2011 Household Sample Redesign Manual

1. Cost Modelling

1.1 Introduction

1.1.1 Overview

[redacted]

The purpose of the cost model is to be combined with the variance model in order to inform the choice of the sample design parameters. The Optimisation documentation gives a description of how this is done (see Chapter 3).

[redacted]

The simulation model involves simulating the work of each interviewer to complete their assigned workload of dwellings and accumulating associated time and travel costs for their activity. It consists of the following steps:

[redacted]

- [redacted]
- [redacted]
- [redacted]
- [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

- [redacted]
- [redacted]

[redacted]

1.1.2 Data Sources

There were a range of data sources for the cost modelling and Section 1.3 gives the actual locations of the datasets used in the various processes. The major sources of data were:

- Calls data [redacted]
- Time and Travel data [redacted]
- Household-level response data [redacted]

[redacted]

[redacted]

[redacted]

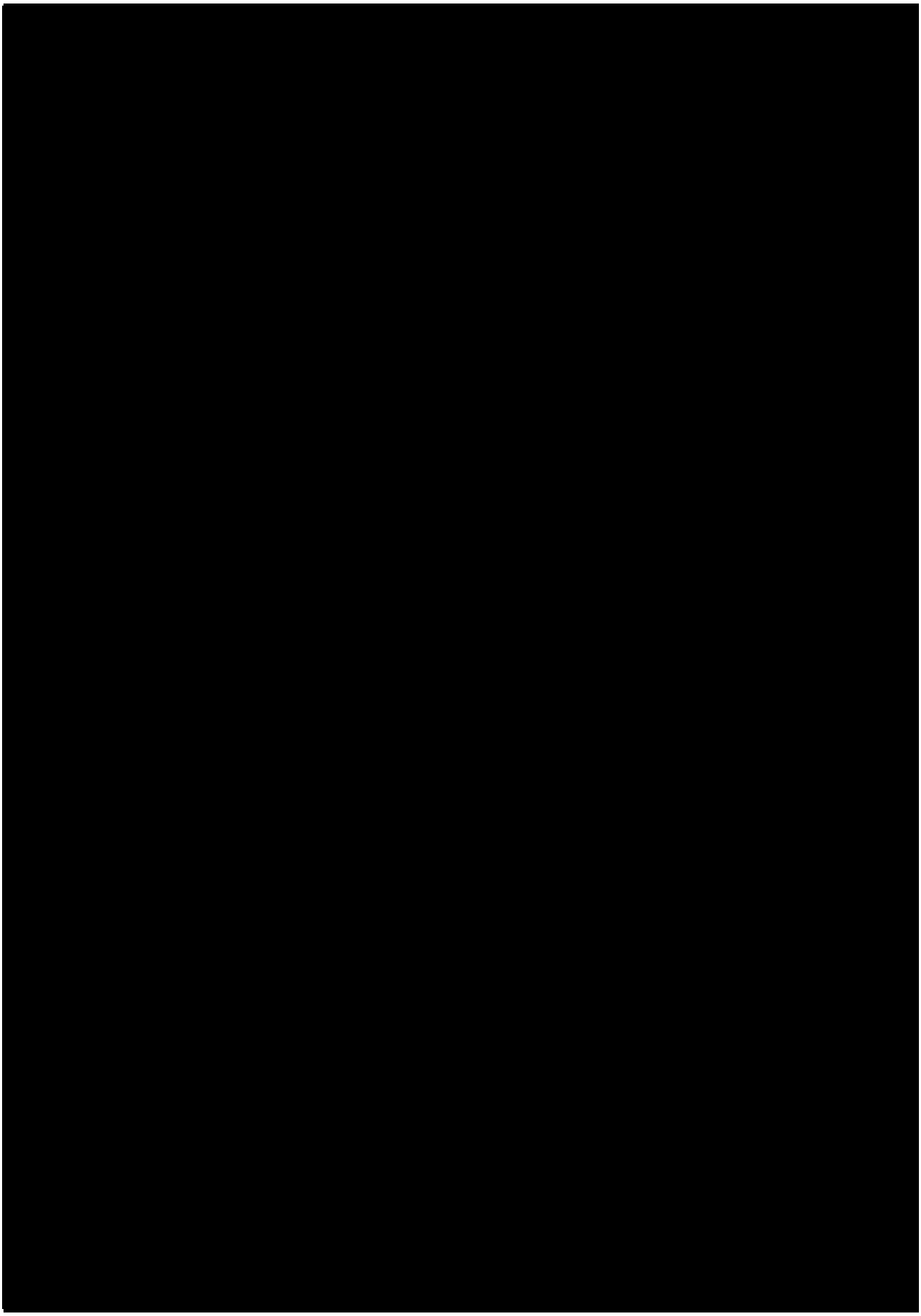
[redacted]

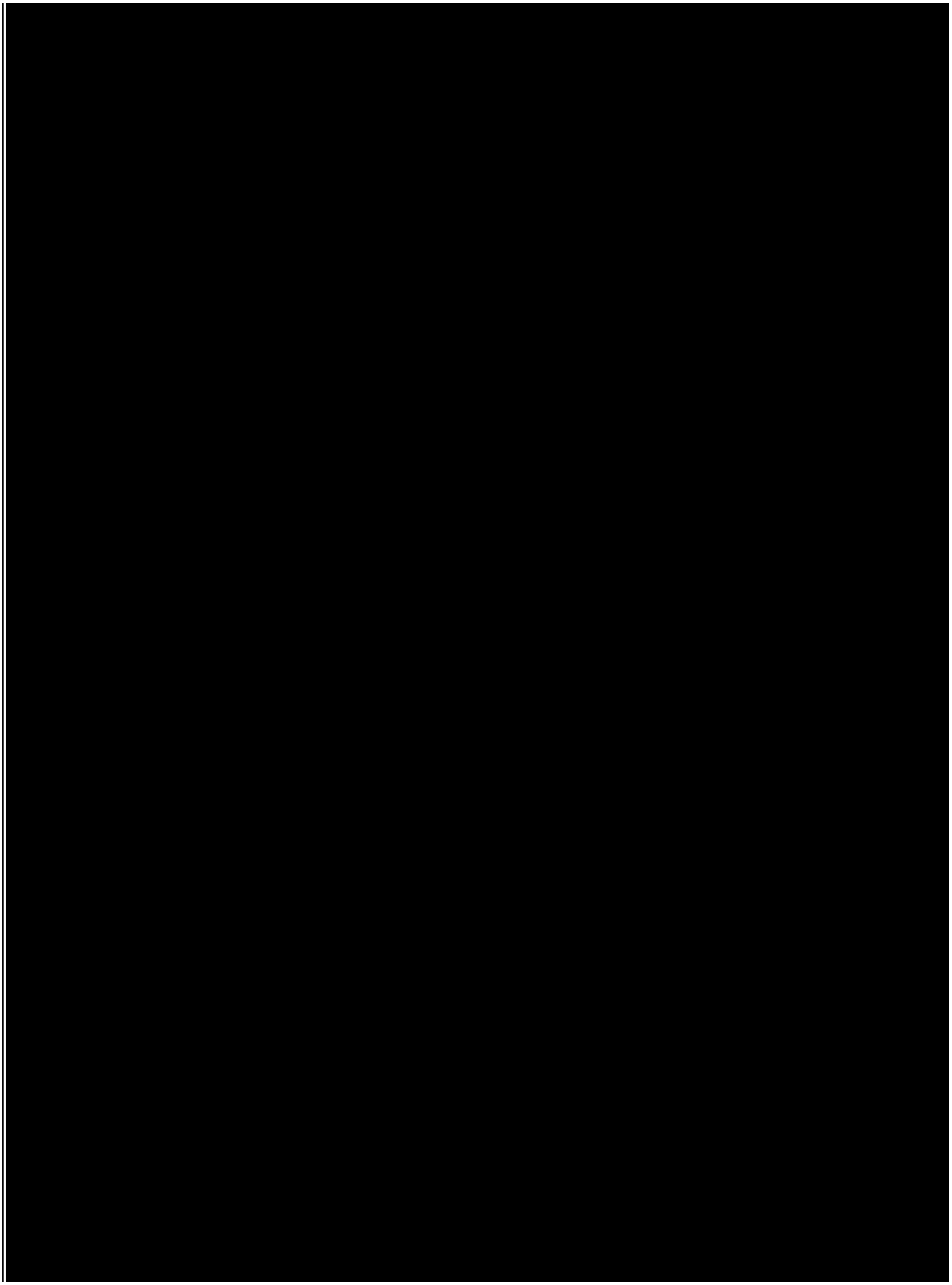
[redacted]

[redacted]

[redacted]

[redacted]





[redacted]

[redacted]

[redacted]

[redacted]

1.3.4 Using the model
See Optimisation Documentation.

[redacted]

[redacted]

2. Variance Modelling

2.1 Overview of 2011 Redesign Variance Modelling

The purpose of variance modelling is to determine explicit variance functions that describe the relationship between the Labour Force Survey's sample variance and its sample design parameters. [redacted] Once such functions have been fully specified, they were used to generate simulated samples which were used in conjunction with cost models describing the Monthly Population Survey's (MPS) cost structure to determine the optimal sample design to implement in order to meet particular cost and variance requirements for the private dwelling component of the 2011 MPS. [redacted]he details of the cost modelling can be found in Section 1 and details of the optimisation process can be found in Section 3.

Section 2.1.1 briefly mentions some features of the LFS that are relevant to the 2011 variance model. Section 2.1.2 discusses the methodology behind the variance model, but a more detailed discussion can be found in Sections 2.2 and 2.3. Section 2.1.3 discusses the features of the 2011 Variance Modelling that are improvements over the work undertaken in 2006 and Section 2.1.4 discusses ways in which the 2016 modelling process could improve upon the work done in 2011.

[redacted]

[redacted]

[redacted]

[redacted]

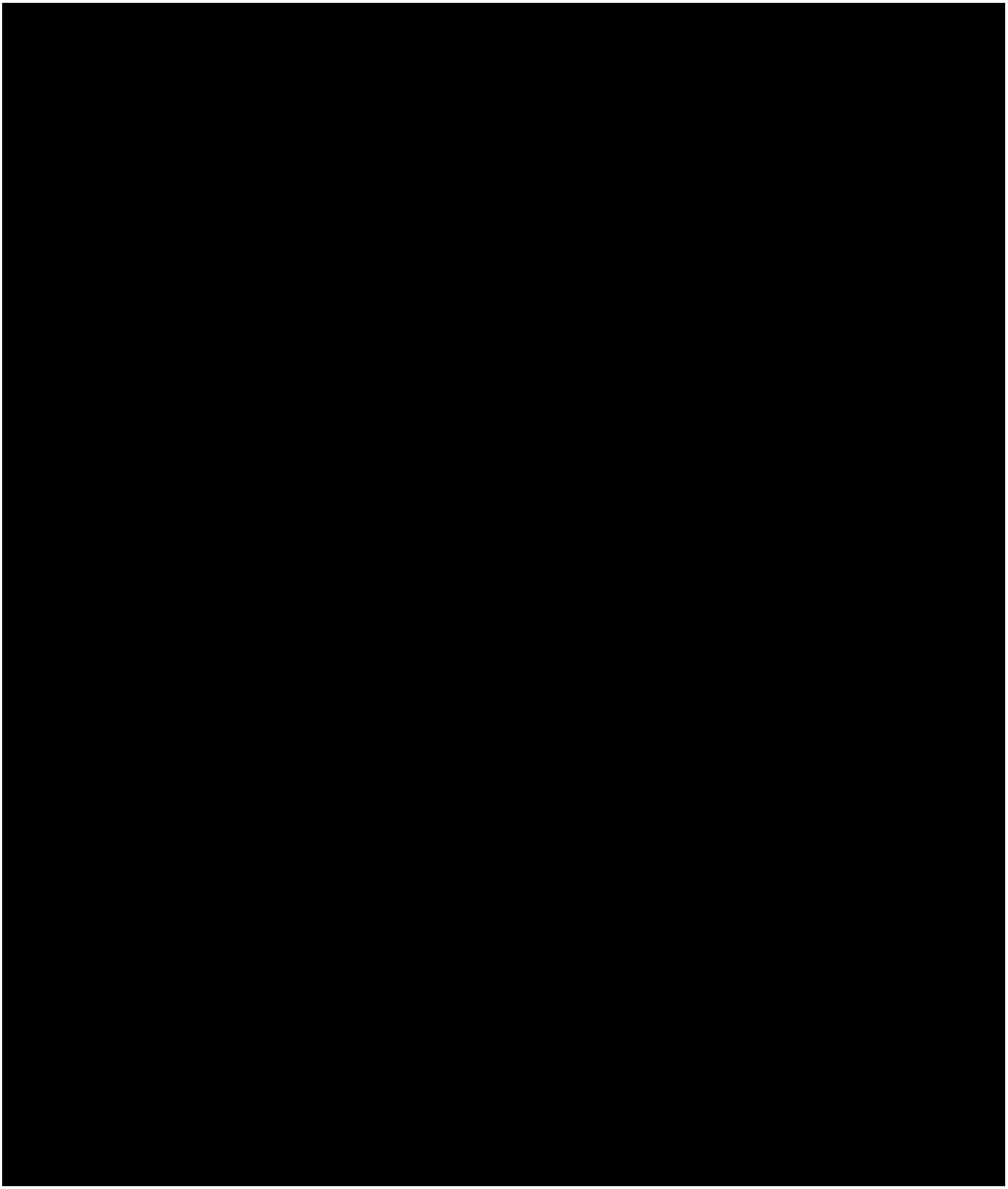
[redacted]

[redacted]

[redacted]

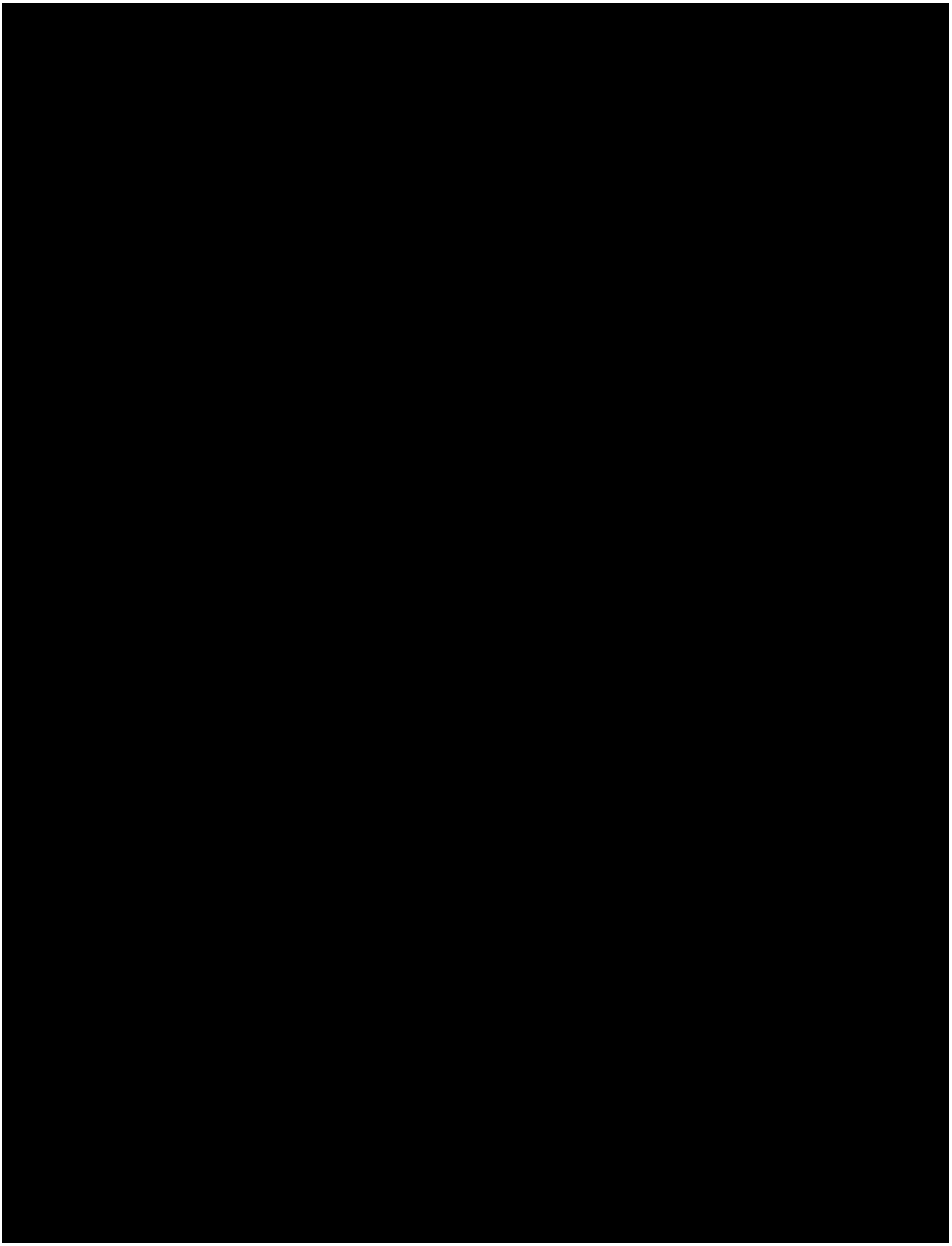
[redacted]

[redacted]



