

NATSEM

National Centre for Social and Economic Modelling
• University of Canberra •

Regional Microsimulation for Improved Service Delivery in Australia: Centrelink's CuSP Model



Anthony King, Jeannie McLellan and Rachel Lloyd

**Paper prepared for the 27th General Conference,
International Association for Research in Income and
Wealth, Stockholm, Sweden, 18-24 August 2002**

About NATSEM

The National Centre for Social and Economic Modelling was established on 1 January 1993, and supports its activities through research grants, commissioned research and longer term contracts for model maintenance and development with the federal departments of Family and Community Services, Health and Aged Care, and Education, Training and Youth Affairs.

NATSEM aims to be a key contributor to social and economic policy debate and analysis by developing models of the highest quality, undertaking independent and impartial research, and supplying valued consultancy services.

Policy changes often have to be made without sufficient information about either the current environment or the consequences of change. NATSEM specialises in analysing data and producing models so that decision makers have the best possible quantitative information on which to base their decisions.

NATSEM has an international reputation as a centre of excellence for analysing microdata and constructing microsimulation models. Such data and models commence with the records of real (but unidentifiable) Australians. Analysis typically begins by looking at either the characteristics or the impact of a policy change on an individual household, building up to the bigger picture by looking at many individual cases through the use of large datasets.

It must be emphasised that NATSEM does not have views on policy. All opinions are the authors' own and are not necessarily shared by NATSEM.

Director: Ann Harding

© NATSEM, University of Canberra and Commonwealth of Australia 2002

National Centre for Social and Economic Modelling
University of Canberra ACT 2601 Australia
170 Haydon Drive Bruce ACT 2617

Phone + 61 2 6201 2750 Fax + 61 2 6201 2751

Email natsem@natsem.canberra.edu.au

Website www.natsem.canberra.edu.au

Abstract

Centrelink is the agency that has responsibility for the delivery of a range of Australian government services including income support payments and associated services. As such, it has a large infrastructure devoted to contact with customers through a variety of access channels – ranging from walk-in offices to e-mail. A constant challenge for Centrelink is the need to ‘tune’ this infrastructure in response to government policy directions, changing customer patterns, changes in customers’ preferred ways of accessing Centrelink, and new modes of access.

In this environment, and with aims of more efficient/effective service delivery and higher customer satisfaction, Centrelink has embarked on a Regional Microsimulation Modelling Project. The modelling work is being undertaken jointly with the National Centre for Social and Economic Modelling (NATSEM) at the University of Canberra.

The purpose of the model under development, the Customer Service Projection (CuSP) Model, is to provide a tool that will assist decision-making through short to medium-term projection of customers and channel use demands at the small area level and under alternative scenarios of customer numbers, customer characteristics, access preferences and opportunities.

The CuSP Model includes a ‘classic’ static tax-transfer microsimulation model, but this is just one element. Other elements include techniques that ‘regionalise’ the model, a projections and ‘what-if’ capability, extensive and detailed benchmarking to administrative data, and a geographical interface for analysis of model output (with GIS specialists assisting with this latter component).

The paper describes the methods and techniques used in the CuSP Model, and highlights a number of particular issues encountered in this major extension of microsimulation modelling. These include issues with the extensive use of administrative data, spatial units and small area projections, customer behaviour and, importantly, the process of model development.

Author note

Anthony King is a Principal Research Fellow and Rachel Lloyd is a Senior Research Fellow with the National Centre for Social and Economic Modelling at the University of Canberra.

NATSEM
University of Canberra
ACT 2601
Australia

Phone +61 2 6201 2759 +61 2 6201 2764
Fax +61 2 6201 2751
Web www.natsem.canberra.edu.au
e-mail anthony.king@natsem.canberra.edu.au
rachel.lloyd@natsem.canberra.edu.au

Jeannie McLellan is Project Manager, Regional Microsimulation Modelling Project, Centrelink.

Centrelink
PO Box 7788
Canberra Mail Centre
ACT 2610
Australia

Phone +61 2 6208 8555
Fax +61 2 6208 8595
Web www.centrelink.gov.au
e-mail jeannie.mclellan@centrelink.gov.au

Acknowledgments

The authors of this paper are but three among the many people involved in development of the CuSP Model. The authors would like to acknowledge the contributions made by other members of the NATSEM team – Ann Harding, Tony Melhuish, The Dinh Phan and Ben Phillips – and other members of the Centrelink team – Kath Brady, Craig Perkins, Martin Robinson, Fran Sankey and Sharon Warman. The contribution from Otto Hellwig of MDS Market Data Systems is also gratefully acknowledged, as is the assistance with mapping from Marcus Blake (NATSEM) and Chris Medlin (GISCA).

General caveat

NATSEM research findings are generally based on estimated characteristics of the population. Such estimates are usually derived from the application of microsimulation modelling techniques to microdata based on sample surveys.

These estimates may be different from the actual characteristics of the population because of sampling and nonsampling errors in the microdata and because of the assumptions underlying the modelling techniques.

The microdata do not contain any information that enables identification of the individuals or families to which they refer.

Contents

Abstract		iii
Author note		iv
Acknowledgments		v
General caveat		v
1 Introduction		1
2 The role of the CuSP Model		2
2.1 Centrelink	2	
2.2 The Regional Microsimulation Modelling Project	4	
2.3 The Customer Service Projection (CuSP) Model	5	
3 The CuSP Model – methods and techniques		6
3.1 The modelling approach - overview	6	
3.2 The 2001 base population	7	
3.3 Regional weights for 2001	10	
3.4 Channel use propensities for 2001	15	
3.5 Projections	17	
4 The output from the CuSP Model		18
4.1 Summary output	18	
4.2 Analytical output	20	
5 Some particular issues		22
5.1 Administrative data	22	
5.2 The spatial unit	23	
5.3 Customer behaviour	24	
5.4 The model development process	25	
6 Conclusions		27
References		28

1 Introduction

Centrelink is the agency that has responsibility for the delivery of a range of Australian government services including income support payments and associated services. As such, it has a large infrastructure devoted to contact with customers through a variety of access channels – ranging from walk-in offices to e-mail. A constant challenge for Centrelink is the need to ‘tune’ this infrastructure in response to government policy directions, changing customer patterns, changes in customers’ preferred ways of accessing Centrelink, and new modes of access.

In this environment, and with aims of more efficient/effective service delivery and higher customer satisfaction, Centrelink embarked in 2000 on a Regional Microsimulation Modelling Project. Centrelink is undertaking this project in partnership with the National Centre for Social and Economic Modelling (NATSEM) at the University of Canberra and the National Key Centre for Social Applications of GIS (GISCA) at the University of Adelaide. The core microsimulation model for the project – the Customer Service Projection (CuSP) Model – has been developed by NATSEM.

The focus of this paper is the CuSP Model. Section 2 sets out the role of the model – and of the broader Regional Microsimulation Modelling Project – in meeting Centrelink’s interests. This contextual material is followed by a description of the methods and techniques used in the CuSP Model in section 3. Particular features of the CuSP Model include its combination of a tax-transfer microsimulation model with regional microsimulation techniques, a projections capacity, extensive benchmarking, and the use of administrative data. The nature of the output from the model is illustrated in section 4.

Section 5 draws attention to a number of particular issues encountered in this major extension of microsimulation modelling. These include issues with the extensive use of administrative data, with spatial units, customer behaviour and the model development process.

2 The role of the CuSP Model

2.1 Centrelink

Centrelink is the government agency that has responsibility for the delivery of a range of Australian government services including income support payments and associated services. Over one in every three adult Australians is a customer of Centrelink, with a total of 6.3 million customers receiving some form of payment or service.

The major payments and services provided are shown in table 2.1, which identifies the larger categories in the customer population used in the CuSP Model, including their grouping into 'customer segments'. While most of the items have self-explanatory names, it is probably useful to point out that Newstart Allowance and Youth Allowance are the main payments for unemployed people looking for work. The four largest payments – Age Pension, Family Tax Benefit, Newstart Allowance and Disability Support Pension – together account for 62 per cent of customers. While table 2.1 separately shows 14 types of benefit (payment or service), the customer data used in the CuSP Model identifies 43 main benefit types¹.

To meet the needs of these 6.3 million customers, Centrelink has a vast physical infrastructure. This includes over one thousand service delivery points Australia-wide (figure 2.1), including:

- around 300 Customer Service Centres (CSCs);
- 28 Call Centres;
- specialist service centres (such as Retirement and Family Service Centres);
- Centrelink Community Agents;
- Visiting Services; and
- Access Points.

