2011 Household Survey Sample Redesign
Project Board Meeting
1 March 2012

Agenda item 5

2011 MPS Sample Design

Purpose of the paper

The purpose of this paper is to present the proposed MPS sample design as part of the 2011 household sample redesign.

Issues for Household Surveys Sample 2011 Redesign Board discussion

1. Does the Board have any comments on the approach taken to the sample design?
2. Does the Board feel that the EPS constraint should be relaxed, noting the trade off of cost savings against likely increases in sampling error for other estimates?
3. Does the Board have any concerns over the slight reduction in overall sample which comes as a result of how the state accuracy targets have been set?
4. Does the Board have any concerns over the increase required to the NT sample?

Note

This paper contains the key information outlining the details of the sample design. Extended information relating to the key design decisions is documented in the Appendicies, however it is not necessary to read these to gain an understanding of the overall design.

2011 MPS Sample Design

1 Introduction

This paper presents the proposed MPS sample design as part of the 2011 household sample redesign. The sample design described in this paper concerns the master sample of areas for the private dwelling component of the MPS sample. ████████████████████ ██ ██

The key parameters specifying the sampling strategy for a multi-stage area sample are the cluster sizes and the number of clusters to select for different areas of Australia. This paper presents the required values of these parameters in order to satisfy output objectives specified by Labour Branch. The parameters have been derived using an optimisation

procedure which aims to provide most cost-effective sampling strategy by trading off the precision (variance) of estimates with the corresponding cost of obtaining these estimates. The sample design presented incorporates the constraint that all dwellings are selected with equal probability within each State (EPS constraint). The impact of removing this constraint on the sample allocation and costs are noted.

2. Sample Design Objectives

The broad sample design objectives were endorsed at the June 2011 Sample Redesign Board meeting ██ ████████. In late 2011 the interpretation of the broad objective of 'no change to the relative priority of the State-level estimates' was further clarified ██ ████ ██ ████████████. It has also been agreed there will be no change to the relative priority for estimates of employment and unemployment.

███
██
██
██
██████████████

██
███
██
███
███
██████████

██
██
██
███
██
████████████████████████████████

3 Sample Design Overview

Scope of sample design
This paper concerns the design of the private dwelling sample for the MPS. The monthly private dwelling sample for the MPS is selected by sub-sampling dwellings from a master sample of fine areas (small aggregations of mesh blocks). Each fine area in the master sample provides a single cluster of dwellings each month over the five-year design period. The sample design can be summarised by the following aspects:
- sampling strata;
- stratum cluster sizes, which are number of dwellings to approach per month within each selected fine area; and
- number of clusters to select in each stratum.

The MPS sample also consists of a special dwelling component selected from a frame of establishments. Importantly, the sample allocation for the private dwelling sample implicitly determines the sample size of special dwellings (assuming maintaining the strategy of selecting persons in special and private dwellings with equal probability).

Stratification
Fine areas are assigned to sampling strata defined by the combination of SA4, the finest geography for routine dissemination of labour statistics, and an 'area type' classification of areas. Dwelling density and remoteness are two key factors for determining area type. The stratification and area type classification are discussed in more detail in Appendix 2.

Equal probability constraint

Advice is sought from the Board on whether the sample design should incorporate the constraint of selecting all dwellings within a State with equal probability (EPS design). Traditionally the MPS design has been EPS.

██
██
██
██
███████████

██
██
██
████████████
██
██
██
██
██
██
████████████████████████████████████

██
████████
██
██
██
██
████████████████████████████
█ ██
██
█ ██
██
████████

██
████████████████

Does the Board feel that the EPS constraint should be relaxed, noting the trade off of cost savings against likely increases in sampling error for other estimates?

Method for choosing cluster sizes

The optimum cluster size for a stratum balances the trade-off between cost and variance which arises when selecting a sample with geographic clustering. ███████████████

███
███
███
███
████████████████████████████████████

█████████████
███
██
███
███
███████████████████████████
██s.

