		−3.293		−0.867	
Other production enterprise	0.762		0.322		8.076***		1.931***		−0.797		0.190	
Mixed enterprise	0.522		0.220		2.380*		0.535		−0.733		−0.079	
Community variables												
Community population size[b]	0.141		0.061		0.613		0.146		0.533		0.176	
Income per capita[b]	0.470		0.289		−0.825		−0.356		1.588		0.380	
Agricultural seasonality	−0.647		−0.288		−1.232		−0.601					
Share income from agriculture	−0.037		−0.689		−0.343		−0.007		−2.676		−0.459	
Benchmark indicators												
Connectivity	−8.202***		−2.885**		−1.336		0.059		−5.425		−1.641	
Infrastructure services	1.828		0.331		−0.239		−0.130		1.351		−0.070	
Business services	2.134**		0.886**		−0.845		0.215		−13.632		−4.114	
Human capital	−4.988		−3.332		−1.630		1.199		−73.193**		−18.765**	
Finance services	−1.681		−0.603		−5.558		−1.799		−9.113		−1.966	
Intercept	7.460**		2.012		15.877		4.345		33.090**		7.645**	
Regression statistics												
Average of dependent variable	6.682		2.368		11.572		2.353		25.605		6.297	
St. dev of comm. RE	2.271		1.055		4.178		1.090		14.840		3.911	
St. dev of enterprise error	4.369		2.048		6.710		1.810		13.282		3.902	
R-squared	0.108		0.066		0.089		0.095		0.081		0.073	
Number of observations	868		868		1113		1113		1177		1177	
Significance of benchmark indicators[c]	0.004***		0.033**		0.716		0.731		0.030**		0.024**	

Source: RIC Surveys.

Notes: a. Weighted random effect regression results.

b. Variable is entered in logarithmic form.

c. p-value of a joint test that none of the benchmark indicators have an effect on the dependent variable.

*** Significant at 1 percent; ** Significant at 5 percent; * Significant at 10 percent.

Table H.7 Opinions About Predictability and Manipulability of Laws

	Laws are unpredictable[a]			Laws cannot be manipulated[b]	
	Nicaragua	Sri Lanka	Tanzania	Sri Lanka	Tanzania
Enterprise characteristics					
Age of enterprise[c]	0.75	−2.09	0.84	0.002	0.015
Education of manager		−5.32***	0.71	0.009	−0.010
Female manager		2.97	−6.50*	−0.014	0.135
Sinhalese manager		3.25		−0.551***	
Household-based enterprise		0.08	−4.24	−0.199*	0.093
Services enterprise	0.09	1.80	0.66	0.092	−0.009
Manufacturing, nonagricultural enterprise	−4.07	3.42	9.82	0.047	0.039
Agricultural processing enterprise	−2.94	11.12	−3.47	−0.210	0.102
Other production enterprise	−13.44	−24.00	−3.03	−0.280	0.245
Mixed enterprise	1.95	−4.57	3.65	−0.087	−0.161
Community & benchmark					
Community population size[c]	5.08	−1.15	−1.49	−0.052	−0.011
Income per capita[c]	−1.16	−4.44	−5.28	−0.028	−0.127
Connectivity	−7.81**	1.46	0.98	−0.112	−0.006
Business services	5.28	0.13	−8.61***	0.048	−0.040
Human capital	−0.59	0.05	10.94***	−0.038	0.119
Regression statistics					
St. dev of comm. RE	18.72***	19.32***	22.42***	0.508***	0.601***
Log-Likelihood	−556.58	−1546.31	−1177.88	−658.13	−1164.44
Number of observations	868	1154	714	609	999

Source: RIC Surveys.

Notes: a. The estimates indicate the change (expressed in percentage points) in the probability that entrepreneur find laws to be unpredictable, as a result of a one-standard-deviation change (continuous variable) or unit change (dummy variable) in the explanatory variable. The model is estimated by random effect weighted probit in Nicaragua and random effects weighted ordered probit in Sri Lanka and Tanzania. For Sri Lanka and Tanzania, the computations for this table aggregate the "fairly," "highly," and "completely," categories that are distinguished in Table 5.2.

b. The statement posed to the entrepreneur is "Laws and regulations can be misinterpreted or manipulated." Responses are coded 0 = *Strongly agree,* 1 = *Agree,* 2 = *Disagree,* 3 = *Strongly disagree*. Estimates indicate the effect of a one-standard-deviation change (continuous variable) or unit change (dummy variable) effect on the tendency to disagree, as estimated with a random effect ordered probit model; the effect is scaled on the range 0–3 for a typical enterprise.

c. Variable enters the model in logarithmic form.

*** Significant at 1 percent; ** Significant at 5 percent; * Significant at 10 percent.

Table H.8 Effect on Business of Kickbacks, by Target: Nicaragua

	Laws	Enforcement	Judge	Credit	Politician
Enterprise characteristics					
Age of enterprise[b]	0.015	−0.004	0.018	0.045	0.049
Services enterprise	−0.111	−0.186	−0.297*	−0.184	−0.331*
Manufacturing, nonagricultural enterprise	0.410**	0.217	−0.075	0.067	−0.121
Agricultural processing enterprise	−0.163	0.047	0.008	0.002	−0.059
Other production enterprise	0.378	0.257	0.317	0.369	0.225
Mixed enterprise	−0.211	−0.239	−0.163	−0.287	−0.392**
Community & benchmark					
Community population size[b]	0.129***	0.113**	0.128***	0.144***	0.180***
Income per capita[b]	−0.154	−0.095	−0.167*	−0.112	−0.063
Connectivity	−0.234**	−0.230**	−0.243**	−0.205*	−0.296**
Business services	0.256***	0.231***	0.225***	0.059	0.210**
Human capital	0.061	−0.022	0.041	0.027	0.029
Regression statistics					
St. dev of comm. RE	0.338***	0.350***	0.296***	0.295***	0.439***
Log-Likelihood	−315.98	−338.21	−318.08	−252.97	−271.06
Number of observations	701	699	693	681	677

Source: RIC Surveys.

Notes: a. EICIs are coded 0 = *No impact,* 1 = *Little impact,* 2 = *Lot of impact.* The estimates indicate the change (expressed in percentage points) in the tendency to believe that kickbacks have an impact, as a result of a one-standard-deviation change (continuous variable) or unit change (dummy variable) in the explanatory variable. The model is estimated by weighted random effect ordered probit.

b. Variable enters the model in logarithmic form.

*** Significant at 1 percent; ** Significant at 5 percent; * Significant at 10 percent.

Table H.9 Determinants of Perceptions About the Legal System[a]

	Reputation is not important[b]			Contracts do not protect[c]			Legal system does not support[d]		
	Nic.	S.L.	Tanz.	Nic.	S.L.	Tanz.	Nic.	S.L.	Tanz.
Enterprise characteristics									
Age of enterprise[d]	0.053*	0.009	−0.014	0.055*	0.046	−0.009	0.078**	0.002	−0.014
Education of manager		0.004	−0.012		0.028	−0.010		0.063	−0.004
Female manager		0.077	0.133*		0.073	0.035		0.152	0.006
Sinhalese manager		−0.107			−0.125			−0.232	
Household-based enterprise		0.000	−0.031		0.000	−0.077		0.075	0.005
Services enterprise	0.124	0.023	−0.020	−0.046	−0.037	−0.060	−0.022	0.042	−0.087
Manufacturing, nonagricultural enterprise	−0.162	0.179*	−0.056	−0.397**	0.164	−0.109	−0.451**	0.262**	−0.077
Agricultural processing enterprise	0.167	0.264	−0.022	−0.058	0.298	0.018	0.178	−0.122	0.330**
Other production enterprise	0.382	−0.248	−0.010	0.048	−0.109	−0.072	0.376	−0.154	0.173
Mixed enterprise	0.199	0.131	0.104	−0.083	−0.073	0.040	0.040	0.048	0.020

Table H.9 (*continued*)

	Reputation is not important[b]			Contracts do not protect[c]			Legal system does not support[d]		
	Nic.	S.L.	Tanz.	Nic.	S.L.	Tanz.	Nic.	S.L.	Tanz.
Community & benchmark									
Community population size[d]	−0.055	0.126*	0.111**	−0.112***	0.073	0.022	−0.031	0.202***	0.051
Income per capita[d]	0.053	0.018	−0.045	0.060	−0.060	−0.133	0.035	−0.031	−0.110
Connectivity	0.056	0.031	0.074	0.095	−0.023	0.093*	0.137*	0.035	−0.011
Business services	0.027	−0.120	−0.006	−0.004	0.025	−0.033	−0.075	−0.112	−0.003
Human capital	−0.018	−0.047	0.066	0.011	−0.041	0.090	−0.058	0.000	0.063
Regression statistics									
St. dev of comm. RE	0.368***	0.428***	0.706***	0.462***	0.594***	0.680***	0.356***	0.534***	0.705***
lnLikelihood	−828.0	−749.1	−1192.1	−820.0	−747.8	−1150.2	−842.5	−676.0	−1191.5
Number of obs	845	892	1134	846	830	1133	829	716	1136

Source: RIC Surveys.

Notes: a. Parameters are estimated with a weighted random effect ordered probit model. Estimates represent the effect of a one-standard-deviation (continuous variable) or one-unit (dummy variable) change on the tendency to agree with the statement at the top of the column-or, semantically more accurately, to disagree with the opposite, i.e., that reputation is important, contracts do protect, and legal system uphold the entrepreneur's contract and property rights. Nicaragua values are scaled on a range of 0–2; Sri Lanka and Tanzania results are scaled on a range of 0–3.

b. The dependent variable reflects disagreement with the statement "I must rely on the reputation of those I enter into agreement with."

c. The dependent variable reflects disagreement with the statement "A contract will protect me from being cheated."

d. The dependent variable reflects disagreement with the statement "The legal system will uphold my contract and property rights in business disputes."

e. Variable enters the model in logarithmic form.

*** Significant at 1 percent; ** Significant at 5 percent; * Significant at 10 percent.

Table H.10 Corruption as an Investment Climate Constraint

	Nicaragua			Sri Lanka			Tanzania		
	number	number		number	number		number	number	
Total enterprises	1174	100	—	1365	100	—	1238	100	—
No response	80	7	—	—	0	—	2	0	—
Governance does not pose a problem[a]	—	—	—	1217	89	—	864	70	—
Total direct response	1094	93	100	148	11	100	372	30	100
of which:									
No obstacle	671	57	61	73	5	45	24	2	6
Minor	171	15	16	53	4	34	66	5	18
Major	241	21	22	14	1	16	274	22	74
Don't know / NA	11	1	1	8	1	5	8	1	2

Source: RIC Surveys.

Note: a. A screening question in Sri Lanka and Tanzania that reduces the total number of responses to the actual question about corruption, resulting in low numbers reporting "no obstacle."

Table H.11 Alternative Ways of Estimating EICO Models: Interest Rate of Loans in Sri Lankaa

	Standard ordered pzrobit			Weighted ordered probit			Weighted random effect ordered probit				
	b	s	t	b	s	t	b	s(V1)	t(V1)	s(V3)	t(V3)
Age of enterprise[b]	−0.030	0.036	−0.83	−0.034	0.045	−0.76	0.000	0.063	0.00	0.096	0.00
Education of manager	−0.018	0.012	−1.47	−0.010	0.015	−0.70	−0.025	0.019	−1.34	0.025	−1.00
Female manager	−0.285	0.086	−3.31	−0.168	0.100	−1.68	−0.184	0.129	−1.43	0.116	−1.58
Sinhalese manager	0.369	0.113	3.27	0.860	0.165	5.20	0.643	0.235	2.74	0.524	1.23
Household-based enterprise	0.043	0.079	0.55	−0.165	0.110	−1.50	−0.174	0.130	−1.34	0.134	−1.29
Industry dummies											
Services enterprise	−0.234	0.105	−2.23	−0.351	0.137	−2.57	−0.392	0.153	−2.57	0.159	−2.47
Manufacturing nonagricultural enterprise	0.058	0.089	0.65	0.172	0.120	1.44	0.273	0.136	2.01	0.128	2.13
Agricultural processing enterprise	−0.239	0.203	−1.18	−0.395	0.209	−1.89	−0.420	0.348	−1.20	0.309	−1.36
Other production enterprise	−0.094	0.288	−0.33	−0.034	0.220	−0.16	−0.371	0.724	−0.51	0.300	−1.24
Mixed enterprise	−0.037	0.126	−0.30	−0.155	0.160	−0.97	0.142	0.204	0.69	0.176	0.81
Community variables											
Community population size[b]	0.226	0.070	3.21	0.286	0.094	3.05	0.257	0.229	1.13	0.353	0.73
Income per capita[b]	−0.102	0.071	−1.43	−0.140	0.085	−1.64	−0.082	0.238	−0.34	0.438	−0.19
Agricultural seasonality	0.045	0.127	0.35	0.216	0.171	1.26	0.026	0.475	0.05	1.408	0.02
Share income from agriculture	0.016	0.061	0.27	0.033	0.068	0.48	0.013	0.128	0.10	0.248	0.05
Benchmark indexes											
Connectivity	0.352	0.275	1.28	0.471	0.348	1.36	0.327	0.586	0.56	1.478	0.22
Infrastructure services	0.064	0.325	0.20	0.216	0.417	0.52	−0.123	0.903	−0.14	3.922	−0.03
Business services	0.168	0.154	1.09	0.373	0.211	1.77	0.529	0.296	1.79	0.637	0.83
Human capital	2.315	0.494	4.69	2.124	0.661	3.21	2.414	1.082	2.23	3.639	0.66
Finance services	−0.198	0.296	−0.67	−0.442	0.393	−1.13	−0.785	0.596	−1.32	1.473	−0.53

Source: RIC Surveys.

Notes: a. Ordered probit parameter estimates, not further scaled.

b. Variable enters the model in logarithmic form.

Annex I. Econometric Analysis of RIC Survey Data

I.1 INTRODUCTION

The Rural Investment Climate Surveys gather information about enterprises and households located in communities. Outcome variables at the enterprise or household level are examined in the light of determinants specific to the individual level (enterprise and/or household) as well as factors at play within the community. This has implications for the econometric approach that should be followed. In particular, acknowledging that observed community variables determine the outcome of interest invites the conclusion that some unobservable factors may also matter. It has been shown that the presence of such factors can seriously bias the statistical inference drawn from basic statistical analyses. The standard solution is to compute the standard errors of the estimated parameters with an adjustment for clustering at the community level, but this approach is not free of critical assumptions either. Alternative methods of analysis borrow from panel-econometric techniques, but these usually do not accommodate sampling weights. The following paragraphs outline the proper methods to follow for parameter estimation in the context of data samples such as those generated with RIC surveys.

In principle, a fixed effects regression is a standard procedure, but weights are typically ignored in the econometric literature because panel data pertain to repeated observations; at most, one would assign weights to the whole time series of a particular observation, not to each separate time period. This situation is different in a clustered cross-section such as the RIC sample. Many of the techniques of panel econometrics apply, but each individual has (or may have) its own weight.

Let communities be denoted by index j, with $j = 1, \ldots, J$, and individuals within the community by i, with $i = 1, \ldots, N_j$, where the number of observations per community may vary. For shorthand notation, define $n = \sum_{j=1}^{J} N_j$. The term *individual* is used loosely here to indicate any kind of economic agent within the community: a household, an enterprise, a person, or an organization such as a bank, cooperative, or politician. The sampling weight associated with this observation is denoted as w_{ji}.

Consider a dependent variable y_{ji} that describes an outcome for individual i in community j. Explanatory variables are a vector X_{ji} consisting of variables that vary with i and perhaps also with j, and a vector Z_j that are community-specific and are the same for all observations within that community. Random influences consist of a community-level component μ_j and an individual component ν_{ji}. Altogether, the empirical model is written as:

$$y_{ji} = X'_{ji}\beta + Z'_j\gamma + \mu_j + \nu_{ji} \qquad (1)$$

We make the following distributional assumptions about the components of the disturbance: μ_j is independently and identically distributed with mean 0 and variance σ^2_μ across communities; ν_{ji} is independently and identically distributed with mean 0 and variance σ^2_μ across all individuals; and μ_j and ν_{ji} are independent of each other. For some estimation approaches, we must also assume that μ_j and ν_{ji} are independent of X and Z. For future reference, define u as

$$u_{ji} = \mu_j + \nu_{ji} \qquad (2)$$

The estimation issues for this model are outlined in the following sections. Section 2 considers the case where the dependent variable y is continuous, and Section 3 addresses the case where y is discrete.

I.2 ESTIMATION APPROACHES FOR CONTINUOUS DEPENDENT VARIABLES

I.2.1 Ordinary Least Squares

The simplest estimation approach is OLS. The parameter estimates are unbiased and consistent, but they are not efficient relative to other estimation methods. The usual standard errors of the parameter estimates are computed under the assumption of independent disturbances, but in this model the disturbance (u_{ji}) of individuals within a given community are clearly correlated: they share μ_j. Thus, the standard errors must be adjusted for this clustering of individuals by community. This is straightforward to do in Stata; see Section A1 of the Appendix.

An important assumption made by the OLS technique is the independence between the explanatory variables and the components of the disturbance term. This assumption is usually made in applied econometric analysis, of course, but in this case the community variables may be inadequately measured where at the same time Z_j represents only some of the relevant variables at the community level. In other words, one may make a case that Z_j and μ_j are often correlated. In that case, the OLS estimator loses its property of being unbiased and consistent.

I.2.2 Fixed Effects Estimation

OLS does not take into consideration the disturbance structure of the model. As a result, the parameter estimates are inefficient. One alternative estimation approach is the fixed effects method; annotated Stata code is provided in Appendix A.2. To understand this method better, order the observations by community and write out the model in matrix form:

$$y = X\beta + Z\gamma + Z_\mu\mu + \nu \tag{3}$$

where Z_μ is a $n \times J$ matrix of ones and zeroes linking the community factors μ_j to each individual. One might write $Z_\mu = diag(\iota_{N_j})$, where ι_{N_j} is a vector of ones of length N_j. Because Z varies only by community, Z is perfectly correlated with Z_μ. Thus, to estimate this model with the fixed effects method, it is necessary to absorb Z into the community component:

$$\mu_j^* = Z_j'\gamma + \mu_j \tag{4}$$

and the regression model becomes

$$y = X\beta + Z_\mu\mu^* + \nu \tag{5}$$

Let W denote the matrix with weights on the diagonal. With sampling weights, the fixed effects estimates are found by minimizing the weighted residual sum of squares:

$$\min_{\{\hat\beta,\hat\mu^*\}} (\hat\nu' W\hat\nu) \tag{6}$$

Usually, a fixed effects model is not estimated in this way, because the dummy variables identifying each group (community) are too numerous and because a simple model transformation eliminates the fixed effects. To investigate the effect of the weights, however, it is still useful to develop the fixed effects estimator in this way. Solving the first order conditions of the maximization process of equation (6), the estimator is written as

$$\begin{pmatrix}\hat\beta\\ \hat\mu^*\end{pmatrix} = \begin{pmatrix}X'WX & X'WZ_\mu\\ Z_\mu'WX & Z_\mu'WZ_\mu\end{pmatrix}^{-1}\begin{pmatrix}X'Wy\\ Z_\mu'Wy\end{pmatrix} \tag{7}$$

From this, using the Frisch-Waugh-Lovell Theorem, it follows that

$$\begin{aligned}\hat\beta_{FE} &= (X'(W - WZ_\mu(Z_\mu'WZ_\mu)^{-1}Z_\mu'W)X)^{-1}\\ &\quad\times X'(W - WZ_\mu(Z_\mu'WZ_\mu)^{-1}Z_\mu'W)y\\ &= (\tilde X'\tilde Q\tilde X)^{-1}\tilde X'\tilde Q\tilde y\end{aligned} \tag{8}$$

where $\tilde Q = (I_n - \tilde Z_\mu(\tilde Z_\mu'\tilde Z_\mu)^{-1}\tilde Z_\mu')$ is an idempotent matrix that creates deviations of $\tilde y = W^{0.5}y$ from the within-community weighted average multiplied by the square root of the weight, which in matrix notation equals $W^{0.5}Z_\mu(Z_\mu'WZ_\mu)^{-1}Z_\mu'Wy$, and similar for $\tilde X = W^{0.5}X$. In other words, the fixed effects estimator with weighted data is found by (*i*) defining the data transformation $y_{ji}^* = w_{ji}^{0.5}(y_{ji} - \breve y_{ji})$, and a similar method is followed for the explanatory variables, where the symbol ˘ denotes a within-community weighted average, and (*ii*) estimate a regression of y^* on X^* by OLS, without an intercept. The fixed effects may be derived from equation (7) as well. Alternatively, going back to equation (6), $\hat\mu_j^*$ follows from the minimizing of $\hat\nu_j'W_j\hat\nu_j$, which is the portion of $\hat\nu'W\hat\nu$ contributed by community j, conditional on $\hat\beta$:

$$\begin{aligned}\hat\mu_j^* &= \left(\sum_{i=1}^{N_j} w_{ji}\right)^{-1}\left(\sum_{i=1}^{N_j} w_{ji}(y_{ji} - X_{ji}'\hat\beta)\right)\\ &= \breve y_j - \breve X_j'\hat\beta\end{aligned} \tag{9}$$

where the symbol ˘ denotes a community-level weighted average.