Interaction between Centrelink and its customers is achieved through this physical infrastructure and a range of access channels. These include telephone, mail, face-to-face appointments, walk-in face-to-face contact, and e-mail. The scale of the operation is indicated by the following aggregate figures (Centrelink 2001).

¹ The CuSP Model at present, however, does not separately distinguish all 43 main benefit types. Initial versions disaggregated customers by customer segment, with the current version further identifying some of the major types of benefit (see section 3).

- Centrelink sends around 100 million letters to customers each year;
- Centrelink receives around 22 million telephone calls from customers each year; and
- Around 650 000 office appointments are booked with customers each month.

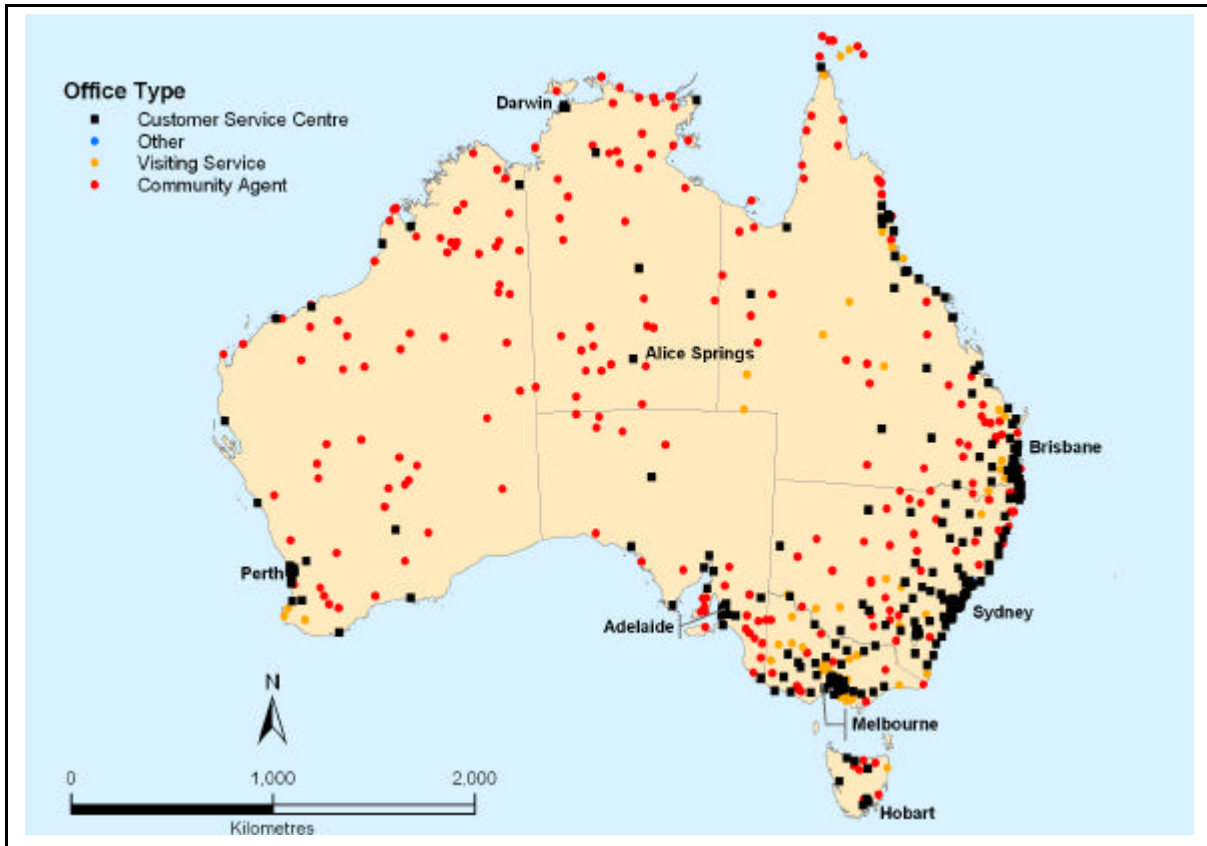
Table 2.1 **Centrelink customers – selected main benefit types^a, June 2001**

Customer segment	Main benefit type	Percentage of customers ^b
Retirement	Age Pension	26.4
	Seniors Health Card	3.5
	Other	0.4
	(Subtotal)	(30.3)
Disability/Carers	Disability Support Pension	9.2
	Other	2.6
	(Subtotal)	(11.8)
Employment	Newstart Allowance	9.3
	Partner Allowance	1.4
	Job Seeker Registration	4.2
	Other	1.4
	(Subtotal)	(16.3)
Youth and Students	Youth Allowance	6.1
	Other	1.3
	(Subtotal)	(7.4)
Families and Children	Parenting Payment (Partnered)	3.1
	Parenting Payment (Single)	6.3
	Child Care Benefit	1.7
	Family Tax Benefit	17.4
	Other	0.1
	(Subtotal)	(28.6)
Other	Health Concession Card Holder	1.7
	Low Income Card	2.2
	Pension Concession Card	1.4
	Other	0.2
	(Subtotal)	(5.5)
Total		100.0

a Where a customer is receiving more than one payment or benefit, 'main benefit type' is assigned according to a ranking that gives priority to base income support payments. For example, someone receiving Newstart Allowance and Family Tax Benefit would be classified here as having Newstart Allowance as their main benefit type.

b The population of customers for this table is the population used in the CuSP Model. This may differ from other figures on the Centrelink customer population due to differences in the scope of customers covered – payments to overseas customers are not included in the CuSP Model for example – and in the treatment of 'non-current' customers.

Figure 2.1 Centrelink service delivery points



Data source: GISCA and Centrelink

2.2 The Regional Microsimulation Modelling Project

A constant challenge for Centrelink is the need to ‘tune’ this infrastructure in response to government policy directions, changing customer patterns, changes in customers’ preferred ways of accessing Centrelink, and new modes of access. This provided the motivation for the Regional Microsimulation Modelling Project.

The objectives of the Project are to provide Centrelink with the capability to assess future property and staffing requirements and service access options at both macro and local level. The capability needs to take into account:

1. the changing demographic and economic environment;
2. decisions about new or changed directions in service delivery, including electronic service delivery;
3. the impact of government policy changes;
4. the impact of Centrelink acquiring responsibility for new groups of customers; and
5. customer requirements and preferences.

The Project commenced in September 2000 with the first year devoted to a 'proof of concept' or feasibility stage and then development of a prototype model. From late 2001 to June 2002, the emphases of the Project have been on:

1. transforming the prototype model to a functioning model (the CuSP Model); and
2. development of a geographic management information interface for analysis of current and projected channel use levels and patterns (with a correspondingly increasing role for GISCA in the Project).

At the time of writing, planning for the next phase of the Project is underway, with a focus on application of the work to priority areas of Centrelink's strategic decision-making.

2.3 The Customer Service Projection (CuSP) Model

The task for the CuSP Model flows directly from the objectives of the broader Project. The model needs to be able to project:

- Centrelink customer numbers and characteristics; and
- associated channel use;

at the small-area level.

The time horizon for the projections is five years into the future, and the projections need to be made under scenarios that can capture alternative settings for those aspects of the future environment listed in section 2.2. That is, the model needs to be able to accommodate changes in the demographic and economic environment, changes in government policy, changes in service delivery directions, and changes in customer behaviour.

How the CuSP Model meets all these requirements is the subject of the next section.