2.2.2 Structure of the document

Section 2.2 covers the First Phase of the variance modelling, which is entirely concerned with the calculation of the variance model parameters based off Census 2006 data, [redacted]



[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

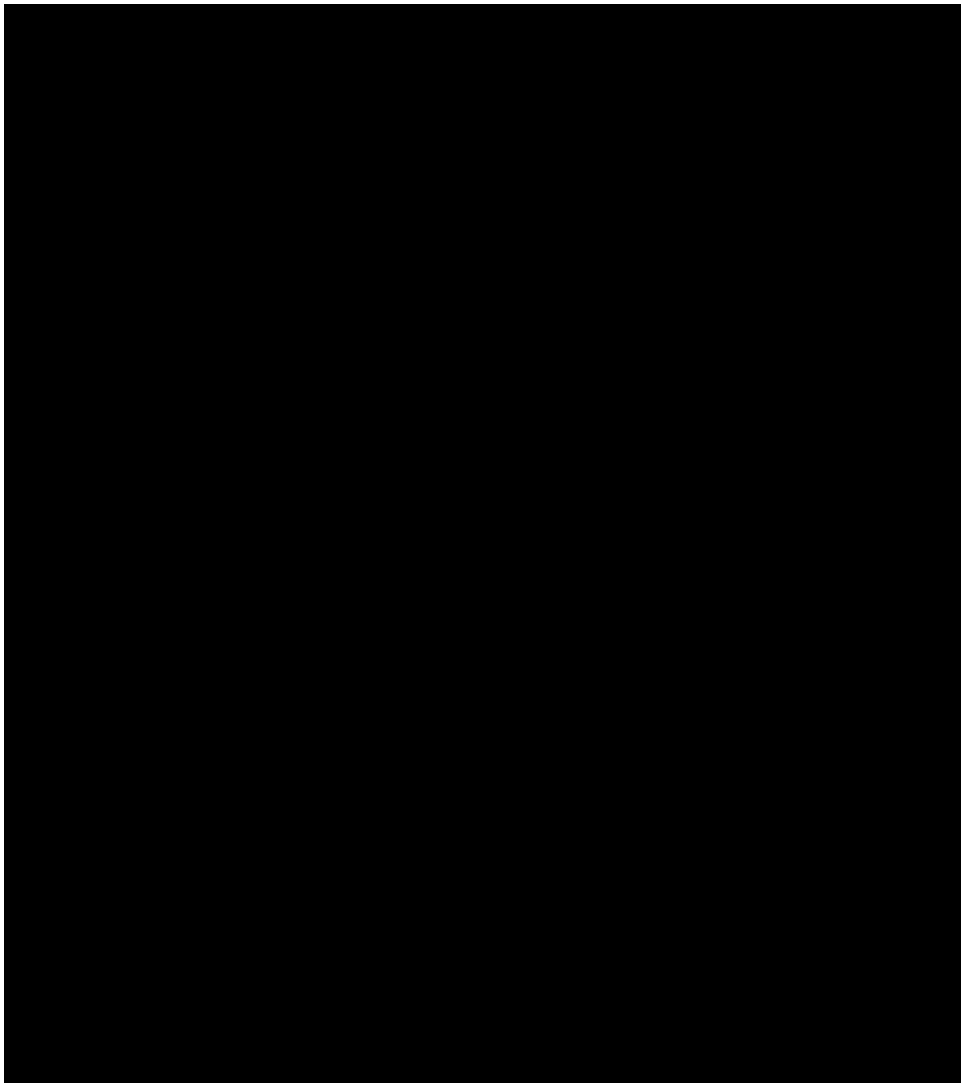
[redacted]

[redacted]

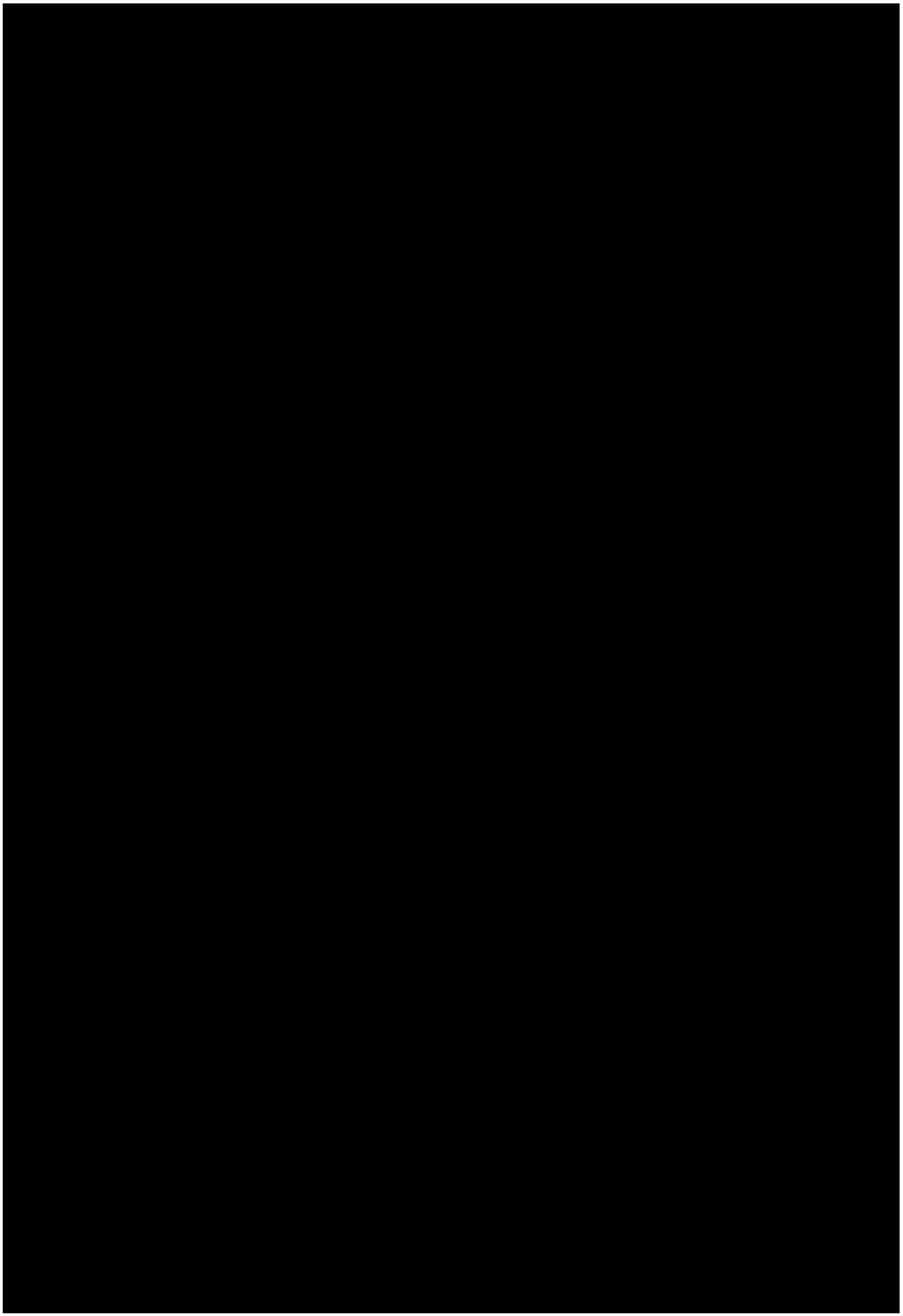
[redacted]

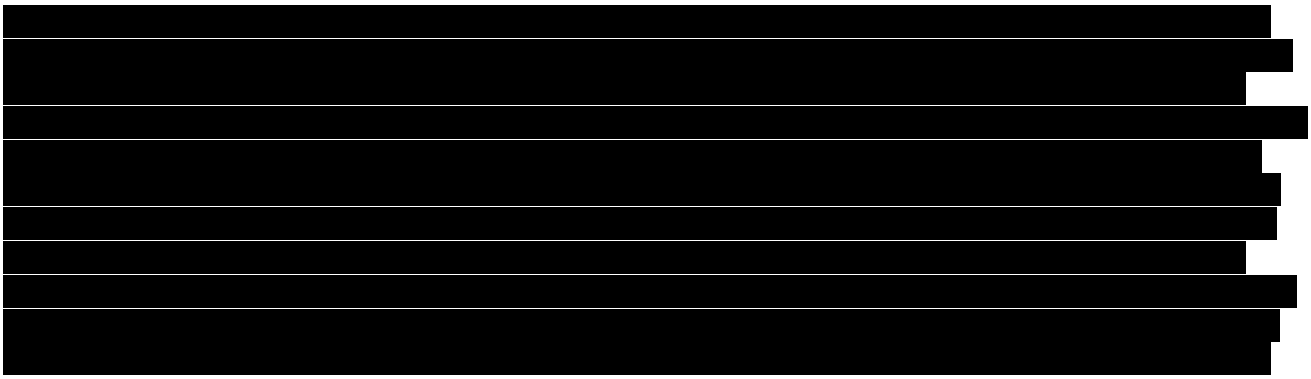
[redacted]

[redacted]



[redacted]





[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]	[redacted]	[redacted]	[redacted]
[redacted]			
[redacted]			
[redacted]			
[redacted]			
[redacted]			
[redacted]			
[redacted]			
[redacted]			
			[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

[redacted]

[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]	
[redacted]	
[redacted]	
[redacted]	
[redacted]	
[redacted]	
[redacted]	
[redacted]	
[redacted]	
[redacted]	
[redacted]	
[redacted]	
[redacted]	
[redacted]	
[redacted]	
[redacted]	

[redacted]

[redacted]

[redacted]

[redacted]

	[redacted]							
[redacted]								
[redacted]		[redacted]						
[redacted]		[redacted]						
[redacted]		[redacted]						
[redacted]		[redacted]						
[redacted]		[redacted]						
[redacted]		[redacted]						
[redacted]		[redacted]						
[redacted]		[redacted]						
[redacted]		[redacted]						
[redacted]		[redacted]						
[redacted]		[redacted]						
[redacted]		[redacted]						
[redacted]	[redacted]							
[redacted]		[redacted]						

[redacted]

[redacted]

[redacted]

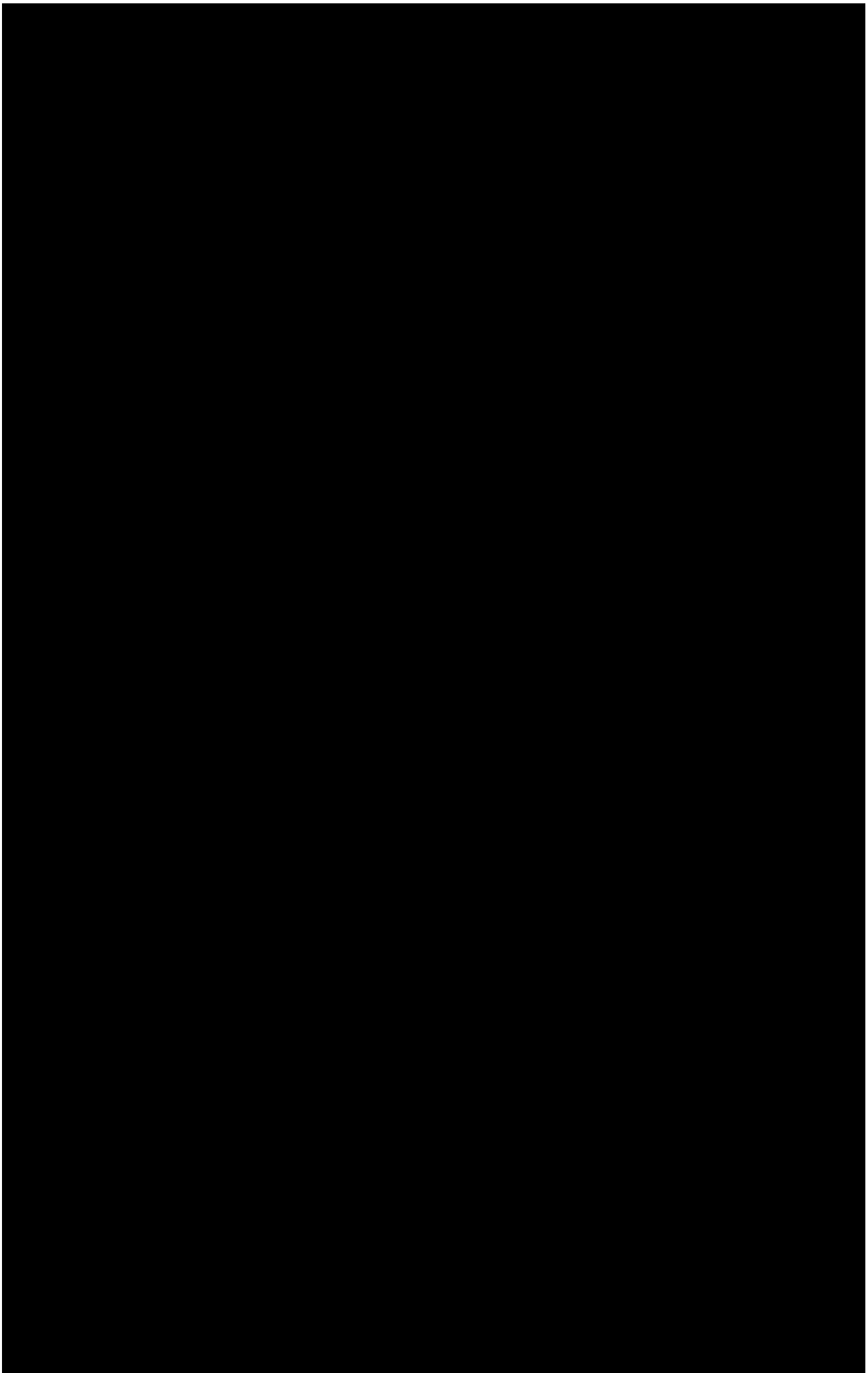
[redacted]

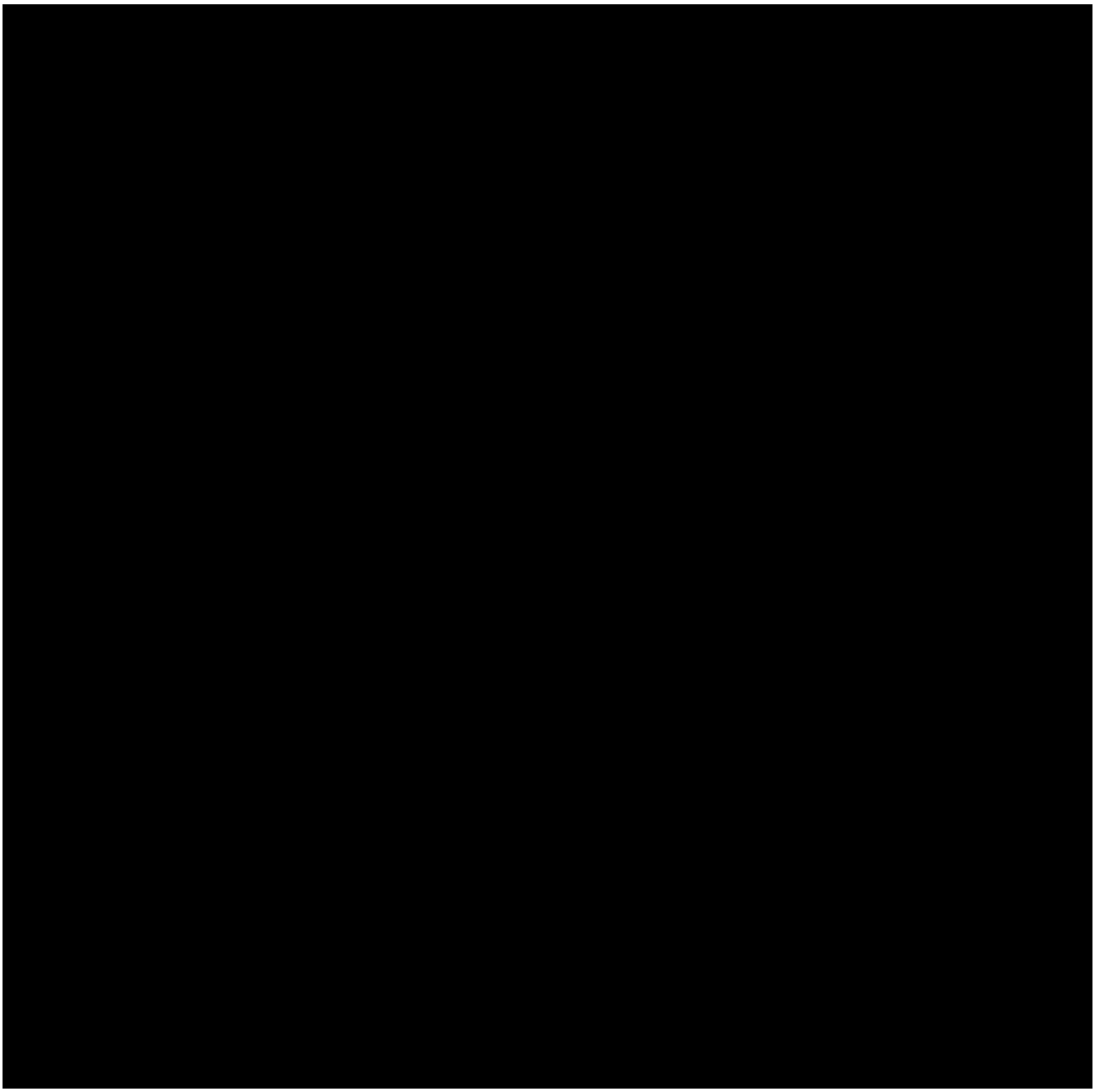
[redacted]