To estimate σ_ν^2, note that the transformation matrix $\tilde{Q}W^{0.5}$ sweeps the community fixed effect out of the equation. Thus consider the inner product of the residuals of the transformed model:

$$\begin{aligned} E[\nu^{*\prime}\nu^*] &= E[\nu' W^{0.5} Q Q W^{0.5} \nu] \\ &= tr(W^{0.5} Q Q W^{0.5} E[\nu\nu']) \\ &= \sigma_\nu^2\, tr(QW) \\ &= \sigma_\nu^2\, tr(W - W^{0.5} Z_\mu (Z_\mu' W Z_\mu)^{-1} Z_\mu' W^{0.5} W) \\ &= \sigma_\nu^2\, (tr(W) - tr((Z_\mu' W Z_\mu)^{-1}(Z_\mu' W^2 Z_\mu))) \\ &= \sigma_\nu^2 \left(\sum_{j=1}^{J} \sum_{i=1}^{N_j} w_{ji} - \sum_{j=1}^{J} \left(\sum_{j=1}^{N_j} w_{ji}^2 \Big/ \sum_{i=1}^{N_j} w_{ji} \right) \right) \end{aligned} \tag{10}$$

Therefore an estimator for σ_ν^2 is found as:

$$\hat{\sigma}_\nu^2 = \hat{\nu}^{*\prime}\hat{\nu}^* \Big/ \left(\sum_{j=1}^{J} \sum_{i=1}^{N_j} w_{ji} - \sum_{j=1}^{J} \left(\sum_{i=1}^{N_j} w_{ji}^2 \Big/ \sum_{i=1}^{N_j} w_{ji} \right) \right) \tag{11}$$

This does not yet yield estimates of γ, of course, which represents the effect of community variables Z_j on individual outcomes y_{ji}. For this, connect the estimated fixed effects with the community error component μ_j using equation (4):

$$\hat{\mu}_j^* = Z_j\gamma + \mu_j + (\hat{\mu}_j^* - \mu_j^*) = Z_j\gamma + \eta_j \tag{12}$$

With $\hat{\mu}_j^*$ estimated and Z_j observed, estimation of this equation yields γ (see Amemiya 1978). The disturbance term is not iid, of course, even if μ_j is, because the variance of $(\hat{\mu}^* - \mu^*)$ equals

$$\begin{aligned} Var(\hat{\mu}^* - \mu^*) = {} & H\Sigma H' - H\Sigma A' X' H' - HXA\Sigma H' \\ & + HXVar\,(\hat{\beta}_{FE}) X' H' \end{aligned} \tag{13}$$

where

$$\begin{aligned} H &= (Z_\mu' W Z_\mu)^{-1} Z_\mu' W \\ A &= (X^{*\prime} W X^*)^{-1} X^{*\prime} W \Sigma^{-0.5} E, \\ E &= I - Z_\mu (Z_\mu' W Z_\mu)^{-1} Z_\mu' W, \\ \Sigma &= \sigma_v^2 I, \\ X^* &= \Sigma^{-0.5} E X. \end{aligned}$$

Thus, equation (13) makes clear that $(\hat{\mu}^* - \mu^*)$ is heteroskedastic (for example, the first term of equation (13) equals $H\Sigma H' = diag(\sigma_v^2(\sum_{i=1}^{N_j} w_{ji}^2)/(\sum_{i=1}^{N_j} w_{ji})^2)$, the elements of which clearly varies with j because of the variation in the weights), and the elements of the vector $(\hat{\mu}^* - \mu^*)$ are correlated. (The fourth term, for example, is a full matrix that depends on the community-averaged values of the explanatory variables, as $HX = \breve{X}$ is a $K \times J$ matrix where the jth row contains the average of X in community j; moreover, the second and third matrices of (13) have non-zero off-diagonal elements as well.)

To implement a GLS estimator of equation (10), the variance of μ_j must first be estimated. Let Z_c be the matrix that combines Z_j at the community level (one row per community, as opposed to repeated rows as in the matrix Z). Then $\hat{\sigma}_\mu^2$ may be derived from the sum of squared OLS residuals of equation (12), corrected for the part of the variance that comes from $(\hat{\mu}_j^* - \mu_j^*)$:

$$\begin{aligned} \hat{\sigma}_\mu^2 = {} & (\hat{\mu}^{*\prime}(I - Z_c(Z_c' Z_c)^{-1} Z_c)\hat{\mu}^* \\ & - tr(\hat{Var}\,(\hat{\mu}^* - \mu^*)))/J \end{aligned} \tag{14}$$

Then, define

$$\hat{Var}(\eta) \equiv \hat{\Sigma}_\eta = \hat{\sigma}_\mu^2 I_J + \hat{Var}\,(\hat{\mu}^* - \mu^*) \tag{15}$$

and the feasible GLS estimator is obtained:

$$\hat{\gamma}_{FGLS} = (Z_c' \hat{\Sigma}_\eta^{-1} Z_c)^{-1} Z_c'\ \hat{\Sigma}_\eta^{-1} \hat{\mu}^* \tag{16}$$

with a variance equal to $Var(\hat{\gamma}_{FGLS}) = (Z_c' \hat{\Sigma}_\eta^{-1} Z_c)^{-1}$.

As a final comment, it should be noted that the fixed effects estimator allows correlation of the community-error component with X and Z (although the correlation with Z becomes irrelevant as it is swept out of the regression model). But this applies to the estimator of β only: the estimator of γ does not permit a correlation between μ and Z, if it is to be consistent.

I.2.3 Random Effects Estimation

To describe the effect of weights on the random effects approach, it is better to write the model in matrix notation at the community level. Thus, let y_j be a $N_j \times 1$ vector of observations in community j, and other variables are defined similarly. Then,

$$\begin{aligned} y_i &= X_j\beta + \iota_{N_j} Z_j\gamma + \iota_{N_j}\mu_j + \nu_j \\ &= X_j\beta + \iota_{N_j} Z_j\gamma + u_j \\ &\equiv T_j\delta + u_j \end{aligned} \tag{17}$$

As is standard,

$$E[u_j u_j'] = \sigma_\mu^2 \iota_{N_j} \iota_{N_j}' + \sigma_\nu^2 I_{N_j} \equiv \Omega_j \tag{18}$$

This means that $\Omega_j^{-0.5} u_j$ is iid with mean 0 and variance 1. Thus, in the absence of weights, the recommended transformation matrix that leads to the GLS estimator is $\sigma_\nu \Omega_j^{-0.5}$; that is, the transformation defines

$$y_{ji}^* = y_{ji} - \theta_j \bar{y}_j \tag{19}$$

where $\theta_j = 1 - (\sigma_\nu/(N_j\sigma_\mu^2 + \sigma_\nu^2))$ and $\overline{y}_j$ is the simple (unweighted) average of observations in community j. After this transformation, the resulting model is estimated with OLS. Stated in another way, in the absence of weights, the random effects (GLS) estimator minimizes the generalized sum of squared residuals $\hat{u}'\Omega^{-1}\hat{u}$.

Application of GLS with weights is equivalent[92] to minimizing the weighted generalized sum of squared residuals $\hat{u}'W^{1/2}\Omega^{-1}W^{1/2}\hat{u} = \hat{u}'\Psi\hat{u}$, where

$$\Omega = diag(\Omega_j) = diag(\tau_j^2\overline{J}_j + \sigma_\nu^2 E_j) \tag{20}$$

where $E_j = I_j - \overline{J}_j$, $\overline{J}_j = W_j^{1/2}\iota_j(\iota_j'W_j\iota_j)^{-1}\iota_j'W_j^{1/2}$, $W_j = diag(w_{ji})$ and $\tau_j^2 = (\iota_j'W_j\iota_j)\sigma_\mu^2 + \sigma_\nu^2$. The slope parameters are estimated as

$$\hat{\delta}_{WGLS} = (T'\Psi T)^{-1}T'\Psi y \tag{21}$$

with a variance equal to

$$Var(\hat{\delta}_{WGLS}) = (T'\Psi T)^{-1}T'\Psi\underline{\Omega}\Psi T(T'\Psi T)^{-1} \tag{22}$$

where $\underline{\Omega} = diag(\sigma_\mu^2\iota_j\iota_j' + \sigma_\nu^2 I_j)$. As should be obvious, the parameters β and γ are both represented in equation (21): the effect of the individual-specific and community-specific factors is estimated at the same time, unlike the situation in the fixed effects approach. Annotated Stata code is provided Appendix A.3.

To implement this procedure, estimates of σ_v^2 and σ_μ^2 are needed. The fixed effects approach yields a suitable estimator of σ_v^2 through equation (27) above, and the nature of the correlation among the u_{ji} inspires the following estimator of σ_μ^2, which is parallel to Baltagi (2005, p. 81):

$$\hat{\sigma}_\mu^2 = \sum_{j=1}^{J}\sum_{i=1}^{N_j}\sum_{\substack{h=1\\h\neq i}}^{N_j} w_{ji}w_{jh}\hat{u}_{ji}\hat{u}_{jh} \Big/ \sum_{j=1}^{J}\sum_{i=1}^{N_j}\sum_{\substack{h=1\\h\neq i}}^{N_j} w_{ji}w_{jh} \tag{23}$$

where $\hat{u}$ is an OLS residual.

As an aside, the weighted GLS procedure is explained above in matrix notation. One might transform the individual variables in a fashion parallel to the standard (unweighted) by means of the following equation:

$$y_{ji}^* = w_{ji}^{0.5}(y_{ji} - \tau_j\breve{y}_j) \tag{24}$$

where once again the symbol ˘ denotes a within-community weighted average, and where τ_j is the same as that defined in equation (20). Because of the weights, however, the variance of the OLS procedure that one might use to regress the transformed variables is incorrect.

I.3 ESTIMATION APPROACHES FOR DISCRETE DEPENDENT VARIABLES

When the dependent variable is discrete, the regression approaches discussed in Section 2 are less appropriate because the distributional assumptions (in particular, that the disturbances have a constant variance) cannot be maintained. The typical solution is found in the maximum likelihood estimation (MLE) method, even though this method requires one to make an arbitrary assumption about the shape of the distribution (such as normal or logistic).

This section discusses two estimation approaches. One applies to situations in which the dependent variable is a simple 0/1 variable. Examples of this abound in the RIC data: whether a household operates an enterprise, whether the enterprise is registered, whether the enterprise pays taxes, and so on. The second approach is suitable for dependent variables with a discrete order 0/1/2/. . . where the order is meaningful but the numerical differences between the values are arbitrary. The primary example in the RIC data are entrepreneurs' responses to questions about the presence of a potential investment climate obstacle and degree of its effect on the operation and growth of their enterprises: the responses of 1 = *No obstacle;* 2 = *Minor obstacle;* 3 = *Moderate obstacle;* 4 = *Major obstacle;* 5 = *Very severe obstacl*e constitute a natural order, but it is impossible to argue that a major obstacle is three times as bad as a minor obstacle just because 4 is three units above 1 and 2 is only one unit above 1. Stated otherwise, instead of 1/2/. . ./5, the responses could be coded 0/1/2/4/8 or 1/10/100/500/1000 without loss of any of the information carried in the entrepreneur's response. This model is estimated with ordered probit.

I.3.1 Weighted Random Effect Maximum Likelihood Estimation

The weighted random effect probit and weighted random effect ordered probit models are special cases of the general weighted random effect MLE approach. In general, the MLE approach starts with a formulation of the density function driving the generation of disturbances. From this, the contribution of observation (ji) to the likelihood function is derived, often written as $l(\theta \mid y_{ji}, X_{ji})$, where θ

comprises all parameters of the model, and the overall likelihood function is the product of these contributions:

$$L = \prod_{j=1}^{J} \prod_{i=1}^{N_j} l(\theta \mid y_{ji}, X_{ji}) \tag{25}$$

In the present case, disturbances are twofold: v_{ji} at the individual level and μ_j at the community level. The contribution of observation (ji) to the likelihood function must therefore be stated conditional on μ_j: $l(\theta \mid y_{ji}, X_{ji}, \mu_j)$. The function l is related to the density specified for v. Let the density of μ_j be written as $g(\mu_j)$, where the function g depends on additional parameters suppressed for simplicity. The contribution of the observations living in community j is then the product of the individual contributions, averaged over all feasible values of μ_j, and the overall likelihood function combines all communities:

$$L = \prod_{j=1}^{J} \left(\int \left(\prod_{i=1}^{N_j} l(\theta \mid y_{ji}, X_{ji}, \mu_j) \right) g(\mu_j)\, d\mu_j \right) \tag{26}$$

The usual way of introducing sampling weights in the common likelihood function is more intuitive when (25) is written in logarithmic form: the log-likelihood function with weights is the weighted sum of the log of the individual contributions:

$$\ln L = \sum_{j=1}^{G} \sum_{i=1}^{N_j} w_{ji} \ln l(\theta \mid y_{ji}, X_{ji}) \tag{27}$$

where it is understood that the weights w_{ji} average to 1. This is equivalent to the following direct modification of (25):

$$L = \prod_{j=1}^{J} \prod_{i=1}^{N_j} l(\theta \mid y_j, X_{ji})^{w_{ji}} \tag{28}$$

If we interpret each $l(\theta \mid y_{ji}, X_{ji})$ as a probability, the likelihood function with weights is therefore a joint probability that is a geometrically weighted combination of individual probabilities with, as stated, weights averaging to 1.

When the model contains a community random effect, it is problematic to raise every $l(\theta \mid y_{ji}, X_{ji}, \mu_j)$ to the power of w_{ji}. In (26), the expression $\int (\prod_{i=1}^{N_j} l(\theta \mid y_{ji}, X_{ji}, \mu_j))\, g(\mu_j)\, d\mu_j$ can be thought of as the joint probability of members of community j, computed as the averaged conditional joint probability. If w_{ji} does not average to 1 within a community, an expression such as

$\int (\prod_{i=1}^{N_j} l(\theta \mid y_{ji}, X_{ji}, \mu_j)^{w_{ji}})\, g(\mu_j)\, d\mu_j$ cannot be interpreted anymore as a joint probability: for one thing, it will now depend on the magnitude of the sum of the weights. Therefore, define $w^*_{ji} = w_{ji}/\overline{w}_j$ as a community-standardized weight, such that w^*_{ji} does indeed average to 1. With this in hand, the likelihood function with weights is defined as

$$L = \prod_{j=1}^{J} \left(\int \left(\prod_{i=1}^{N_j} l(\theta \mid y_{ji}, X_{ji}, \mu_j)^{w^*_{ji}} \right) g(\mu_j)\, d\mu_j \right)^{\overline{w}_j} \tag{29}$$

such that the community-level joint probability created by the integration over μ_j is itself geometrically averaged to obtain the overall joint probability that constitutes the likelihood function.

The integral in equation (29) may be difficult to evaluate analytically. This is the case in particular with the probit and ordered probit applications discussed below. As a substitute, the integral may be simulated as the average of R random draws of μ_j^r from the distribution that is characterized by g:

$$\int \left(\prod_{i=1}^{N_j} l(\theta \mid y_{ji}, X_{ji}, \mu_j)^{w^*_{ji}} \right) g(\mu_j)\, d\mu_j \approx \frac{1}{R} \sum_{r=1}^{R} \prod_{i=1}^{N_j} l(\theta \mid y_{ji}, X_{ji}, \mu_j^r)^{w^*_{ji}} \tag{30}$$

This is an application of the GHK simulation estimation method (Vijverberg 1997).

I.3.2 Probit

As a context for the probit model, consider households that may or may not operate an enterprise. Let entrepreneurship be observed as $y_{ji} = 1$ if the household operates an enterprise and $y_{ji} = 0$ if it does not. It is explained by household variables X and community variables Z. Random factors at the household and community level are incorporated with disturbances ν_{ji} and μ_j, respectively, which are identically and independently normally distributed. These factors are combined into a latent variable y^*_{ji} that measures the tendency towards entrepreneurship:

$$y^*_{ji} = X'_{ji}\beta + Z_j\gamma + \mu_j + \nu_{ji} \tag{31}$$

Then, $y_{ji} = 1(0)$ if $y^*_{ji} \geq (<)0$ and therefore $\nu_{ji} \geq (<) -X'_{ji}\beta - Z'_j\gamma - \mu_g$. As the distribution of ν_{ji} is symmetric, the range of ν_{ji} can be stated in concert with the outcome variable y_{ji} in a simple

way: $\nu_{ji} \leq (2y_{ji} - 1)(X'_{ji}\beta + Z'_j\gamma + \mu_j)$. Let us denote the normal cdf as Φ and pdf as ϕ. Then,

$$l(\theta \,|\, y_{ji}, X_{ji}, \mu_j) = \Phi[(2y_{ji} - 1)(X'_{ji}\beta + Z'_j\gamma + \mu_j)] \tag{32}$$

Substitution into equation (30) provides the weighted random effect probit model. Parameters are estimated subject to parameter standardization, which typically amounts to setting $Var(\nu_{ji})$ equal to 1. Annotated Stata code in Appendix A.4 may be used to estimate a weighted random effect probit model.

I.3.3 Ordered Probit

In the annotated Stata code in Appendix A.5, the ordered probit model assumes that categories are coded as 0/1/2/3/4. Thus, y_{ji} takes on one of these outcomes. A latent variable y^*_{ji} measures the tendency to fall into one of these categories and is related to explanatory variables in the same way as in the probit model; see equation (31). Then,

$$y_{ji} = c \quad \text{iff} \quad \alpha_c \leq y^*_{ji} < \alpha_{c+1} \quad \text{for} \quad c = 0, 1, \ldots, 4 \tag{33}$$

or

$$y_{ji} = c \quad \text{iff} \quad \alpha_c - X'_{ji}\beta - Z'_j\gamma - \mu_j \leq \nu_{ji} < \alpha_{c+1} - X'_{ji}\beta - Z'_j\gamma - \mu_j \quad \text{for} \quad c = 0, 1, \ldots, 4 \tag{34}$$

where the α's are parameters to be estimated, where $\alpha_0 = -\infty$ and $\alpha_5 = \infty$, and where standardization requires one other α to be fixed: $\alpha_1 = 0$. From this, the function l is written as:

$$l(\theta \,|\, y_{ji} = c, X_{ji}, \mu_j) = \Phi[\alpha_{c+1} - X'_{ji}\beta - Z'_j\gamma - \mu_j] - \Phi[\alpha_c - X'_{ji}\beta - Z'_j\gamma - \mu_j] \tag{35}$$

I.3.4 Estimating Standard Errors of the Estimates

Writing $l(\theta \,|\, y_{ji}, X_{ji})$ as l_{ji}, derivatives of the common likelihood function (25) are used to compute the variance of the estimate $\hat{\theta}$ of the parameter vector θ (see, for example, Greene 2008, p. 495).

$$EstVar(\hat{\theta}) = \left(-\sum_{j=1}^{J} \sum_{i=1}^{N_j} \frac{\partial^2 \ln l_{ji}(\hat{\theta})}{\partial \theta \, \partial \theta'} \right)^{-1} \equiv V_1 \tag{36}$$

or

$$EstVar(\hat{\theta}) = \left(\sum_{j=1}^{J} \sum_{i=1}^{N_j} \left(\frac{\partial \ln l_{ji}(\hat{\theta})}{\partial \theta} \right) \left(\frac{\partial \ln l_{ji}(\hat{\theta})}{\partial \theta} \right)' \right)^{-1} \equiv V_2 \tag{37}$$

The matrices V_1 and V_2 are the same in expectation, that is, when samples become large. In small samples, the two matrices may well deviate, and V_1 is the preferred one (Greene 2008, p. 496).

With weights, using equation (27), the parallel expressions for V_1 and V_2 are:

$$EstVar(\hat{\theta}) = \left(-\sum_{j=1}^{J} \sum_{i=1}^{N_j} w_{ji} \frac{\partial^2 \ln l_{ji}(\hat{\theta})}{\partial \theta \, \partial \theta'} \right)^{-1} \equiv V_1 \tag{38}$$

$$EstVar(\hat{\theta}) = \left(\sum_{j=1}^{J} \sum_{i=1}^{N_j} w_{ji}^2 \left(\frac{\partial \ln l_{ji}(\hat{\theta})}{\partial \theta} \right) \left(\frac{\partial \ln l_{ji}(\hat{\theta})}{\partial \theta} \right)' \right)^{-1} \equiv V_2 \tag{39}$$

It is straightforward to see that V_1 and V_2 can no longer be the same, since (38) is linear in the weights and (39) is quadratic. Instead, the variance of $\hat{\theta}$ is found by a Taylor expansion of the gradient of the likelihood function:

$$\frac{\partial \ln L(\hat{\theta})}{\partial \theta} = 0 = \frac{\partial \ln L(\theta)}{\partial \theta} + \frac{\partial^2 \ln L(\theta)}{\partial \theta \, \partial \theta'} (\hat{\theta} - \theta) \tag{40}$$

Thus

$$\hat{\theta} - \theta = -\left(\frac{\partial^2 \ln L(\theta)}{\partial \theta \, \partial \theta'} \right)^{-1} \frac{\partial \ln L(\theta)}{\partial \theta} \tag{41}$$

and therefore

$$Var(\hat{\theta}) = E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)'] = \left(\frac{\partial^2 \ln L(\theta)}{\partial \theta \, \partial \theta'} \right)^{-1} \left(\frac{\partial \ln L(\theta)}{\partial \theta} \frac{\partial \ln L(\theta)}{\partial \theta'} \right) \left(\frac{\partial^2 \ln L(\theta)}{\partial \theta \, \partial \theta'} \right)^{-1} \tag{42}$$

The estimated variance is obtained by evaluating the first and second order derivatives at $\hat{\theta}$:

$$EstVar(\hat{\theta}) = \left(\frac{\partial^2 \ln L(\hat{\theta})}{\partial \theta \, \partial \theta'} \right)^{-1} \left(\frac{\partial \ln L(\hat{\theta})}{\partial \theta} \frac{\partial \ln L(\hat{\theta})}{\partial \theta'} \right) \left(\frac{\partial^2 \ln L(\hat{\theta})}{\partial \theta \, \partial \theta'} \right)^{-1} = V_1 V_2^{-1} V_1 \equiv V_3 \tag{43}$$

This expression is not sensitive to the scaling of the weights.

I.3.5 A finding About Standard Errors with RIC Data

The Stata programs given in Appendices A.4 and A.5 automatically estimate the standard errors of the estimated parameters with equation (43). In estimating the ordered probit model for Tanzania, however, enterprise weights are absent, and thus w_{ji} is set equal to 1. One would have expected that

V_3 would be roughly the same as V_1 since $V_2^{-1}V_1 = I$. In practice, the standard errors showed significant differences.