3 The CuSP Model – methods and techniques

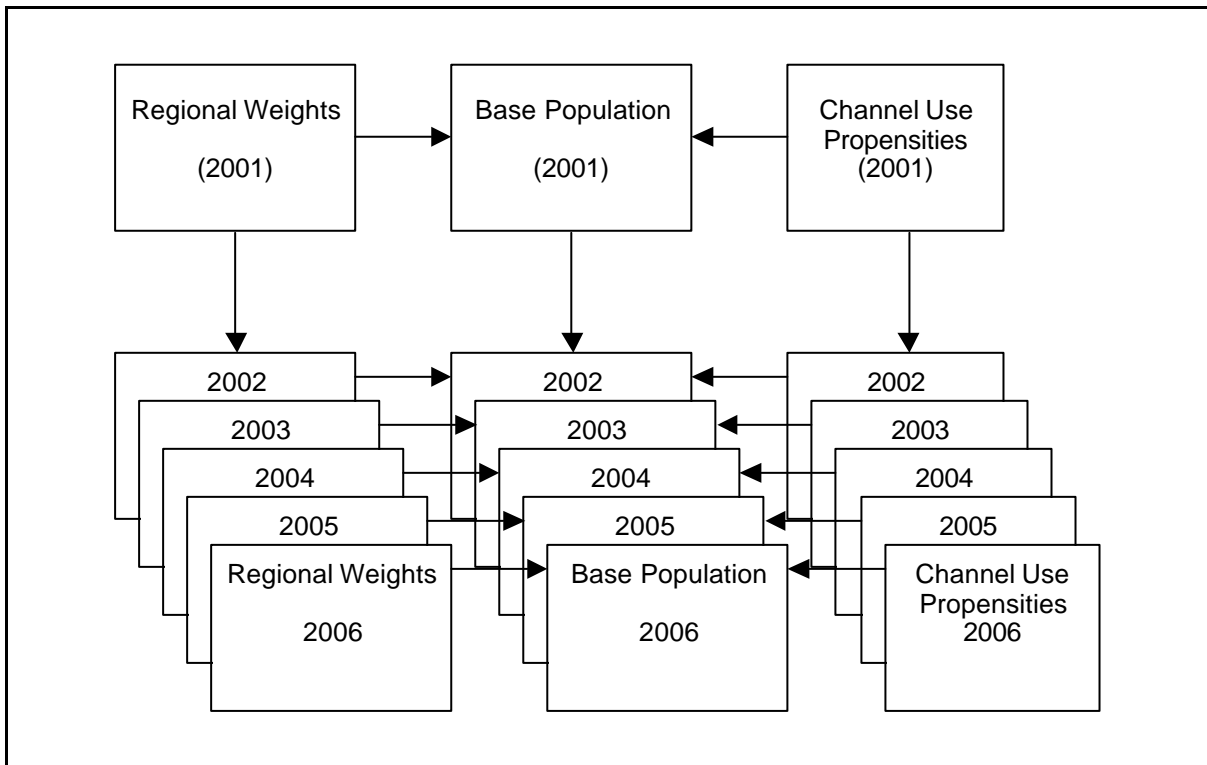
3.1 The modelling approach - overview

The core of the CuSP Model is a combination of a standard static microsimulation model, regional microsimulation techniques, and benchmarking to administrative data. Essentially, the model involves the generation and linking of three datasets (figure 3.1):

1. a base population file (which includes a pool of detailed records for individuals);
2. small area weights (which specify the prevalence of each individual in the pool in each small area); and
3. channel use propensities (which specify expected channel use given a person's characteristics).

Application of the weights to the base population file gives the estimate of the numbers and characteristics of Centrelink customers in each small area. Application of the channel use propensities to the customer estimates then gives the estimate of channel use. These three datasets are produced for the base year (currently 2001) and then each is updated for the projection years (currently 2002 to 2006).

Figure 3.1 **CuSP Model – overall structure**



The CuSP Model incorporates elements of two other models:

- STINMOD – a tax-transfer microsimulation model; and
- MarketInfo – a regional microsimulation model

and the following major data inputs from the Australian Bureau of Statistics (ABS), Centrelink and other sources:

- small area data from the 1996 Australian Census of Population and Housing;
- unit record data from the 1998-99 ABS Household Expenditure Survey (HES);
- detailed Centrelink customer data;
- detailed Centrelink channel use data; and
- various data for the construction of projections scenarios.

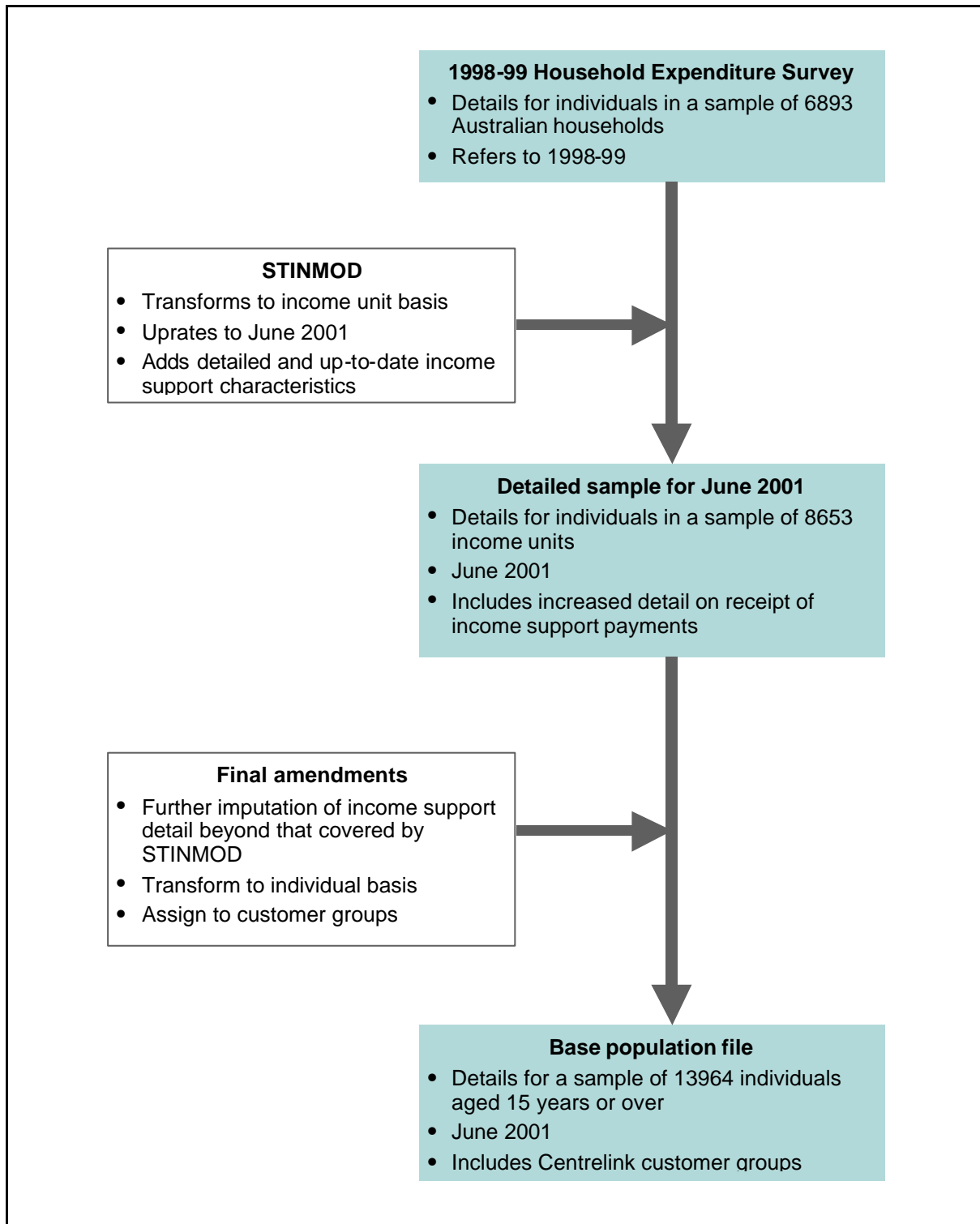
The way in which these model elements and data inputs are brought together in the CuSP Model is described below – looking firstly at the generation of each of the three parts of the model for 2001 (the base population, regional weights, and channel use propensities) and then covering the projections of these three datasets.

3.2 The 2001 base population

Centrelink customers within the context of the population at large are represented by a base population file that includes information about a range of characteristics for a sample of individuals. A series of weights is subsequently attached to this dataset to reflect the composition of the population in each postal area.

The base population file is derived from the confidentialised unit record file from the ABS 1998-99 Household Expenditure Survey (ABS 2000), using the steps shown in figure 3.2. The Household Expenditure Survey (HES) is a national sample survey that provides detailed information on the characteristics, incomes and expenditures of the individuals aged 15 years and over in almost 7000 Australian households².

² While the current version of the model uses the 1998-99 HES as the basis for the base population file, there are possible alternatives. These include unit record data from the ABS Survey of Incomes and Housing Costs and the 1% sample from the Australian census. Each has its strengths and weaknesses in terms of population coverage and variables and, importantly, in terms of timing. At the moment, the 1998-99 HES provides the most up-to-date suitable unit record dataset.

Figure 3.2 **Generation of the 2001 base population file in the CuSP Model**

While the HES includes a considerable level of detail on people's sources of incomes, the CuSP Model requires additional detail on the receipt of particular payments. Also, we need characteristics that reflect circumstances in June 2001, not 1998-99, and there have been important changes in income support payments over that period – in particular, the tax-transfer changes associated with the introduction of a GST in 2000.