Interviewer activity simulation cost model

For the 2011 sample design project a new procedure for modelling interviewer costs has been used for the sample optimisation task. The costs of alternative sample designs have been compared using a simulation model of interviewer activity. The simulations measure the time spent by interviewers making telephone and face-to-face call attempts, travelling and interviewing, as well as measure distances travelled (to account for motor vehicle allowance payments). █████████████████████████████████

██
██
███████████████████████████

4 Optimal cluster size choices

███
███
███

████████████████████████████
███

███

██████████████	██████	██████	██████	██████	██████	██████
██						
██						
███████████████████████████████████						
██						
██						
███						
█████████████████████████████████						

5. Properties of sample with the recommended cluster sizes

This section presents the State-level distribution of sample for the proposed design, which produces the RSEs indicated in Table 2.1. Comment is sought from the Board on the allocation, in particular the high sample allocation in NT. Based on the design outlined above, the expected numbers of dwellings selected over the coming years are as follows:

Does the Board have any concerns over the slight reduction in overall sample which comes as a result of how the state accuracy targets have been set?

Advice is sought from the Board on the sample allocation in NT, in light of the sample allocation and expected sampling skip for the proposed design. ████████████████

Does the Board have any concerns over the increase required to the NT sample?

6 Impact on cost due to changes to sample clustering

Appendix 1

Appendix 2   Stratification and Area type classification

Stratification scheme

The stratification of the area frame for the 2011 sample redesign is the combination of SA4 and area type class. The number of strata defined within an SA4 is the number of distinct area types which defined within it. Including the SA4 dimension ensures the sample design controls the amount of sample selected for each SA4. Incorporating area type facilitates adoption of the most sampling scheme within each of the variety of areas contained in each SA4. There are a small number of special strata (Pre-determined growth, Indigenous and secure apartment building), which span the SA4s within the State.

Area type classification

The area type classes, presented in Table 1 below, primarily aim to combine geographic areas with similar costs for data collection and possibly similar levels of geographic clustering of labour characteristics. The area type classes may also divide distinguish areas in which are more homogenous with respect to the labour characteristics of the persons within. For example, the demographics of persons in the "Capital City Past Growth" class, which largely contains outer suburbs, are likely to differ from the demographics of persons living in the inner suburbs.

The capital city area types are defined for areas within SA4s which are define the capital city boundary of the State. The variety of area type classes distinguish areas of high dwelling density, areas with recent high growth (typically outer suburbs), pockets with many secure apartment buildings and outlying rural areas within the capital city boundary. A key distinction between the area types outside of the capital cities are Self Representative Areas (SRA) and non-SRA. The distinction is based on dwelling density, and in non-SRA areas a modified sampling procedure is used to increase sample clustering. This ensures interviewer workloads can be created from sample selections which are relatively close together. The Remoteness and Urban/Rural classifications defined on the ASGC were used as the basis for classifying areas to the SRA rural area type and the different non-SRA area types.

Table 2 presents the distribution of dwellings across the area types in each State, providing perspective on the relative size of the classes. Nationally, the only area type classes with more than 2.5% of the population of dwellings are "Capital City Settled", "Capital City Inner City", "Capital City Past Growth", "Rest of State SRA Major Cities", "Rest of State SRA Medium towns", "Rest of State SRA Small Towns" and "Rest of State non-SRA Least Remote".

**Table 1**: Area types for the 2011 MPS stratification

Number	Area Type	Explanation
1	Capital City Settled	Default Capital city areas
2	Capital City Inner City	Population density > 3125 persons per sq km CHECK
3	Capital City Secure Apartment Buildings	SA1s with a high proportion of secure apartment building *
4	Capital City Past Growth	growth of more than 10% since 2006
5	Capital City Rural	Rural outskirts of capital cities
6	Rest of State SRA Major Cities	Population 100k +
7	Rest of State SRA Medium towns	Population 12-100k
8	Rest of State SRA Small Towns	Population <12k
9	Rest of State SRA Rural	Rural SRA areas
10	Rest of State non-SRA Least Remote	
11	Rest of State non-SRA Remote	
12	Rest of State non-SRA Very Remote	
13	Pre-determined growth	Capital city** greenfield areas with large growth expected in the next 5 years
14	Indigenous	

Notes:
* SA1s with at least 100 apartments OR 3 or more buildings containing a total of 75 or more apartments
** NSW contains pre-determined growth SA1s which are outside the capital city