This led to a more general check against the data from Nicaragua and Sri Lanka as well from the three enterprise investment climate outcome variables (cost of finance, road quality, and electricity). Table I.1 summarizes the findings, showing ratios of standard errors averaged over the three estimated models in each country. For enterprise and industry variables, the standard errors obtained from V_1 and V_2 are roughly the same: overall, across the three countries and all available ratios, the ratio averages out at 0.999. Among parameters of community variables, however, the ratio averages only 0.616: standard errors obtained through V_2 are typically much smaller. The only plausible explanation for this difference would have to be the number of enterprises (over 1,000) as compared to the number of communities (between 100 and 150): the former is large enough to roughly equalize V_1 and V_2, but the latter falls significantly short.

This finding implies that V_1, V_2 and V_3 are very different estimators of the variance of $\hat{\theta}$. Columns 1 through 3 of Table I.2 illustrate the difference between V_1 and V_3. Where V_2-based standard errors for community-level variables are much smaller than those derived from V_1, the V_3-based standard errors are much larger.

As a further test, for one Sri Lanka variable (obstacles related to electricity), the *t*-statistics of one of the benchmark indicators (infrastructure

Table I.1 Ratio of Standard Errors of Parameter Estimates Obtained Through V_2 and V_1

Variable	Nicaragua	Sri Lanka	Tanzania	Average
Enterprise and industry variables				
Age of enterprise#	1.046	1.022	0.974	
Education of manager		1.000	0.975	
Female manager		1.019	1.009	
Sinhalese manager		0.858		
Household-based enterprise	1.049	0.970	1.022	
Services enterprise	1.013	1.007	0.929	
Nonagricultural manufacturing enterprise	1.014	1.010	0.943	
Agricultural processing enterprise	1.019	1.111	0.898	
Other production enterprise	1.063	1.081	0.972	
Mixed enterprise	1.000	1.038	0.884	
Average	1.029	1.012	0.956	0.999
Community and benchmark variables				
Community population size#	0.623	0.656	0.490	
Income per capita#	0.735	0.584	0.527	
Agricultural seasonality	0.700	0.592		
Share income from agriculture	0.730	0.807	0.513	
Connectivity	0.701	0.645	0.498	
Infrastructure services	0.665	0.684	0.546	
Business services	0.617	0.634	0.556	
Human capital	0.689	0.634	0.519	
Finance services	0.629	0.641	0.516	
Intercept	0.737	0.612	0.495	
Average	0.682	0.649	0.518	0.616
Other model parameters				
St.dev of community random error (σ_μ)	0.862	0.877	0.712	0.817
α_2	1.035	1.007	1.004	
α_3		1.015	0.988	
α_4		1.002	0.962	
Average across α's	1.035	1.008	0.985	1.009

Source: Authors' calculation from RIC survey data.

Note: #Variable enters in logarithmic form.

Table I.2 Ratio of Standard Errors of Parameter Estimates Obtained Through V_3 and V_1

	Unweighted			Weighted[a]		Ratio of weighted over unweighted[a]	
Variable	**Nicaragua**	**Sri Lanka**	**Tanzania**	**Nicaragua**	**Sri Lanka**	**Nicaragua**	**Sri Lanka**
Enterprise and industry variables							
Age of enterprise[b]	0.979	0.995	1.078	1.094	1.477	1.117	1.485
Education of manager		1.019	1.164		1.481		1.453
Female manager		1.001	1.056		1.321		1.319
Sinhalese manager		1.247			2.459		1.971
Household-based enterprise	0.970	1.066	1.024	0.994	1.323	1.025	1.241
Services enterprise	1.003	1.018	1.219	1.143	1.325	1.140	1.301
Nonagricultural manufacturing enterprise	1.002	1.010	1.123	1.119	1.280	1.117	1.267
Agricultural processing enterprise	0.999	0.917	1.192	1.117	1.014	1.118	1.106
Other production enterprise	0.954	0.949	1.098	1.073	0.644	1.125	0.678
Mixed enterprise	1.016	0.991	1.197	1.127	1.340	1.110	1.352
Average	0.989	1.021	1.128			1.107	1.318
Community and benchmark variables							
Community population size[b]	1.733	1.588	2.406	2.038	2.194	1.176	1.381
Income per capita[b]	1.468	1.790	2.117	1.634	2.437	1.113	1.361
Agricultural seasonality	1.497	1.756		1.583	2.875	1.058	1.637
Share income from agriculture	1.427	1.435	2.213	1.664	2.281	1.166	1.590
Connectivity	1.503	1.630	2.232	1.639	2.182	1.091	1.339
Infrastructure services	1.569	1.510	2.070	1.680	3.009	1.071	1.993
Business services	1.711	1.709	1.985	1.792	2.318	1.048	1.357
Human capital	1.527	1.678	2.209	1.634	2.491	1.070	1.484
Finance services	1.655	1.634	2.150	1.750	2.195	1.057	1.343
Intercept	1.542	1.740	2.377	1.729	2.184	1.122	1.255
Ave com char	1.563	1.647	2.195			1.097	1.474
Other model parameters							
St. dev of com. random error (σ_μ)	1.322	1.224	1.630	1.306	1.946	0.989	1.590
α_2	0.988	1.002	1.014	1.096	1.386	1.109	1.383
α_3		0.999	1.044		1.367		1.368
α_4		1.031	1.094		1.284		1.245
Average across α's	0.988	1.011	1.051			1.109	1.332

Source: Authors' calculation from RIC survey data.

Notes: a. Tanzania's survey does not contain sampling weights.

b. Variable enters in logarithmic form.

services) was compared to the likelihood ratio test value computed if this benchmark indicator were to be omitted. The p-value of the likelihood ratio was 0.1288; the *t*-statistics and p-values were −1.518 and 0.1290 when the V_1-based standard errors are used, −2.062 and 0.0392 when the V_2-based standard errors are used, and −1.088 and 0.2768 when the V_3-based standard errors are used. This illustrates that V_1 is the best estimator of the variance of $\hat{\theta}$.

In principle, this difference would not be problematic, since V_1 is the recommended estimator of the variance anyway if there are no sampling weights. With weights, however, V_3 is the theoretically correct (asymptotic) estimator of the variance of $\hat{\theta}$, and the divergence of V_1 and V_2 starts to bite into the implications of the analysis. Table I.2 illustrates this point in columns 4 and 5 with the ratio of the standard errors derived from V_3 over those derived from V_1 when the model is estimated with sampling weights. For Nicaragua, V_3-based standard errors are about 70 percent larger than the V_1-based ones, and for Sri Lanka the V_3-based standard errors are about two and a half times as large—and *t*-statistics are correspondingly much lower.

But, because V_2 apparently is a poor estimator of what it is supposed to measure, should V_3 actually be considered as an estimator of the variance of $\hat{\theta}$? One could argue that the ratio of the standard errors with weights set to 1, as shown in columns 1 through 3 of Table I.2, reflects the consequence of using RIC samples with relatively small numbers of communities. Furthermore, it is arguable that the ratio of standard errors using actual sampling weights (as shown in columns 4 and 5 of Table I.2) deviates from 1 both because of the small number of communities and because of the variation in the sampling weights. It may be concluded then that the difference in those ratios with unit and actual weights—measured additively or, as in columns 6 and 7 of Table I.2, proportionally—is the actual effect of using weights on the standard errors.

Stated otherwise, in Nicaragua, the effect of sampling weights on standard errors is a correction of V_1-based standard errors of 10.7 percent for enterprise and industry variables (and also the α's) and of 9.7 percent for the community variables and benchmark indicators. In Sri Lanka, the use of sampling weights necessitates a correction of V_1-based standard errors of 31.8 percent for enterprise and industry variables (and also the α's), of 47.4 percent for the community variables and benchmark indicators, and of 59 percent for $\hat{\sigma}_\mu$.

These rules of thumb are likely to generate standard errors and *t*-statistics that are more accurate when sampling weights are used. They appear to differ between countries, and these particular values are suggested only for the ordered probit models with which they were computed.

In principle, the validity of this correction to V_1-based standard errors may be confirmed with a Monte Carlo study. Alternatively, standard errors may be computed with bootstrap methods, but, as with a Monte Carlo study, this is a very time-intensive undertaking that will be left for future research.

I.4 THE IMPORTANCE OF SAMPLING WEIGHT AND RANDOM EFFECTS IN RIC ANALYSIS

The techniques discussed in this Annex aim to offer econometric solutions to two characteristics of RIC databases: (i) observations have unequal sampling weights, and (ii) observations are clustered in communities. The question arises how much difference it makes to account for weights and random effects in the actual empirical analysis of RIC data. This section considers representative examples of enterprise performance, entrepreneurship selection, and EICO models estimated in alternative ways, using data from Nicaragua and Sri Lanka.

Table I.3 starts off this comparison with enterprise performance models, referring to both sales and net value added. The models contain enterprise characteristics, industry dummies, benchmark indicators, and other community characteristics. The estimates of these models therefore correspond with those found in column 3 of Table C.4 to Table C.7. The first column of Table I.3 consists of unweighted OLS estimates with common *t*-statistics labeled as "t(u)" and *t*-statistics adjusted for clustering by community ("t(a)"). The second set of estimates use sampling weights. The third and fourth columns present unweighted and weighted random effect models, respectively. The comparisons yield the following conclusions: (i) accounting for sampling weights affects estimates more dramatically than does incorporating random effects; (ii) the estimated effects of community variables (including benchmarks) tend to be more sensitive to the application of weights or random effects than are enterprise variables; (iii) adjusting *t*-statistics of OLS estimates for clustering reduces many but not all *t*-statistics; and (iv) if the weighted random effect procedure is not available, weighed OLS with *t*-statistics adjusted for clustering is the best substitute estimation technique.

Next, Table I.4 employs the same strategy in the context of entrepreneurship selection. Once again, the specification of the estimated model is typical of those reported in Chapter 4, containing household characteristics, benchmark indicators, and other community characteristics. The first column contains standard probit estimates without weights with common *t*-statistics labeled as "t(u)" and *t*-statistics adjusted for clustering by community ("t(a)"). The second set of estimates use sampling weights. The third and fourth columns present unweighted and weighted random effect probit estimates. The table reports estimated slopes, rather than standardized effects on the probability of entrepreneurship as in Table F.2. The most important conclusions are similar to those above: (i) accounting for sampling weights generates more substantial changes in parameter

estimates than does incorporating random effects, and (ii) the estimated effects of community variables (including benchmarks) tends to be more sensitive to the application of weights or random effects than are household characteristics. Furthermore, (iii) adjusting standard errors for clustering is more important for community variables and benchmark indexes than for household characteristics, and (iv) if the weighted random effect probit procedure is not available, weighed probit with *t*-statistics adjusted for clustering is the best substitute estimation technique. In the case of Nicaragua, the weighted random effect probit model yields a vanishing random effect and thus automatically simplifies to a weighted standard probit model.

As the last example, Table I.5 examines two enterprise investment climate outcomes (EICOs) prominent in all countries: interest charged on loans and issues surrounding electricity. The specified models include enterprise characteristics and the selection of community variables (benchmark and other) that in the weighted random effect ordered probit version appeared to contribute most to the explanation of the particular EICO. Thus, the estimated models are similar to those reported in Table H.4 of Annex H. But unlike that table, Table I.5 reports estimates of slopes rather than standardized effects on EICO outcomes. Conclusions are as follows: (i) accounting for sampling weights has a greater effect than does incorporating random effects; (ii) estimated slopes of community variables (including benchmarks) are especially affected; (iii) *t*-statistics must be adjusted for clustering, especially for community variables where unadjusted *t*-statistics are significantly upwardly biased; and (iv) if weighted random effect ordered probit procedure is not available, no alternative procedure is clearly second-best, as estimates of the unweighted random effect ordered probit and weighted standard ordered probit models both deviate in substantial ways from the theoretically preferred weighted random effect ordered probit estimates.

In sum, the evidence suggests it is mandatory to account for weights, and it is strongly advisable to account for clustering. In particular, the effect of benchmark indicators and community characteristics is sensitive to the specification of the model in regard to application of sampling weights and the use of random effects.

Table I.3 Effect of Sampling Weights and Random Effect Specification on Enterprise Performance Regression Results

A1: Nicaragua: Dependent variable is ln(sales)

	Ordinary Least Squares (unweighted)			Ordinary Least Squares (weighted)			Random Effects (unweighted)		Random Effects (weighted)	
	b	t(u)	t(a)	b	t(u)	t(a)	b	t	b	t
Factor and nonfactor inputs										
Total labor input#	−0.330	−2.68	−2.16	−0.232	−1.67	−1.54	−0.311	−2.53	−0.231	−5.07
Total capital input#	−0.071	−0.66	−0.66	−0.158	−1.36	−1.45	−0.086	−0.80	−0.165	−4.07
lnVLL (lnL*lnL)	0.062	4.42	3.51	0.046	2.78	2.61	0.061	4.30	0.048	9.13
lnVLC (lnL*lnC)	−0.009	−0.51	−0.50	0.010	0.51	0.51	−0.008	−0.43	0.010	1.46
lnVCC (lnC*lnC)	0.026	3.50	3.06	0.024	2.80	2.59	0.027	3.56	0.024	8.79
Depreciation#	0.057	2.29	2.14	0.050	1.77	1.70	0.058	2.30	0.060	6.29
Nonfactor Cost#	0.083	5.58	4.67	0.087	5.27	4.73	0.079	5.16	0.077	12.96
Other enterprise characteristics										
Age of enterprise#	0.115	3.36	3.13	0.105	2.61	2.83	0.112	3.31	0.098	7.63
Registration	−0.051	−0.62	−0.54	−0.040	−0.38	−0.35	−0.048	−0.59	−0.045	−1.46
Industry dummies										
Services enterprise	−0.160	−1.82	−1.72	−0.160	−1.62	−1.66	−0.135	−1.54	−0.116	−3.48
Manufacturing nonagricultural enterprise	0.086	0.73	0.71	0.144	1.04	1.08	0.077	0.66	0.114	2.61
Agricultural processing enterprise	−0.097	−0.99	−0.99	−0.022	−0.20	−0.19	−0.074	−0.76	0.000	0.00
Other production enterprise	−0.197	−1.12	−1.23	−0.120	−0.91	−0.82	−0.178	−1.01	−0.072	−1.10
Mixed enterprise	0.105	1.15	1.29	0.189	1.88	2.08	0.120	1.32	0.183	5.29
Community characteristics										
Agricultural seasonality	0.149	1.86	1.84	−0.010	−0.12	−0.10	0.141	1.39	−0.042	−0.56
Enterprise density	0.001	1.71	1.37	0.001	1.67	1.28	0.001	1.51	0.001	2.36
Community population size#	0.031	0.95	0.84	0.047	1.30	0.99	0.038	0.91	0.068	2.16
Agricultural land per capita	0.014	0.31	0.33	−0.005	−0.11	−0.10	0.001	0.01	−0.048	−1.16
Illiteracy	−0.001	−0.47	−0.43	0.000	−0.06	−0.06	−0.001	−0.33	0.000	0.15
Benchmark indicators										
Connectivity	0.555	2.01	1.99	0.490	1.71	1.48	0.601	1.68	0.673	2.48
Infrastructure services	−0.106	−0.44	−0.43	−0.040	−0.17	−0.14	−0.162	−0.54	−0.112	−0.51
Development services	0.110	1.23	1.36	0.048	0.48	0.52	0.142	1.18	0.079	0.85
Governance	0.354	1.20	1.20	0.303	0.94	1.03	0.313	0.82	0.212	0.75
Human capital	−0.047	−0.09	−0.10	−0.441	−0.87	−0.74	−0.023	−0.03	−0.590	−1.26
Finance services	−0.042	−0.29	−0.31	−0.092	−0.62	−0.55	−0.044	−0.22	−0.163	−1.04
Intercept	5.339	9.88	9.73	5.293	8.79	8.18	5.286	8.96	5.212	15.20
Regression statistics										
Number of observations	846			846			846		846	
R^2	0.422			0.431			0.439		0.428	
sv[a]	0.887			0.891			0.864		0.286	
smu[a]	0.000			0.000			0.241		0.247	

(*continued on the next page*)

Table I.3 Effect of Sampling Weights and Random Effect Specification on Enterprise Performance Regression Results (*continued*)

A2: Nicaragua: Dependent variable is ln(net value added)

	Ordinary Least Squares (unweighted)			Ordinary Least Squares (weighted)			Random Effects (unweighted)		Random Effects (weighted)	
	b	t(u)	t(a)	b	t(u)	t(a)	b	t	b	t
Factor and nonfactor inputs										
Total labor input#	−0.359	−2.62	−2.37	−0.240	−1.62	−1.58	−0.345	−2.52	−0.242	−5.16
Total capital input#	0.057	0.45	0.36	−0.046	−0.27	−0.27	0.050	0.40	−0.024	−0.54
lnVLL (ln*ln)	0.077	4.84	4.16	0.060	3.38	3.19	0.076	4.81	0.065	12.03
lnVLC (lnL*lnC)	−0.031	−1.47	−1.31	−0.016	−0.60	−0.64	−0.031	−1.47	−0.022	−3.09
lnVCC (ln*ln)	0.033	3.61	2.39	0.035	3.17	2.40	0.033	3.63	0.035	11.69
Other enterprise characteristics										
Age of enterprise#	0.111	2.87	2.74	0.094	2.05	2.32	0.110	2.86	0.090	6.72
Registration	0.079	0.92	0.73	0.100	0.87	0.78	0.084	0.98	0.108	3.57
Industry dummies										
Services enterprise	−0.175	−1.75	−1.74	−0.178	−1.64	−1.66	−0.158	−1.59	-0.142	-4.06
Manufacturing nonagricultural enterprise	0.026	0.20	0.21	0.069	0.47	0.50	0.017	0.13	0.036	0.80
Agricultural processing enterprise	−0.089	−0.81	−0.80	−0.006	−0.05	−0.05	−0.077	−0.71	−0.006	−0.15
Other production enterprise	−0.385	−1.97	−2.02	−0.284	−1.75	−1.67	−0.365	−1.87	−0.231	−3.45
Mixed enterprise	0.144	1.42	1.54	0.249	2.14	2.50	0.152	1.51	0.238	6.68
Community characteristics										
Agricultural seasonality	0.155	1.71	1.69	0.006	0.06	0.05	0.147	1.34	−0.018	−0.20
Enterprise density	0.001	1.83	1.43	0.001	1.89	1.33	0.001	1.67	0.001	2.48
Community population size#	0.034	0.92	0.82	0.047	1.20	0.90	0.040	0.90	0.067	1.84
Agricultural land per capita	0.025	0.48	0.47	−0.005	−0.08	−0.07	0.013	0.22	−0.046	−0.96
Illiteracy	−0.003	−1.23	−1.12	−0.002	−0.72	−0.66	−0.003	−1.02	−0.002	−0.83
Benchmark indicators										
Connectivity	0.912	2.95	2.91	0.805	2.43	2.01	0.929	2.43	0.930	2.93
Infrastructure services	−0.274	−1.02	−0.98	−0.194	−0.70	−0.52	−0.321	−1.00	−0.252	−0.98
Development services	0.041	0.41	0.44	−0.017	−0.15	−0.17	0.071	0.55	0.021	0.19
Governance	0.545	1.64	1.71	0.488	1.45	1.48	0.469	1.15	0.357	1.07
Human capital	−0.157	−0.27	−0.30	−0.609	−1.11	−0.90	−0.087	−0.12	−0.725	−1.32
Finance services	−0.072	−0.43	−0.45	−0.125	−0.74	−0.67	−0.071	−0.33	−0.191	−1.04
Intercept	5.124	8.44	8.28	5.084	7.30	7.13	5.089	7.82	4.970	12.67
Regression statistics										
Number of observations	814			814			814		814	
R^2	0.321			0.339			0.340		0.336	
sv[a]	0.982			0.980			0.959		0.292	
smu[a]	0.000			0.000			0.238		0.295	

Table I.3 Effect of Sampling Weights and Random Effect Specification on Enterprise Performance Regression Results (*continued*)

B1: Sri Lanka: Dependent variable is ln(sales)