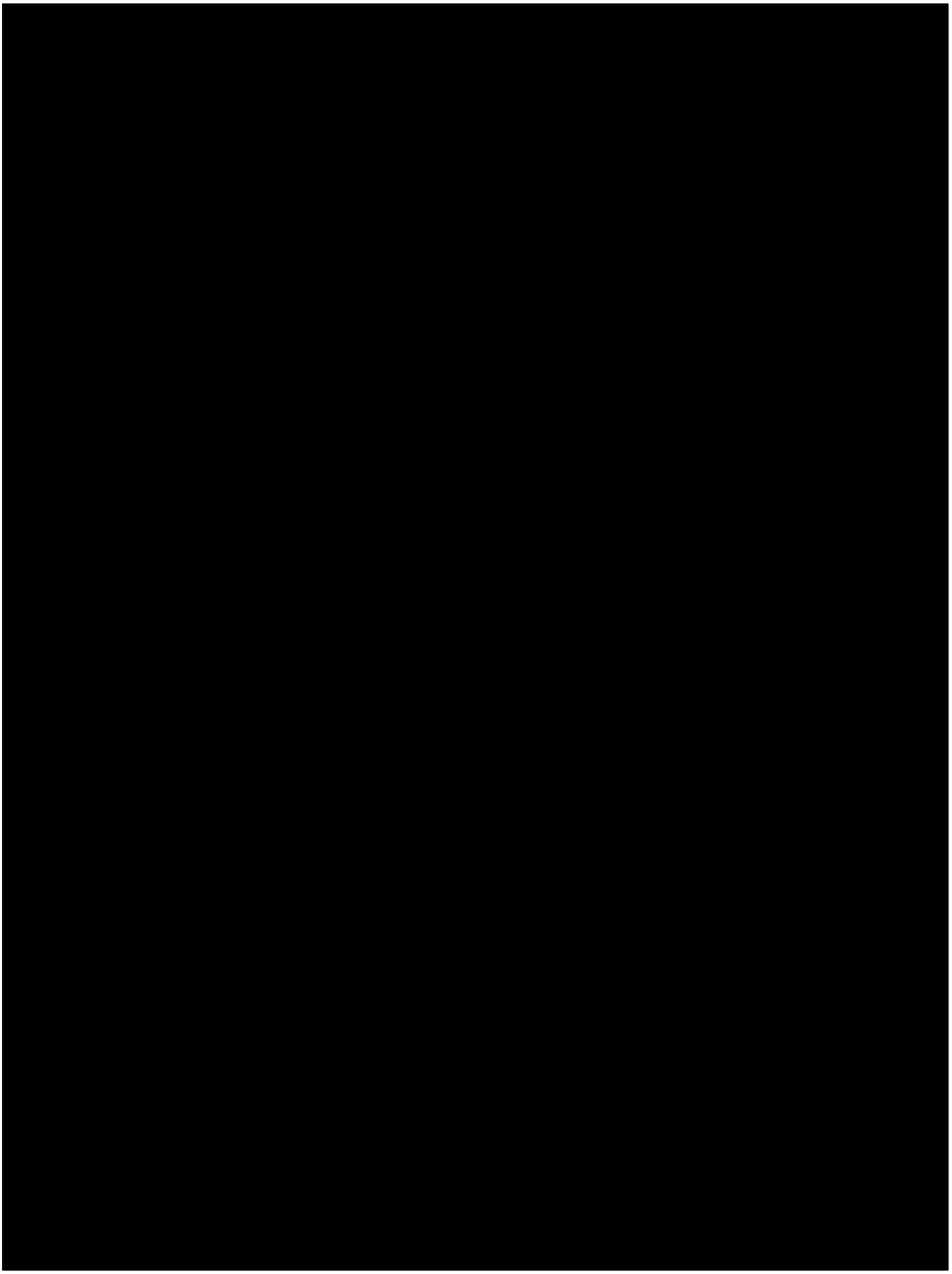
[redacted]

[redacted]

[redacted]







[redacted]

3.1 Introduction

3.1.1 Overview

The aim of the optimisation is to determine the optimal number of clusters to select and optimal cluster sizes in each state by area type in order to minimise costs whilst meeting variance targets.

[redacted]

[redacted]

[redacted]

[redacted]

Note that the number of people is used instead of the number of dwellings. [redacted]

[redacted]

[redacted]

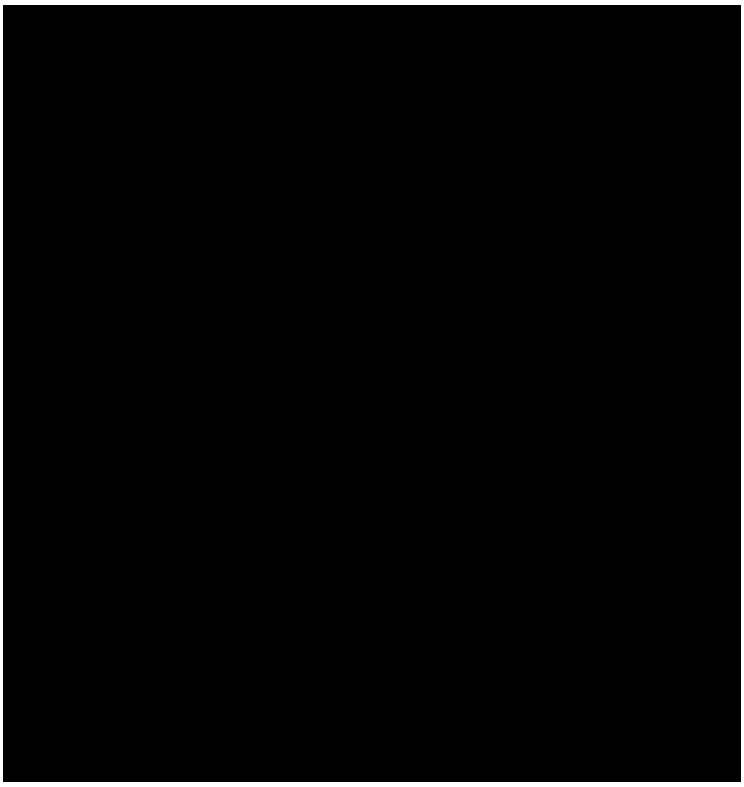
[redacted]

[redacted]

[redacted]

[redacted]

[redacted]



[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]
[redacted]
[redacted]
[redacted]
[redacted]
[redacted]
[redacted]
[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

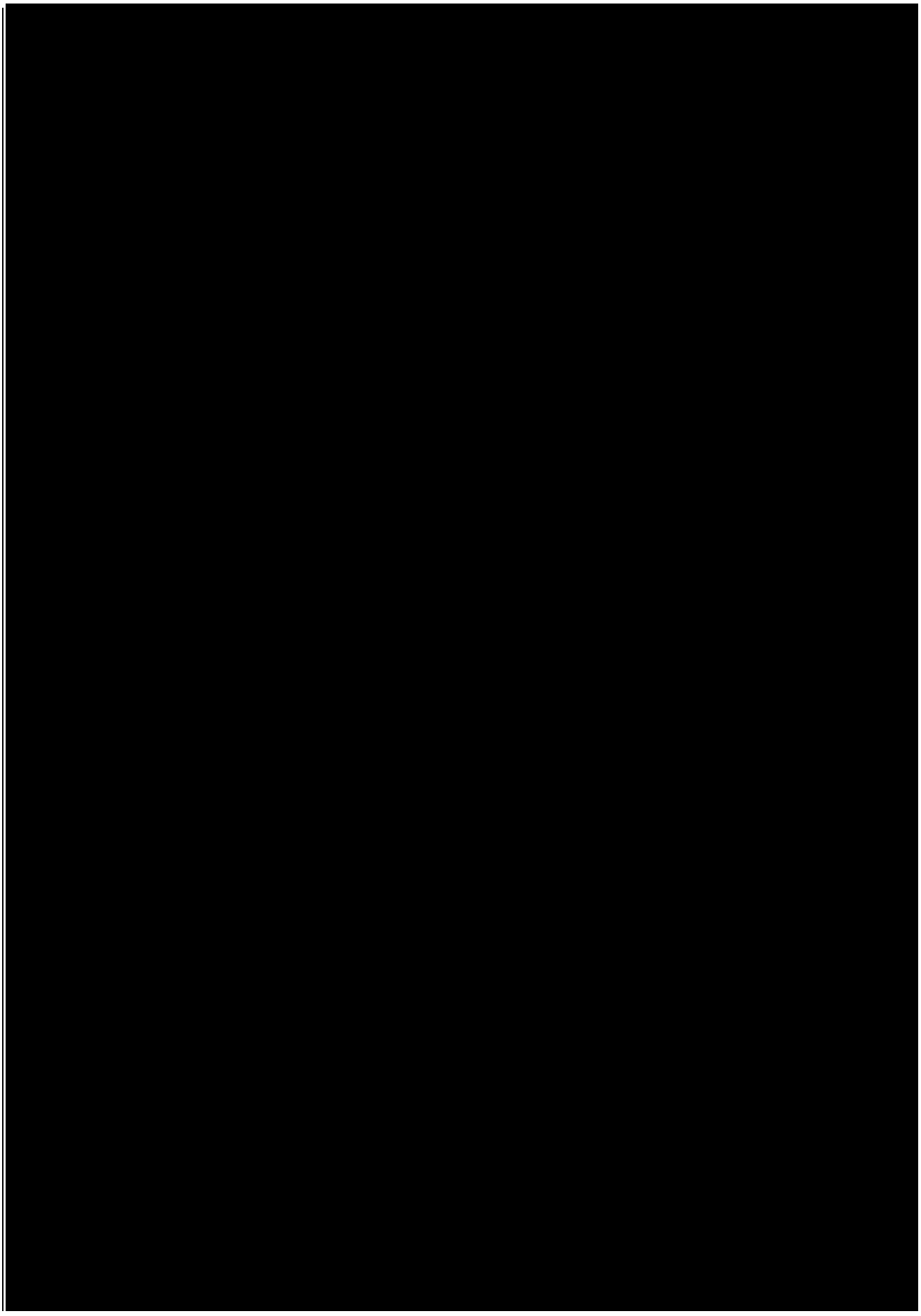
[redacted]

[redacted]

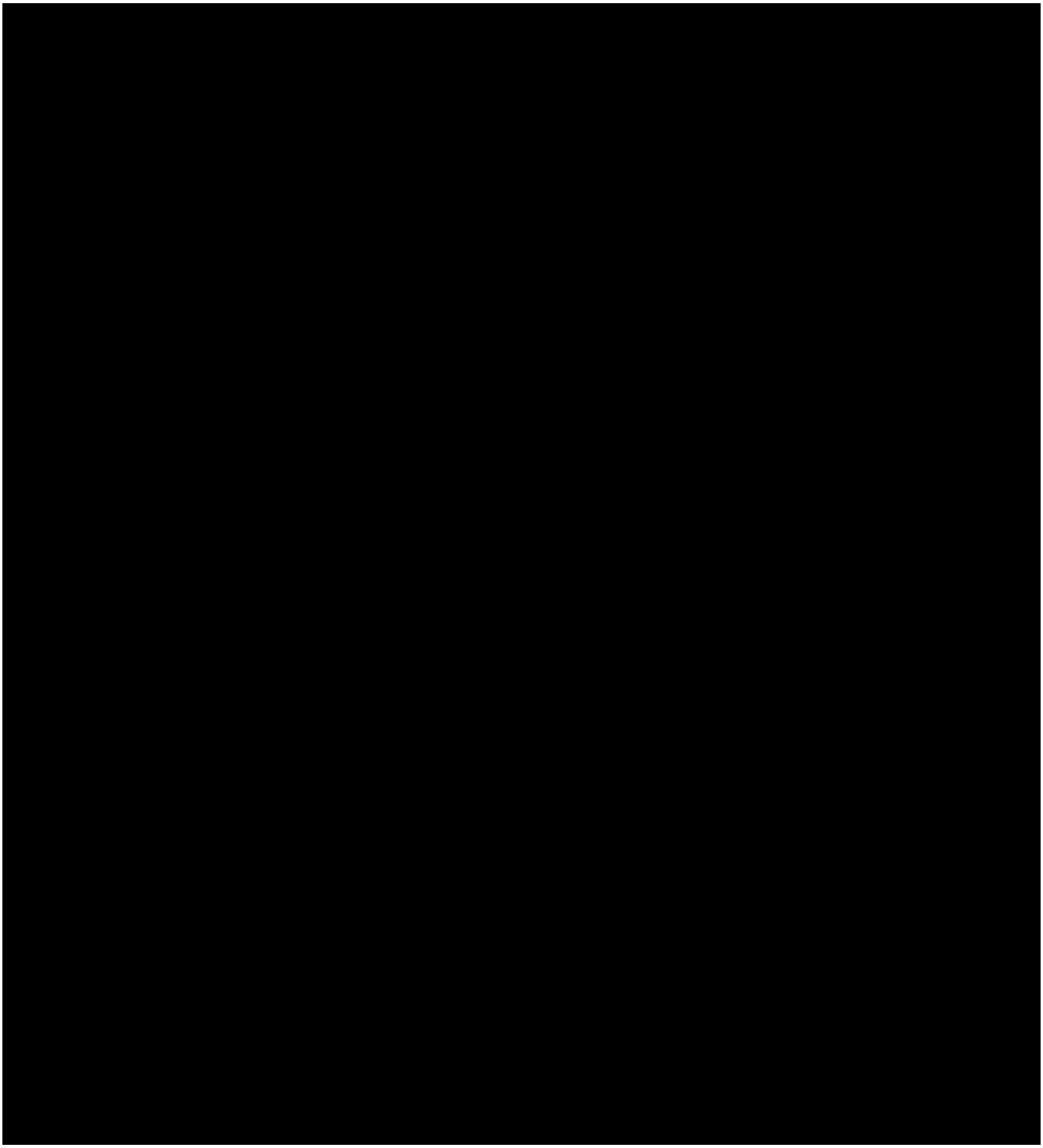
[redacted]

[redacted]

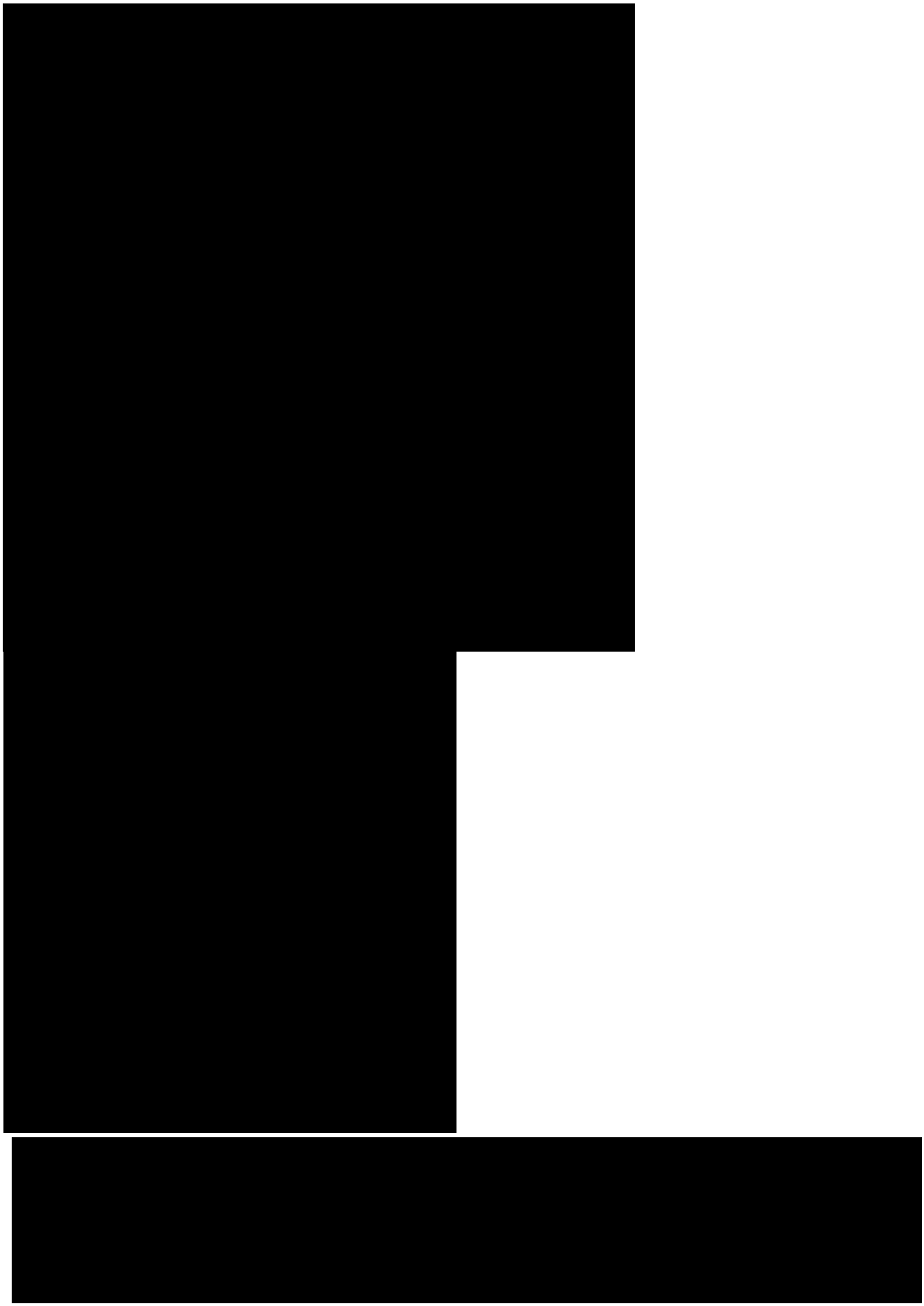
[redacted]