To address these shortcomings, the algorithms in NATSEM's well-established static microsimulation model, STINMOD, are applied to the 1998-99 HES dataset.

STINMOD

STINMOD (Lambert et al 1994) has primarily been designed for the up-to-date and detailed analysis of the distributional and budgetary effects of income support and taxation policy. The standard base data for STINMOD is the unit record data from the latest ABS Survey of Incomes and Housing Costs (SIHC) – though it can also run on the HES data as is the case here. The core of STINMOD then comprises two parts:

1. the application of reweighting and uprating techniques to bring the survey data up to date (SIHC data is typically about two years old); and
2. algorithms to impute income support entitlements and income tax liabilities.

STINMOD is regularly updated to maintain its currency, and as more recent base datasets become available. The specific version of STINMOD used here was STINMOD-01B, which includes the income support system as at June 2001. In this case, the application of STINMOD to the 1998-99 HES data involved the following steps:

1. Uprating of private income characteristics – such as earnings and investment incomes – from 1998-99 terms to June 2001.
2. Conversion of the household structure of the HES data (individuals nested within households) to the income unit structure (individuals nested within income units), which is needed to model income support entitlements.
3. Imputation of certain characteristics required by STINMOD that are not present in the HES – for example, single-year ages for some age groups.
4. Imputation of income support entitlements as at June 2001 according to the rules in STINMOD-01B.

Note that we are not concerned with the reweighting element of STINMOD, which in a normal application would be used to account for demographic and labour market change since the time when the survey was conducted. The task here is to create an up-to-date sample with a range of characteristics – including detailed information on the receipt of income support – which will then be subject to reweighting at the small area level (covered in the following section on regional weights).

Supplementary imputations

The result of the application of STINMOD to the 1998-99 HES data is generation of a unit record dataset with a sample of almost 14 000 individuals in around 8500 income units, with their characteristics updated to June 2001 terms and the addition of greater detail on the receipt of income support payments. While the raw HES data would have allowed us to distinguish 11 major income support payments that cover about 80 per cent of Centrelink customers, the use of STINMOD results in identification of 18 payments that cover about 85 per cent of the population – as well as bringing the dataset into line with the current income support system.

Still, there are a number of significant benefit types that are not covered by the standard STINMOD routines. These primarily cover benefits that involve no payment but provide an entitlement to free or discounted services – such as the Seniors Health Card. These customers have been identified through supplementary imputations with reference to eligibility conditions, the known characteristics of these customers, and the receipt of other benefits. After these supplementary imputations have been undertaken, 31 of the 43 main benefit types identified in Centrelink data are explicitly covered, with these types of benefit accounting for over 99 per cent of customers.

The final steps in generation of the base population file are transformation of the dataset to an individual basis (from one with individual details on income unit records) and assignment of Centrelink customer status. Customer status identifies whether or not someone is a Centrelink customer and, if so, the type of customer. Benefit type is generally grouped into ‘customer segment’ (see table 2.1), though with some additional identification of particular benefit types.

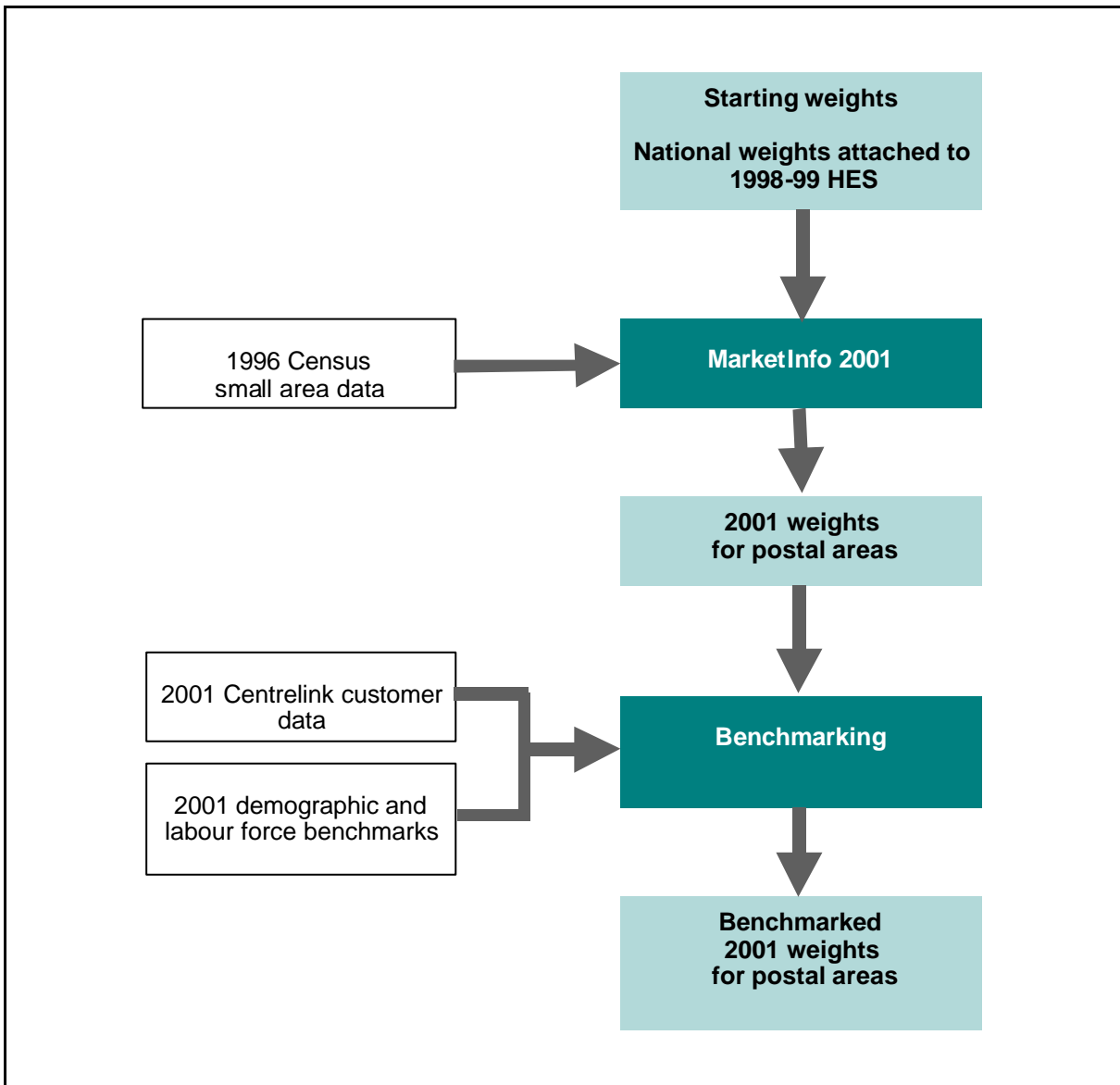
3.3 Regional weights for 2001

The second element of the CuSP Model is the set of postal area weights for 2001. These weights provide the link between the pool of cases in the 2001 base population file and the characteristics of each postal area. The ‘postal area’ is an ABS approximation of the postcodes used by Australia Post. There are 2388 postal areas across the country with an average population of about 8000 people, though with considerable variation in their size – the largest includes a population of around 90 000 people. As a spatial unit for this model, postal areas are not without their difficulties – which are discussed in section 5.2 – but they have been adopted as the spatial unit due to current constraints on available geographic data.

The derivation of the postal area weights for 2001 is shown in figure 3.3. There are two steps:

1. the generation of regional weights using MarketInfo 2001; and
2. the benchmarking to Centrelink administrative data and other key characteristics.

Figure 3.3 **Generation of the 2001 regional weights in the CuSP Model**



Initial regional weights

Regional weights are devised using MarketInfo 2001 – a model for the analysis of small area household expenditure, which is a joint commercial product of NATSEM and MDS Market Data Systems³. The MarketInfo techniques have an established reputation in commercial applications – such as market identification – and have also increasingly been used in socio-economic analyses (Lloyd, Given and Hellwig 2000; Lloyd, Harding and Hellwig 2001). MarketInfo 2001 updates the small area data from the 1996 Australian census to 2001 and links it with the unit record data from the ABS 1998-99 HES to produce a set of weights for the HES for each Census Collection District (CCD). There are 34 410 CCDs across the country, with an average size of around 200 households.

The link in MarketInfo between the small area Census data – the tabular socio-demographic profiles provided in the ABS product CDATE – and HES data is provided by a selection from the limited set of household and personal characteristics that are common to both datasets. The characteristics used – mostly single variables, but some multivariate characteristics – are listed in table 3.1.