	Ordinary Least Squares (unweighted)			Ordinary Least Squares (weighted)			Random Effects (unweighted)		Random Effects (weighted)	
	b	t(u)	t(a)	b	t(u)	t(a)	b	t	b	t
Factor and nonfactor inputs										
Total labor input#	−0.256	−1.20	−0.89	−0.294	−0.80	−0.79	−0.256	−1.20	−0.325	−1.09
Total capital input#	0.451	3.61	2.30	0.100	0.44	0.39	0.451	3.61	0.091	0.53
lnVLL (ln*ln)	0.091	3.88	2.54	0.080	2.01	1.86	0.091	3.88	0.081	2.50
lnVLC (lnL*lnC)	−0.097	−4.14	−2.80	−0.075	−1.93	−1.85	−0.097	−4.14	−0.074	−2.24
lnVCC (ln*ln)	0.033	4.64	2.77	0.047	2.37	2.45	0.033	4.64	0.047	4.13
Depreciation#	−0.053	−2.08	−1.87	−0.041	−1.01	−0.93	−0.053	−2.08	−0.039	−1.16
Nonfactor cost#	0.468	27.15	16.28	0.477	14.18	12.88	0.468	27.15	0.473	20.61
Other enterprise characteristics										
Age of enterprise#	0.069	2.53	2.26	0.115	3.15	3.01	0.069	2.53	0.116	3.18
Experience of manager#	0.098	3.15	2.63	0.082	2.17	2.02	0.098	3.15	0.075	1.79
Gender of manager	0.043	0.69	0.63	0.104	1.10	1.13	0.043	0.69	0.117	1.45
Education of manager	0.019	2.15	2.22	0.026	2.03	2.14	0.019	2.15	0.026	2.20
Registration	0.011	0.18	0.17	0.095	1.11	1.01	0.011	0.18	0.094	1.17
Industry dummies										
Services enterprise	0.277	3.47	2.92	0.239	2.30	2.48	0.277	3.47	0.215	2.08
Manufacturing, nonagricultural enterprise	−0.019	−0.28	−0.29	0.008	0.09	0.09	−0.019	−0.28	0.020	0.22
Agricultural processing enterprise	0.064	0.41	0.46	0.089	0.49	0.48	0.064	0.41	0.085	0.42
Other production enterprise	−0.177	−0.85	−1.31	−0.152	−0.51	−1.14	−0.177	−0.85	−0.097	−0.43
Mixed enterprise	0.030	0.30	0.28	−0.075	−0.70	−0.59	0.030	0.30	−0.122	−0.91
Community characteristics										
Agricultural seasonality	−0.064	−0.62	−0.75	−0.142	−1.14	−1.27	−0.064	−0.62	−0.131	−0.83
Enterprise density	−0.001	−0.63	−0.70	0.000	0.17	0.19	−0.001	−0.63	0.000	0.28
Community population size#	0.016	0.28	0.30	0.001	0.02	0.02	0.016	0.28	−0.006	−0.07
Agricultural land per capita	0.067	0.38	0.34	0.414	1.81	1.94	0.067	0.38	0.412	1.63
Illiteracy	0.012	0.05	0.05	−0.089	−0.31	−0.32	0.012	0.05	−0.064	−0.17
Main market in										
Neighboring communities	0.178	2.02	1.71	0.378	3.36	3.17	0.178	2.02	0.369	2.61
Commercial center	0.142	1.56	1.57	0.120	0.96	1.10	0.142	1.56	0.108	0.76
Nearest city	−0.006	−0.09	−0.09	0.038	0.42	0.48	−0.006	−0.09	0.036	0.31
Main community income from:										
Wages	0.001	0.01	0.02	0.162	2.03	2.17	0.001	0.01	0.166	1.66
Self-employment	0.091	0.97	1.29	0.167	1.44	1.92	0.091	0.97	0.156	1.03
Benchmark indicators										
Connectivity	0.214	0.93	0.87	−0.126	−0.42	−0.43	0.214	0.93	−0.094	−0.27
Infrastructure services	−0.664	−2.70	−2.45	−0.201	−0.71	−0.68	−0.664	−2.70	−0.198	−0.53
Development services	0.210	1.69	2.03	0.430	3.14	3.65	0.210	1.69	0.438	2.19
Governance	−0.767	−1.08	−1.14	0.248	0.26	0.30	−0.767	−1.08	0.190	0.18
Human capital	0.257	0.68	0.75	0.182	0.46	0.54	0.257	0.68	0.126	0.21
Finance services	0.280	1.25	1.47	0.356	1.32	1.47	0.280	1.25	0.349	1.03
Intercept	2.117	2.44	2.24	2.384	2.01	1.96	2.117	2.44	2.634	2.05
Regression statistics										
Number of observations	1018			1018			1018		1018	
R^2	0.762			0.750			0.769		0.750	
sv[a]	0.775			0.735			0.771		0.737	
smu[a]	0.000			0.000			0.000		0.210	

(*continued on the next page*)

Table I.3 Effect of Sampling Weights and Random Effect Specification on Enterprise Performance Regression Results (*continued*)

B2: Sri Lanka: Dependent variable is ln(net value added)

	Ordinary Least Squares (unweighted)			Ordinary Least Squares (weighted)			Random Effects (unweighted)		Random Effects (weighted)	
	b	t(u)	t(a)	b	t(u)	t(a)	b	t	b	t
Factor and nonfactor inputs										
Total labor input#	−0.401	−1.29	−1.29	−0.862	−1.90	−1.84	−0.459	−1.49	−0.946	−2.22
Total capital input#	0.868	4.82	5.55	0.932	4.06	4.10	0.873	4.88	0.953	3.96
lnVLL (ln*ln)	0.161	4.64	5.25	0.209	4.23	4.45	0.166	4.82	0.218	4.55
lnVLC (lnL*lnC)	−0.177	−5.00	−5.18	−0.207	−4.15	−4.35	−0.179	−5.09	−0.217	−4.33
lnVCC (ln*ln)	0.047	4.29	2.98	0.055	2.83	2.87	0.048	4.36	0.059	3.21
Other enterprise characteristics										
Age of enterprise#	0.054	1.33	1.27	0.088	1.99	1.87	0.061	1.51	0.095	1.83
Experience of manager#	0.065	1.41	1.34	0.068	1.30	1.17	0.060	1.30	0.060	1.01
Gender of manager	−0.010	−0.11	−0.09	−0.007	−0.06	−0.07	0.002	0.03	0.025	0.21
Education of manager	0.043	3.17	3.49	0.049	3.03	3.58	0.042	3.14	0.050	2.90
Registration	0.123	1.40	1.45	0.076	0.79	0.75	0.106	1.20	0.064	0.57
Industry dummies										
Services enterprise	0.334	3.04	2.93	0.177	1.45	1.34	0.323	2.97	0.148	1.07
Manufacturing, nonagricultural enterprise	0.005	0.05	0.05	−0.095	−0.83	−0.74	0.014	0.15	−0.085	−0.71
Agricultural processing enterprise	0.277	1.21	1.25	0.145	0.51	0.51	0.262	1.15	0.106	0.38
Other production enterprise	−0.005	−0.02	−0.01	−0.036	−0.10	−0.08	0.053	0.17	0.086	0.27
Mixed enterprise	0.186	1.30	1.06	−0.188	−0.91	−0.96	0.154	1.08	−0.207	−1.08
Community characteristics										
Agricultural seasonality	−0.232	−1.52	−1.66	−0.166	−0.99	−1.04	−0.260	−1.40	−0.166	−0.71
Enterprise density	0.001	0.39	0.34	0.002	1.15	1.04	0.001	0.38	0.002	0.99
Community population size#	0.035	0.43	0.46	0.041	0.47	0.43	0.032	0.32	0.036	0.28
Agricultural land per capita	0.179	0.64	0.53	0.515	1.53	1.62	0.154	0.47	0.491	1.26
Illiteracy	0.549	1.43	1.44	−0.001	0.00	0.00	0.650	1.40	0.113	0.20
Main market in:										
Neighboring communities	0.213	1.68	1.35	0.526	3.36	2.81	0.206	1.32	0.542	2.64
Commercial center	0.068	0.51	0.54	0.257	1.69	2.00	0.072	0.45	0.294	1.41
Nearest city	−0.107	−1.00	−1.01	0.078	0.62	0.72	−0.099	−0.75	0.097	0.56
Main community income from:										
Wages	−0.043	−0.44	−0.44	0.043	0.38	0.37	−0.046	−0.39	0.043	0.29
Self-employment	0.079	0.57	0.65	0.175	1.17	1.54	0.091	0.54	0.166	0.74
Benchmark indicators										
Connectivity	−0.054	−0.16	−0.14	−0.085	−0.22	−0.21	−0.039	−0.10	0.017	0.03
Infrastructure services	−0.504	−1.41	−1.45	−0.305	−0.68	−0.68	−0.499	−1.17	−0.337	−0.62
Development services	0.225	1.24	1.65	0.578	2.67	3.36	0.185	0.83	0.587	1.99
Governance	0.426	0.41	0.40	0.286	0.24	0.24	0.422	0.34	0.108	0.07
Human capital	0.622	1.14	1.01	0.318	0.57	0.55	0.564	0.85	0.319	0.36
Finance services	0.665	2.04	2.03	0.438	1.21	1.33	0.689	1.76	0.464	0.93
Intercept	0.960	0.77	0.73	2.219	1.45	1.24	1.190	0.86	2.506	1.36
Regression statistics										
Number of observations	841			841			841		841	
R^2	0.590			0.523			0.605		0.522	
sv[a]	1.029			0.955			1.003		0.950	
smu[a]	0.000			0.000			0.275		0.323	

Source: Authors' calculation from RIC data.

Note: a. "sv" is the standard deviation of the enterprise-specific disturbance. "smu" is the standard deviation of the community random effect.

Table I.4 Effect of Sampling Weights and Random Effect Specification on Entrepreneurship Selection Results

A: Nicaragua

	Standard Probit (unweighted)			Standard Probit (weighted)			Random Effect Probit (unweighted)		Random Effect Probit (weighted)	
	b	t(u)	t(a)	b	t(u)	t(a)	b	t	b	t
Household characteristics										
Number of male adults	0.095	2.25	2.37	0.112	2.24	2.32	0.096	2.26	0.112	2.24
Number of female adults	0.103	2.32	2.42	0.116	2.30	2.29	0.102	2.31	0.116	2.30
Average age	0.016	3.15	3.22	0.017	2.88	2.85	0.016	3.13	0.017	2.89
Human capital index	0.052	0.15	0.16	0.229	0.59	0.63	0.052	0.15	0.229	0.59
Head female	0.127	1.41	1.44	0.197	1.81	1.71	0.131	1.45	0.197	1.81
Head parents entrepreneur	0.067	0.82	0.89	0.077	0.80	0.83	0.069	0.84	0.077	0.80
ln(Other income)	−0.052	−3.03	−3.55	−0.045	−2.21	−2.67	−0.052	−3.02	−0.045	−2.21
ln(Remittances)	−0.087	−5.22	−5.02	−0.104	−5.71	−4.92	−0.088	−5.28	−0.104	−5.71
ln(Assets)	0.084	2.94	2.82	0.108	3.56	3.53	0.085	2.95	0.108	3.57
Benchmarks										
Connectivity	−0.137	−0.36	−0.55	−0.105	−0.23	−0.29	−0.123	−0.31	−0.105	−0.23
Infrastructure services	0.315	0.95	1.39	0.403	1.10	1.35	0.318	0.92	0.403	1.10
Business services	−0.126	−1.08	−1.62	−0.157	−1.16	−1.32	−0.132	−1.08	−0.157	−1.16
Governance	−0.299	−0.77	−1.18	0.029	0.07	0.08	−0.318	−0.78	0.029	0.06
Human capital	0.171	0.20	0.29	2.467	2.45	2.28	0.150	0.17	2.467	2.45
Finance services	−0.163	−0.88	−1.54	−0.186	−0.85	−1.21	−0.167	−0.85	−0.186	−0.85
Community variables										
ln(Community size)	−0.044	−1.25	−1.69	0.015	0.36	0.39	−0.044	−1.18	0.015	0.36
ln(Income per capita)	0.170	1.74	2.44	0.146	1.36	1.49	0.174	1.72	0.146	1.36
Agricultural seasonality	0.030	0.28	0.52	0.175	1.40	1.43	0.030	0.28	0.175	1.40
Enterprise openness	−0.032	−0.25	−0.42	0.390	2.48	2.87	−0.039	−0.30	0.390	2.48
ln(Male wage)	−0.095	−0.43	−0.70	−0.090	−0.35	−0.47	−0.091	−0.40	−0.090	−0.35
Intercept	−1.363	−2.44	−3.11	−4.763	−7.75	−8.60	−1.374	−2.37	−4.763	−7.74
Regression statistics										
Number of observations	1163			1163			1163		1163	
Log-likelihood	−689.48			−535.83			−692.24		−535.83	
smu[a]	0.000			0.000			0.106		0.000	

(*continued on the next page*)

Table I.4 Effect of Sampling Weights and Random Effect Specification on Entrepreneurship Selection Results (*continued*)

B: Sri Lanka

	Standard Probit (unweighted)			Standard Probit (weighted)			Random Effect Probit (unweighted)		Random Effect Probit (weighted)	
	b	t(u)	t(a)	b	t(u)	t(a)	b	t	b	t
Household characteristics										
Number of male adults	0.048	0.87	0.84	0.102	1.29	1.23	0.048	0.87	0.114	1.38
Number of female adults	0.072	1.29	1.19	−0.087	−0.99	−0.92	0.073	1.30	−0.087	−0.94
Average age	−0.012	−1.79	−1.57	0.004	0.31	0.27	−0.012	−1.78	0.004	0.32
Human capital index	0.338	0.85	0.76	0.185	0.32	0.31	0.336	0.84	0.227	0.38
Head female	−0.421	−3.28	−3.10	−0.280	−1.07	−1.00	−0.425	−3.29	−0.296	−1.08
Not Sinhalese	0.022	0.13	0.15	−0.103	−0.38	−0.49	0.019	0.11	−0.163	−0.51
Head parents entrepreneur	0.291	2.89	2.60	0.047	0.30	0.27	0.295	2.90	0.071	0.42
ln (Assets)	0.239	4.57	4.33	0.265	3.36	3.30	0.240	4.56	0.279	3.18
Benchmarks										
Connectivity	0.506	1.37	1.41	0.736	1.27	1.39	0.525	1.39	0.879	1.04
Infrastructure services	−0.559	−1.27	−1.29	−0.726	−1.17	−1.09	−0.584	−1.30	−0.924	−1.13
Business services	0.367	1.58	1.73	0.457	1.39	1.41	0.373	1.57	0.483	1.08
Governance	−0.135	−0.11	−0.13	3.210	1.77	2.16	−0.160	−0.13	3.358	1.40
Human capital	0.028	0.04	0.04	0.981	0.79	0.90	0.040	0.06	1.282	0.81
Finance services	−0.109	−0.29	−0.38	−0.858	−1.26	−1.38	−0.111	−0.28	−0.961	−1.02
Community variables										
ln(Community size)	0.106	1.09	1.22	0.228	1.37	1.44	0.109	1.10	0.277	1.25
ln(Income per capita)	0.022	0.25	0.26	0.052	0.38	0.37	0.023	0.25	0.058	0.31
Agricultural seasonality	−0.031	−0.17	−0.17	−0.087	−0.30	−0.31	−0.021	−0.11	−0.106	−0.27
Enterprise openness	−0.017	−0.23	−0.25	−0.047	−0.43	−0.43	−0.014	−0.19	−0.032	−0.20
ln(Male wage rate)	0.190	0.80	0.95	0.337	0.88	1.12	0.191	0.79	0.411	0.78
Intercept	−3.696	−2.06	−2.35	−8.998	−3.21	−3.19	−3.737	−2.04	−10.139	−2.62
Regression statistics										
Number of observations	849			849			849		849	
Log-likelihood	−542.20			−455.80			−542.66		−451.39	
smu[a]	0.000			0.000			0.102		0.328	

Source: Authors' calculation from RIC data.

Note: a. "smu" is the standard deviation of the community random effect.

Table I.5 Effect of Sampling Weights and Random Effect Specification on Selected EICO Regression Results

A1: Nicaragua: Dependent variable is "Interest rate of loan"

	Standard Ordered Probit (unweighted)			Standard Ordered Probit (weighted)			Random Effect Ordered Probit (unweighted)		Random Effect Ordered Probit (weighted)	
	b	t(u)	t(a)	b	t(u)	t(a)	b	t	b	t
Enterprise characteristics										
Age of enterprise[b]	0.022	0.50	0.47	0.010	0.20	0.17	−0.004	−0.08	−0.012	−0.21
Household-based enterprise	−0.519	−3.46	−3.53	−0.500	−3.42	−3.19	−0.499	−3.07	−0.489	−3.06
Services enterprise	−0.021	−0.19	−0.25	0.017	0.13	0.18	0.029	0.23	0.080	0.57
Manufacturing, nonagricultural enterprise	−0.043	−0.28	−0.29	−0.060	−0.37	−0.31	0.002	0.02	−0.019	−0.10
Agricultural processing enterprise	−0.183	−1.39	−1.73	−0.137	−0.94	−1.12	−0.175	−1.24	−0.137	−0.89
Other production enterprise	0.069	0.32	0.34	0.204	0.93	0.91	0.036	0.16	0.238	1.03
Mixed enterprise	0.070	0.59	0.60	0.045	0.33	0.33	0.101	0.79	0.090	0.63
Community characteristics										
Time to nearest city	−0.482	−2.65	−1.73	−0.242	−1.09	−0.76	−0.324	−1.20	−0.201	−0.46
Cost of transportation to city	2.190	3.30	2.13	1.204	1.52	1.07	2.292	2.15	1.424	0.84
Insurance service	0.111	1.12	0.58	0.170	1.56	0.88	0.088	0.40	0.202	0.33
Human capital benchmark	−0.379	−0.63	−0.36	−0.802	−1.20	−0.73	−0.591	−0.61	−0.995	−0.67
Number of banks	−0.081	−0.24	−0.13	−0.146	−0.43	−0.22	−0.554	−0.81	−0.654	−0.53
Number bank services	0.149	0.38	0.20	0.260	0.64	0.34	0.703	0.91	0.804	0.54
Intercept[b]							0.162	0.55	0.351	0.86
Regression statistics										
smu[a]	0.000			0.000			0.571	6.24	0.544	5.05
Cut point 1[b]	−0.264	−1.23	−0.93	−0.407	−1.81	−1.29				
Cut point 2[b]	0.153	0.71	0.52	0.010	0.04	0.03	0.476	12.33	0.473	11.72
Number of observations	893			893			893		893	
Log-likelihood	−842.897			−858.751			−812.135		−829.213	

(*continued on the next page*)

Table I.5 Effect of Sampling Weights and Random Effect Specification on Selected EICO Regression Results (*continued*)

A2: Nicaragua: Dependent variable is "Electricity"

	Standard Ordered Probit (unweighted)			Standard Ordered Probit (weighted)			Random Effect Ordered Probit (unweighted)		Random Effect Ordered Probit (weighted)	
	b	t(u)	t(a)	b	t(u)	t(a)	b	t	b	t
Enterprise characteristics										
Age of enterprise[b]	−0.054	−1.21	−1.36	−0.078	−1.63	−1.84	−0.026	−0.54	−0.033	−0.64
Household-based enterprise	−0.555	−3.75	−3.38	−0.474	−3.31	−2.72	−0.527	−3.32	−0.455	−3.05
Services enterprise	0.098	0.85	0.77	0.072	0.53	0.49	0.129	1.05	0.088	0.65
Manufacturing, nonagricultural enterprise	0.213	1.41	1.35	0.268	1.60	1.44	0.191	1.17	0.250	1.30
Agricultural processing enterprise	0.132	1.01	0.96	0.107	0.76	0.69	0.109	0.79	0.081	0.52
Other production enterprise	−0.113	−0.50	−0.42	−0.132	−0.59	−0.48	−0.108	−0.44	−0.146	−0.60
Mixed enterprise	0.377	3.23	2.93	0.450	3.29	2.94	0.393	3.18	0.438	2.97
Community characteristics										
Percent households with electricity	−0.925	−5.74	−3.68	−0.883	−4.93	−3.32	−1.227	−4.86	−1.129	−2.79
Information technology	−0.183	−2.03	−1.34	−0.071	−0.67	−0.44	0.012	0.08	0.167	0.57
Human capital benchmark	2.693	4.63	3.26	2.725	4.25	3.13	2.556	2.82	2.598	1.72
Intercept							0.471	1.63	0.329	0.86
Regression statistics										
smu[a]	0.000			0.000			0.535	7.48	0.559	5.92
Cut point 1[b]	−0.387	−1.74	−1.34	−0.299	−1.32	−1.00				
Cut point 2[b]	0.239	1.08	0.79	0.300	1.31	0.97	0.702	14.58	0.681	13.41
Number of observations	892			892			892		892	
Log-likelihood	−817.17			−818.71			−793.29		−792.34	

Table I.5 Effect of Sampling Weights and Random Effect Specification on Selected EICO Regression Results (*continued*)

B1: Sri Lanka: Dependent variable is "Interest rate of loan"

	Standard Ordered Probit (unweighted)			Standard Ordered Probit (weighted)			Random Effect Ordered Probit (unweighted)		Random Effect Ordered Probit (weighted)	
	b	t(u)	t(a)	b	t(u)	t(a)	b	t	b	t
Enterprise characteristics										
Age of enterprise[b]	−0.027	−0.76	−0.72	−0.041	−0.94	−0.85	−0.022	−0.58	0.010	0.19
Education of manager	−0.015	−1.28	−1.20	−0.001	−0.10	−0.08	−0.019	−1.49	−0.022	−1.32
Female manager	−0.287	−3.41	−3.09	−0.174	−1.77	−1.52	−0.370	−4.04	−0.219	−1.83
Sinhalese manager	0.349	3.39	2.26	0.733	4.80	3.70	0.295	2.29	0.623	2.42
Household-based enterprise	0.033	0.42	0.39	−0.198	−1.84	−1.91	0.029	0.34	−0.154	−1.17
Services enterprise	−0.228	−2.22	−2.38	−0.326	−2.44	−2.77	−0.266	−2.42	−0.378	−2.76
Manufacturing, nonagricultural enterprise	0.040	0.46	0.49	0.157	1.35	1.61	0.113	1.19	0.239	1.83
Agricultural processing enterprise	−0.220	−1.09	−1.08	−0.307	−1.41	−1.25	−0.335	−1.54	−0.353	−1.26
Other production enterprise	−0.038	−0.14	−0.22	0.148	0.65	0.81	−0.236	−0.74	−0.348	−1.16
Mixed enterprise	0.000	0.00	0.00	−0.100	−0.64	−0.63	0.151	1.12	0.129	0.71
Community characteristics										
Time to nearest city	0.059	0.40	0.24	−0.166	−0.94	−0.58	−0.024	−0.09	−0.383	−0.66
Cost of transportation to city	−0.208	−1.13	−0.74	0.039	0.18	0.12	−0.275	−0.86	−0.082	−0.11
Insurance service	0.124	1.74	1.07	0.206	2.24	1.48	0.069	0.54	0.239	0.73
Human capital benchmark	2.088	4.41	2.79	1.718	2.68	1.88	2.052	2.35	2.269	0.95
Number of banks	0.917	2.85	1.93	1.795	4.52	3.26	0.843	1.55	2.271	1.73
Number bank services	−0.369	−1.28	−0.80	−1.177	−3.27	−2.09	−0.228	−0.48	−1.379	−1.20
Intercept[b]							−0.799	−2.52	−1.553	−2.29
Regression statistics										
smu[a]	0.000			0.000			0.568	8.94	0.677	6.07
Cut point 1[b]	0.984	4.59	2.90	1.518	5.46	4.06				
Cut point 2[b]	1.155	5.37	3.40	1.669	5.97	4.46	0.197	8.86	0.181	6.45
Cut point 3[b]	1.508	6.97	4.39	2.018	7.16	5.43	0.602	16.04	0.596	11.98
Cut point 4[b]	2.407	10.89	6.63	3.129	10.44	7.57	1.609	25.03	1.869	19.86
Number of observations	1146			1145			1145		1145	
Log-likelihood	−1500.05			−1343.71			−1457.94		−1291.59	