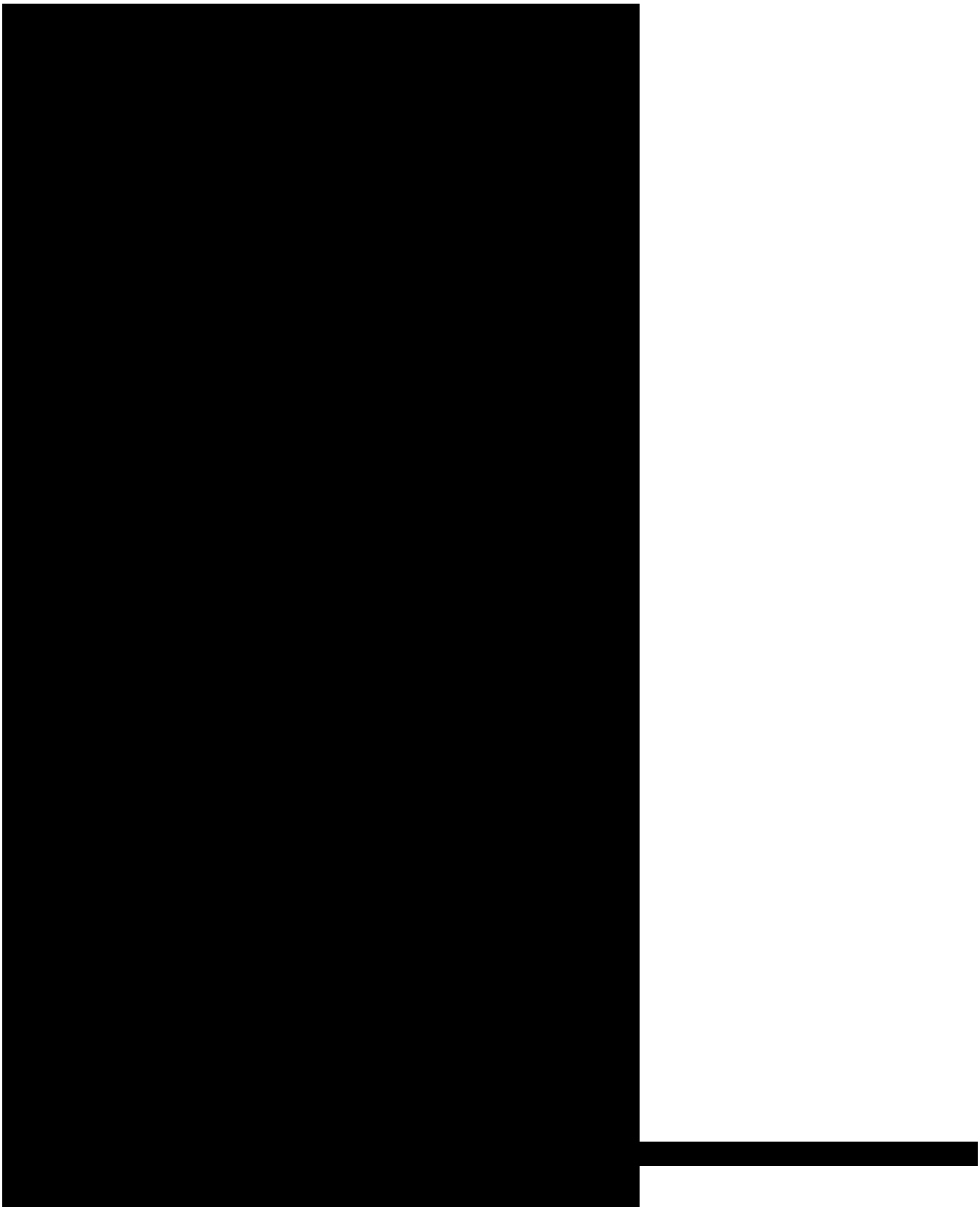
[redacted]

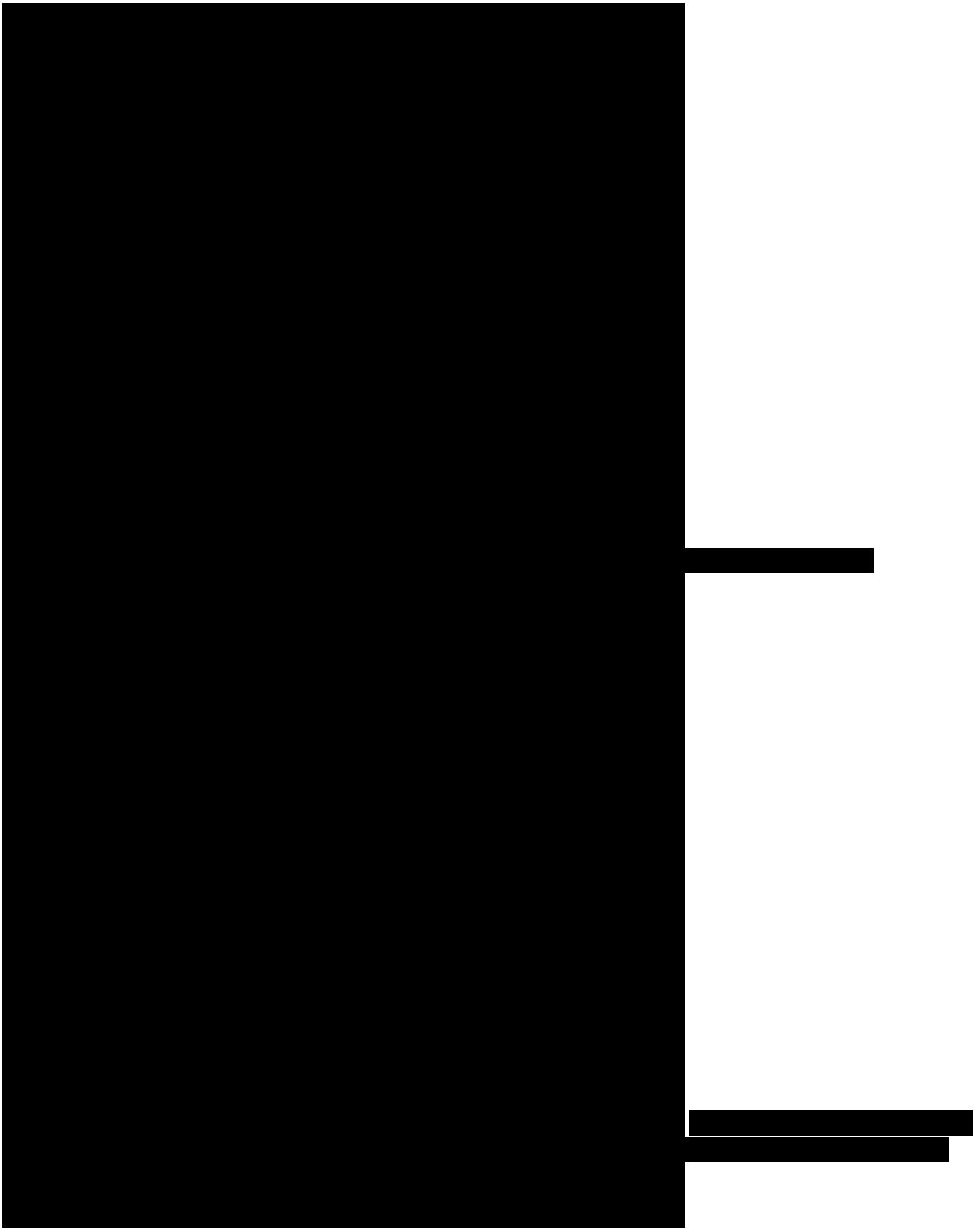


[redacted]



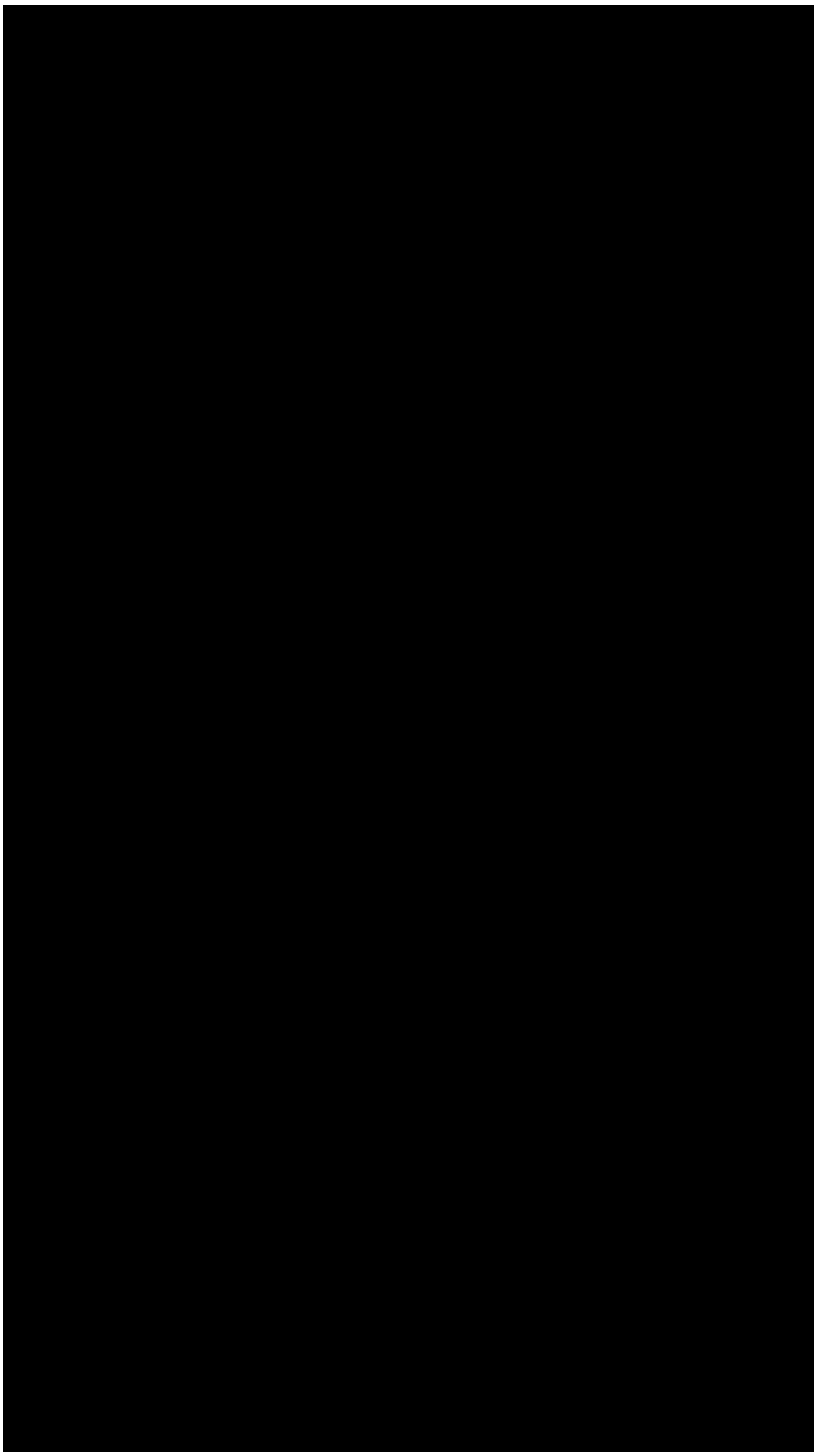
[redacted]





[redacted]

[redacted]



4.1 Overview

The complete area of Australia is partitioned into geographic sampling strata for the MPS sample design. The selection units on the MPS area frame are MPS First Stage Units (FSUs), which are geographic areas created by combining mesh blocks. The FSUs are actually an amalgamation of Base Frame Units (BFUs), which themselves consist of one or more mesh blocks. All dwellings belonging to the same FSU are assigned to the same stratum. The stratification used for the MPS design is essentially geographic, because the stratum assignment for an FSU is essentially determined by its location and the geographic distribution of dwellings in the proximity of the FSU.

The roles of the stratification for the MPS sample design are:

- to reduce sampling variance by providing control over the sample distribution across strata (exploiting any small correlation between labour force characteristics and stratum);
- to provide control over the sample size in each State and region for which labour statistics are disseminated;
- to provide control over the sample distribution across areas which are difficult or more expensive to enumerate;
- to enable choice of the degree of sample clustering which is most efficient for a type of region (considering the region's dwelling density and other factors related to enumeration cost).

In general, the strata are formed by combining SA4 and "area type" class. The SA4 of an area unit is defined as part of the Australia Statistical Geography Standard (ASGS). [redacted]

[redacted]

The 2011 MPS sample design stratification is similar to that used for the 2006 design. The most significant changes are a consequence of the selection units being built-up from areas defined under the ASGS instead of the Australian Standard Geographic Classification (ASGC). In particular:

- SA4 replaces the "sample design regions" created for the previous MPS master samples from the ASGC boundaries within each State.
- the classification of an area into belonging to a either "capital city area" or the "balance of State" is simply determined by its SA4. Previously the split between "capital city" and "balance of State" was defined by the region's Statistical Division, as defined under the ASGC.

There were also some minor changes to the area type class definitions. More detail on the changes from 2006 are discussed in the Appendix.

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

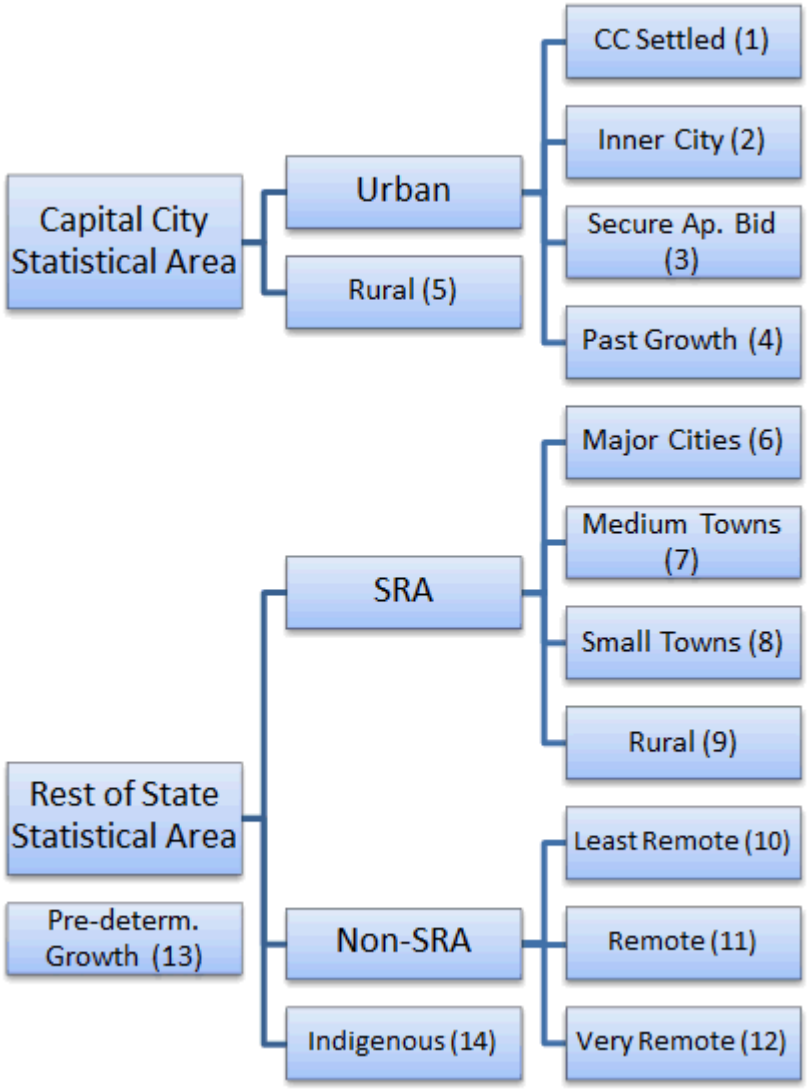
[redacted]			

4.3.2 Overview of area type class properties and derivation

Chart 4.1 presents the area type classes in a hierarchy, illustrating relationships and common properties for the area type classes. Entries in the chart containing a number within parentheses are an area type class, with the numbers representing the area type number assigned. Boxes which do not include a number in parentheses are properties of the classes. Section 4.3.3 describes the classifications which were used to define the area type classes.

[redacted]

Chart 4.1: Relationship between area type classes



4.3.3 Classifications used for area type assignment

[redacted]

A variety of geographic classifications defined by different sources are combined to derive the area type classes:

- ASGS boundaries - Greater Capital City Statistical Area (GCCSA)
- ABS Geography classifications defined in ASGS - Remoteness area (RA), Section of state (SoS) Urban centre or locality (UCL).
- HSM - Self representing Area (SRA) / non-SRA (based on estimated population density).