Table 3.1 Characteristics used in MarketInfo 2001 reweighting

Characteristic

Age
 Total individual income
 Marital status
 Labour force status & gender
 Country of birth
 Occupation
 Family type
 Student status
 Level of highest qualification
 Age & income market segments
 Housing type
 Housing tenure
 Household size
 Number of motor vehicles
 Level of mortgage repayments
 Level of rent payments

Source: MarketInfo

³ See www.natsem.canberra.edu.au/research/marketinfo/marketinfo.html

From the characteristics listed in table 3.1, 68 reweighting target variables are defined – for example: unemployed female, age 45-54 years, private rental housing. The link then proceeds through the following steps.

1. Some amendments are made to both datasets to ensure consistent variable definitions and categories for the linkage characteristics.
2. The linkage variables in the 1996 Census data are updated to 2001 using cross-sectional time series regressions.
3. The updated Census variables for each CCD are used as targets for reweighting the 1998-99 HES unit record dataset. Given the number of variables involved and the diversity of small area populations, this reweighting task is a major exercise. It involves use of a suite of reweighting algorithms that has been devised and tailored over the course of the development of the MarketInfo model.

The output from this process is a separate set of HES weights for 2001 for each of the 34410 CCDs. While the CCD is the basic unit of analysis for MarketInfo, the CuSP Model operates with the larger unit of postal areas. There is an exact concordance between CCDs and postal areas and the CCD weights are accordingly aggregated to the postal area level.

Benchmarking

While MarketInfo 2001 provides us with sets of weights for small areas, that step of the reweighting does not include the level of detail on Centrelink customers that is required in the CuSP Model. The MarketInfo 2001 weights will give us a good estimate of the number of Centrelink customers in each postal area, but will not necessarily give a good estimate of the number and characteristics of particular groups of customers. Accordingly, a second reweighting step is applied to benchmark the data to detailed Centrelink administrative customer data.

The Centrelink customer dataset is a cross-section of Centrelink customers in June 2001. It covers 6.5 million customers and provides the number of customers classified by a range of customer characteristics including their postcode (table 3.2). There is thus scope to apply very detailed benchmarking for Centrelink customer numbers and characteristics. At this stage, however, we are only benchmarking by postcode, main benefit type, age and sex. For this purpose, main benefit type is generally grouped into 'customer segment' (see table 2.1), though with some additional identification of particular benefit types.

Table 3.2 Main customer characteristics included in the June 2001 Centrelink customer dataset

Customer characteristic
Postcode
Age group
Sex
Marital status
Main benefit type
Country of birth
Preferred language
Rent Assistance flag
Student flag
Presence of dependents flag
Indigenous flag

The geographic indicator in the customer dataset is postcode. This is an Australia Post unit that is close to, but far from identical to, the ABS postal areas that provide the current spatial unit for the CuSP Model. The issue of spatial units is given particular attention in section 5.2, while two ramifications of the mismatch between the two units need to be mentioned here.

1. Customers in postcodes that do not correspond directly to an ABS postal area are excluded from the model. This affects less than one per cent of customers.
2. Those postal areas where the mismatch between postcode and postal areas appears to be most serious are amalgamated with neighbouring postal areas. This results in the number of small areas in the model being reduced from 2388 to 1803.

Besides benchmarking to Centrelink administrative data, this step is also used to benchmark to the demographic data series that is used in the projections part of the CuSP Model. While both MarketInfo-2001 and the CuSP Model projections use data from the ABS, the need for this demographic benchmarking stems from some differences in the two demographic data sources.

This benchmarking step is accomplished by nesting the Centrelink customer benchmarks within the ABS demographic targets (specified by postal area, age and sex), and then the simple application of cell-based or matrix reweighting. In an earlier version of the model, using different input data, this step also included benchmarking against regional labour force data. In that case, multi-vector reweighting was undertaken using the CALMAR package developed at INSEE (Deville and Särndal 1992). One vector included postal area, age, sex and Centrelink customer status, while the second vector included postal area, age, sex and labour force status (with distinction between full-time employment, part-time employment, unemployment and being out of the labour force).

3.4 Channel use propensities for 2001

Channel use propensities measure the average likelihood of a Centrelink customer with given characteristics using a particular access channel over a 12-month period. Besides the detailed cross-section of Centrelink customers, the second major administrative dataset used in the CuSP Model is a channel use dataset. This dataset covers all instances of use of certain channels over the 12 months of 2001. Alongside each instance of channel use are recorded the same customer characteristics as those collected for the customer cross-section (see table 3.2). The dataset is reduced to a matrix format – giving the number of instances of each channel use and the number of customers for each combination of the classifying characteristics – though is still a sizeable file with 4.7 million records.

The following channels are covered by the channel use dataset:

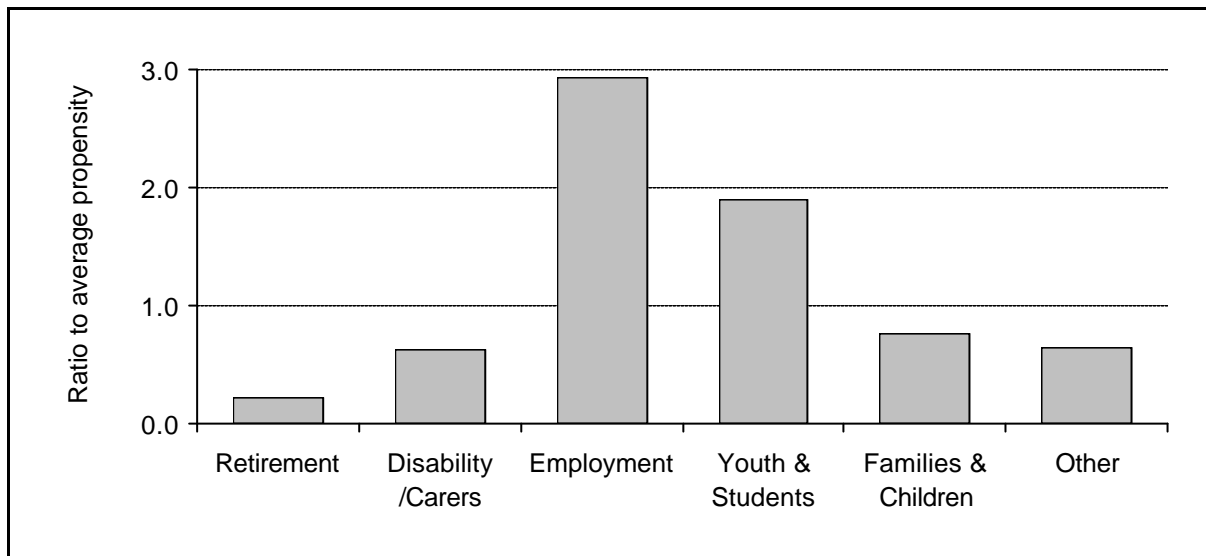
- outgoing mail (from Centrelink to customers);
- incoming telephone calls (from customers to Centrelink Call Centres);
- hosted appointments – internal (appointments held in Centrelink offices);
- hosted appointments – external (for example, at the customer's home);
- hosted appointments – seminars;
- hosted appointments – callback (arranged telephone calls from Centrelink to customers); and
- secure internet messages (incoming e-mail from customers).

These channels are those for which data can be extracted from the Centrelink data system, but do not cover all channels. Notable omissions are incoming mail, outgoing phone calls, and 'walk-ins' – that is, cases where a customer simply comes into a Centrelink office without an appointment. Walk-ins are understandably very common, as well as being an expensive form of channel use for Centrelink, so their absence from the channel use dataset is a serious gap. In response, Centrelink have initiated a sample data collection of walk-ins and incoming mail, and the initial data from these collections is being incorporated in the CuSP Model.

Channel use propensities vary considerably between different groups of customers. For a start, they are affected by the different reporting requirements attached to each payment. For example, Newstart customers – people looking for work – generally need to submit fortnightly applications for continuation of payment, while Age Pensioners only need to contact Centrelink if there is a change in their family or financial circumstances. Channel use propensities are also influenced by customers' channel use preferences and constraints. For example, those with easy access to a Centrelink office are more likely to choose to visit the office than are those with more difficult access.

The basic variation in channel use propensities by customer segment is illustrated in figure 3.4, which shows the propensities for internal appointments by customer segment. There is clearly considerable variation across customer segments in the propensity to use this channel.