(*continued on the next page*)

Table I.5 Effect of Sampling Weights and Random Effect Specification on Selected EICO Regression Results (*continued*)

B2: Sri Lanka: Dependent variable is "Electricity"

	Standard Ordered Probit (unweighted)			Standard Ordered Probit (weighted)			Random Effect Ordered Probit (unweighted)		Random Effect Ordered Probit (weighted)	
	b	t(u)	t(a)	b	t(u)	t(a)	b	t	b	t
Enterprise characteristics										
Age of enterprise[b]	−0.039	−1.08	−1.19	0.002	0.04	0.04	−0.027	−0.72	0.022	0.41
Education of manager	0.035	2.93	2.76	0.040	2.45	2.40	0.025	1.90	0.032	1.74
Female manager	−0.051	−0.61	−0.62	−0.181	−1.73	−1.51	−0.048	−0.54	−0.211	−1.81
Sinhalese manager	−0.110	−1.17	−0.90	−0.143	−1.10	−0.91	−0.095	−0.78	−0.102	−0.39
Household-based enterprise	−0.228	−2.95	−2.63	−0.350	−3.24	−3.01	−0.279	−3.27	−0.369	−2.92
Services enterprise	0.307	3.05	3.15	0.396	2.88	2.97	0.340	3.16	0.420	2.91
Manufactruing, nonagricultural enterprise	0.378	4.29	3.86	0.559	4.39	4.04	0.489	5.09	0.737	5.43
Agricultural processing enterprise	0.017	0.08	0.09	0.522	2.67	2.43	0.127	0.61	0.618	3.04
Other production enterprise	−0.547	−1.48	−1.25	−0.340	−0.88	−0.77	−0.780	−1.93	−0.549	−1.17
Mixed enterprise	0.334	2.86	2.47	0.347	2.12	1.82	0.327	2.50	0.345	1.87
Community characteristics										
Percent households with electricity	−0.812	−5.75	−3.24	−0.804	−3.98	−3.35	−0.939	−4.28	−0.816	−1.29
Information technology	0.048	0.47	0.50	0.163	1.36	1.06	0.112	0.58	0.203	0.59
Human capital benchmark	−0.384	−0.83	−0.58	−1.149	−1.71	−1.23	−0.661	−0.91	−1.788	−1.21
Intercept[b]							0.650	2.30	0.744	0.97
Regression statistics										
smu[a]	0.000			0.000			0.484	8.67	0.526	7.12
Cut point 1[b]	−0.424	−2.05	−1.81	−0.577	−1.87	−2.00				
Cut point 2[b]	−0.102	−0.49	−0.44	−0.262	−0.85	−0.90	0.358	12.39	0.356	7.97
Cut point 3[b]	0.252	1.22	1.07	0.153	0.50	0.52	0.749	18.31	0.824	12.19
Cut point 4[b]	0.853	4.10	3.68	0.786	2.50	2.85	1.400	24.08	1.521	16.92
Number of observations	1152			1152			1152		1152	
Log-likelihood	−1540.68			−1518.46			−1510.81		−1483.24	

Source: Authors' calculation from RIC data.

Note: a. "smu" is the standard deviation of the community random effect.
b. The standard ordered probit and random effect ordered probit models use a different standardization, which is immaterial to the estimated values of slope parameters.

APPENDIXES TO ANNEX I. STATA PROCEDURES

Appendix A1. OLS Estimation

Let the dependent variable of the empirical model be called Y. Let X1 and X2 be the explanatory variables at the individual level and let Z1 and Z2 be the explanatory variables at the community level; of course, there may be more variables at each level, or there may be just one. Let WGT be the weight of an observation and let GRP indicate the community. Stata estimates this model with the statement:

```
regress Y X1 X2 Z1 Z2 [pweight=WGT],cluster(GRP)
```

Appendix A2. Fixed Effects Estimation

In the following program statements, the weight WGT is standardized such that the sum of WGT over the sample equals the number of observations in the sample. This standardized weight is called wgt; note that Stata is case-sensitive and that WGT and wgt are distinct variables.

```
quietly su WGT,meanonly
g wgt=WGT/r(mean)
```

The fixed effects estimates of β are found with the following few statements.

```
g swgt=sqrt(wgt)
egen wgt_g=total(wgt),by(GRP)
foreach x of varlist Y X1 X2 {
  egen m`x'=total(`x'*wgt),by(GRP)
  quietly replace m`x'=m`x'/wgt_g
  g d`x'=swgt*(`x' - m`x')
}
reg dY dX1 dX2,noconst
```

The estimate of the variance of the individual-specific disturbance ν, namely σ^2_ν, is found as follows. The result is stored in a scalar named s2v.

```
predict vtilhat,resid
g vtilhat2=vtilhat*vtilhat
g wgt2=wgt*wgt
g wgt2stan=total(wgt*wgt/wgt_g)
quietly su vtilhat2
scalar den=r(sum)
quietly su wgt
scalar den1=r(sum)
quietly su wgt2stan
scalar den2=r(sum)
scalar s2v=num/(den1-den2)
display "Estimated sigv2 = " s2v;
```

The fixed effects $\hat{\mu}$ are found with the following statements. They appear as a new variable Muhat in the database, having the same value for all observations in a given community.

```
matrix b=e(b)
g Muhat = mY - b[1,1]*mX1 - b[1,2]*mX2
```

In order to estimate the effect of Z, one must first estimate σ^2_μ:

```
matrix V=e(V)
foreach x of varlist Y X1 X2 {
  quietly replace d`x'=d`x'*swgt
```

```
}
matrix accum XWXstar = dX1 dX2,noconstant
egen totwgt2j=total(wgt*wgt),by(GRP);
g wgtterm = (wgt*wgt - (wgt*totwgt2j/totwgtj))/totwgtj;
foreach x of varlist X1 X2 {
  quietly replace d`x'=d`x'/swgt
  quietly replace d`x'=d`x'*sqrt(s2v)*wgtterm
  egen t`x'=total(d`x'),by(GRP)
}
keep if GRP ~= GRP[_n-1]
mkmat mX1 mX2,matrix(Xbar)
mkmat tX1 tX2,matrix(Xt)
g varmu=s2v*totwgt2j/(totwgtj*totwgtj)
mkmat varmu,matrix(varmuv)
matrix varmu=diag(varmuv)-Xbar*invsym(XWXstar)*Xt' - /*
    */  Xt*invsym(XWXstar)*Xbar'+Xbar*V*Xbar'
scalar J_com = rowsof(varmu)
g iota=1
mkmat Z1 Z2 iota,matrix(Z)
mkmat Muhat,matrix(muhat)
matrix lambda=muhat-Z*syminv(Z'*Z)*(Z'*muhat)
matrix sll=lambda'*lambda
scalar tvarmu=trace(varmu)
scalar sigmu2=(sll[1,1]-tvarmu)/J_com
display "Estimated sigmu2 = " sigmu2
```

The parameter γ that measures the effect of Z on y is now quickly found:

```
matrix SIG = sigmu2*I(J_occ) + varmu
matrix invSIG=syminv(SIG)
matrix bgls = syminv(Z'*invSIG*Z)*(Z'*invSIG*muhat)
matrix vgls = syminv(Z'*invSIG*Z)
matrix stdgls=vecdiag(vgls)'
matrix tgls = bgls
local j=1
while `j' <= rowsof(stdgls) {
  matrix stdgls[`j',1]=sqrt(stdgls[`j',1])
  matrix tgls[`j',1]=tgls[`j',1]/stdgls[`j',1]
  local j=`j'+1
}
display "Impact of Z:"
matrix result=(bgls,stdgls,tgls)
matrix colnames result = bAMEMIYA stdev tstat
matrix list result
```

Appendix A3. Random Effects Estimation

In the following program code, the weight WGT is standardized such that the sum of WGT over the sample equals the number of observations in the sample. This standardized weight is called wgt; note that Stata is case-sensitive and that WGT and wgt are distinct variables.

```
quietly su WGT,meanonly
g wgt=WGT/r(mean)
```

The first task is to compute the variance σ_ν^2 of the individual idiosyncratic disturbance ν_{ji}. This estimator derives from a fixed effects regression and so is similar to what was provided in Appendix A.2, with one improvement that helps to get an estimator of σ_μ^2 more likely to be positive. The first step is to compute a matrix labeled XQWQX, of which the trace is needed later on. During the process of the computation, all but the first observation in a group (or community) are removed. Therefore, the existing dataset is stored and subsequently retrieved when the computation is finished.

```
quietly save temptrash,replace
foreach x of varlist X1 X2 {
  g w05`x'=`x'*sqrt(`wgt')
  g w`x'=`x'*`wgt'
  egen tw`x'=total(w`x'),by(`GRP')
  egen tww`x'=total(w`x'*`wgt'),by(`GRP')
  local w05x = "`w05x' w05`x'"
  local wx = "`wx' w`x'"
  local twx = "`twx' tw`x'"
  local twwx = "`twwx' tww`x'"
}
matrix accum XW1X = w05X1 w05X2,noconst
matrix accum XW2X = wX1 wX2,noconst
quietly drop if `GRP' == `GRP'[_n-1]
mkmat twX1 twX2,matrix (ZWX)
mkmat twwX1 twwX2,matrix(ZW2X)
mkmat totwgt_g,matrix(ZWZ)
matrix ZWZ = diag(ZWZ)
matrix iZWZ = invsym(ZWZ)
mkmat totwgt2_g,matrix(ZW2Z)
matrix ZW2Z = diag(ZW2Z)
matrix XQWQX = XW2X - (ZW2X')*iZWZ*ZWX - (ZWX')*iZWZ*ZW2X + /*
  */ (ZWX')*iZWZ*ZW2Z*iZWZ*ZWX

use temptrash,clear
```

At this point, the program proceeds with a regression on variables in deviation that generates residuals from which the estimate of σ_ν^2 is computed.

```
g swgt=sqrt(wgt)
egen wgt_g=total(wgt),by(GRP)
foreach x of varlist Y X1 X2 {
  egen m`x'=total(`x'*wgt),by(GRP)
  quietly replace m`x'=m`x'/wgt_g
  g d`x'=swgt*(`x' - m`x')
}
reg dY dX1 dX2,noconst
predict vtilhat,resid
g vtilhat2=vtilhat*vtilhat
g wgt2=wgt*wgt
g wgt2stan=total(wgt*wgt/wgt_g)
quietly su vtilhat2
scalar den=r(sum)
quietly su wgt
scalar den1=r(sum)
quietly su wgt2stan
```

```
scalar den2=r(sum)
scalar s2v=num/(den1-den2)
display "Estimated sigv2 = " s2v;
```

The second task is to compute the variance σ_{μ}^2 of the community disturbance μ_j. In this program, we first run an OLS regression to retrieve the residuals that are the basis of the estimator.

```
regress Y X1 X2 Z1 Z2 [pw=WGT],cluster(GRP)
predict uhat,resid
egen sumwgtu=total(wgt*uhat),by(GRP)
g sumwgtu2=sumwgtu*sumwgtu
quietly su sumwgtu2,meanonly
scalar num1=r(mean)
g wgtu2=wgt*wgt*uyr*uyr
quietly su wgtu2,meanonly
scalar num2=r(mean)
egen sumwgt=total(wgt),by(GRP)
g sumwgt2=sumwgt*sumwgt
quietly su sumwgt2,meanonly
scalar den1=r(mean)
g wgt2=wgt*wgt
quietly su wgt2,meanonly
scalar den2=r(mean)
scalar sigmu2 = (num1-num2)/(den1-den2)
display "Estimated sigmu2 = " sigmu2
```

The third step in the program is to compute the estimated parameters δ.

```
by GRP: g Nj = _N
g tau2j = sumwgt*sigmu2 + s2v
foreach x of varlist Y X1 X2 Z1 Z2 {
  egen tm`x'=total(`x'*wgt),by(GRP)
  g m`x'=tm`x'/sumwgt
  g s`x'=(`x' - (1-sqrt(s2v/tau2j)) * m`x')/sqrt(s2v)
  quietly replace tm`x'=0 if GRP == GRP[_n-1]
}
g intercpt = (1 - (1-sqrt(s2v/tau2j)) * 1)/sqrt(s2v)
reg  sY sX1 sX2 sZ1 sZ2 intercpt [pw=WGT],noconst
drop tmlnwage - intercpt

**Now do by hand, in order to get correct standard errors**
display "** estimate deltaGLS **"
g twgt1=sqrt(wgt/s2v)
g twgt2=sqrt(sigmu2/(tau2j*s2v))
g intercpt = 1
foreach x of varlist Y X1 X2 Z1 Z2 intercpt {
  g s`x'=`x'*twgt1
  egen tm`x'=total(`x'*wgt),by(GRP)
  quietly replace tm`x'=0 if GRP == GRP[_n-1]
  quietly replace tm`x'=tm`x'*twgt2
}
matrix accum tbatb1 = sY sX1 sX2 sZ1 sZ2 sintercpt,noconst
```

```
matrix accum tbatb2 = tmY tmX1 tmX2 tmZ1 tmZ2 tmintercpt,noconst
scalar tK=rowsof(tbatb1)
matrix tat=tbatb1[2..tK,2..tK] - tbatb2[2..tK,2..tK]
matrix tay=tbatb1[2..tK,1] - tbatb2[2..tK,1]
matrix itat = invsym(tat)
matrix deltaGLS = itat*tay
drop twgt1 twgt2 slnwage - tmintercpt
```

The fourth step in the program secures the estimated covariance matrix of the estimated parameters δ.

```
display "** estimate Var(deltaGLS) **"
g twgt1=wgt/sqrt(s2v)
g twgt2=sqrt(sigmu2/(tau2j*s2v))
foreach x of varlist X1 X2 Z1 Z2 intercpt {
  g s`x'=`x'*twgt1
  egen t1`x'=total(`x'*wgt),by(GRP)
  quietly replace t1`x'=0 if GRP == GRP[_n-1]
  quietly replace t1`x'=t1`x'*twgt2
  egen t2`x'=total(`x'*wgt*wgt),by(GRP)
  quietly replace t2`x'=0 if GRP == GRP[_n-1]
  quietly replace t2`x'=t2`x'*twgt2
}
matrix accum tbatb1 = sX1 sX2 sZ1 sZ2 sintercpt,noconst
matrix accum tbatb2 = t1X1 t1X2 t1Z1 t1Z2 t1intercpt     /*
       */             t2X1 t2X2 t2Z1 t2Z2 t2intercpt,noconst
scalar tK=rowsof(tbatb1)
matrix tavat=tbatb1 - tbatb2[(tK+1)..(2*tK),1..tK]       /*
       */           - tbatb2[1..tK,(tK+1)..(2*tK)]
egen twgt3a=total(wgt*wgt),by(GRP)
g twgt3=sqrt(sigmu2 + (sigmu2*sigmu2*twgt3a/s2v))/tau2j
foreach x of varlist X1 X2 Z1 Z2 intercpt {
  quietly replace t1`x'=t1`x'*twgt3/twgt2
}
matrix accum tbatb3 = t1X1 t1X2 t1Z1 t1Z2 t1intercpt,noconst
matrix tavat = tavat + tbatb3
matrix VardeltaGLS = itat*tavat*itat
```

Finally, report the results: the estimated parameters, their standard errors, and the t-statistics.

```
matrix stdev=vecdiag(VardeltaGLS)'
matrix tstat=J(tK,1,0)
local j=1
while `j' <= tK {
  matrix stdev[`j',1]=sqrt(stdev[`j',1])
  matrix tstat[`j',1]=deltaGLS[`j',1]/stdev[`j',1]
  local j=`j'+1
}
matrix result=(deltaGLS,stdev,tstat)
matrix colnames result = deltaGLS stdev tstat
matrix list result
drop sumwgtu - twgt3
```

Appendix A4. Weighted Random Effect Probit Estimation

The weighted random effect probit model is estimated by maximum likelihood. In Stata, this means the researcher must provide a program called by the Stata's built-in ml procedure. In the following Stata code, this program is called wrep. This ml procedure requires some data set-up, which is provided below.

The program wrep defines a temporary variable `mu' which is initialized at 0 but subsequently filled with random draws of a standard normal distribution. These draws must be reused for every iteration of the maximum likelihood search routine. For that reason, the random draws are inserted into a matrix ZZ prior to the start of the maximum likelihood estimation. Furthermore, the program uses an antithetical simulation method, in which random draws of `mu' are reused in a different form to raise the precision of the simulated probability: here, the sign of the random draws are switched. At the end of the program, the variables gradient1 and gradient2 are computed for use in computations when the estimation is completed.

```
*** Define the weighted RE probit likelihood function ***;

capture program drop wrep;
program define wrep;
version 9.0;

  args todo b lnf g;
  tempvar xb lntheta theta wgtmean wgt mu xbmu p;
  tempvar lnLgi lnLg elnLg Lg lnfgi arg g1 g2;
  mleval `xb'      = `b',eq(1);
  mleval `theta' = `b',eq(2);

  /*****************************************************
  *  ML_y1 is the categorical variable, with two outcomes
  *  ML_y2 is the group indicator
  *  ML_y3 is the weight of each individual
  *  lntheta is the log-stdev of mu
  *  theta will be the standard deviation of mu
  *****************************************************/

  quietly su `theta'; if (r(min) <= 0) {; scalar `lnf' = .; exit; };
  egen `wgtmean'=mean($ML_y3),by($ML_y2);
  g `wgt'=$ML_y3/`wgtmean';

  quietly g double `xbmu'=0;
  quietly g double `mu'=0;
  quietly g double `p'=0;
  quietly g double `Lg'=0;
  quietly g double `lnLgi'=0;
  quietly g double `elnLg'=0;
  quietly g double `arg'=0;
  quietly g double `g1'=0;
  quietly g double `g2'=0;
  local jr=1;
  local JR = colsof(ZZ);
  local Ngroups = rowsof(ZZ);
  while `jr' <= `JR' {;
    *display "jr = `jr' ";
    local jg=1;
```

```
    quietly replace `mu'=0;
    while `jg' <= `Ngroups' {;
      quietly replace `mu' = `mu' + ($ML_y2 == `jg')*ZZ[`jg',`jr'];
      local jg=`jg'+1;
    };
    **Work the regular draw**;
    quietly replace `xbmu' = `xb'+`theta'*`mu';
    quietly replace `p' = normal(`xbmu');
    quietly replace `p' = 1-`p' if $ML_y1 == 0;
    quietly su `p'; if (r(min) == 0) {; scalar `lnf' = .; exit; };
    quietly replace `lnLgi' = `wgt'*ln(`p');
    quietly replace `lnLgi' = -999 if `lnLgi' == .;
    quietly egen `lnLg' = total(`lnLgi'),by($ML_y2);
    quietly replace `elnLg' = exp(`lnLg');
    quietly su `elnLg'; if (r(min) == 0) {; scalar `lnf' = .; exit; };
    quietly replace `Lg' = `Lg' + `elnLg';
    quietly replace `arg' = (2*$ML_y1 - 1) * normalden(`xbmu') / `p';
    quietly replace `g1' = `g1' + `elnLg'*`arg';
    quietly replace `g2' = `g2' + `elnLg'*`arg'*`mu';
    drop `lnLg';
    **Work the antithetical draw**;
    quietly replace `xbmu' = `xb'-`theta'*`mu';
    quietly replace `p' = normal(`xbmu');
    quietly replace `p' = 1-`p' if $ML_y1 == 0;
    quietly su `p'; if (r(min) == 0) {; scalar `lnf' = .; exit; };
    quietly replace `lnLgi' = `wgt'*ln(`p');
    quietly replace `lnLgi' = -999 if `lnLgi' == .;
    quietly egen `lnLg' = total(`lnLgi'),by($ML_y2);
    quietly replace `elnLg' = exp(`lnLg');
    quietly su `elnLg'; if (r(min) == 0) {; scalar `lnf' = .; exit; };
    quietly replace `Lg' = `Lg' + `elnLg';
    quietly replace `arg' = (2*$ML_y1 - 1) * normalden(`xbmu') / `p';
    quietly replace `g1' = `g1' + `elnLg'*`arg';
    quietly replace `g2' = `g2' - `elnLg'*`arg'*`mu';
    drop `lnLg';
    local jr=`jr'+1;
  };
  quietly replace `Lg' = `Lg'/(2*`JR');

  quietly g double `lnfgi' = `wgtmean'*ln(`Lg') * ($ML_y2 ~= $ML_y2[_n-1]);
  mlsum `lnf' = `lnfgi';

  if (`todo'==0 | `lnf' >= .) exit;
  tempname d1 d2;
  quietly replace gradient1 = $ML_y3*`g1'/(`Lg'*(2*`JR'));
  quietly replace gradient2 = $ML_y3*`g2'/(`Lg'*(2*`JR'));
  mlvecsum `lnf' `d1' = gradient1, eq(1);
  mlvecsum `lnf' `d2' = gradient2, eq(2);
  matrix `g' = (`d1',`d2');

end ;
```

The following program wreprobit is a driver for weighted random effect probit estimation. After some initializations, it creates sequential group (which, in the case of RIC data, are community) numbers. It permits the user to specify a few comparison models such as ordinary probit, weighted probit (without random effects), and unweighted random effect probit. The latter is used to generate starting values for the ML estimation of the weighted random effect probit model.

The next steps are: generating a matrix ZZ of standard normal random draws, executing the ml procedure with a specific reference to the wrep program provided above, and computing corrected standard errors and t-statistics. Relative to the discussion in Sections 3.3 and 3.4, Stata's matrix V stands for the variance matrix V_1; Stata's matrix G stands for V_2^{-1}; and Stata's matrix Varb stands for V_3. The program only reports standard errors based on V_3, but alternatives are easily inserted.