The class values are described in (i)-(v) below.

[redacted]

i) Greater Capital City Statistical Area (GCCSA)

The GCCSA classification divides each State or Territory (except ACT) into the capital city area (CC) and the balance, referred to as Rest of State areas (RoS). The second letter of the GCCSA code ("A", "G" or "R") was used to denote the CC and RoS areas:

- CC areas: "A" (for ACT) and "G" (other states).
- RoS areas: "R" (except ACT).

The classification is defined at the SA4 level (meaning all dwellings within the same SA4 have the same GCCSA).

In some States the size of area classified as CC under the GCCSA is noticeably larger than the size of area which was classified as "capital city" or "Met" using the Capital City Statistical Division classification from the ASGC. This is because the GCCSA CC areas include regional commuter zones and because the new CC areas have been designed to cater for expected growth of the capital city areas.

ii) Section of State

The remoteness classification is based on ARIA (Accessibility/Remoteness Index of Australia), which measures how far a region is from a service centre of varying sizes and thus availability of services such as health and education. The ASGC remoteness classes were:

0 - Major Cities of Australia
1 - Inner Regional Australia
2 - Outer Regional Australia
3 - Remote Australia
4 - Very Remote Australia

The area type assignment used a collapsed version of this ASGC classification. Values 0, 1 and 2 were combined to give a "Least Remote" class. Areas in the "Least Remote" class are primarily capital city areas or towns in the rest of state.

iii) Remoteness

The remoteness classification is based on ARIA (Accessibility/Remoteness Index of Australia) which measures how far a region is from a service centre of varying sizes and thus availability of services such as health and education. The ASGC remoteness classes were:

0 - Major Cities of Australia
1 - Inner Regional Australia
2 - Outer Regional Australia
3 - Remote Australia
4 - Very Remote Australia

The area type assignment used a collapsed version of this classification. Values 0, 1 and 2 were combined to give a "Least Remote" class. Areas in the "Least Remote" class are primarily capital city areas or towns in the rest of state.

iv) Urban Centre or Locality

The UCL code is unique for every urban town or locality (very small town or village). Rural areas are not assigned a UCL code and hence this classification does not cover all of Australia.

The area type classification used geographic regions which are a cross-classification between SA2 and UCL code, referred to as an "SA2 locality". Since an individual city or town is covered by a single UCL code, there is typically no distinction between "SA2" and "SA2-locality" in capital cities and large towns. Rural areas not assigned a UCL code were assigned the same code of 0, so rural areas in the same SA2 were considered to belong to the same "SA2-locality".

v) Self Representing Areas (SRA)

To ensure viable interviewer workloads can be formed in areas with low dwelling density, a further classification based on population density is needed for the area type stratification. This classification divides Australia into Self Representing Areas (SRA) and non-SRA. SRA include the CC areas and areas with medium-to-high population density. All non-SRA areas are in RoS, and cover the low population density parts of Australia ("the Outback"). In non-SRA regions a "higher level" stage of area selection is adopted, so as to ensure the MPS sample in these regions is sufficiently clustered to provide practical workloads for interviewers which can be enumerated at reasonable cost. Whereas the classifications described in (i)-(iv) are formal ABS geographic classifications, the SRA classification was derived for the purpose of the sample design.

An SA2 is classified as a self-representing area (SRA) if it is located within a compact region with at least moderate population density. [redacted]

4.3.4 Assignment of classifications to SA2-localities

The SoS and UCL classifications defined on the 2011 ASGS were released in October 2012, and the RA classification for the ASGS was released in January 2013. [redacted]

[redacted]

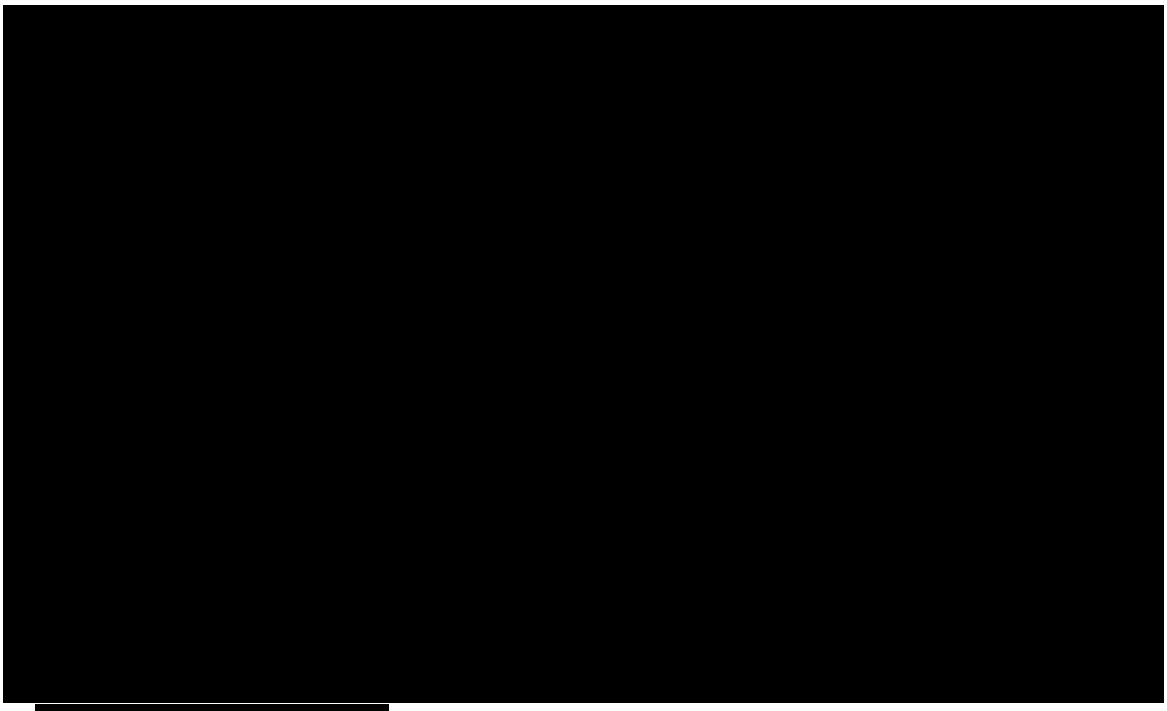
[redacted]

[redacted]

[redacted]

[redacted]

[redacted]



[redacted]

[redacted]

[redacted]

[redacted]

[redacted]	
	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	

[redacted]

[redacted]

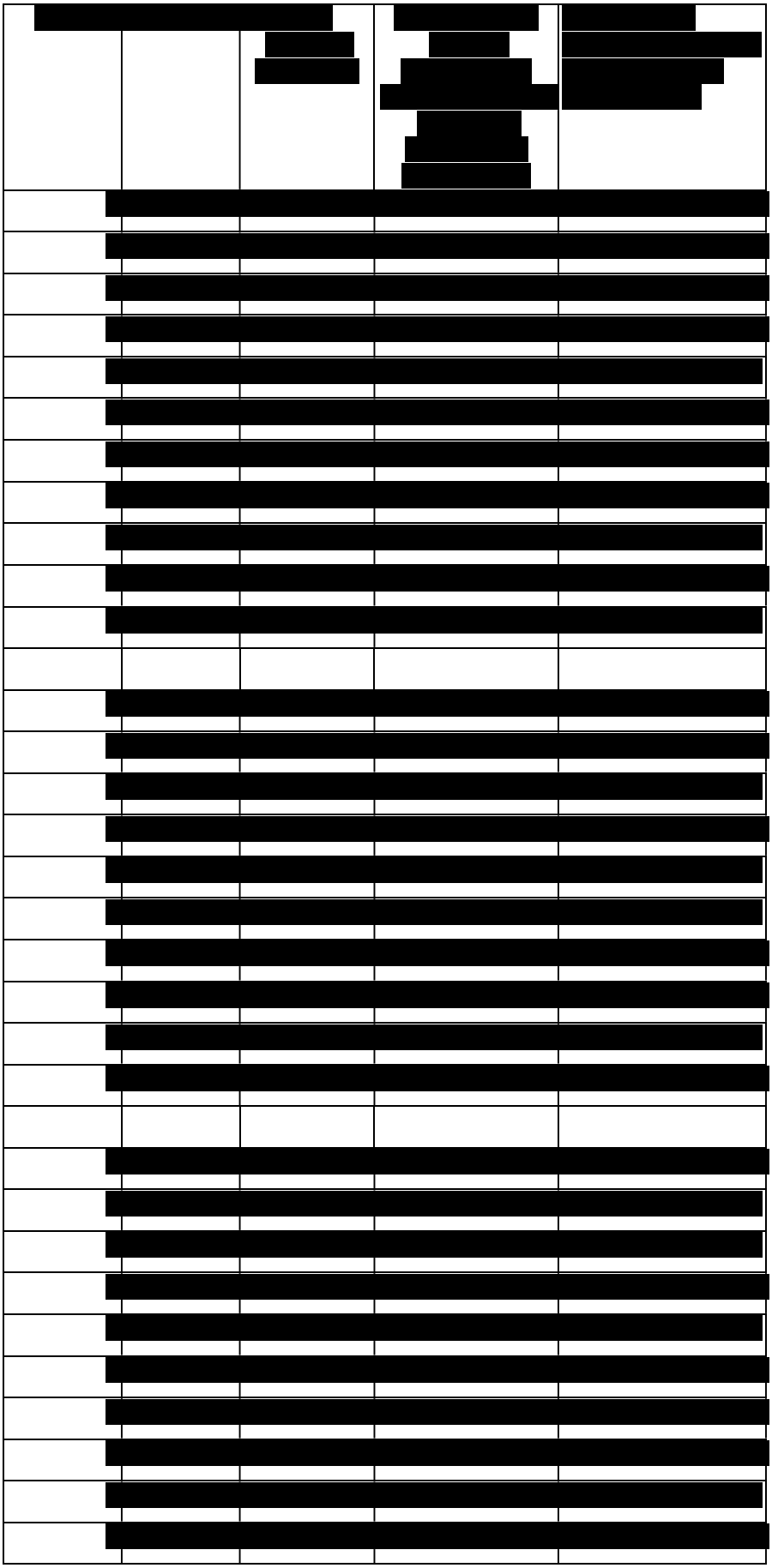
[redacted]

	[redacted]		
			[redacted]
[redacted]		[redacted]	[redacted]
[redacted]		[redacted]	[redacted]
[redacted]		[redacted]	[redacted]
[redacted]		[redacted]	[redacted]
[redacted]		[redacted]	[redacted]
[redacted]		[redacted]	[redacted]
[redacted]		[redacted]	[redacted]
[redacted]			[redacted]
[redacted]			[redacted]

[redacted]

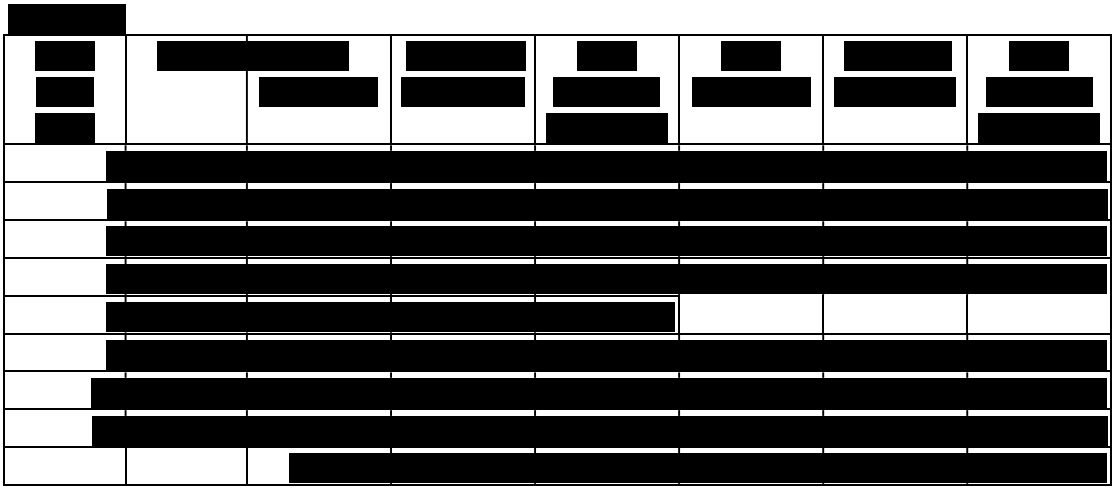
[redacted]

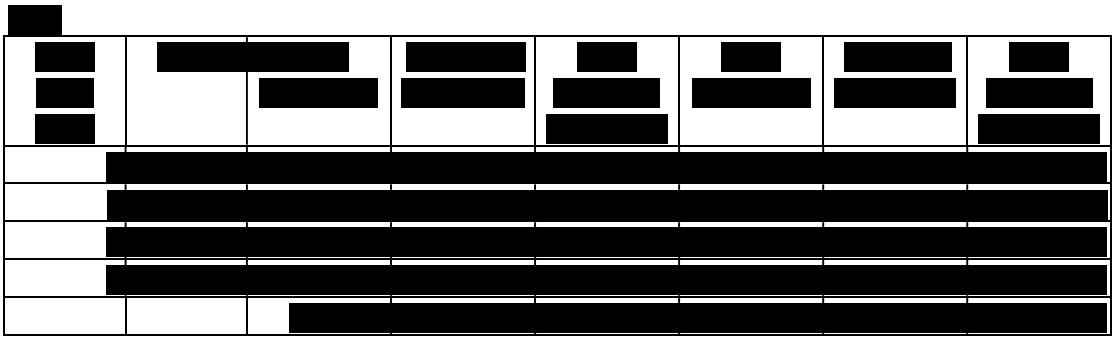
[redacted]



[redacted]

[redacted]





[redacted]

Creation of 2011 area sampling frame

5.1.1 Introduction

In recent Monthly Population Survey (MPS) redesigns, Census Collection Districts (CDs) have been used as the basic geographic building block for the area frame. In 2006, CDs were the second smallest geographic area defined in the Australian Standard Geographical Classification (ASGC). Additionally, in 2006 there were about 38,200 CDs throughout Australia and its other territories.

[redacted]

For the 2011 MPS redesign, we have created a new sampling geography that will support these initiatives. [redacted]

[redacted]

Section 5.1.4 describes the procedure undertaken to create the area sampling units,
Section 5.1.5 discusses the steps undertaken to finalise the area sampling frame, and
Section 5.1.6 provides some descriptive statistics of the area units on our new frame.

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

5.1.2.3 Description of final geographic strata

Table 5.1.2.3.1 provides a summary of the number of strata in 2011 compared with the 2006 redesign. There are about 50 fewer strata across Australia in 2011 compared with 2006. Most of the decrease is from areas outside of the capital city SA4s.

[redacted]

5.1.3 Private dwelling population counts

In order to form the area sampling frame, counts of the number of MPS private dwellings in each mesh block were required. Both occupied and unoccupied private dwellings were included in these counts, 'scoped' to the Population Survey Operations (PSO) definition of a private dwelling (see below). These counts were then used in conjunction with the optimal state by area type design cluster sizes to form the BFUs, FSUs and FSU-2s.

For the geographic stratification and the cost-variance optimisation process, pre-Census data was used. The pre-Census data was based on a 2011 total dwelling estimate constructed for the enumeration of the 2011 Census. This was initially calculated in 2009 based on building approvals, and updated as part of the 2011 Census Collector Workload design. These counts included all private (occupied and unoccupied) as well as non-private (including special dwellings) dwellings.

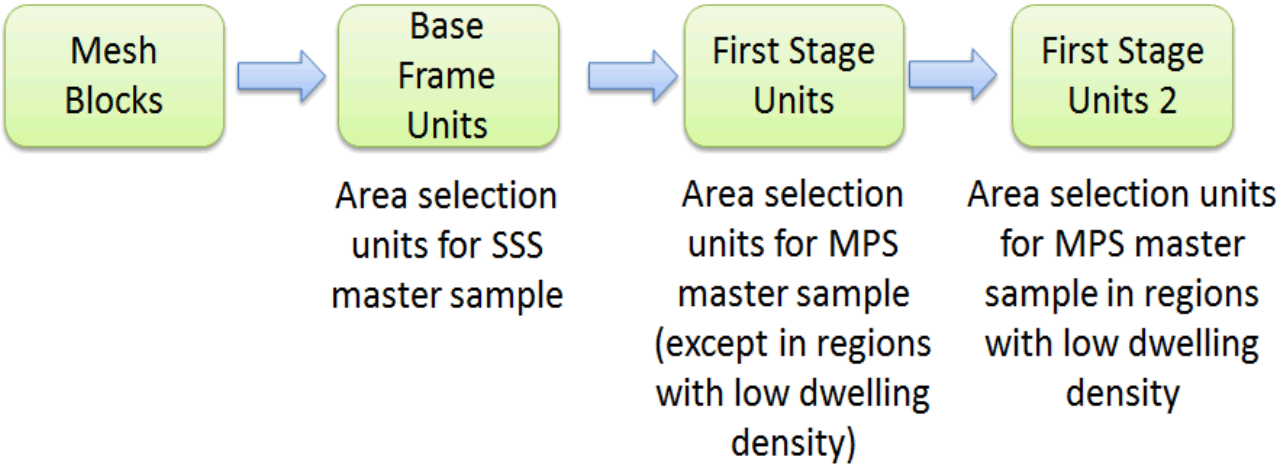
For the formation of our area sampling units, we made use of preliminary 2011 total private dwelling Census counts. These private dwelling counts were for occupied and unoccupied dwellings, adjusted to account for differences between the Census and MPS definitions of private dwellings. This adjustment was made at the mesh block level using

5.1.4 Area unit formation

5.1.4.1 Criteria for forming area sampling units

Figure 5.1.4.1.1 depicts the hierarchy of the various types of area units on the area sampling frame. As can be seen, the mesh block geography on the ASGS is used as the basic building block to form BFUs. Whole BFUs are then pooled together to form FSUs. In remote areas, FSUs are used in a final stage of pooling to form FSU-2s.