Figure 3.4 Propensities for office appointments in 2001: Centrelink customers by segment (propensity as ratio of average propensity across all customers, average = 1.0)



Data source: Centrelink 2001 channel use dataset.

Analysis of the channel use dataset has revealed further systematic variation in the channel use propensities according to a range of customer characteristics. Based on this analysis, the propensities in the basic current version of the CuSP Model are specified for groups defined by:

- customer segment (with further identification of specific payments in some applications);
- sex;
- age; and
- region (generally a broad distinction between metropolitan, other urban, and other areas, but with finer regional specification for some applications).

At this stage, we have created the three elements of the 2001 model (see figure 3.1). The first element, the 2001 base population, is the enhanced unit record dataset that provides a pool of 13 964 cases for the model. The second element is a set of weights for each of these cases for each of the 1803 postal areas (or amalgamation of postal areas). The third element is the channel use propensities. The fourth part of the model involves the projection of each of these datasets for future years.

3.5 Projections

The elements of the CuSP Model for 2001 shown in figure 3.1 have accounted for most of the modelling effort, but really amount to only the foundation for the model. The main role for the CuSP Model is to provide short to medium-term projections of Centrelink customers and channel use under alternative scenarios. The projections modules of the CuSP Model involve creating projected versions of the three model elements, with opportunities in each case to vary the assumed future environment.

Base populations for the projection years are generated by repeating the steps described in section 3.2 for creation of the 2001 base population, though with the addition of the STINMOD ‘outyears’ facility. This involves the specification of income support parameters and the rates of growth in components of private income, such as earnings or investment incomes. It also provides the opportunity to impute entitlements under alternative structures for the income security system.

The sets of *regional weights* are projected through adjustment in line with demographic and labour force projections. At present, the demographic projections used are ABS postal area projections of the population by age and sex⁴. Labour force projections are currently applied with distinction between full-time employment, part-time employment, unemployment and being out of the labour force. The basis for these future demographic and labour force scenarios is discussed further in section 5.2.

Finally, the channel use *propensities* for future years can be amended to incorporate views about possible changes to access channels and requirements, or changes in customer behaviour.

In summary, the CuSP Model allows projections of customer numbers and channel use to be made under future scenarios that vary according to alternative specifications of:

- small area demographics;
- small area labour market structure;
- private incomes;
- the income security system; and
- channel use propensities.

⁴ This is a specific set of projections prepared by the ABS according to assumptions agreed to by NATSEM.

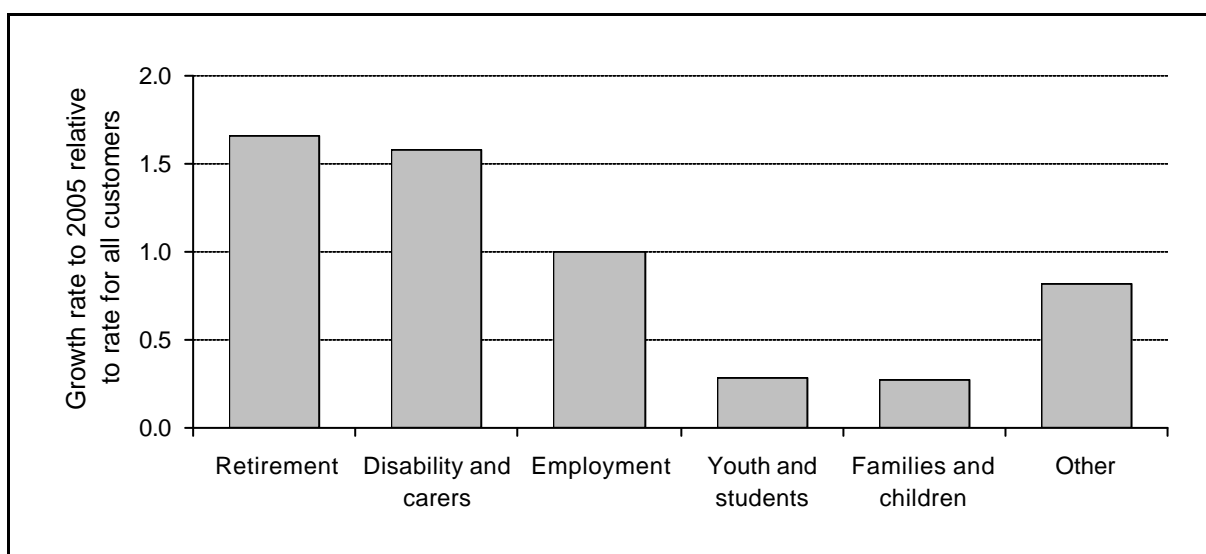
4 The output from the CuSP Model

The CuSP Model can clearly generate a considerable amount of output – detailed data on customers and channel use for each small area and for each time point in the projection period, with projections generated under alternative scenarios about the future environment. Here, the nature of CuSP Model output is illustrated with some results that move from summary model output to a discussion of the more detailed analytical output.

4.1 Summary output

Through aggregation of the detailed results, the CuSP Model can be used to generate the type of broad summary outcomes illustrated in figure 4.1. These are projected growth rates for the Australia-wide number of customers by customer segment over the period from 2001 to 2005. The scenario underlying these projections was one in which the only future aspect varied was population size by age and sex at the small area level. The projection produced an overall growth in customer numbers due to increasing population size, but different growth rates for customer segments due to the gradual ageing of the population. The growth rates in figure 4.1 are expressed in terms of their relativity to the overall customer growth rate. The highest growth rates are for customers in the ‘retirement’ and ‘disability/carers’ segments. The lowest growth rates are for customers in the ‘youth and students’ and ‘families and children’ segments.

Figure 4.1 **Illustrative CuSP Model simulation. Projected growth in customer numbers by segment: June 2001 to June 2005 (growth rate relative to growth rate for all customers, growth rate for all customers = 1.0)**

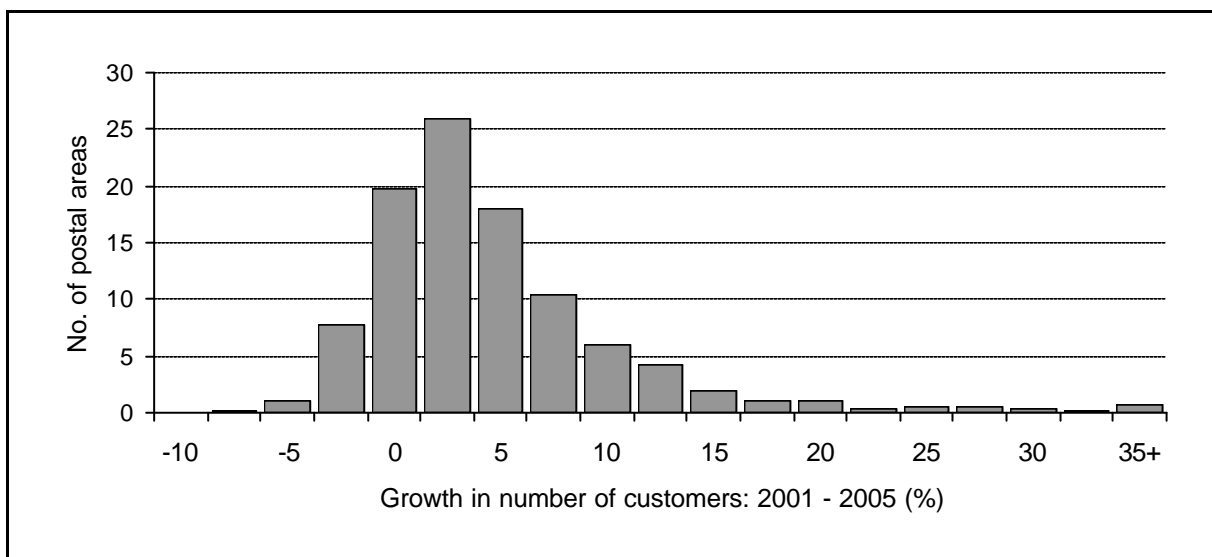


Data source: Centrelink CuSP Model illustrative simulation

These differences in growth rates by customer segment have a corresponding impact on the structure of the population of Centrelink customers – in particular, increased shares for the ‘retirement’ and ‘disability & carers’ segments and reduced shares for the ‘youth & students’ and ‘families and children’ segments.