At the very bottom of the program, commented code briefly explains the manner in which wreprobit should be called by the user's do-file.

```
*** Define weighted RE probit as a regression command ***;

capture program drop wreprobit;
program define wreprobit;
  version 9.0;

  syntax varlist(min=3)[,trace simnum(integer 25) compareUOP compareWOP
compareURP trialrun];
  gettoken y rest : 0 /*,parse(" ,")*/;
  gettoken g rest : rest ,parse(" ");
  gettoken wgt rest : rest ,parse(" ");
  gettoken xzvar rest : rest ,parse(",");
  display "Dependent variable is      `y'";
  display "Group variable is          `g'";
  display "Weight variable is         `wgt'";
  display "Explanatory variables are `xzvar'";
  display "Number of draws used by simulator = `simnum' (used twice: +/-)";
  display "  ";

** rescale weight to an average of 1 **;

  quietly su `wgt',meanonly;
  g `wgt'_stan=`wgt'/r(mean);
  label variable `wgt'_stan "Scaled weight at individual level";
   display "Variation in weights rescaled to an average of 1, at individual
level";
  su `wgt'_stan,detail;
  scalar stdvwgt = r(sd);

** Determine the number of groups **;
  sort `g';
  quietly g `g'_nummer=1 if _n==1;
  quietly replace `g'_nummer=`g'_nummer[_n-1] + (`g' ~= `g'[_n-1]) if _n > 1;
  label variable `g'_nummer "Numeric group number";
  di "  ";
  quietly su `g'_nummer;
  local Ngroups = r(max);
  display "Number of groups =  `Ngroups'";
  tempvar gsize;
```

```
  sort `g'_nummer;
  by `g'_nummer: g `gsize' = _N;
  quietly su `gsize';
  display "Smallest group size = " r(min);
  display "Average group size =  " r(mean);
  display "Largest group size =  " r(max);
  di "  ";
  *di "Tabulation of group sizes for each group";
  *tab `g'_nummer;
  di "  ";

** Estimate weighted ordinary probit as a trial run for the specification **;
  if "`trialrun'" ~= "" {;
    display "  ";
    display " Estimate weighted ordinary probit as a trial run for the specifi-
cation ";
    probit `y' `xzvar' [pw=`wgt'_stan];
  };
  else {;  /*begin of nontrial segment*/

** Estimate unweighted models for comparison, as requested through options **;
  if "`compareUOP'" ~= "" | "`trialrun'" ~= "" {;
    display "  ";
    display ">> Unweighted Ordinary Probit Model, for comparison only <<";
    probit `y' `xzvar';
    display ">> Note: without applying weights or random effects, estimates may
be biased and inconsistent";
    display ">> End of comparison <<";
    display "  ";
  };

** Estimate weighted models for comparison, as requested through options **;
  if "`compareWOP'" ~= "" {;
    display ">> Weighted Ordinary Probit Model, for comparison only <<";
    probit `y' `xzvar' [pw=`wgt'_stan];
    display ">> End of comparison <<";
    display "  ";
  };
  else {;
    quietly probit `y' `xzvar' [pw=`wgt'_stan];
  };
  matrix b=get(_b)';

** Estimate weighted models for comparison, as requested through options **;
  if "`compareURP'" ~= "" {;
   display ">> Unweighted Random Effects Probit Model, for comparison only <<";
    xtprobit `y' `xzvar',re i(`g'_nummer);
    display ">> End of comparison <<";
    display "  ";
  };
  else {;
    quietly xtprobit `y' `xzvar',re i(`g'_nummer);
```

```
  };
  matrix bre=get(_b)';
  scalar thet = bre[rowsof(bre),1;
  if thet < 0.05 {;
    matrix theta=0.05;    };
  else {;
    matrix theta=thet;
  };
  matrix bb=(b*0.95)\theta;
  matrix rownames bb = `xzvar' intercept stdevmu;
  matrix list bb;

  preserve;

** Compute a matrix of random draws for simulation **;
  quietly drop if _n > `Ngroups';
  g id=_n;
  set seed 19901;
  matrix ZZ=J(`Ngroups',1,0);
  local jr=1;
  while `jr' <= `simnum' {;
    g z=invnorm(uniform());
    mkmat z;
    matrix ZZ=ZZ,z;
    drop z;
    local jr=`jr'+1;
  };
  matrix ZZ=ZZ[1...,2...];
  *matrix list ZZ;
  restore;

  if stdvwgt > 0 {;
  ** Start ML estimation of random effect probit model **;
    di "** Start ML estimation of weighted random effect probit model **";
    g gradient1=0;
    g gradient2=0;
    ml model d1 wrep
       (eq1:`y' `g'_nummer `wgt'_stan=`xzvar')
       (eq2:);
    ml init bb,copy;
    ml search;
    if "`trace'" ~= "" {;
      ml max,trace grad;
    };
    else {;
      ml max;
    };
** organize the outerproduct of the gradient for computation of correct
   stdev of b **;
    local ggvar = "";
    local j = 0;
    foreach aa of varlist `xzvar' {;
```

```
      local j = `j' + 1;
      g ggg`j' = gradient1*`aa';
      local ggvar = "`ggvar' ggg`j'";
      display "Adding to local ggvar: `j' equals `aa'";
    };
    g gggintercept = gradient1;
    g gggtheta = gradient2;
    local ggvar = "`ggvar' gggint gggtheta";
    display "Complete local ggvar equals `ggvar'";
    matrix accum G = `ggvar',noconst;
    matrix b_est = e(b)';
    matrix V = e(V);
    matrix list G;
    matrix list V;
    matrix VARb = V * G * V;
    matrix SDb = cholesky(diag(vecdiag(VARb)));
    matrix tstatb = syminv(SDb)*b_est;
    matrix SDb = vecdiag(SDb)';
    matrix result = b_est,SDb,tstatb;
    matrix colnames result = b stdev tstat;
    matrix list result;
    drop ggg*;
  };
  else {;
    di "** Weights are constant across observations";
    di "** Use random effect probit without weights";
    xtprobit `y' `xzvar',re i(`g'_nummer);
  };
  drop `g'_nummer `wgt'_stan gradient*;
  };  /*end of nontrial segment*/
end;

/*** Information on wreprobit routine ************************************************
*   Depvar: y
*   Group indicator: group
*   Weight: weight
*   Explanatory variables, individual level: x1a x1b x1c
*   Explanatory variables, group level: x2a x2b
*
*   The call to the weighted fixed effect regression command must first list
*     the depvar groupvar and weightvar, and then the explanatory variables.
*   The explanatory variables may be in any order, mixing individual level and
*     group level variables in random order. The program will sort them into
*     a set of individual level variables and a set of group level variables.
*   The option 'simnum' must specify an integer, indicating the number
*     of random draws that the simulated random probit estimation uses. A higher
*     number yields greater precision but lengthens the computation time more or
*     less proportionately.  Default is 25.
*   The option 'trace' asks for more detail during the iterative search.
*   The option 'compareUOP' will run a unweighted ordinary probit model for comparison.
*   The option 'compareWOP' will run a weighted ordinary probit model for comparison.
*   The option 'compareURP' will run a unweighted random effects probit model for comparison.
*   The option 'trialrun' will only estimate a weighted ordinary probit model to
*     test the specification of a preliminary model.
```

```
*
*   The output will appear to yield estimates of two equations.  The first equation
*   represents the parameters of the explanatory variables (and intercept).  The
*   second equation shows the estimate of the standard deviation of the
    random effect
*   (under the label of '_cons').
*   The code of the program is simply modified to estimate the natural log of the
*   standard deviation of the random effect (log-stdev) rather than the standard
*   deviation itself.  An estimate of the standard deviation (stdevmu)
*   is obtained by taking the antilog of the log-stdev.  The standard
*   error of stdevmu is found by multiplying the standard error of the
*   estimate of the log-stdevmu with the estimated value of stdevmu.
*
*   Examples of how to call the program:
*
* wreprobit y group weight x1a x2a,trace simnum(6);
* wreprobit y group weight x1a x2a x1b x1c x2b,compareUOP compareWOP;
* wreprobit y group weight x1a x2a x1b x1c x2b,trialrun;
*
*************************************************************************/;
```

It should be noted that wreprobit does not tolerate missing values. Observations with missing values must be removed before calling wreprobit. Moreover, whereas many Stata regression routines will automatically remove variables that are perfectly collinear, such incidences will generate a fatal error in wreprobit. For this reason, the user ought to inspect his model specification with the trialrun option, which will only estimate an ordinary probit model. If a variable is omitted in this estimation, the user must examine the specification of the regression model more carefully.

Appendix A5. Weighted Random Effect Ordered Probit Estimation

The estimation of the random effect ordered probit model proceeds along a line similar to that of the random effect probit model. The program wreop defines the likelihood function for the random effect ordered probit model for use by Stata's ml routine. The program wreoprobit is the driver program that sets up the estimation and delivers the estimates. For information on how to use this program in a Stata do-file, see the instructions at the end of the text.

It should be emphasized that this version of the wreoprobit routine estimates an ordered probit where the dependent variable takes on five responses. The program does not automatically adjust to the number of distinct responses in the dependent variable. If the number of responses is four instead of five, for example, the program must be adjusted accordingly by omitting the estimation of `alpha4` and adjusting comparisons of `$ML_y1` with potential response values.

```
*** Define the weighted RE ordered probit likelihood function ***;

capture program drop wreop;
program define wreop;
version 9.0;

  args todo b lnf g;
  tempvar xb lntheta theta alpha2 alpha3 alpha4 wgtmean wgt mu xbmu p;
  tempvar lnLgi lnLg elnLg Lg lnfgi arghi arglo g1 g2 g3 g4 g5;
  mleval `xb'      = `b',eq(1);
  mleval `theta'   = `b',eq(2);
  mleval `alpha2'  = `b',eq(3);
  mleval `alpha3'  = `b',eq(4);
```

```
  mleval `alpha4'  = `b',eq(5);

  /*****************************************************
  *  ML_y1 is the categorical variable, with two outcomes
  *  ML_y2 is the group indicator
  *  ML_y3 is the weight of each individual
  *  lntheta is the log-stdev of mu
  *  theta will be the standard deviation of mu
  *  alpha_j are cutpoints
  *    The program is designed for number of alternatives equal to 5,
  *       ranging from 0,...,4
  *    By constraint, alpha_0 = -inf
  *                   alpha_1 = 0
  *                   alpha_2 .. alpha_4 are estimated parameters
  *                   alpha_5 = +inf
  *    If y == j, then alpha_j < ystar <= alpha_{j+1}
  *****************************************************/

** Check consistency of parameters provided **;
  quietly su `theta';
    if (r(min) <= 0) {; scalar `lnf' = .; exit; };
  quietly su `alpha2'; scalar m2=r(min);
    if m2 <= 0 {; scalar `lnf' = .; exit; };
  quietly su `alpha3'; scalar m3=r(min);
    if m3 <= m2 {; scalar `lnf' = .; exit; };
  quietly su `alpha4'; scalar m4=r(min);
    if m4 <= m3 {; scalar `lnf' = .; exit; };
  egen `wgtmean'=mean($ML_y3),by($ML_y2);
  g `wgt'=$ML_y3/`wgtmean';

  quietly g double `xbmu'=0;
  quietly g double `mu'=0;
  quietly g double `arglo'=0;
  quietly g double `arghi'=0;
  quietly g double `p'=0;
  quietly g double `Lg'=0;
  quietly g double `lnLgi'=0;
  quietly g double `elnLg'=0;
  quietly g double `g1'=0;
  quietly g double `g2'=0;
  quietly g double `g3'=0;
  quietly g double `g4'=0;
  quietly g double `g5'=0;

  local jr=1;
  local JR = colsof(ZZ);
  local Ngroups = rowsof(ZZ);
  while `jr' <= `JR' {;
    *display "jr = `jr' ";
    local jg=1;
    quietly replace `mu'=0;
    while `jg' <= `Ngroups' {;
```

```
      quietly replace `mu' = `mu' + ($ML_y2 == `jg')*ZZ[`jg',`jr'];
      local jg=`jg'+1;
    };
    **Work the regular draw**;
    quietly replace `xbmu' = `xb'+`theta'*`mu';
    quietly replace  `arglo' = -`xbmu' + ($ML_y1 == 2)*`alpha2' +
($ML_y1 == 3)*`alpha3' + ($ML_y1 == 4)*`alpha4';
    quietly replace  `arghi' = -`xbmu' + ($ML_y1 == 1)*`alpha2' +
($ML_y1 == 2)*`alpha3' + ($ML_y1 == 3)*`alpha4';
    quietly replace  `p' = ($ML_y1 == 4) + ($ML_y1 < 4) * normal(`arghi') -
($ML_y1 > 0) * normal(`arglo');
    quietly su `p'; if (r(min) <= 0) {; scalar `lnf' = .; exit; };
    quietly replace `lnLgi' = `wgt'*ln(`p');
    quietly replace `lnLgi' = -999 if `lnLgi' == .;
    quietly egen `lnLg' = total(`lnLgi'),by($ML_y2);
    quietly replace `elnLg' = exp(`lnLg');
    quietly su `elnLg'; if (r(min) == 0) {; scalar `lnf' = .; exit; };
    quietly replace `Lg' = `Lg' + `elnLg';
    if (`todo'~= 0) {;
      quietly replace `arglo' = (($ML_y1 > 0) * normalden(`arglo')) / `p';
      quietly replace `arghi' = (($ML_y1 < 4) * normalden(`arghi')) / `p';
      quietly replace `g1' = `g1' - `elnLg'*(($ML_y1 < 4)*`arghi' -
($ML_y1 > 0)*`arglo');
      quietly replace `g2' = `g2' - `elnLg'*(($ML_y1 < 4)*`arghi' -
($ML_y1 > 0)*`arglo')*`mu';
      quietly replace `g3' = `g3' + `elnLg'*(($ML_y1 == 1)*`arghi' -
($ML_y1 == 2)*`arglo');
      quietly replace `g4' = `g4' + `elnLg'*(($ML_y1 == 2)*`arghi' -
($ML_y1 == 3)*`arglo');
      quietly replace `g5' = `g5' + `elnLg'*(($ML_y1 == 3)*`arghi' -
($ML_y1 == 4)*`arglo');
    };
    drop `lnLg';
    **Work the antithetical draw**;
    quietly replace `xbmu' = `xb'-`theta'*`mu';
    quietly replace  `arglo' = -`xbmu' + ($ML_y1 == 2)*`alpha2' +
($ML_y1 == 3)*`alpha3' + ($ML_y1 == 4)*`alpha4';
    quietly replace  `arghi' = -`xbmu' + ($ML_y1 == 1)*`alpha2' +
($ML_y1 == 2)*`alpha3' + ($ML_y1 == 3)*`alpha4';
    quietly replace  `p' = ($ML_y1 == 4) + ($ML_y1 < 4) * normal(`arghi') -
($ML_y1 > 0) * normal(`arglo');
    quietly su `p'; if (r(min) == 0) {; scalar `lnf' = .; exit; };
    quietly replace `lnLgi' = `wgt'*ln(`p');
    quietly replace `lnLgi' = -999 if `lnLgi' == .;
    quietly egen `lnLg' = total(`lnLgi'),by($ML_y2);
    quietly replace `elnLg' = exp(`lnLg');
    quietly su `elnLg'; if (r(min) == 0) {; scalar `lnf' = .; exit; };
    quietly replace `Lg' = `Lg' + `elnLg';
    if (`todo'~= 0) {;
      quietly replace `arglo' = (($ML_y1 > 0) * normalden(`arglo')) / `p';
      quietly replace `arghi' = (($ML_y1 < 4) * normalden(`arghi')) / `p';
```

```
      quietly replace `g1' = `g1' - `elnLg'*(($ML_y1 < 4)*`arghi' -
($ML_y1 > 0)*`arglo');
      quietly replace `g2' = `g2' + `elnLg'*(($ML_y1 < 4)*`arghi' -
($ML_y1 > 0)*`arglo')*`mu';
      quietly replace `g3' = `g3' + `elnLg'*(($ML_y1 == 1)*`arghi' -
($ML_y1 == 2)*`arglo');
      quietly replace `g4' = `g4' + `elnLg'*(($ML_y1 == 2)*`arghi' -
($ML_y1 == 3)*`arglo');
      quietly replace `g5' = `g5' + `elnLg'*(($ML_y1 == 3)*`arghi' -
($ML_y1 == 4)*`arglo');
    };
    drop `lnLg';
    local jr=`jr'+1;
  };
  quietly replace `Lg' = `Lg'/(2*`JR');

  quietly g double `lnfgi' = `wgtmean'*ln(`Lg') * ($ML_y2 ~= $ML_y2[_n-1]);
  *su `lnfgi' `g1' `g2' `g3' `g4' `g5';
  mlsum `lnf' = `lnfgi';

  if (`todo'==0 | `lnf' >= .) exit;
  tempname d1 d2 d3 d4 d5;
  **Note: $ML_y3 contains the full weight**;
  **Note: gradient1 and gradient2 are variables made available in wreprobit,
          to help compute the correct stdev**;
  quietly replace gradient1 = $ML_y3*`g1'/(`Lg'*(2*`JR'));
  quietly replace gradient2 = $ML_y3*`g2'/(`Lg'*(2*`JR'));
  quietly replace gradient3 = $ML_y3*`g3'/(`Lg'*(2*`JR'));
  quietly replace gradient4 = $ML_y3*`g4'/(`Lg'*(2*`JR'));
  quietly replace gradient5 = $ML_y3*`g5'/(`Lg'*(2*`JR'));
  mlvecsum `lnf' `d1' = gradient1, eq(1);
  mlvecsum `lnf' `d2' = gradient2, eq(2);
  mlvecsum `lnf' `d3' = gradient3, eq(3);
  mlvecsum `lnf' `d4' = gradient4, eq(4);
  mlvecsum `lnf' `d5' = gradient5, eq(5);
  matrix `g' = (`d1',`d2',`d3',`d4',`d5');

end ;



*** Define weighted RE ordered probit as a regression command ***;

capture program drop wreoprobit;
program define wreoprobit;
  version 9.0;

  syntax varlist(min=3)[,trace simnum(integer 25) compareUOP compareWOP];
  gettoken y rest : 0 /*,parse(" ,")*/;
  gettoken g rest : rest ,parse(" ");
  gettoken wgt rest : rest ,parse(" ");
```

```
  gettoken xzvar rest : rest ,parse(",");
  display "Dependent variable is      `y'";
  display "Group variable is          `g'";
  display "Weight variable is         `wgt'";
  display "Explanatory variables are `xzvar'";
  display "Number of draws used by simulator = `simnum' (used twice: +/-)";
  display "  ";

** rescale weight to an average of 1 **;

  quietly su `wgt',meanonly;
  g `wgt'_stan=`wgt'/r(mean);
  label variable `wgt'_stan "Scaled weight at individual level";
  display "Variation in weights rescaled to an average of 1, at individual
level";
  su `wgt'_stan,detail;

** Estimate unweighted models for comparison, as requested through options **;
  if "`compareUOP'" ~= "" {;
    display "  ";
    display ">> Unweighted Ordinary Probit Model, for comparison only <<";
    oprobit `y' `xzvar';
    display ">> Note: without applying weights or random effects, estimates
may be biased and inconsistent";
    display ">> End of comparison <<";
    display "  ";
  };


** get starting values from a weighted ordinary probit **;
  if "`compareWOP'" ~= "" {;
    display ">> Weighted Ordinary Probit Model, for comparison only <<";
    oprobit `y' `xzvar' [pw=`wgt'_stan];
    display ">> End of comparison <<";
    display "  ";
  };
  else {;
    quietly oprobit `y' `xzvar' [pw=`wgt'_stan];
  };
  matrix b=get(_b)';
  scalar bK=rowsof(b);
  ** oprobit sets the intercept to zero and estimates all cutpoints **;
  ** this program estimates the intercept and sets the first cutpoint to zero **;
  matrix b_b = b[1..(bK-3),1];
  matrix b_alpha = b[(bK-2)..bK,1] - (b_b[bK-3,1]*J(3,1,1));
  matrix theta=0.1;
  matrix bb=(b_b*0.95)\theta\b_alpha;
  matrix rownames bb = `xzvar' intercept stdevmu a2 a3 a4;
  matrix list bb;

** Determine the number of groups **;
  sort `g';
```

```
  quietly g `g'_nummer=1 if _n==1;
  quietly replace `g'_nummer=`g'_nummer[_n-1] + (`g' ~= `g'[_n-1]) if _n > 1;
  label variable `g'_nummer "Numeric group number";
  di "  ";
  quietly su `g'_nummer;
  local Ngroups = r(max);
  display "Number of groups =  `Ngroups'";
  di "  ";
  *di "Tabulation of group sizes for each group";
  *tab `g'_nummer;
  di "  ";

  preserve;

** Compute a matrix of random draws for simulation **;
  quietly drop if _n > `Ngroups';
  g id=_n;
  set seed 19901;
  matrix ZZ=J(`Ngroups',1,0);
  local jr=1;
  while `jr' <= `simnum' {;
    g z=invnorm(uniform());
    mkmat z;
    matrix ZZ=ZZ,z;
    drop z;
    local jr=`jr'+1;
  };
  matrix ZZ=ZZ[1...,2...];
  *matrix list ZZ;
  restore;

  ** Start ML estimation of random effect probit model **;
    di "** Start ML estimation of weighted random effect probit model **";
    g gradient1=0;
    g gradient2=0;
    g gradient3=0;
    g gradient4=0;
    g gradient5=0;
    ml model d1 wreop
       (eq1:`y' `g'_nummer `wgt'_stan=`xzvar')
       (eq2:) (eq3:) (eq4:) (eq5:);
    ml init bb,copy;
    ml search;
    if "`trace'" ~= "" {;
      ml max,trace grad;
    };
    else {;
      ml max;
    };
** organize the outerproduct of the gradient for computation of correct stdev
of b **;
```

```
    local ggvar = "";
    local j = 0;
    foreach aa of varlist `xzvar' {;
      local j = `j' + 1;
      g ggg`j' = gradient1*`aa';
      local ggvar = "`ggvar' ggg`j'";
      display "Adding to local ggvar: `j' equals `aa'";
    };
    g gggintercept = gradient1;
    g gggtheta = gradient2;
    g gggcut2 = gradient3;
    g gggcut3 = gradient4;
    g gggcut4 = gradient5;
    local ggvar = "`ggvar' gggint gggtheta gggcut2 gggcut3 gggcut4";
    display "Complete local ggvar equals `ggvar'";
    matrix accum G = `ggvar',noconst;
    matrix b_est = e(b)';
    matrix V = e(V);
    matrix list G;
    matrix list V;
    matrix VARb = V * G * V;
    matrix SDb = cholesky(diag(vecdiag(VARb)));
    matrix tstatb = syminv(SDb)*b_est;
    matrix SDb = vecdiag(SDb)';
    matrix result = b_est,SDb,tstatb;
    matrix colnames result = b stdev tstat;
    matrix list result;
    drop ggg*;
  drop `g'_nummer `wgt'_stan gradient*;

end;

/** Information on wreoprobit routine ********************************************
*   Depvar: y
*   Group indicator: group
*   Weight: weight
*   Explanatory variables, individual level: x1a x1b x1c
*   Explanatory variables, group level: x2a x2b
*
*   The call to the weighted random effect ordered probit command wreoprobit must first
*     listthe depvar groupvar and weightvar, and then the explanatory variables.
*   The explanatory variables may be in any order, mixing individual level and
*     group level variables in random order.  The program will sort them into
*     a set of individual level variables and a set of group level variables.
*   The option 'simnum' must specify an integer, indicating the number
*     of random draws that the simulated random probit estimation uses.  A higher
*     number yields greater precision but lengthens the computation time more or
*     less proportionately.  Default is 25.
*   The option 'trace' asks for more detail during the iterative search.
*   The option 'compareUOP' will run a unweighted ordered probit model for comparison.
*   The option 'compareWOP' will run a weighted ordered probit model for comparison.
*
*   NOTE: the program is set up for a y variable with FIVE categories.  It is
*     easily modified to handle a different number of categories, but at the moment
```

```
*     the program does not automatically accommodate an arbitrary number of
*     categories.
*
*   The output will appear to yield estimates of five equations.  The first equation
*   represents the parameters of the explanatory variables (and intercept).  The
*   second equation shows the estimate of the standard deviation of the random effect
*   (under the label of '_cons').  The third, fourth, and fifth equation shows the
*   estimated second, third, and fourth cutpoints.  The first cutpoint is standardized
*   to be equal to 0 a priori.
*
*   The code of the program is simply modified to estimate the natural log of the
*   standard deviation of the random effect (log-stdev) rather than the standard
*   deviation itself.  An estimate of the standard deviation (stdevmu)
*   is obtained by taking the antilog of the log-stdev.  The standard
*   error of stdevmu is found by multiplying the standard error of the
*   estimate of the log-stdevmu with the estimated value of stdevmu.
*
*   Examples of how to call the program:
*
* wreoprobit y group weight x1a x2a,trace simnum(6);
* wreoprobit y group weight x1a x2a x1b x1c x2b,compareUOP compareWOP;
*
*******************************************************************************/
```