Figure 5.1.4.1.1: Area sampling frame hierarchy



[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]	[redacted]	[redacted]
[redacted]		
[redacted]		
[redacted]		
[redacted]		
[redacted]		
[redacted]		
[redacted]		
[redacted]		
[redacted]		

[redacted]			
[redacted]			
[redacted]			
[redacted]			

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

5.2.2.5 Implementation in 2011 Redesign

The change in statistical geography from the ASGC to the ASGS gave the 2011 Redesign team the opportunity to re-evaluate the method used for determining the width of the latitude bands used to sequester FSUs in different parts of the stratum. There was one conceptual change and a change in the rounding parameter value used.

It was decided that simply using the 'average' area of an FSU in the bandwidth calculation, regardless of how that 'average' area was calculated, was not the right approach. Area is a two dimensional quantity - it's units are the the square of some one-dimensional quality's units, which in this case are distance and kilometres. The latitude bands have a length and width but the width is the only quantity that needs to be estimated; the length of the band is determined by the distance between the left-most border of the stratum/FSU-2 and the right-most border of the stratum/FSU-2 at certain longitude co-ordinates. Therefore, rather than using than area, a two-dimensional quality, in the rounding formula for the latitude bands a one dimensional quality was used instead. This resulted in the median of the average area of FSUs within the stratum/FSU-2 being replaced by the *square root* of the median average area of FSUs within the stratum/FSU-2.

Also, as the make up of the first stage sampling units had changed from the 2006 Redesign the rounding parameter applied last time might not have been the right value to use in the current Redesign. After some trial and error it was decided that the rounding parameter value that produced the bands that fit the best was 125. Other values considered were 30, 60 and 250.

[redacted]

[redacted]

5.3 Private Dwelling Selections

5.3.1 Introduction

The purpose of this documentation is to give explanations of the methodology and outputs for the second phase of the private dwelling (PD) selections undertaken for the 2011 MPS Redesign. The private dwelling component of the MPS has a multistage sample design: in Self-Representing areas (SRAs), where the population density is sufficiently high, the first stage of selection involves selecting a number of FSUs probability proportional to size (PPS) without replacement from a stratum of FSUs with similar characteristics. In more remote areas, called non-Self-Representing Areas (non-SRAs) FSUs are first pooled together to form First Stage Unit-2s (FSU-2s) and a small number of these FSU-2s are selected PPS without replacement from within each stratum. FSUs are then selected PPS without replacement only from the selected FSU-2s. The addition of this extra stage in the more remote areas increases the clustering of the sample, thereby reducing costs. The selection of a FSU also simultaneously implies a selection of a Base Frame Unit (BFU). The final stage of selection is the systematic selection of a cluster of more or less equally dispersed dwellings from within each selected BFU.

[redacted]

[redacted]

5.3.2 Background

5.3.2.1 Selection in Self-Representing Areas

The selection mechanism for the MPS in Self-representing area (SRA) strata can be described as a three stage selection process, in which FSUs are selected from each stratum, BFUs are selected from within each selected FSU, and one cluster of dwellings is selected from within each selected BFU. This description can cause confusion initially, because the separate stages of selection are not performed independently in the way we might expect, based on our previous experience of multistage selection processes. The random start that is used to select FSUs also simultaneously identifies which BFU and which cluster are to be selected; independent random starts are not chosen for each successive stage of selection. The independence of selection at each stage depends on the assumption that the method of ordering BFUs within selected FSUs and ordering clusters within selected BFUs are both random, an assumption that may not be entirely true in practice.

To understand the selection procedure in SRA strata, it is useful to think of the selection process in terms of how the whole process achieves the selection of clusters of dwellings. Given population dwelling counts and cluster sizes for the FSUs, it is possible to determine how many clusters each FSU has. Each stratum of FSUs can therefore be considered to form a list of clusters of the population dwellings. We do not know what dwellings are actually in each cluster, but we know approximately how many clusters there are (and the closer our frame counts are to the true population when the sample arrives in the field, the more accurate this measure will be - this is in fact why we spend so much time trying to come up with accurate population counts for the private dwelling frame). Given such a list of clusters and a skip (the state skip for the state to which the stratum belongs), it is easy to choose a systematic sample. For example, if the random start for the stratum was 32 and the state skip was 250, then we would select clusters 32, 282, 532 and so on.

This is all well and good, except that we don't know which dwellings belong to these clusters. This is where FSUs and BFUs comes into play. FSUs and BFUs are essentially used as a mechanism for homing in on the area each selected cluster should come from, before finally splitting that local area up into true clusters of dwellings on the ground. This avoids having to try and split the whole of Australia up into clusters of dwellings on the ground, which is not possible, and is the whole reason why household surveys are selected from an area based frame instead of a list based frame in the first place.

Given the number of clusters in each FSU and a fixed ordering of the FSUs (serpentine ordering), the FSUs containing the selected clusters can be identified. For example, cluster 32 above may be the 32nd cluster of the first FSU. Cluster 282 may be the 16th cluster of the 8th FSU, and so on. As FSUs are made by pooling BFUs (which in turn are made by pooling MBs) not only does each FSU contain a defined number of clusters, but each BFU within it also contains a defined number of clusters. Therefore the BFUs containing the selected clusters can also be identified. For example, cluster 32 above may be in the 4th cluster of the 2nd BFU in the first FSU and cluster 282 may be the 16th cluster of the 1st BFU in the 8th FSU.

Occasionally a BFU will be too big (geographically) to skip through it to form clusters, so blocks within the BFU must be formed. Each block is defined by a clearly defined boundary, so that every dwelling that exists or could subsequently be built in the future in the BFU will fall exactly one block. PSO get an up-to-date count of the BFU and each block so that each block consists of a known number of dwellings and therefore a known number of clusters (note that the number of clusters in each block is calculated in such a way as to take into account differences between the BFU-level counts on the frame and in the field. That is, so as to preserve consistency between the number of clusters allocated to the BFU on the frame and the number of clusters allocated to all blocks in the BFU out in the field). Again, once the blocks within the selected BFU are formed into an ordered list, it is possible to identify exactly which block each of the selected clusters belongs to.

Once the selected BFU (or in rare cases, BFUs) within a FSU have been identified, we finally get to the point of identifying which dwellings belong to each cluster. PSO form an ordered list of the dwellings in the selected BFU or block. The clusters are formed by running a skip through the list, so that the dwellings in each cluster are more or less equally distributed around the BFU/block, producing a heterogeneous BFU/block sample. PSO then select the required cluster from this list, cluster three in our example above.

5.3.2.2 Selection in Non-Self Representing Areas

Selections in non-SRA strata, the least remote, remote and very remote strata, have a fourth stage of selection. In these areas, the FSUs in each stratum are pooled together to form First Stage Units-2 (FSU-2s). Each FSU-2 is a group of FSUs located reasonably close together geographically. They are formed so as to be large enough to produce enough dwelling selections to form a reasonable workload for a single interviewer, while covering a small enough area for a single interviewer to be able to enumerate within the enumeration period. Another condition of their design was that they are also formed so that, in so far as possible, each of them could represent the whole stratum. Thus, in a non-self representing Area type the first-stage selection unit does not only represent itself but other selection units within the same stratum. More description of how FSU-2s were formed can be found in Section 4.2.4.

FSU-2s are selected from each stratum and then FSUs, and BFUs and clusters, are selected from the selected FSU-2s. The idea is to select a small number of FSU-2s from each stratum, and then select a larger than normal number of FSUs/BFUs/clusters from within each of those selected FSU-2s. As the FSU-2s are small geographic areas, this means that instead of selecting clusters in FSUs spread throughout the whole stratum, we restrict our attention to a small number of areas, in which we take an increased amount of

sample. This increases the clustering of the sample in the more remote areas. However, it also cuts down on the number of interviewers required to enumerate the sample than would be required if the sample was spread out geographically, thereby reducing costs.

The FSU-2s in each stratum are ordered by increasing density and then selected from each stratum PPS without replacement. Each FSU-2 within the stratum is given a size measure specifying the number of workloads in the FSU-2. A random start is selected, and the state skip is used to select the FSU-2s based on which FSU-2s the state skip lands in according to the size measure. Once a FSU-2 is selected, the remaining three stages of selection are undertaken within the FSU-2, in the same manner as they are undertaken within SRA strata. However, the FSU-2 stage is truly independent of the remaining stages of selection, with a new random start chosen to select FSUs from within each selected FSU-2s, independent of the random start chosen to select the FSU-2s. More detail of how FSU-2s selections are undertaken is given below in Section 5.3.4.1.

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

5.3.4 Methodology

The FSU-2, FSU and BFU stages of selection are performed by HSM. The resulting selections were then provided to PSO, who count and block the selected FSUs and BFUs in the field and complete the final two stages of the selection of dwellings for the private dwelling component of the MPS, using the selection information provided.

5.3.4.1 Notation

There are *S* states, and the state *s* can take any value in the set $\{1,2,...,S\}$. Here S = 8 and s = 1, 2, 3, 4, 5, 6, 7, 8 is equivalent to s = NSW, VIC, QLD, SA, WA, TAS, NT, ACT;

There are A_s area types within state *s*, and the area type *a* in state *s* can take any value in the set $\{1,2,...,A_s\}$;

There are H_a area types within area type *a,* and the stratum *h* in area type *a* can take any value in the set $\{1,2,...,H_a\}$;

There are J_h FSU-2s within stratum *h,* and the FSU-2 *j* in stratum *h* can take any value in the set $\{1,2,...,J_h\}$. Note that SRA strata do not contain any FSU-2s but in this notation the SRA stratum will also act a FSU-2; and

There are I_j FSUs within stratum/FSU-2 *j,* and the FSU *i* in stratum/FSU-2 *j* can take any value in the set $\{1,2,...,I_j\}$.

A prime in the superscript means that the value represented by the variable in the base is an adjustment to the value represented by that variable without a prime in the superscript. Similarly, two primes in the superscript means that the value of the variable is an adjustment of the value represented by the variable with only one prime in the superscript;

An asterisk in the superscript means that the value of the associated variable was not adjusted at any further stage during the PD selections. In other words, if there was a cluster count c, and if there were variables c' , c'' and c* then c' would be an adjustment of c, c'' would be an adjustment of c' and c* would be an adjustment of c'';

5.3.4.2 FSU-2 Stage of Selection

This stage of selection was only used in non-SRA (least remote, remote and very remote) strata. Generally, one FSU-2 was selected per stratum. However, if the expected number of selections from the stratum (defined to be the number of dwellings in the stratum divided by the state skip) was between 50 and 66, then two FSU-2s were selected. If the expected number of selections from the stratum was between 67 and 100, then three FSU2-s were selected and so on. Three workloads containing on average 100 dwellings between them was seen to be a valid workload size by PSO.

FSU-2s were sorted by increasing dwelling density within each stratum. Using this sort, a heterogeneous sample was obtained. FSU-2s were then selected probability proportional to size without replacement. A size measure was assigned to each FSU-2 depending on the number of MPS private dwellings in the FSU-2 and the number of FSU-2s to be selected and this size measure was used to derive an unconditional probability of selection for the FSU-2. Based off this unconditional probability and the FSU-2s' past usage a conditional probability was then derived that minimised overlap between the 2006 MPS sample and the 2011 MPS sample. It was this conditional probability that was used to select FSU-2s. An independent random start was determined for each stratum so the FSU-2 selections were therefore independent from stratum to stratum. The state skip was then run through each stratum from the start point to select FSU-2s PPS. A more detailed explanation of the selection process for FSU-2s is given below.

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

5.3.4.3 FSU Stage of Selection

FSUs were selected from selected FSU-2s in non-SRA strata and directly from the stratum in SRA strata. [redacted] FSUs on the frame were sorted by serpentine ordering within each stratum/FSU-2. Using this sort, the FSU selections were distributed geographically around the stratum/FSU-2, producing a more heterogeneous sample.

FSUs were selected probability proportional to size without replacement. A size measure was assigned to each FSU depending on the number of MPS private dwellings in the FSU and the optimal cluster size for the stratum to which the FSU belonged. An independent random start is determined for each stratum/FSU-2 so that FSU selections were independent from stratum to stratum and FSU-2 to FSU-2. A skip was then run through each stratum from the start point to select FSUs PPS.

The only difference between selection in SRA and non-SRA strata was the skip used to skip through the clusters in the selected FSU-2s and strata to select FSUs. In SRA areas, all FSUs were listed in serpentine order within strata and FSUs were simply selected using a random start and skipping through the cumulative clusters using the state skip. The process for selecting FSUs within FSU-2s was the same, except that the number of FSU-2 clusters in the FSU-2 was used as the skip. This ensured that an increased amount of sample was taken from the small number of selected FSU-2s, resulting in the desired overall stratum sample size clustered in small geographic areas. Using the state skip would produce less sample at the stratum level than required, because that would mean focusing in on a small number of FSU-2s.

As the process for selecting FSUs within FSU-2s was the same, except for the skip, selected FSU-2s in non-SRA areas and strata in SRA areas were treated the same in the algorithm and code, with only the skip changing. [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

6.1 Background

Determining a framework for sampling Australia's Indigenous community population (see section 6.3 for a definition) presented unique challenges due to the particularities of this sub-population compared to the population at large. Key reasons for having a special sampling framework for the Indigenous community population included:

I. Needing to control burden/load on communities within and across surveys;
II. Explicitly controlling the number and types of communities selected;
III. Managing special operational requirements prior to selection;
IV. Meeting survey scope and coverage requirements where they relate to community status;
V. Tailoring sampling mechanism to facilitate complex operational requirements, in particular the need to approach main communities to enumerate outstations; and
VI. Controlling survey cost/accuracy outcomes and ensuring cost-efficient sampling for this particular population;
VII.

With regards to the first and second points, stabilising and controlling the number of Indigenous community dwellings in sample would reduce the volatility in LFS estimates as well as hopefully curtail the historically high rates of sample loss and non-response from these communities. This is important because the labour force characteristics of the Indigenous population living in communities are significantly different to the rest of the population. With regards to the third point, these requirements include ensuring that the sample design reflects the appropriate sensitivity to community needs such as allowing interviewers to readily obtain approval to access communities.
I.
[redacted]

[redacted] Thus, the sampling framework remains relatively similar to previous redesigns. Under the new proposal, the ICF would be removed and the PDF would once again [redacted] cover all areas. There would only be one sampling frame for private dwellings (the PDF). While there have been no fundamental structural changes some major improvements or changes to note include:

- [redacted] Sampling of Indigenous communities would occur under the PDF and be facilitated by:

    I. **Defining a set of Indigenous Community Primary Sampling Units (ICPSUs) across Australia**. As per the PDF these PSUs would be defined as area units, in this case an aggregate of mesh blocks, would each ideally contain a single main community and seek to incorporate linked outstations as best as possible. These PSUs would be similar in nature to those on the current ICF and similar sampling procedures would be applied within them.

    II. **Defining Indigenous community sampling strata on the PDF**. These strata would be composed completely of ICPSUs where possible but in those cases where there were not enough ICPSUs to form a viable sampling stratum (for example in some states with small numbers of communities) they would join standard BFUs in some other relevant stratum. Despite the conceptual difference, this is practically the same as the previous approach except that communities formerly excluded from the ICF (and selected under the PDF) are now flagged up front. The problems of small strata are removed and it allows communities to be identified and treated (e.g. removed from scope) at the individual ICPSU level.

-

A common problem with the ICF was that states with fewer communities (in particular NSW) were not incorporated into the ICF since ICF sampling strata in those states were often too small to select a viable sample. Communities not on the ICF therefore received their chance of selection under the PDF. This had on occasion resulted in some confusion around issues such as Indigenous community scope and sampling/enumeration requirements due to an assumption that the ICF exhaustively covered all Indigenous communities.

- The Primary Sampling Units were purely area based, being defined by one or more mesh block area units rather than a particular main community. This was possible because mesh blocks are typically much smaller than CDs. One major benefit of area based sampling units was that the coverage of the ICPSUs would probably be much improved over the Indigenous sets from the 2006 design, and in particular it would be easier to deal with non-communities and communities alike not listed on the frame suddenly coming to attention.

- Improved main community/outstation links. This was noted by the 2006 Redesign team as much needed area of improvement. The 2011 Redesign team received a data set of inter-community links sourced explicitly from Census 2011 where possible and current ILDB links in the remaining cases where much effort was made in the lead up to the 2011 Census to update the link information.