While total customer numbers are projected to increase under this illustrative simulation, the projected outcomes are very different across postal areas (figure 4.2). Not all areas can expect a growth in customers. Around 10 per cent of postal areas are projected to experience a decline in customer numbers, and about 20 per cent to see no marked change in numbers. Among those postal areas projected to see a growing number of customers, a significant number of them (around 10 per cent of all areas) can expect to see an increase in customer numbers of more than 10 per cent over the four-year period. This diversity of outcomes is evident across all customer segments. It reflects the concentrations of customers – and of particular types of customer – in different areas, as well as the areas’ differing demographic paths.

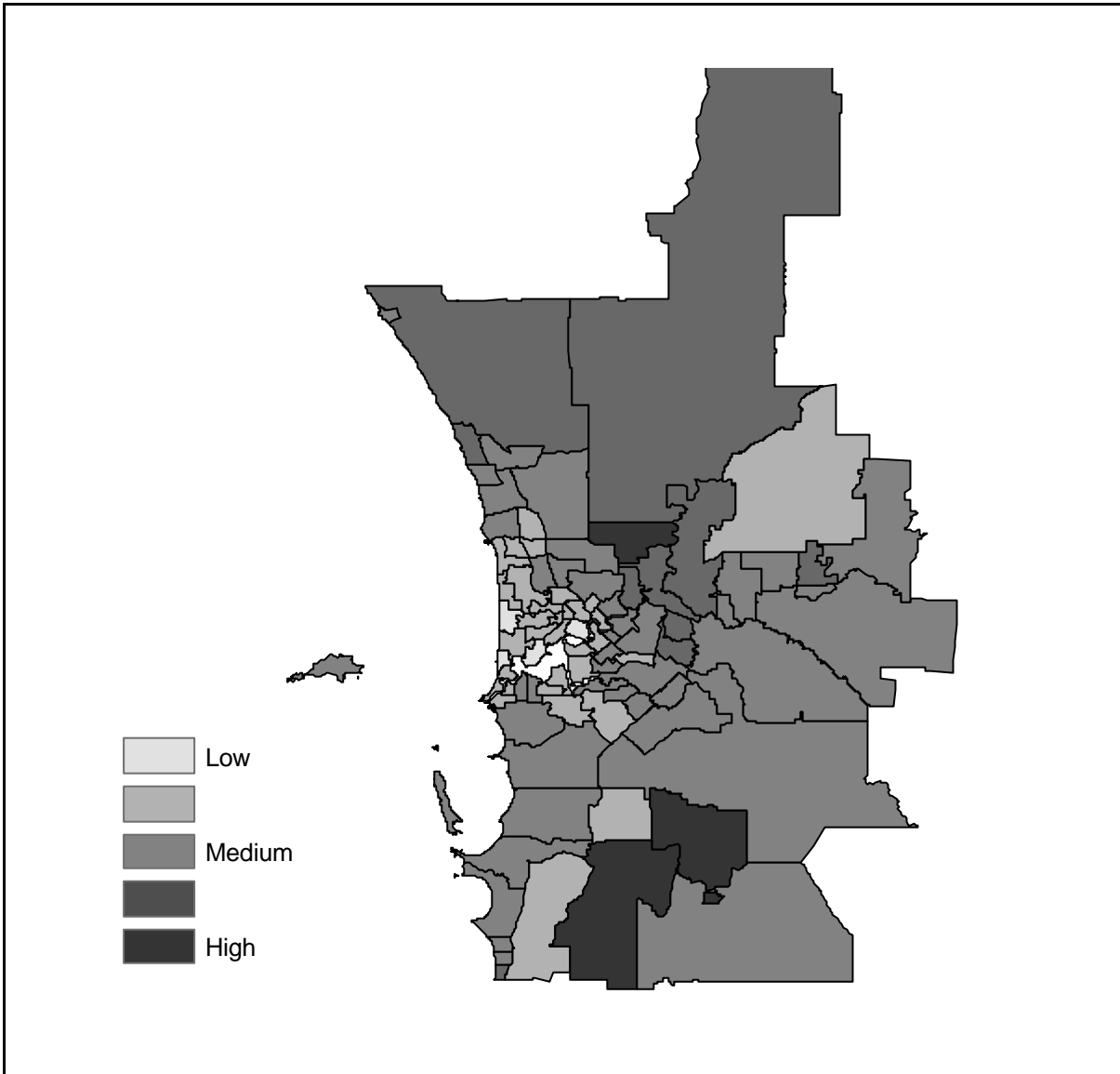
Figure 4.2 **Illustrative CuSP Model simulation. Distribution across postal areas of projected growth in customer numbers: June 2001 to June 2005**



Data source: Centrelink CuSP Model illustrative simulation

Similar summary output can of course be produced for projected growth in channel use and its distribution across postal areas. Given the range of outcomes across the country, the next question is to identify those areas with high projected growth rates and those with low projected growth rates. The best way to do this is through mapping the output, as is illustrated in figure 4.3. Figure 4.3 shows the projected growth rate to 2005 for one particular access channel for the postal areas within the Perth metropolitan area of Western Australia. This illustrative output is taken from a simulation that incorporated demographic change, labour market change, and change in the way Centrelink was delivering a particular service.

Figure 4.3 **Illustrative CuSP Model simulation. Projected growth in channel use June 2001 to June 2005: postal areas in metropolitan Perth.**



Data source: Centrelink CuSP Model illustrative simulation

4.2 Analytical output

While the type of output illustrated above can give a good feel for the broad terms and patterns in the output from a CuSP Model simulation, it is still some way away from being in a form that meets those specific project objectives set out in section 2.2. These objectives called for the capability to assess future property and staffing requirements and service access options at both macro and local level. A means to handle the extensive detailed output from the model and to place this in the context of the local infrastructure for Centrelink service delivery is required.

This is where the other major part of the Regional Microsimulation Modelling Project comes into play; namely, a spatial interface. This interface is being developed by Centrelink and GISCA, with input from NATSEM, and allows current and projected future customer numbers and channel use to be mapped with overlays of other aspects, such as the catchment areas for the Centrelink Customer Service Centres or particular population characteristics.

5 Some particular issues

5.1 Administrative data

The development of the CuSP Model is one example of the growing emphasis on the use of administrative data in microsimulation modelling activities at NATSEM. It is true that the base data for the model is derived from the dataset from a sample survey of the Australian population, but administrative data are used extensively in benchmarking the model – through reweighting – and in the calculation of channel use propensities. As such, we have encountered the common issues with the use of administrative data.

First, administrative datasets have a tendency to be large – with millions rather than thousands of records. Not so long ago this would have amounted to a real constraint on our use of such datasets in modelling. But with today’s computing power, the cost of this data coverage and detail is simply that things run a little more slowly. The size of the datasets is no longer a real issue.

On the other hand, the second issue – concerns to assure the confidentiality of the data – do remain a real and serious issue. In the development of the CuSP Model, one option at the outset was for the modelling work by NATSEM to be undertaken within Centrelink. Such working arrangements have been used in other NATSEM modelling using administrative data. For practical reasons, however, that was not the preferred option. Instead, NATSEM has access to the administrative data on its own premises. This was made possible by:

- the project operating under strict provisions for data confidentiality and security; and
- administrative data being generally provided in detailed matrix format, rather than as unit records, with identifiers removed.

A third issue with administrative data, as its name suggests, is that it is typically collected for operational rather than analytical reasons. The form of the ‘raw’ administrative data can be a long way from the required analytical dataset, and the Centrelink administrative datasets were no exception. Behind the clean Centrelink datasets received by NATSEM, there has been a huge effort in their compilation. Data has needed to be merged from a variety of datasets within the vast array of information held by Centrelink, and then the variables need to be made consistent. For example, because different benefits have different eligibility and entitlement rules, the same information is not collected from all customers, and what might seem to be the same information often turns out to be different information. Transforming

the administrative data into a useable dataset for the CuSP Model has involved a great deal of work and an intimate understanding of the intricacies of the administrative data system.

This effort by Centrelink has, however, had a major and unanticipated bonus besides providing the necessary input for the CuSP Model. The creation of the customer dataset and particularly the channel use dataset, and their annual updating, has unlocked information of considerable value to the organisation. These datasets alone can provide Centrelink with a clear and detailed picture of current customers and channel use at the small area level. Previously, only parts of this picture could be readily obtained. Given the usefulness of these datasets, an element of the Regional Microsimulation Modelling Project that has grown in importance has been the development of the user-friendly spatial interface for presentation of these data.