Annex J. Employment and Income Estimates in the Surveyed Communities

Table J.1 Nicaraguan Communities

	Unweighted		
	Entire Sample	Communities	
	(n = 98 communitites)	Smaller (n = 46)	Larger (n = 48)
1. Employment Shares			
Total rural nonfarm	89%	85%	92%
Total rural nonfarm self-employed	55%	53%	58%
Total wage workers /1	41%	42%	42%
Agriculture /2	30%	40%	19%
Labor force participation ratios	74%	76%	72%
2. Income Shares			
Rural nonfarm /3	89%	89%	92%
Agriculture /4	11%	11%	8%
3. Average Per Capita and Per Worker Incomes ($)			
Gross household income/capita /5	490	388	571
Household "earned" income/capita /6	395	317	482
Household earnings/cap. fm RNF activity /7	449	354	534
Average *per rural nonfarm worker*	1048	839	1272

Source: RIC Survey, Nicaragua.

/1 includes part-time and seasonal farm labor.

/2 includes all household members engaged in agricultural operations; ***viz.*** full-time self-employed, part-time, and seasonal workers.

/3 = Σ (enterprise income + wage earnings/HH "earned" income [cf. ff 6 below]) * 100; ***N.B.,*** a small portion of wage earnings likely derives from agriculture, thus the estimate contains moderate upwards bias.

/4 possible moderate downwards bias — cf. ff 3 above.

/5 = Σ (enterprise & farm incomes + wage earnings + other cash inflows [e.g. interest, remittances, etc.])/HH size.

/6 = Σ (enterprise & farm incomes + wage earnings)/HH size.

/7 = Σ (enterprise income and wage earnings); ***n.b.*** a small portion of wage earnings likely derives from agriculture, thus the estimate contains moderate upwards bias.

Table J.2 Sri Lankan Communities

	Unweighted		
	Entire Sample	Communities	
	(n = 147 communitites)	Smaller (n = 73)	Larger (n = 73)
1. Employment Shares			
Total rural nonfarm	86%	81%	92%
Total rural nonfarm self-employed	42%	40%	44%
Total wage workers /1	49%	46%	52%
Agriculture /2	35%	42%	27%
Labor force participation ratios	62%	63%	61%
2. Income Shares			
Rural nonfarm /3	74%	69%	79%
Agriculture /4	26%	31%	21%
3. Average Per Capita and Per Worker Incomes ($)			
Gross household income/capita /5	406	369	448
Household "earned" income/capita /6	190	172	210
Household earnings/cap. fm RNF Activity /7	259	252	268
Average *per rural nonfarm worker*	518	484	552

Source: RIC Survey, Sri Lanka.

/1 includes part-time and seasonal farm labor.

/2 includes all household members engaged in agricultural operations; ***viz.*** full-time self-employed, part-time, and seasonal workers.

/3 = Σ (enterprise income + wage earnings/HH "earned" income [cf. ff 6 below]) * 100 ; ***N.B.,*** a small portion of wage earnings likely derives from agriculture, thus the estimate contains moderate upwards bias.

/4 possible moderate downwards bias — cf. ff 3 above.

/5 = Σ (enterprise & farm incomes + wage earnings + other cash inflows [e.g. interest, remittances, etc.])/HH size.

/6 = Σ (enterprise & farm incomes + wage earnings)/HH size.

/7 = Σ (enterprise income and wage earnings); ***n.b.*** a small portion of wage earnings likely derives from agriculture, thus the estimate contains moderate upwards bias.

Table J.3 Tanzanian Communities

	Unweighted		
	Entire Sample	Communities	
		Smaller	Larger
	(n = 154 communitites)	(n = 71)	(n = 73)
1. Employment Shares			
Total rural nonfarm	42%	42%	43%
Total rural nonfarm self-employed	38%	38%	38%
Total wage workers /1	15%	16%	17%
Agriculture /2	89%	90%	90%
Labor force participation ratios	90%	90%	90%
2. Income Shares			
Rural nonfarm /3	65%	63%	69%
Agriculture /4	35%	37%	31%
3. Average Per Capita and Per Worker Incomes ($)			
Gross household income/capita /5	149	126	157
Household "earned" income/capita /6	124	102	128
Household earnings/cap. fm RNF Activity /7	77	59	85
Average *per rural nonfarm worker*	391	301	453

Source: Calculated from RIC household surveys.

/1 includes part-time and seasonal farm labor.

/2 includes all household members engaged in agricultural operations; ***viz.*** full-time self-employed, part-time and seasonal workers.

/3 = Σ (enterprise income + wage earnings/HH "earned" income [cf. ff 6 below]) * 100; ***N.B.,*** a small portion of wage earnings likely derives from agriculture, thus the estimate contains moderate upwards bias.

/4 possible moderate downwards bias — cf. ff 3 above.

/5 = Σ (enterprise & farm incomes + wage earnings + other cash inflows [e.g. interest, remittances, etc.])/HH size.

/6 = Σ (enterprise & farm incomes + wage earnings)/HH size.

/7 = Σ (enterprise income and wage earnings); ***n.b.*** a small portion of wage earnings likely derives from agriculture, thus the estimate contains moderate upwards bias.

Annex K. International Benchmarking of the Nonmetropolitan/Rural Investment Climate

See Table K.1 on the following pages.

Table K.1 International Benchmarking of the Nonmetropolitan/Rural Investment Climate

	Bangladesh	Nicaragua	Sri Lanka	Pakistan	Tanzania	Bangladesh: by geographic segment		
						Peri-urban	Town	Village
	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean
Index 1: Connectivity	**0.45**	**0.34**	**0.40**	**0.38**	**0.20**	**0.55**	**0.50**	**0.43**
Subindex 1 Time taken by main means of transportation to the nearest major city	0.44	0.32	0.47	0.32	0.26	0.57	0.55	0.40
Subindex 2 Cost of transportation to the nearest major city (by public transportation)	0.18	0.09	0.66	0.25	0.08	0.18	0.14	0.19
Subindex 3 Time taken by main means of transportation to the main market	0.78	0.33	0.49	0.51	0.52	0.86	0.86	0.74
Subindex 4 Cost of transportation to the main market (by bus)	0.52	0.07	0.61	0.39	0.08	0.58	0.52	0.51
Subindex 5 Distance to the post office	0.67	0.47	0.69	0.67	0.26	0.82	0.68	0.63
Subindex 6 Rail stop in walking distance	0.57		0.12	0.34	0.23	0.64	0.60	0.55
Subindex 7 Percentage of households with fixed telephone lines	0.04	0.59	0.08	0.38	0.03	0.08	0.10	0.04
Subindex 8 Percentage of households with cell phones	0.42	0.51	0.10	0.20	0.13	0.68	0.54	0.39
Index 2: Infrastructure Services	**0.44**	**0.51**	**0.35**	**0.58**	**0.19**	**0.54**	**0.50**	**0.40**
Subindex 1 Percentage of households that use electricity	0.70	0.67	0.69	0.87	0.12	0.88	0.80	0.71
Subindex 2 Availability of electricity index	0.00	0.62	0.55	0.46	0.31	0.00	0.00	0.00
Subindex 3 Percentage of households with access to protected water source	0.98	0.62	0.80	0.72	0.47	1.00	0.99	0.97
Subindex 4 Percentage of households with fixed telephone lines	0.04	0.59	0.08	0.38	0.03	0.08	0.10	0.04
Subindex 5 Percentage of households with cell phones	0.42	0.51	0.10	0.20	0.13	0.68	0.54	0.39
Subindex 6 Sewage channels in the community	n.a.	0.08	0.11	0.57	0.07	n.a.	n.a.	n.a.
Subindex 7 Garbage collection or disposal service in the community	n.a.	0.49	0.12	0.34	0.13	n.a.	n.a.	n.a.
Subindex 8 Most common road surface (internal road) is concrete or asphalt	0.39	n.a.	0.33	0.88	0.22	0.60	0.56	0.31

Index 3: Business Services	**0.05**	**0.24**	**0.15**	**0.40**	**0.07**	**0.03**	**0.21**	**0.02**
Subindex 1 Engineering services available for businesses in the community	0.05	0.23	0.06	0.42	0.04	0.04	0.20	0.02
Subindex 2 Management consulting services available for businesses in the community	0.04	0.22	0.07	0.26	0.15	0.04	0.20	0.01
Subindex 3 Marketing services available for businesses in the community	0.06	0.19	0.08	0.37	0.13	0.04	0.32	0.02
Subindex 4 Accounting services available for businesses in the community	0.02	0.27	0.11	0.29	0.02	0.00	0.16	0.00
Subindex 5 Legal services available for businesses in the community	0.06	0.31	0.26	0.55	0.07	0.08	0.20	0.03
Subindex 6 Insurance services available for businesses in the community	0.09	0.21	0.39	0.46	0.03	0.04	0.28	0.07
Subindex 7 Information services available for businesses in the community	0.02	0.24	0.10	0.40	0.03	0.00	0.12	0.01
Index 4: Governance	**0.62**	**0.68**	**0.67**	**0.72**	**0.50**	**0.68**	**0.67**	**0.60**
Subindex 1 General policy and institutional constraints	0.83	0.79	0.75	0.94	0.85	0.85	0.78	0.84
Subindex 2 Infrastructure and services	0.21	0.50	0.49	0.74	0.01	0.27	0.29	0.19
Subindex 3 Dealing with government agencies	0.91	0.94	1.00	0.86		0.89	0.95	0.90
Subindex 4 Rule of law	.		0.13	0.57	0.49	.	.	.
Subindex 5 Conflict resolution and contract enforcement	0.77	0.59	0.98	0.62	0.66	0.78	0.82	0.76
Index 5: Human Capital	**0.24**	**0.21**	**0.33**	**0.21**	**0.21**	**0.29**	**0.28**	**0.22**
Index 6: Finance Services*	**0.19**	**0.17**	**0.50**	**0.39**	**0.18**	**0.27**	**0.33**	**0.15**
Subindex 1 Number of financial and insurance institutions that offer services in the community, weighted by distance from the community	0.24	0.20	0.56	0.73	0.24	0.34	0.42	0.19
Subindex 2 Number of financial and insurance services provided in the community by formal institutions, weighted by distance from the community	0.23	0.15	0.47	n.a.	0.12	0.34	0.41	0.16
Subindex 3 Access to loans	0.11		0.47	0.05	0.19	0.12	0.16	0.10

Source: Data for Nicaragua, Sri Lanka, and Tanzania, from RIC2; for Bangladesh and Pakistan, from World Bank (2008a; forthcoming).

Note: * See note to Table E.6.

Notes

EXECUTIVE SUMMARY

1 The survey methodology has been addressed in "Rural Investment Climate Assessment: Implementation Manual" (World Bank 2007c).
2 For Nicaragua and Sri Lanka, gaps in weights required that some observations be discarded as well.
3 Related to this, consistent database design should be another objective. The RIC databases for Nicaragua, Sri Lanka, and Tanzania followed entirely different formats. Considerable effort was necessary to reconcile the analysis of one country with data from another. Among other undesirable consequences, lack of consistency and even of common definitions both muddled and discouraged the prospects for robust cross-country analysis.

INTRODUCTION

4 Perhaps, more accurately, we should say that the growth of small informal enterprises—which predominate in rural areas—has lagged behind growth of the larger "formals," which are largely urban-based.
5 Studies offering some relevant information were based on household-based surveys without particular focus on the conditions of nonfarm enterprises and thus provided limited evidence on investment climate constraints. Probably the most elaborate study to date to use household survey data is that by Vijverberg et al. (2006), compiled using the Vietnam Household Living Standards Survey data of 2006.
6 The RIC Implementation Manual developed by the Bank's Agriculture and Rural Development Department is being used by task teams throughout the regions, as are the prototype questionnaires. The manual backstopped a training course delivered in 2007 to participants of a multinational workshop convened in Tanzania.
7 Haggblade et al. 2007, p. 400.
8 Though an intersect exists, the orientation of the PICS has largely been toward IC constraints endured by registered, larger, mainly urban-based enterprises. The rural nonfarm enterprises (RNFEs), in contrast, are relatively smaller and, for the most part, informally organized.

CHAPTER 2

9 Using household data from The Gambia, Chavas, Petrie, and Roth (2005) examine the determinants of efficiency in households that generate farm and nonfarm income. They take a two-stage approach that first generates efficiency measures from a joint farm-nonfarm household production model based on nonparametric linear programming techniques; these subsequently serve as dependent variables in a statistical model designed to get at the underlying determinants of efficiency. The authors conclude that household incomes could be improved in The Gambia if additional household resources were allocated to nonfarm activities. They suspect that imperfections in local labor and capital markets are the chief constraints preventing households from doing so. In a study based on data from China taken during a period of factor-market liberalization, Yang (2004) finds that education is a key determinant of the rate at which households reallocated assets to more profitable nonfarm activities.
10 These findings apply to other regions as well. Similar results were observed in notes prepared for *The Rural Investment Climate: It Differs and It Matters* (World Bank 2006), using survey data from the three pilots under review in this report (Nicaragua, Sri Lanka, and Tanzania). A main problem for investment is that firms with profitable investment projects often cannot use external funds to finance them.
11 Noting that the LSS approach captures in the main sole proprietorships and partnerships, it was evident that even some of these enterprises had not been captured in the LSS databases. Meanwhile, incorporated enterprises, joint ventures and cooperatives, and parastatals remained completely outside the LSS sampling design. Consequently, both a broader enterprise and community survey and complementary surveys of household-owned enterprises were recommended as essential for more rigorous assessments of the rural investment climate. These findings guided the scope and design of the prototype questionnaire employed for the RIC pilot surveys.
12 An additional resource, not available when the pilot surveys were being designed, is the draft WDR (2008b), which documents and emphasizes the importance of spatial considerations. The RIC Implementation Manual (World Bank 2007c) suggests ways to address this issue in the context of the connectivity of rural-urban RNFE markets, although, unfortunately, the surveys did not really address spatial considerations, except for a parallel sample of larger and better-connected rural enterprises in the Sri Lanka RIC assessment. This led to indirect inferences, as the larger enterprises tend to serve markets beyond the borders of their communities.
13 It also contradicts the traditional assumption that earnings from labor migration exceed those from local nonfarm activities (Haggblade 2007, p. 122).
14 Ibid., p. 123.
15 Ibid., p. 138–39.
16 Examples are indicators for governance (World Bank 2007a), investment climate (World Bank 2004), cost of doing business (World Bank 2007b), and political rights and civil liberties (Freedom House 2008), http://www.freedomhouse.org/template.cfm?page=15 .

17 See the RIC Implementation Manual for details (World Bank 2007c).
18 To illustrate, suppose that a community has 185 small enterprises typically employing 1.5 workers and 15 large ones employing 10 workers on average. Thus, the population proportion of large enterprises is 7.5 percent (15 of 200). A sample of 10 enterprises is drawn, consisting of 5 large ones (a sampling rate of 33.3 percent) and 5 small ones (a sampling rate of 2.7 percent). The simple sample proportion of large enterprises is 50 percent, which of course greatly exceeds the actual population proportion it seeks to estimate. Sample weights would equal 3 (= 15/5) for large enterprises and 37 (= 185/5) for small enterprises. A weighted sample proportion of large enterprises equals $3 \times 5 / (3 \times 5 + 37 \times 5) = 0.075$, which is the correct estimate. The unweighted estimate of the number of employees equals 5.75 per enterprise, which is completely unrealistic since the population includes so many small enterprises; the weighted mean equals $(3 \times 5 \times 10 + 37 \times 5 \times 1.5) / (3 \times 5 + 37 \times 5) = 2.14$.
19 The dependent variable is continuous in Chapter 3, binary in Chapter 4, and ordered-categorical in Chapter 5. Each implies its own variation of the same econometric structure.
20 This is not to say that a successful investment climate necessarily raises entrepreneurship rates. It may equally well stimulate enterprise growth, which soaks up labor in the community and gives some entrepreneurs an incentive to abandon their own enterprise and seek employment in a larger, more successful business—thus driving the entrepreneurship rate down.

CHAPTER 3

21 The questionnaires did not formally define *enterprise* or *self-employment* nor did it set a clear minimum threshold for enterprise size. With a sharper definition of enterprise some of these cases would not have been included, although the various sources of income would still have been recorded as household income.
22 This is because much more variation relates to enterprise size than to community variables.
23 This finding mirrors the conclusion in Haggblade et al. (2007) reported above in Chapter 2.

CHAPTER 4

24 Haggblade et al. (2007) mention a range of 20 to 50 percent.
25 This suggests—or, rather, reaffirms from the stratum of rural households—that as GDP/capita increases (Table 3.1) the significance of agricultural employment declines somewhat monotonically.
26 When describing household incomes it is important to note that the Sri Lanka questionnaire contained a flaw that caused understating of income components other than wages and salaries and farming. In addition, the Tanzania statistics did not use sampling weights, and the stratification is explicitly based on a feature (entrepreneurship) related to income.
27 In addition to farm income, income from wages and salaries are lumped together, and expenses related to wage earning are deducted from the gross wage income to estimate net wage income. Households also receive other income in the form of remittances and gifts from friends and relatives, interest income from savings and dividends, retirement benefits, and lottery winnings.
28 A precise definition of entrepreneurship appears in the next section. The definitions for Tanzania and Nicaragua are the same; that for Sri Lanka differs.
29 These variables relate to the productivity of household members in their own enterprises and to the attractiveness of alternative employment and therefore determine the choice between them.
30 Among the explanatory variables, household assets and community income levels might have been endogenous. Fortunately, a test of endogeneity of these two variables gave no evidence that this was the case (Annex F).
31 The distribution of assets is highly skewed. For both countries, the change in assets is also equivalent to the difference between the seventy-fifth and about the ninety-second percentiles.
32 It is noted that this estimated effect is causal and not merely a reflection of reverse causation where prosperous entrepreneurs are accumulating household wealth: as mentioned above, assets were found to be exogenous relative to entrepreneurship choice.
33 A portion of the difference between Nicaragua and Sri Lanka derives from the difference in definition: Nicaragua uses the income-based definition and Sri Lanka a work-based definition. Thus, the percentage actually engaged in farm *work* rises to 35 percent in Nicaragua. By comparison, the percentage of rural households in the 2001 Living Standards Measurement Study data that reported any agricultural activity was 35.6 percent. Additionally, it may be the case that a significant number of wage workers are actually agricultural laborers; the questionnaires did not differentiate wage employment by sector.
34 The choices are probably interdependent, and some efficiency gain might be derived from joint estimation. A multinomial logit model incorporates all alternatives simultaneously but requires that individuals or households make exclusive choices. As seen in Table 4.1, households frequently participate in several activities, contrary to the logic of a multinomial logit application.
35 These conditions might provide incentives to start one's own business instead, but column 1 of the table reveals no such effect.
36 The effect of management consulting and marketing services is strong and statistically significant and in a plausible direction—but may have to be taken with a grain of salt, as the number of communities with such services (9 and 8 respectively, with an overlap of 6 between them, out of 117) is small.
37 The precise timing of the RIC survey matters as well, since the period of start-up measurement covers calendar years only. This recognition leads to an upward revision of the estimated start-up rate in Nicaragua to 4.0 to 4.5 percent (see Annex F).
38 If p measures the annual proportion of enterprises that exit and are replaced by new start-ups, the two-year ratio of start-up enterprises relative to the total number of enterprises equals $p(1 - p) + p$, which then may be equated to the two-year percentage q of start-up enterprises in the sample. The survival rate equals $1 - p$ and solves to $(1 - q)^{0.5}$. The assumption of a stationary enterprise population is crucial in identifying the survival rate.