6.2 Purpose of document

The purpose of this documentation is to give explanations of the methodology for the construction of the frame of the first stage sampling units for Indigenous communities, referred to as ICPSUs. There is a particular focus on the difficulties faced in not only constructing these ICPSUs as well as effort required in determining their probabilities of selection. Furthermore, given that it was a two-stage selection process, there is a description of how both the first-stage was selected and the second-stage sample was constructed respectively. The programs used in the course of this project are also explained in the appendices at the end of the document.

Section 6.3 sets up some definitions regarding the indigenous population that will be used throughout the rest of the document. Section 6.4 lists the main data sources that were used to construct ICPSUs, Section 6.5 details the methodology behind that construction as well as the formation of an ICPSU-level frame and Section 6.6 outlines the procedures used to populate the ICPSUs with clusters and then sample ICPSUs from the frame and then take clusters from the selected ICPSUs.

6.3 Some Definitions

Indigenous Community - An Indigenous community is defined as a physical, geographic location bounded by physical or cadastral (legal) boundaries, that is predominantly inhabited by Indigenous people. These are are usually, but not always, located in the remoter regions of Australia. It should be noted that "Indigenous community" does not refer to the Indigenous population at large (most of which is dispersed thinly across densely populated areas) or any cultural notion of community that transcends geography. Indigenous persons not residing in communities are assumed to be the same as the general population with respect to required enumeration procedures and appropriate collection instruments.

For the 2006 Redesign there were essentially only two types of Indigenous communities but with the development of the ILDB there are now seven different types, with most of the new ones coming from a tightening of the definitions from 2006. [redacted] For the rest of this document, the two main types of communities that will be referenced are 'Main Communities' and 'Outstations'.

Main Community/Outstation - An Indigenous community is classified as a 'main community' if it has a main community identifier on the ILDB or the estimated population of the community is 40 or more persons and does not have a link to a parent community. [redacted]

6.4 Data sources

There were a number of data sources used to construct the ICS and undertake selections. The main data sources used to develop the frame as well as develop size measures for sample selection are listed below. [redacted] The data sources are loosely grouped by their purpose in the process of forming ICPSUs.

[redacted]

6.5 Formation of Indigenous Community Primary Sampling Units (ICPSUs)

6.5.1 Designation of ICPSUs

Individual Indigenous communities were selected for ABS household surveys via the selection of the mesh blocks which contain them. The set of mesh blocks containing communities will be referred to below as the "Indigenous community mesh blocks". Instead of Indigenous community mesh blocks being selected individually, they were selected in (mostly) contiguous groups referred to as "Indigenous Community Primary Sampling Units" (ICPSUs) that fully covered the Indigenous community sampling stratum within a state. The ICPSUs aggregated mesh blocks according to the cultural links between the communities within the mesh blocks. It should be noted that a mesh block containing one or more main communities without links to communities in other mesh blocks falls under the definition of an ICPSU. This strategy of grouping communities for selection ensured culturally appropriate procedures could be applied for contacting and coordinating the enumeration of Indigenous communities.

Any ICPSUs that is selected will be selected within an Indigenous community sampling stratum. Some Indigenous communities were assigned to non-Indigenous strata (see Section 6.4.2 for further details of stratification) because there were not enough communities to form a viable sampling strata. In Indigenous strata, the ICPSUs are the first-stage selection unit. The Indigenous areas that were not within an Indigenous stratum were treated as Base Frame Units (BFUs) in their stratum and were flagged up front for containing Community dwellings.

6.5.2 Stratification of ICPSUs

The MPS design assigns every mesh block to a stratum for the purpose of sample selection. [redacted]

[redacted]

[redacted]

[redacted]

6.5.3 Basic sampling strategy

The basic sampling strategy for the ICS was a two-stage selection process where a first-stage sample of ICPSUs was selected and then a second-stage sample of cluster of dwellings from within the selected ICPSU was selected. It should be noted that each ICPSU had the potential to have three different types of dwellings, namely:

Community dwellings: All the private dwellings associated with each community within an ICPSU were called 'community dwellings';
Non-community private dwellings: All private dwellings within a MB that were not assigned to an Indigenous community; and
Special Dwellings: Non-private dwellings said to be located within a MB. This data was given at the MB level;

The two-stage approach was used due to the the perceived cost and difficulty of approaching child communities independently of the parent community they are associated with. Firstly there are significant cost savings in enumerating child communities only when their parent community is selected since this avoids extra costly journeys to these individual child communities, and secondly it can be culturally inappropriate to approach (and very difficult to obtain any useful data from) a child community without contacting the parent community it is associated with.

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

Rounding process

When dwelling counts were pro-rated to mesh blocks or communities it was often the case that this new dwelling count was not a whole number. In those cases the dwelling counts for all mesh blocks within a CLW or all communities within a mesh block were adjusted in such a way that:

- all mesh blocks/communities in a CLW/mesh block had an integer number of dwellings; and
- the mesh blocks'/communities' dwelling counts summed up to the CLW/mesh block dwelling count;

[redacted]

Seasonal dwelling counts

It may have been the case that some communities that have low Census population counts, which is not uncommon for 'outstation' communities because of the seasonal nature of occupancy of the communities. [redacted]

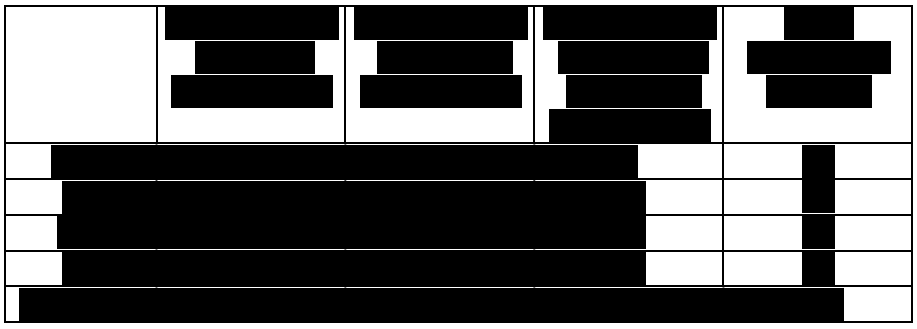
[redacted]

[redacted]

[redacted]

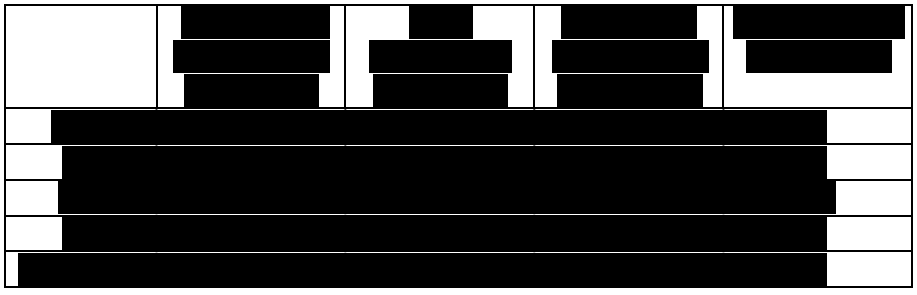
[redacted]

[redacted]



[redacted]

[redacted]



[redacted]

6.6 Sampling Method

The 2011 ICS sampling method was exactly the same as that employed for the 2006 ICF, and uses a two-stage sample design. The first stage was a systematic probability proportional to size (PPS)

sample of ICPSUs where the ICPSUs were ordered geographically within each state x size cell. The second stage sample was a selection of 1 cluster of dwellings from the selected ICPSUs.

This is basically the same approach that was taken for the PDF if ICPSUs were viewed as being equivalent to FSUs, except that the PDF sample design includes an extra BFU stage of selection and an extra FSU-2 stage of selection in non-SRA areas. Another essential difference is the nature of the clusters within ICPSUs, which can include private dwellings and non-private dwellings whereas the PDF clusters include only private dwellings.

6.6.1 First stage sample (ICPSUs)

The first stage sample consists of a sample of Indigenous community primary sampling units or 'ICPSUs' taken from an ICPSU-level frame.

6.6.1.1 Ordering of ICPSUs

There were two variables used to sort ICPSUs within a state: a size measure and serpentine rank. The serpentine rank was calculated by the same method that had been used previously in the HSR when required. The size measure was based on the number of clusters in the ICPSU. Immediately before systematic selection the ICPSUs for a given state were first ordered by this size measure and within each size class the ICPSUs were then sorted by the state-level serpentine rank.

Given the way the ICPSUs were constructed, continually grouping together mesh blocks until there was the latest set of mesh blocks added did not have parent communities in different mesh blocks, the size and location of the ICPSUs was not known before this pooling of mesh blocks had finished. Early in the planning for the sample design in the Indigenous community strata there was discussion about creating separate strata in the NT for large and small communities. The rationale was to ensure that the sample has an equal representation of communities of different size. In the end it was decided that rather than pursue this through stratification it would be done by sorting prior to selection. [redacted]

The frame of ICPSUs was then ordered using geographic serpentine ordering within each size class. To facilitate this, each ICPSU was assigned the latitude and longitude coordinates belonging to its 'Greatest Parent'.

Serpentine ordering orders the ICPSUs in a lateral snake like fashion moving from north to south within each state by ICPSU size cell. The rationale behind serpentine ordering is to obtain a sample with greatest possible spread, with an absence of the clumping that could occur if a simple left-to-right approach is used. Read Section 5.5 of the Redesign Manual for more explanation of serpentine ordering in the context of the PDF selections.

6.6.1.2 Assigning a number of clusters to each ICPSU

Following the ordering and the stratification of the ICPSUs, a PPS systematic sample was selected from the ordered ICPSUs where an ICPSU's size measure is the number of clusters in the ICPSU. In order to do this each ICPSU needed to be assigned set a number of clusters. It would have been ideal that for each cluster within the ICS strata to have as similar size as possible since this will reduce sample variability. Therefore, an average cluster size for each ICS stratum was calculated and this cluster size was then used to determine the number of clusters within each ICPSU.

6.6.1.2.1 Constriction of dwelling counts in ICs

Historically there have been high rates of sample loss from indigenous communities enumerated as part of the LFS, especially from 'outstation' communities, and this might be partly due to an overstatement of the number of dwellings that were expected to provide sample. The dwelling count of an Indigenous community determines the number of clusters that is gets assigned. Section 6.5.4.2 detailed the complicated method of estimating the number of dwellings were in a community, given that the fine-level count was unvalidated and the only validated dwelling counts came from a much broader level (the Census Collector Workload). While a great deal of effort was expended trying to obtain reliable data for Indigenous communities in the run up to as well as during the 2011 Census HSM was still wary of the potential for high rates of sample loss from these communities.

HSM decided that the population of persons at Census time should be used in conjunction with the Census dwelling count to estimate the number of dwellings likely to be occupied, as opposed to just using Census dwelling counts which would also include unoccupied dwellings. The dwelling count that would be used as the measure of size in an Indigenous community was the minimum of $N'_{D,IC}$, the calibrated number of dwellings in the community, and $\frac{N_{POP,IC}}{2}$, half of the number of the people in the community at the time of the Census. This adjustment was used because it was deemed to be sufficiently conservative in its impact on low population communities as compared to other suggested methods to deal with the unreliable community-level dwelling counts.

This estimate of dwellings likely to be occupied in a community was used to determine the number of clusters. It should be noted that this approach would cause an overall reduction in dwelling counts across communities, which in turn would lead to fewer community selections; this new method would improve operational efficiency by reducing the likelihood of high sample loss from communities and reduce expected levels of sample loss per community selection.

In order to reduce the risk that when community is enumerated for the LFS and the interviewer finds that in the field there is significantly more population in the ICPSU than they were expecting, any reduction of dwellings expected in a community was adjusted to ensure that the revised total community dwellings in an ICPSU must not be less than 80% of the original total number of expected community dwellings. The formula for this capping procedure was:

[redacted]

[redacted]

6.6.2 Second stage sample (dwellings)

The second stage sample is a single cluster of dwellings which can include dwellings from any of the three set components:

- Community dwellings
- Non-community dwellings
- Special Dwellings

[redacted] The type of clusters formed will change on a case-by-case basis depending on the distribution of dwellings between the three type of dwellings as well as the links of the communities within the ICPSU to other communities. [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

Automatic rotation

If a sufficiently large proportion of the ICPSU had been previously selected, it is simply replaced by the next available ICPSU on the geographically serpentine ordered list. At the time of writing the current threshold for rotation was if the ICPSU contained 2 or less clusters worth of dwellings that had not been selected for the 2006 MPS. This replacement method of course has the disadvantage of opening up the sample to bias, depending on how different the replacement is with respect to the variables of interest.

Decision

It was decided that, [redacted] the automatic rotation method would be favoured [redacted] due to its simplicity. At the time of writing an initial selections had been made and had been provided to PSP for them to investigate the level of past usage in those selections. After receiving their feedback the Redesign team would alter the selections accordingly. Recent developments have meant that these new selections might not be put out into the field until mid-2014, so the selection from the 2006 Redesign will continue being rolled over until then.

6.6.5 ICS selections

The final MPS 2011 Indigenous selections were carried out according to the methodology above [redacted]

[redacted]

Community Definitions

The community definitions used on the ILDB have been updated from the definitions used to classify indigenous communities in the 2006 Redesign. While these definitions were not necessary in the body of the documentation they are given below for reference.

- **Discrete Indigenous Community**: A geographic location, bounded by physical or cadastral (legal) boundaries, and inhabited or intended to be inhabited predominantly (i.e. greater than 50% of usual residents) by Aboriginal or Torres Strait Islander peoples, with housing or infrastructure (power, water, sewerage) that is managed on a community basis. Discrete communities have populations of (but not limited to) 50 or more Aboriginal or Torres Strait Islander People. Services such as schools, health clinics, shops, and council depots are usually present.
- **Outstation**: A small community of mostly Aboriginal and Torres Strait Islander peoples in a permanent or semi-permanent residence with a water supply and some form of accommodation. These sites generally do not have community essential services but rely on house based systems (eg a water tank, generators and septic tanks). Generally administered and serviced by a larger discrete community, outstations usually have a population of (but not limited to) less than 50 people. Services such as schools, shops, health clinics and council depots are limited and in most cases, not provided.
- **Regional Centre**: A regional, rural or remote town that exists outside of the major metropolitan areas of Australia which has a significant Aboriginal and Torres Strait Islander population (over 20%). These centres may also act as a parent to surrounding outstations
- **Town Camp/Reserve**: A discrete community of mostly Aboriginal and Torres Strait Islander peoples situated within or adjacent to an urban area. Generally found on the fringes of towns and cities.
- **Cultural/ Seasonal Camping Areas**: Generally homeland locations inhabited by Aboriginal and Torres Strait Islander peoples on a seasonal or temporary basis. Can include areas of cultural significance (used for ceremony), or areas of temporary stay such as stock camps waterholes and bore sites. No permanent dwellings are present.
- **Pastoral Stations**: A pastoral station that is owned by and falls within Aboriginal homelands. Such pastoral properties are generally inhabited by traditional owners and viewed as their homeland, however these pastoral stations can be leased back to caretakers to continue running the property as a pastoral business.
- **Other Known Population**: An ‘other known’ Aboriginal and Torres Strait Islander population that cannot be categorised into any other Community Type. To be used only as a last resort.
-

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

6.8 Final Selections

There was a significant delay in the introduction of the ICPSU sample compared to the rest of the 2011 MPS design sample. [redacted]

There were four main steps in this process:

1. Run code to select initial ICPSUs

[redacted]

2. Negotiate with PSO which ICPSUs will be selected

As there was extensive past sample usage in Indigenous communities, largely due to NATSIHS/NATSINPAS, much discussion had to be had with PSO about which Indigenous Communities would be OK to use (and therefore which ICPSUs we could select). [redacted]

3. Form clusters within selected ICPSUs

The aim of this was to split the main communities, outstations, non-community dwellings (split by meshblock) and special dwellings amongst the clusters in the ICPSU in a way that resulted in clusters of roughly equal size and minimal travelling distance whilst still maintaining a reasonable spread of dwelling types in each cluster.

- We roughly followed the cluster formation strategy defined in section 6.9
- Occasionally it was more convenient to treat a large outstation as a main community (only if it didn't have a parent community in that ICPSU) or a small main community as an outstation.

- In some ICPSUs the number of non-community dwellings was so large that if we spread the community dwellings across all clusters the number of community dwellings in each cluster would be too small. Therefore some clusters had no community dwellings. (Theoretically such ICPSUs shouldn't exist, as an area should have only been assigned to the Indigenous Community stratum if at least 75% of its dwellings were in Indigenous communities, so there were presumably differences between the counts used for stratification and the final counts.)
- [redacted]

4. Produce output files

After the clusters were formed in a spreadsheet, that spreadsheet was read into SAS and the final files were produced. [redacted]

6.9. Cluster Formation Strategy

0. Background

As part of the indigenous stratum of the MPS Redesign a new set of first-stage sampling units called Indigenous Community Primary Sampling Units (ICPSUs) were created. These ICPSUs are the analogues of the Indigenous Community Sets created as part of the 2006 MPS Redesign process. Those 'Sets' were groups of indigenous communities that shared cultural ties and/or geographical proximity while ICPSUs are groups of mesh blocks (MBs) that contain indigenous communities such that the MBs within the ICPSU contain as many communities as possible that share cultural ties and/or geographical proximity without the ICPSU becoming operationally unfeasible.