5.2 The spatial unit

The small area unit in the CuSP Model is the ‘postal area’ – an ABS approximation of the postcodes used by Australia Post. Besides its suitable size, the primary reason for adoption of this unit was the ready availability of postcode as a geographic identifier in the Centrelink customer data. It also has the advantage that the catchment areas for Centrelink Customer Service Centres tend to be defined in terms of postcodes. However, because postal areas and postcodes are not identical, linking data for postcodes (such as Centrelink customer numbers) and data for postal areas (such as ABS population numbers) presents some difficulties as described in section 3.3. A further disadvantage with postal areas is that they are a relatively uncommon unit. For example, while the ABS does produce demographic projections for postal areas, small area data that could be used for benchmarking the CuSP Model and for projections scenarios is more typically provided for other units such as ‘statistical local areas’.

With the spatial unit for the model being constrained by the spatial information in the Centrelink administrative data, we are currently limited to the use of postal areas. An alternative is, however, in prospect. This stems from current Centrelink work on geocoding customer addresses – a major task that has commenced but is still some way from completion. Geocoded customer data would allow us to assign customers to the very fine spatial units of Census Collection Districts (CCDs). With an average of around 200 households in each CCD, these units are much smaller than required for most analytical purposes, but their strength lies in their flexibility. The CCD is the basic building block for Australian geographic classifications (ABS 2001) and its use would allow aggregation to different spatial units for different purposes. That will be possible when geocoded customer data become available. In the meantime, we continue to use postal areas, with work underway to minimise the

associated difficulties by developing a better concordance between postcodes and postal areas.

The spatial unit is clearly an issue in setting the scenarios for projections using the CuSP Model. Demographic projections for small areas are readily available, being produced by the ABS and also by State and Territory governments. As noted above, these are generally produced for units such as 'statistical local areas' rather than for postal areas. Labour market projections for small areas, on the other hand, are not readily available. To date, the CuSP Model has generally been run using labour market scenarios specified at a broad level – for example, an assumed change in the unemployment rate applying across the whole country. To fill this gap, Centrelink is currently seeking to contract the generation of small area labour market projections from an economic forecasting organisation.

5.3 Customer behaviour

Another important element of projection scenarios for the CuSP Model is customer behaviour, as measured by their propensities to use different access channels. Take the case, for example, of the propensity to use e-mail. At present, these propensities are highest for young people and for people receiving family payments, with this pattern being a likely reflection of the degree of adoption of the technology and access to e-mail (many people receiving family payments are working, for example). Now, if we run a straight demographic projection with the CuSP Model, the ageing of the population sees relatively low growth rates for these groups of customers and a consequently low growth rate for the e-mail channel. Is this plausible? Most of us would expect the opposite outcome – for the use of e-mail to show high growth as the rate of adoption of the technology outweighs any demographic effects. The point is to highlight the importance of giving explicit attention to channel use propensities in the projections.

To some extent, customers' channel use is constrained by their reporting obligations and more generally by Centrelink's approach to service delivery. But to a large extent they also have choices in how often and in what way they contact Centrelink. The element of future channel use propensities that is related to basic service delivery procedures can be incorporated in scenarios in a reasonably straightforward manner. The real questions arise where there is customer choice. Which channels are customers likely to use, given their preferences and the encouragement or degree of access provided by Centrelink, and what would be the effect of change in one channel on use of the other channels? Research in this area is fundamental for the CuSP Model though has been limited to date, using some information that can be gleaned from the various customer surveys undertaken by Centrelink. In response to this need, Centrelink has begun new research designed to elicit information on these

issues. Specific research into the pattern of adoption of internet technology is also in the planning stage.

5.4 The model development process

A distinctive feature of the CuSP Model has been the process of model development – in particular, the model development strategy and the nature of project management.

Centrelink's Regional Microsimulation Modelling Project is ambitious, as are the capabilities required of the CuSP Model. Indeed, at the outset it appeared a somewhat daunting task, particularly given the demands to be made on administrative data. While development of the model has been conceived as a 4 or 5-year project, it has, however, been split into discrete stages with project continuation dependent on successful completion of each stage. The work thus began with a feasibility stage – which revealed the availability of considerably better administrative data than anticipated – and this was followed by development of a prototype model, leading to the current work on the model proper. Development continues to proceed in an incremental manner, with the basic model having been brought to a robust and useable point prior to the addition of further elements and detail in incremental steps. The ability to produce early output from the modelling project has been an important feature. The overall design is for a regularly updated core model, with add-on modules designed to address specific issues.

With regard to project management, a project management team within Centrelink brings together the three organisations involved in the project – Centrelink, NATSEM and GISCA – in what becomes virtually continuous project monitoring and planning. This, together with detailed quarterly reporting against milestones, means that the project has a high administrative overhead – for all parties. The benefits of this overhead are, however, manifest in a number of ways. For a start, the work has proceeded much more as a partnership between the three organisations, rather than as a straight client-contractor relationship. This in turn has led to development of a good understanding of the requirements and capabilities of the respective parties. For example, Centrelink clearly understands NATSEM's data requirements for model development, while NATSEM clearly understands the difficulties involved in extracting certain administrative data items. This approach has also contributed to a better shared understanding of and focus on the objectives and outcomes of the project.

A second valuable feature of the approach to project management has been the flexibility it entails. This type of model development proceeds with a degree of uncertainty about how long particular tasks will take. It has, however, been possible

to revise the project plan in the light of matters such as data not becoming available when originally expected. At a broader level, one role of the project management team has been to engage with the various areas of Centrelink that have an interest in the project. As a result, the model development has seen some shifts in emphases to reflect the priorities of the model users. Finally, the early exposure of the model to its users has ensured that their needs are taken into account and is also providing a means of user validation of the CuSP Model.

6 Conclusions

From a modelling perspective, the main feature of the CuSP Model is its combination of particular techniques and data in order to meet the needs of this extended application of microsimulation – Centrelink service delivery at the small area level. The model brings together a tax-transfer microsimulation model, regional microsimulation techniques, a projections capacity, extensive benchmarking, and the use of administrative data. Much of the background modelling development had thus already been undertaken – with the many years of previous development of STINMOD and the MarketInfo techniques – but establishing the CuSP Model amounted to much more than simply bolting these elements together in a new framework. In particular, the incorporation of administrative data has presented some challenges including the sheer task of extracting these data and associated issues such as the spatial unit.

The plan for the CuSP Model includes regular updating – probably annually – with the updates to incorporate up-to-date administrative data as well as new population data. For example, while the small area population data is currently derived from the 1996 census data, appropriate 2001 census data will become available in late 2002. When incorporating the new census data, the opportunity will be taken to revisit the techniques used to generate weights for small areas. The reweighting techniques used in benchmarking and projections will also be reviewed as the number of variables and detail increases. The addition of new variables and more detail will allow the incorporation in the model of further characteristics that can have a bearing on customers' channel use. These include matters such as: duration of payment (channel use is likely to be higher at the start of a period of payment), a customer's position with respect to means-testing (channel use is likely to be higher for those with private incomes than for those without), and aspects that emerge from the new research being conducted into customer behaviour. Development of the spatial interface for analysis of the model output continues, while the other major development in prospect is the shift to use of geocoded customer data.

The CuSP Model has already reached the stage where it is providing useful output, though there is clearly considerable scope for further improvement. While the broad course of further developments of the model can be plotted, the actual path taken is likely to be flexible – continuing to respond to data availability and the needs and priorities of the model users within Centrelink.

References

- Australian Bureau of Statistics 2000, *1998-99 Household Expenditure Survey, Australia: Confidentialised Unit Record File (CURF) Technical Paper*, ABS Cat. No. 6544.0.30.001, ABS, Canberra.
- Australian Bureau of Statistics 2001, *Statistical Geography Vol. 1. Australian Standard Geographical Classification (ASGC) 2001*, ABS Cat. No. 1216.0, ABS, Canberra.
- Centrelink 2001, *Centrelink Annual Report 2001 – 2002*, Centrelink, Canberra.
- Deville, J. and Särndal, C. 1992, 'Calibration estimators in survey sampling', *Journal of American Statistics*, vol. 87, pp. 367–82.
- Lambert, S., Percival, R., Schofield, D. and Paul, S. 1994, *An Introduction to STINMOD: A Static Microsimulation Model*, STINMOD Technical Paper No. 1, NATSEM, University of Canberra, Canberra.
- Lloyd, R., Given, J. and Hellwig, O. 2000, 'The Digital Divide: Some Explanations', *Agenda*, vol. 7 no. 4, pp. 354-58.
- Lloyd, R., Harding, A and Hellwig, O. 2001, 'Regional Divide? A Study of Incomes in Regional Australia', *Australasian Journal of Regional Studies*, vol. 6, no. 3, pp. 271-292.