CHAPTER 5

39 Entrepreneurs' responses on the importance of governance and corruption as possible IC constraints are addressed in the first section of Annex H and recorded in Table H.10.

40 In Nicaragua, different enterprise questionnaires were administered to household-based and stand-alone enterprises. Enumerators apparently skipped questions about labor EICOs for household-based enterprises. Another difference among the countries is that the Sri Lanka and Tanzania questionnaires use screen questions to uncover general problems with, for example, finance or infrastructure; if affirmative, they proceed with the detailed EICOs. If the respondent reports no general problems, the researcher is forced to conclude there are also no problems with detailed EICO aspects. This may have led to some underreporting of EICO barriers in Sri Lanka and Tanzania relative to Nicaragua.
41 A detailed check of data by enterprise suggests that this is not due to a "halo" effect caused by entrepreneurs who, disgruntled about a few barriers, flagged all EICOs across the board: their responses varied by EICO.
42 The precise wording is "In many countries, companies give informal payments (kickbacks, bribes, or other exemptions) to government officials to gain advantages in the writing of laws, decrees and regulations, etc. In this country, to what extent do these practices have a direct impact on your establishment?"
43 In Sri Lanka, the questions about influencing officials or drafting legislation are answered only by entrepreneurs employing at least three workers.
44 The remaining benchmark indicator, governance, is not used since it is constructed out of community-averaged entrepreneur responses, some of which are among the EICOs and EICIs these models seek to explain.
45 When studying whether entrepreneurs list postal service as an obstacle, for example, it is useful to know the distance to the nearest post office. This is merely one of eight elements of the connectivity benchmark.
46 In the case of variables entered in logarithmic form, the mean and standard deviation of the regular variable are considered in determining the magnitude of the standardized effect.
47 These estimates are presented in Table H.3.
48 Ibid. Table H.3, section C.
49 Ibid. Table H.3 section B. Note that the dummy for trade is omitted in the model.
50 A mixed enterprise is active in more than one sector.
51 Table H.5 reports the estimates of the impact of size, as measured by the log of sales, and productivity, as measured by the log of the ratio of net value added over total factor cost (V/C). The reported values are not scaled estimates as in the other tables; rather, they are unscaled estimates of the parameters of the weighted random effect ordered probit model. In Nicaragua, many of the two effects are statistically significant; in Sri Lanka, the effect is statistically relevant for only one EICO, that is, telecommunication; and in Tanzania, five of the twenty parameter estimates are statistically significant. It is notable that the size effect is positive for 23 of the 30 estimates, that the productivity effect is negative for 24 of the 30 estimates, and that the effects have the same sign only five times. Taken at face value, this implies that operators of larger enterprises tend to complain more often about investment climate conditions and that entrepreneurs overseeing more productive businesses tend to view the environment in less problematic terms. But, as mentioned, these effects are probably biased downward: the size effect is probably more strongly positive, and the productivity effect less negative.
52 The average percentage of households in a community with access to protected water sources is higher in Sri Lanka (80%) than Nicaragua (62%) and Tanzania (48%), suggesting the possibility of a threshold effect. But even in Sri Lanka some communities have a very low proportion of households with this amenity.
53 For Tanzania, this EICO did not score in the top ten, and it is therefore not included in the table. In a regression run for Tanzania not reported here, marketing services reduce the obstacle of low market demand, as does the human capital benchmark. Connectivity, however, had no effect.
54 Annex L contains a broader summary of governance and corruption as perceived investment climate constraints.
55 There is no dummy for traders; the parameters compare entrepreneurs with traders.
56 Unfortunately, standard estimation routines in statistical software such as Stata and SAS allow either sampling weights or random effects, but not both. Annex I describes Stata programs that allow both features to be included in the estimation.
57 As described in the analysis above, including enterprise performance in the EICO equations creates biased estimates. To the degree that these biased parameter estimates are still reliable, it appears that enterprise size raises perceptions of investment climate constraints, and enterprise profitability reduces perceived constraints. But one may question the direction of causality, since the EICO models in this chapter do not confront this issue: are managers of profitable enterprises better able to circumvent investment climate barriers, or are such enterprises more profitable because barriers are lower?

CHAPTER 6

58 Community size varies greatly, however, especially in Nicaragua, but also in Tanzania, such that the smallest community surveyed in Nicaragua has only slightly more than one-half the population of the smallest Sri Lankan village, though the size of its largest community exceeded Sri Lanka's by more than 16 times.
59 The percentage with debt (formal and informal sources together) is 21, 42 and 23 percent respectively in Nicaragua, Sri Lanka and Tanzania.
60 Moreover, the specification of the outcome variables is not always optimal from a theoretical point of view because of limitations imposed by the database.
61 Measured by the average tendency of enterprises to interact with suppliers and clients outside the community.
62 Despite a number of relevant variables in the regression model, the F-statistic is extremely low. This may occur in an equation suffering from substantial heteroskedasticity. The standard errors of the parameter estimates are computed to allow for heteroskedasticity, but the computed p-value of the F-statistic could be seriously biased. The ratio variable here may be subject to wild swings because of measurement error in costs of production.

CHAPTER 7

63 Thus, comparisons of benchmarks between countries in this study are illustrative but tentative.
64 With the loss of weights, as in the case of Tanzania, much of the value of the RICS is lost. For Nicaragua and Sri Lanka some observations had to be discarded because of gaps in information on weights.
65 The benefits from utilizing the pilot assessments in Nicaragua and Tanzania is less clear, owing to delays in completing the assessments, lukewarm government response (in Tanzania), and recurrent staffing changes both on the country teams and among counterparts.

ANNEX A

66 Incomplete data on benchmark indicators causes a loss of 167 enterprises. Unavailability of other community variables reduces the sample size by 53 observations. Many of these enterprises are located in the war-torn northeast region of the country.

67 The questionnaires included no formal definition of *enterprise* and set no minimum threshold size.

68 Other common reasons might be added, such as response error and data entry mistakes.

69 Determination of the threshold was arbitrary. Taking 25 percent of an adult's employment per year and assuming 240 working days per year at a net income of US$0.25 per day gives a threshold of US$60 per year. In the analysis of household income and employment in Annex D no enterprises were excluded.

ANNEX C

70 The Cobb-Douglas model augmented with quadratic and interactive labor and capital variables is a partial implementation of the translog model, which would add quadratic interactive terms involving nonfactor inputs and depreciation. As the NVA model omits the latter, it may be seen as a complete translog specification.

71 Following common practice, there is no depreciation charge for capital.

72 Although variables for effective access to bank loans and maximum amount of informal credit obtained are available, they were not selected because of endogeneity. Banks and informal lenders presumably extend credit based on their evaluations of enterprise performance.

73 As a result, the registration dummy variable may behave as a dummy for larger enterprises. Size is already measured elaborately, however, with the input variables of labor, capital, nonfactor costs, and depreciation. The measured impact of the registration dummy is therefore likely to measure the pure effect of registration, rather than the effect of enterprise size differential.

74 Note that these dummy variables do not reflect differences in technology, which would imply a different structural relationship between inputs and output by sector. Theoretically, differences in sectoral technologies are entirely plausible, but the sectoral subsamples are too small to permit estimation of relationships by sector. Moreover, one subsample consists of enterprises that cannot be clearly located in a single sector, causing analytical problems as well. It should also be noted, however, that the (semi)translog specification of the input-output specification alleviates the problem caused by sample pooling.

75 Among the benchmark indicators in the regression model is a human capital index that summarizes the community's skills that are accumulated through education and experience. Illiteracy is therefore more specific than having little human capital: there is room for both variables in the regression model. The two variables also describe different parts of the labor force, and it is an empirical question which parts the rural non-farm enterprises draw from.

76 Self-selection effects may pertain to many other variables, including the use of household labor and enterprise capital, and the sector of activity. Correcting for self-selection is often done by means of the so-called Heckit method, which is appropriate when disturbances are identically and independently distributed among observations. The clustering by community and the community random effect this entails complicates the Heckit method greatly. Moreover, sampling weights for enterprises differ from those for households, even if a given enterprise belongs to a particular household. Sorting out these econometric issues is left for the future.

77 The prices of some items, in particular Coca-Cola, appeared to be significant in several of the draft regression specifications.

78 This is statistically relevant, but it may not be unique to rural enterprises. No comparable data for PICS are available, but it is not unlikely that for urban enterprises as well enterprise characteristics dominate other variables.

79 In simplified models similar to variant (1), population size and enterprise density have significant positive impacts, but these effects vanish in the more elaborate specifications of variants (4) and (5).

80 In variants (3) and (4) without quadratic terms (not reported in the table), the parameter estimates of log-labor and log-capital are 0.25 and 0.09, respectively, both highly significant; in variant (5) without quadratic terms, the estimates equal –0.04 (imputed labor, not significant), 0.11 (paid labor), 0.04 (imputed capital), and 0.07 (paid capital). Thus, a positive relation exists between input use and net value added, but these parameter estimates add up to much less than 1, which suggests decreasing returns to scale.

81 That is, $100(e^{1.34} - 1)$ percent.

ANNEX D

82 The exchanges rates used in this paper are 15.9372 cordobas per dollar for Nicaragua, 1089.33 shillings per dollar for Tanzania, and 96.7699 rupees per dollar for Sri Lanka.

83 According to human capital theory, the wage rate represents payment for the rental of a worker's human capital. Thus, a wage w equals the product of the rental rate r and the human capital stock h. Since r is unobservable but the same for all participants in the labor market, h is determined up to a multiplicative scalar constant, found as an intercept in the $\ln(h)$ equation. For now, this intercept in the $\ln(h)$ equation is set to 0. Later on, a further scaling is introduced that effectively modifies the scaling applied here.

84 As a concept, Bils and Klenow (2000) also include a more involved version of the same model, which allows for nonlinear effects of education as well as intergenerational transfers from parents and teachers to children. Nonlinear effects proved to be less important empirically, and intergenerational transfers are difficult to establish with the data at hand.

85 These assumptions are inspired by the evidence in the data, as the completion of a given level of schooling is typically achieved after a set number of years, which is frequently reported by respondents.

86 Households reported on nonwage, nonfarm income (including enterprise income) only if they held any wage job. A later section, provides more information.

87 For example, if the respondent offered information about the plot size in both acres and perches, the ratio of reported perches over reported acres should vary around 65. Alternatively, if the total plot size should be the sum of the two reported values, one would expect that the number of perches would be consistently below 65. Moreover, whatever interpretation is followed, the price per unit of land should also correspond roughly to that recorded for households using only one measure.

88 According to information provided at http://www.sizes.com/units/manzana.htm.

ANNEX F

89 A logical alternative to these definitions is to rely on the RICS enterprise sample, which is linked to the household sample: a household would be entrepreneurial if it is linked to an enterprise in the enterprise sample. Due to a design feature of the early RICS data, however, if the household owned and operated a stand-alone enterprise, it was not coded as operating a household enterprise. Thus, the rate of entrepreneurship based on the link in the RICS data reflects only household-based entrepreneurship and would understate the overall rate of entrepreneurship among rural households. Examining income receipt and work activity is a way around this design flaw.

90 Note that these are unweighted counts. Thus, Tanzania's sample evidence is valid here too. In Tanzania, 16.7 percent of the households received enterprise income but reported no "work on own account," and 10.2 percent had household members working in a household enterprise but reported no income from a nonfarm enterprise. In other words, Tanzanian data suffer from a more serious consistency problem. In this regard, RICS data are not unique: for similar evidence in the context of Living Standards Survey data in Ghana, Guatemala, Kyrgyz Republic, and Vietnam, see Vijverberg (2005). The lesson from this is that when collecting survey data such as RICS, survey managers must be on guard against common inconsistencies of this nature. Proper training of enumerators is a first step.

ANNEX H

91 The first insertion makes the grammar appropriate for the context of the text here; the second insertion is implied by the context of the screening question, but the actual text *in loco* does not specifically refer to the operation and growth of the enterprise.

ANNEX I

92 This follows from deriving the (pseudo-)likelihood function and retaining the portion that contains the parameters of the regression equation.

References

Amemiya, Takeshi. 1978. "A Note on a Random Coefficients Model." *International Economic Review*, Vol. 19, No. 3 (Oct., 1978), pp. 793–796.

Baltagi, Badi H. 2005. *Econometric Analysis of Panel Data*, Third edition. Chichester: Wiley.

Barrett, C. B., T. Reardon, and P. Webb. 2001. "Nonfarm Income Diversification and Household Livelihood Strategies in Rural Africa: Concepts, Dynamics, and Policy Implications." *Food Policy* 26 (4): 315–31.

Batra, G., D. Kaufmann, and A. H. W. Stone. 2003. *Investment Climate Around the World: Voices of the Firms from the World Business Environment Survey*. Washington, DC: IBRD.

Bigsten, A., P. Collier, S. Dercon, M. Fafchamps, B. Gauthier, J. W. Gunning, A. Oduro, R. Oostendorp, C. Pattillo, M. Söderbom, F. Teal, and A. Zeufack. 2003. "Credit Constraints in Manufacturing Enterprises in Africa." *Journal of African Economies* 12 (1): 104–25.

Bigsten, A., P. Collier, S. Dercon, B. Gauthier, J. W. Gunning, A. Isaksson, A. Oduro, R. Oostendorp, C. Pattillo, M. Söderbom, M. Sylvain, F. Teal, and A. Zeufack. 1999. "Investment in Africa's Manufacturing Sector: A Four Country Panel Data Analysis." *Oxford Bulletin of Economics and Statistics* 61 (4): 489–512.

Bigsten, A., P. Kimuyu, and K. Lundvall. 2004. "What to Do with the Informal Sector?" *Development Policy Review* 22 (6): 701–15.

Bigsten, A., and M. Soderbom. 2006. "What Have We Learned from a Decade of Manufacturing Enterprise Surveys in Africa?" *World Bank Research Observer* 21 (2): 241–65.

Bils, Mark, and Peter J. Klenow. 2000. "Does Schooling Cause Growth?," *American Economic Review,* vol. 90(5), pages 1160–1183, December.

Chavas, Jean-Paul, Ragan Petrie, and Michael Roth. 2005. "Farm Household Production Efficiency: Evidence from The Gambia." *American Journal of Agricultural Economics* 87 (1): 160–179.

Deininger, Klaus, Songqing Jin, and Mona Sur. 2007. "Sri Lanka's Rural Non-Farm Economy: Removing Constraints to Pro-Poor Growth." *World Development* 35 (12): 2056-2078.

De Soto, Hernando. 1989, reissued 2002. *The Other Path.* New York: Basic Books.

Eifert, B., A. Gelb, and V. Ramachandran. 2005. "Business Environment and Comparative Advantage in Africa: Evidence from the Investment Climate Data." Policy Research Working Paper, World Bank, Washington, DC.

Freedom House. 2008. *Freedom in the World 2007,* http://www.freedomhouse.org/template.cfm?page=15.

Greene, William H. 2008. *Econometric Analysis*. Upper Saddle River, New Jersey: Pearson Prentice Hall.

Haggblade, Steven, Peter Hazell, and James Brown. 1989. "Farm-Nonfarm Linkages in Rural Sub-Saharan Africa." *World Development* 17 (8): 1173–1201.

Haggblade, Steven, Peter B. R. Hazell, and Thomas Reardon, eds. 2007. *Transforming the Rural Nonfarm Economy: Opportunities and Threats in the Developing World*. Baltimore: The Johns Hopkins University Press for the International Food Policy Research Institute.

Hayami, Yujiro, ed. 1998. *Towards a Rural-Based Development of Commerce and Industry: Selected Experiences from East Asia.* Washington, DC: Economic Development Institute of the World Bank.

Lanjouw, Peter. 2001. "Nonfarm Employment and Poverty in Rural El Salvador." *World Development* 29 (3): 529–47.

Lanjouw, Jean O., and Peter Lanjouw. 2001. "The Rural Non-Farm Sector: Issues and Evidence from Developing Countries." *Agricultural Economics* 26 (1): 1–23.

Larson, Donald F., and Yyannu Cruz-Aguayo. 2008. "On the Role of Self-Employment and Specialization in Family Farms and Businesses: Evidence from Nicaragua." Unpublished manuscript.

Maloney, W. F. 2004. "Informality Revisited." *World Development* 32 (7): 1159–78.

Moock, Peter, Philip Musgrove, and Morton Stelcner. 1990. "Education and Earnings in Peru's Informal Nonfarm Family Enterprises." Living Standard Measurement Study Working Paper 64, World Bank, Washington, DC.

Perry, Guillermo, W.F. Maloney, O.S. Arias, P. Fajnzylber, A.D. Mason, and J. Saavedra-Chanduvi. 2007. *Informality: Exit and Exclusion.* Washington, DC: World Bank Latin American and Caribbean Studies and IBRD.

Reardon, T., J. A. Berdegué, and G. Escobar. 2001. "Rural Nonfarm Employment and Incomes in Latin America: Overview and Policy Implications." *World Development* 29 (3): 395–409.

Sundaram-Stukel, Reka, Klaus Deininger, and Songqing Jin. 2006. "Fostering Growth of the Rural Non-Farm Sector in Africa: The Case of Tanzania." Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Long Beach, California, July 23–26. Washington, DC: IBRD-DECRG.

Van der Sluis, Justin, Mirjam van Praag, and Wim Vijverberg. 2005. "Entrepreneurship Selection and Performance: A Meta-analysis of the Impact of Education in Developing Countries." *World Bank Economic Review* 19 (2): 225–61.

Vijverberg, Wim P.M. 1990. "Non-Farm Self-Employment and the Informal Sector in Côte d'Ivoire: A Test of Categorical Identity." *Journal of Developing Areas* 24 (4): 523–42.

———. 1997. "Monte Carlo Evaluation of Multivariate Normal Probabilities." In: *Journal of Econometrics*, January 1997, 76: 1–2, pp. 281–307.

———. 1998. "Nonfarm Household Enterprises in Vietnam." In *Household Welfare and Vietnam's Transition,* ed. David Dollar, Paul Glewwe, and Jennie Litvack, 137–78. Washington, DC: World Bank.

———. 1999. "The Impact of Schooling and Cognitive Skills on Income from Non-Farm Self-Employment." In *The Economics of School Quality Investments in Developing Countries: An Empirical Study of Ghana,* ed. Paul Glewwe, 206–52. New York: St. Martin's Press.

———. 2005. "Rural Investment Climate and Non-Farm Entrepreneurship in Living Standards Surveys." Unpublished manuscript, ARD, World Bank, Washington, DC.

Vijverberg, Wim, and Jonathan Haughton. 2004. "Household Enterprises in Vietnam: Survival, Growth, and Living Standards." In *Economic Growth, Poverty and Household Welfare: Policy Lessons from Vietnam,* ed. David Dollar and Paul Glewwe, 95–132. Washington, DC: The World Bank.

Vijverberg, Wim, with Hoang Thi Thanh Huong, Nguyen Chien Tang, Nguyen Ngoc Que, Nguyen The Quan, Phung Duc Tung, and Vu Thi Kim Mao. 2006. "Non-Farm Household Enterprises in Vietnam: A Research Project using Data from VHLSS 2004, VHLSS 2002 and AHBS 2003." The World Bank, Hanoi Office.

World Bank. 2003. *Reaching the Rural Poor: A Renewed Strategy for Rural Developmemt.* Washington, DC: World Bank.

———. 2005. *A Better Investment Climate for Everyone.* World Development Report 2005. Washington, DC: World Bank and IBRD.

———. 2006. *The Rural Investment Climate: It Differs and It Matters.* Washington, DC: IBRD.

———. 2007a. *Governance Matters 2007: Worldwide Governance Indicators, 1996–2006.* Washington, DC: World Bank.

———. 2007b. *Doing Business 2008.* Washington, DC: World Bank.

———. 2007c. "Rural Investment Climate Assessments: Implementation Manual" (draft). Washington, DC: World Bank.

———. 2007d. *World Development Indicator Database.* Washington, DC: World Bank. www.worldbank.org/data/onlinedatabases/onlinedatabases.html

———. 2007e. *Tanzania: Pilot Rural Investment Climate Assessment: Stimulating Non-farm Microenterprise Growth.* Report No. 40108-TZ. Washington, DC: World Bank.

———. 2008a (forthcoming). *Bangladesh Second Investment Climate Assessment: Harnessing Competitiveness for Stronger Growth.* Washington, DC: World Bank.

———. 2008b. *Agriculture for Development*. World Development Report. Washington, DC: IBRD.

———. 2009 (forthcoming). *Reshaping Economic Geography.* World Development Report. Washington, DC: IBRD.

World Bank and Asian Development Bank. 2005. *Sri Lanka – Improving the Rural and Urban Investment Climate.* Washington, DC: World Bank; Manila: Asian Development Bank. http://www.worldbank.lk/WBSITE/EXTERNAL/COUNTRIES/SOUTHASIAEXT/SRILANKAEXTN/0,,contentMDK:20569317~menuPK:287057~pagePK:141137~piPK:217854~theSitePK:233047,00.html.

World Bank/RUTA. 2008 (forthcoming). *Nicaragua: Rural Investment Climate.* Washington, DC: World Bank.

Yang, Dennis Tao. 2004. "Education and Allocative Efficiency: Household Income Growth during Rural Reforms in China." *Journal of Development Economics* 74: 137–62.