An ICPSU would have been considered 'operationally unfeasible' on the basis of the estimated number of dwellings that would be within each proposed ICPSU; there were minimum thresholds in place for the number of dwellings that an ICPSU was allowed to have. [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

1. Introduction

Having determined the number of clusters within each ICPSU the next step of the process is populate the clusters within ICPSUs that have been selected for the MPS or the SSS master samples with dwellings. That is, cluster formation is based on the frame dwelling counts. It should be noted that is what is currently proposed. There is the alternative of attempting to form clusters shortly before enumeration after an attempt to assess the likely size of the community within the ICPSU.

For the purposes of creating clusters there are four community types that contain dwellings:

1. "Main Community" : Communities listed with a Discrete Community, Regional Centre or Town Camp/Reserve identifier on the ILDB
   OR
   The estimated population of the community is > 40 & the community does not have a link to a parent
2. "Outstation" with links : Communities that are not "Main Communities" and have a link to a parent community within the same ICPSU
3. "Outstation" without links : Communities that are not "Main Communities" and do not have link to a parent community within the same ICPSU
4. Non-Community : Dwellings in a mesh block that are not assigned to an indigenous community

(Note: a Pastoral Station can in general be treated as a main community, however it can't be a parent and should be enumerated in one cluster only.)

The aim of this document is to outline the strategies that the Redesign team will use to create clusters within ICPSUs (selected for surveys in the 2011 Redesign Period). The definition of 'cluster' used by the Redesign team is:

> "A cluster of dwellings is a set of dwellings in the same ICPSU which will be enumerated in the same month. A single cluster can be composed of community dwellings from one main community, multiple outstations as well as non-community dwellings and special dwelling establishments. The process of "creating clusters" for the selected ICPSUs refers to defining which parts of an ICPSU will provide sample to each cluster formed within an ICPSU, subject to the constraint of the number of clusters assigned to the ICPSU at the time of selection. It will not be until dwelling lists are prepared for the areas of the ICPSU that specific dwellings can be assigned to a cluster"

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

3. Other Issues

- If there are many outstations in the ICPSU that contain [redacted] then they will be evenly divided among the clusters [redacted]
- For MBs that contain Special Dwellings the SDs as going to be divided evenly as possible amongst the clusters of the ICPSU, given a minimum sample of [redacted] from a SD. If not all clusters will get some SD people, the SD people will be allocated to the geographically closest clusters. (PSO may wish to alter this if roads etc. make things more difficult.)

7 Special Dwelling Frame and Selections

Contents

7.1 Introduction

7.1.1 Overview

There are two frames for the Monthly Population Survey (MPS) - the Private Dwellings (PD) Frame (which includes the Indigenous communities strata) and the Special Dwellings (SD) Frame. The SD Frame consists of a list of SDs, so any dwelling that is not listed on the SD frame is therefore part of the PD frame. Clusters (groupings of rooms or beds within an SD) are selected from SDs with the same probability of selection as clusters in the PD frame. The selection method is a systematic selection, with a few tweaks.

7.2 Special Dwelling Frame Construction

A list of all special dwellings (SDs), including out of scope SDs, is produced by Population Survey Program section ( ).

The occupancy values on the frame were derived from a range of occupancy values that Census Branch determined for each SD prior to the 2011 census or collected during the census. These values were then converted by PSP to produce their best estimate of the likely occupancy of the SDs when approached for MPS.

The frame from PSP came with a SD Type. This can be expressed as either a 1 digit type (broad categories) or a 2 digit type (smaller categories).

A variety of variables (being SA1, SA4, FSU-2 ID, 2006 remoteness classification (0-4), area type, Indigenous flag, state skip, FSU-2 probability of selection, FSU-2 selection flag) are merged on from FSU and FSU-2 selection information and a couple of other sources.

[redacted]

[redacted]

[redacted]

[redacted]

There are three reasons why a SD should be permanently selected:

It contains more clusters than the effective skip that goes through it, therefore there would always be at least one selection and sometimes may be two or more selections in it, so it is logical to permanently select it and even out the amount of sample taken from it each month;

It contains enough clusters that it will be selected for a large portion of the life of the design (and therefore also the previous design and the subsequent design for SDs in SRA), so it is simpler to just permanently select it from the perspective of workload management and the relationship with the manager of the SD;

It contains enough occupants that it could support a permanent selection and the more permanent selections there are, the less volatility there will be in the SD sample, leading to higher quality LFS movement estimates.

[redacted]

[redacted]

7.3 Selection Methodology

SDs which are permanently selected are separated from those with a random chance of selection, and the selections are done separately for the two groups.

7.3.1 Permanent Selections

The first SD is given a random rotation group between [redacted] and [redacted], with each number having an equal chance of selection. The SD skip is set to be the State skip multiplied by the SD probability of inclusion as these SDs are selected with certainty. Selection identifier is set to [redacted]. The rotation group is then incremented for each SD, and when it reaches [redacted] is set back to [redacted]. Selection identifier is incremented after every [redacted] SD. [redacted] Selection identifier is used to determine the value of the [redacted] character.. A random cluster start is assigned to each SD selected with certainty. When the rotation group in which the SD is contained rotates out of the sample the next cluster is selected and so on, so that each cluster will remain in the sample for at most 8 months.

7.3.2 Random Selections Process

As the selection process and program are reasonably complicated a detailed description of the process follows here.

At the beginning of each redesign, SD selections are made for the entire length of the new redesign. So the SD selections need to last for 5 years or until the next redesign is done. Enough selections are made which should last for 6 years ie 72 months. Each sample is in the survey for 8 months and then rotates out. Therefore, we need 9 * 8 months or 9 samples. But we also want to select SDs which could be used for the SSS master sample for the few SSSs in which SDs are included so we need 2*9 = 18 samples. Therefore, a selection skip is calculated, equal to [redacted].

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

Once the permanent and random selections are done, the replicate group identifier is assigned. This is done in selection order, with all [redacted] clusters with the same value of [redacted] receiving the same replicate group identifier. The assignment doesn't strictly follow a [redacted] and repeat pattern, as we would like all replicate groups to be approximately equally represented in each rotation group.

[redacted]

[redacted]

[redacted]

[redacted]

Appendix SD Types

(a) TYPE 1 - Hostels for the Homeless, Night Shelters and Refuges (PSOTYPE 11)

Institutions providing cheap or free residential accommodation on a nightly or short-term basis, to homeless or destitute adults. Institutions providing emergency residential accommodation for homeless and destitute persons. Includes night shelters, flop houses, Salvation Army shelters, Matthew Talbot hostels, men's, women's and youth refuges.

Refuges that provide predominantly long term accommodation (ie periods greater than 2 months) should not be included in this category. They should be treated as PDs.

(b) TYPE 2 - Hotels-Motels and other Short-Term Accomodation Establishments (PSOTYPE 21)

This category includes all establishments which predominantly provide short-term accomodation (ie for periods of less than 2 months) and are available to the general public.

Included in this category are establishments such as:

- Hotels with facilities, licensed to operate public bars,
- Motels, licensed and unlicensed with facilities (not licensed to operate a public bar),
- Hotel-motels,
- Most holiday units,
- Serviced apartments,
- Fitness centres,
- Conference centres, and
- Ski lodges

according to the conditions set out below.

In some cases it is difficult to decide whether a particular establishment is a motel. A motel will be defined as an establishment:

- Where 'motel' is included in the name; or
- Offering short term occupancy. It may be necessary to contact the management to determine this.

Establishments which do not have 'motel' included in the name and which offer predominantly long term accommodation are to be included in the PD sample. Persons will be included or excluded on coverage. Establishments with combinations of short term and long term accommodation, with predominantly short term accommodation, are to be included in the SD sample.

Private hotels are not included in this category. They are more like boarding houses in that the occupants tend to be longer term residents. Instead they are included in the boarding house/guest house category, ie Type 6, Subtype 2.

Managers quarters in hotels/motels are included in the SD, even when there are no guests.

Holiday units, in most cases, are treated as hotels/motels since they provide short-term accommodation (ie for periods of less than 2 months). There are exceptions to this, for example the case where small blocks of holiday flats or houses are located in streets which include normal PDs, these should be treated as private dwellings. Holiday units that are grouped together into a resort complex should be treated as a hotel/motel.

Serviced apartments (ie self-contained apartments where the management provides certain services, such as cleaning, washing, etc) which provide predominantly short-term accommodation (ie periods of less than 2 months) should also be included in this category.

(c) TYPE 3 Hospitals and Homes

Staff quarters attached to a hospital or home are included with the institution to which they belong.

Subtype 1 - Hospitals - General (PSOTYPE 31)

This category includes all hospitals which provide a general range of medical care, that is not confined to one area of treatment such as mental health or geriatrics. If a special purpose hospital is a component of a general hospital then it should be included with the general hospital in the 'Hospitals - General' category. General hospitals restricted to a single group of people (eg Children's hospital) belong in this category.

Subtype 2 - Hospitals - Other (PSOTYPE 32)

This category is intended for any special purpose hospitals providing care for a particular area of illness eg mental hospitals, geriatric hospitals. Hospitals for aged people such as Convalescent Hospitals, should only be included in this category if the majority of patients are expected to return to another usual place of residence. If this is not the case, then the establishment should be included in the "Homes for the Aged" category (Subtype 3).

Subtype 3 - Homes for the Aged (PSOTYPE 33)

This category includes only establishments which are for the care of old people eg convalescent homes. The only exception to this is retirement homes which are in a separate category (Subtype 5). Convalescent hospitals should be included in the 'Homes for the Aged' category if the majority of the patients are expected to remain there.

Subtype 4 - Homes - Other (PSOTYPE 34)

This category is intended to include all institutions of the hospital-home type not included in Subtypes 1, 2 and 3 above. For example, it will include establishments such as orphanages, children's institutions, unmarried mother's homes and government welfare homes. If the institution is being combined with or run by a religious institution or group (eg the Salvation Army), the establishment is still to be included in the 'Homes - Other' category. Children's home - reformatory type institutions are not to be included in this category. If its sole purpose is detention, the establishment should be included in the 'Prisons (Minors)' category. Otherwise all children's home type institutions are to be included in the 'Homes - Other' category.

Establishments (private or otherwise), providing medical and/or residential care for handicapped people, with an average occupancy of 6 or more should also be included in this category. Includes homes for the deaf, dumb, blind, crippled, intellectually handicapped, mentally retarded children or adults. Establishments of this nature with average occupancy less than 6 should be included in the PD sample.

Subtype 5 - Retirement Homes (PSOTYPE 35)

Retirement homes are generally some combination of hospitals, nursing homes, hostels, cottage homes or staff quarters. The whole complex is usually run by a central administration. Retirement homes are a separate category for the following reasons:

- Confusion regarding the type of dwellings in retirement homes. Whether they should be in the PD or SD sample is avoided by putting the whole complex in the SD sample.
- List maintenance is facilitated by treating the retirement home as a whole. If the component parts of retirement homes were treated separately then separate lists would have to be maintained. Different parts of the retirement home would also have to be visited for listing and interviewing purposes.
- A central administration makes it easier to conduct the survey there.

An establishment is to be regarded as a retirement home if:

- 'Retirement Home' or similar wording appears in title; and
- The establishment includes a hostel and/or nursing home component offering nursing care such as: at call nursing care, 24 hour 'paid' medical care, full-time care, etc.

The establishment could also provide funded beds, whether it is whole or part funding, private or public.

Establishments that cater for people over, for example, 50 years of age which have no nursing home component and/or hostel should not be confused as retirement homes. These establishments should be included in the PD sample.

The basic rule is:
If an establishment has the appropriate title and includes a hostel and/or nursing care then it would be classed as an SD (Psotype 35), otherwise it would be put into the PD sample.

If in doubt whether an establishment is classed as a Retirement Home, treat it as an SD (Psotype 35)

Note:
- There is no justification to class establishments as SDs solely on the fact that they contain communal facilities such as billiard rooms, restaurants, etc within its boundary.
- Residents of Type 3 and 5 SDs are classed as institutionalised so no interviews should be conducted. Only demographic information of the residents should be collected from the management.

(d) TYPE 4 - Religious and Educational Institutions

Subtype 1 - Religious Institutions (PSOTYPE 41)

This category includes all religious institutions such as convents and monasteries which provide accommodation only. Where such an institution is attached to a hospital or educational institution, it should be included with that institution.

Subtype 2 - Educational Institutions (PSOTYPE 42)

This category includes all institutions providing educational classes or providing accommodation for persons attending an educational institution. Included in this category are school hostels, boarding colleges whether for primary, secondary or tertiary students, university halls of residence and staff quarters attached to educational institutions. In the case of institutions concerned with religious teaching, such as theological colleges, these also should be included in the educational category. Where such institutions are combined with a religious institution (eg a convent or monastery) the whole establishment is still to be included in the educational category.

(e) TYPE 5 - Prisons

Staff quarters attached to the prison are included with the prison to which they belong.

Subtype 1 - Prisons (Minors) (PSOTYPE 51)

This category includes all juvenile, remand and detention centres. Reformatories will also be included provided they are solely for detention of juveniles. In cases where the establishment is partly for the detention of juveniles and partly a children's home, the establishment should not be included in this category. It should be included in the 'Homes - Other' category. In the case of a prison for minors being part of a prison for adults, the establishment should be included in the 'Prisons (Adults)' category.

Subtype 2 - Prisons (Adults) (PSOTYPE 52)

This category includes all institutions for the detention of adults, such as prisons, prison farms and remand centres. In the case of small local lock-ups these are not to be included in the prisons category as they are part of the PD sample.

(f) TYPE 6 - Boarding Houses and Other

Subtype 1 - Staff Quarters (PSOTYPE 61)

This category includes all general staff quarters. Also included are construction camps with more than 100 occupants. Construction camps with less than 100 occupants should be specified as a Type 0 and included in the PD sample. The exceptions to this are:

- Staff quarters attached to another type of SD eg hospitals, prisons, schools. These are included with the appropriate institution to which they belong. While they do not have to be in the same building, they should be adjacent or on the same block of land.
- Habitable buildings which are only occupied for a short period each year, for example, shearers' quarters. These buildings are included in the PD sample.

Subtype 2 - Guest Houses and Other Long-Term Accomodation Establishments (PSOTYPE 62)

This category includes establishments which can provide longer term accommodation than the hotel-motel type of SD, ie more live-in type establishments. It includes all guest houses, boarding houses, private hotels with facilities (not licensed to operate a public bar) and hostels subject to the following condition: They have an average occupancy of 15 or more or are clearly identifiable as SDs (eg establishments which have the following included in their name : 'guest house', 'boarding house', 'private hotel' or 'hostel').

Establishments which offer predominantly long term accommodation and which cannot be readily identified as a PSOTYPE 62 are to be included in the PD sample, and persons are included/excluded on coverage.

Please note that an establishment with an average occupancy of less than 15 and which is not identifiable from the street as a PSOTYPE 62 should not be considered as a PSOTYPE 62.

(g) TYPE 8 - Short-Term Caravan Parks, Youth Camps and Camping Grounds (PSOTYPE 81)

This category comprises any recognisable area set aside for camping and/or caravan use, with toilet and/or shower facilities and/or which provide powered and/or unpowered sites for caravans on a predominantly short-term basis.

Long-term caravan parks are occupied by a majority of people who consider the park as their usual place of residence, ie people who live there permanently (ie for a period of 6 months or more). The long-term caravan park should have at least 75 per cent or more sites that are occupied by usual residents.

These long-term parks are to be placed in the PD sample.

If in doubt as to whether a caravan park should be classed as an SD or a PD, especially if the caravan park offers both long and short-term accomodation, then the caravan park should be included in the SD sample.

Note: Other long-term parks, such as manufactured home estates, relocatable home parks or villages, which predominantly cater for long-term residents should be included in the PD sample. The same treatment as for long-term caravan parks should be followed.

Also included in this category are camps which provide dormitory style accomodation for short periods of time, eg Youth Camps.

A camping ground should be classified as a Type 8 if:

- There should be a recognisable area set aside for camping. This could be areas designated just for camping which are bound in by some physical form, eg by pine logs. If there is no bounding of any sport, then it can still be considered a camping ground if it is visibly so, eg if there is an honesty box, or sign-posts; and
- The area should be listed in a guide of some form, eg. Parks and Conservation Guide;and
- It is of a 'reasonable' size. Small camping areas should not be put on the SD list as these are costly to enumerate if selected.

A camping ground does not need to have any services/ facilities, though it could have pit toilets, places to barbeque, etc.

Please note that conference and fitness centres which provide short-term accomodation are to be included in the SD sample as PSOTYPE 21.

The main reason for transfering long-term caravan parks from the SD to the PD sample was that it was increasingly found by states that people in caravan parks considered such establishments as their usual place of residence and it was felt that it would be more appropriate to treat the people living in these caravan parks as usual residents.

For some types of SDs in some states, it is more convenient to maintain lists separate from the normal list. This is the case for aboriginal communities in NT and QLD.

In selecting SDs, the clear definition of the SD boundary is important. If boundaries are vague it could lead to some PDs or components of SDs having no chance of selection or a double chance of selection. This mainly occurs when an SD has some component which appears to be of a PD nature (eg Retirement Home). The boundaries of the SD must be clearly set so the dwellings in question are clearly included as part of either the PD or SD sample.

As the subtype code is mainly used in ordering the list for selection, a minor misclassification is not important. While type is also used in ordering the list, its main use is in defining institutionalised persons. The main emphasis is therefore placed on distinguishing types 3 and 5 from other types.

(h) MARINA