Table 2: Distribution of State's dwellings across area type classes (% points)
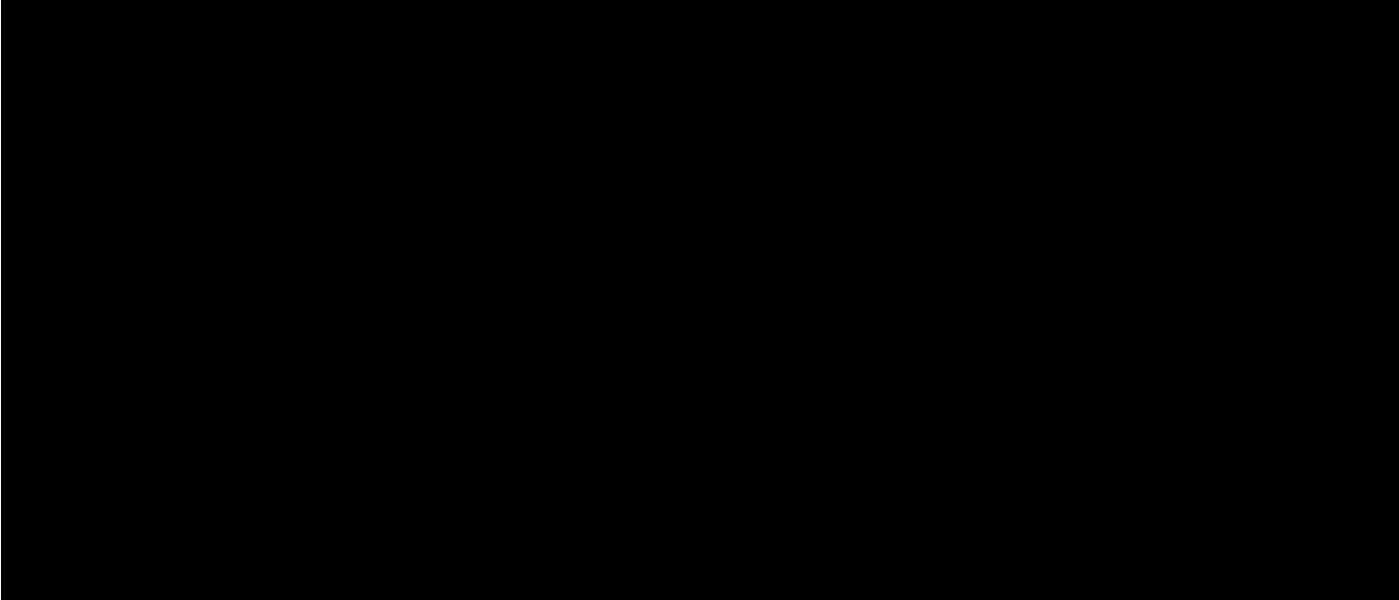
Summary of method for area type assignment

In general area types have been assigned to areas at the SA2 level. Situations in which multiple area types can exist in an SA2 are:
- the SA2 contains multiple localities, the localities in the SA2 may belong to different area types

- the SA2 contains SA1s belonging to the secure apartment building, pre-determined growth or Indigenous area types

██████████████████████

██
██
███████

███

██
███

██

████	████████████	███████████
	████ ████	█████████
██████████	███████████	████████████
██████████	███████████	████████████
██████████	███████████	████████████
██████████	███████████	████████████
██████████	███████████	████████████
██████████	███████████	████████████
██████████	███████████	████████████
██████████	███████████	████████████
		████████████████

██
██
██
████████████
  ███
  ███

██

██		████ ████	███████████████
	█████████████████	█████████████	█████████████
		█████████████	███████████████
			████
██
████████████		██████████████████████████████████
██
██
██
██
██
██
██
██
██

Appendix 4   Overview of interviewer activity simulations and assumptions

Simulation model

For the 2011 MPS sample redesign the cost of alternative sample designs have been compared using a simulation model of the activity undertaken by interviewers. The purpose of the models is to measure the relative rather than absolute PSO costs for sample design alternatives. The simulation model measures the cost of *interviewer time* and *motor vehicle allowance* costs associated with travel. The interviewer time includes time for:
- travelling to workloads,
- travelling within workloads,
- making call attempts by phone and in the field,
- conducting interviews

The simulations include all of the activity for callbacks. The simulation model does not measure indirect costs, which should be similar for the alternative sample designs.

██████████████████████████████████████

██

██
██
██████████████████████████████

██
███

██
██
██

██
██
██
███████

██
██
██
██

██