1351.0.55.057

Australian Bureau of Statistics

Research Paper

Constructing an Experimental Household-Level Socio-Economic Index of Disadvantage Using GSS Data

www.abs.gov.au

New Issue

Research Paper

Constructing an Experimental Household-Level Socio-Economic Index of Disadvantage Using GSS Data

Courtney Williamson, Summer Wang and Anil Kumar

Analytical Services Branch

AUSTRALIAN BUREAU OF STATISTICS

EMBARGO: 11.30 AM (CANBERRA TIME) WED 05 AUG 2015

ABS Catalogue no. 1351.0.55.057

Produced by the Australian Bureau of Statistics

INQUIRIES

The ABS welcomes comments on the research presented in this paper. For further information, please contact Ms Sybille McKeown, Program Manager, Analytical Services Branch, on Brisbane (07) 3222 6400 or email <analytical.services@abs.gov.au>.

CONSTRUCTING AN EXPERIMENTAL HOUSEHOLD-LEVEL SOCIO-ECONOMIC INDEX OF DISADVANTAGE USING GSS DATA

Courtney Williamson, Summer Wang and Anil Kumar
Analytical Services Branch

EXECUTIVE SUMMARY

This paper explored the feasibility of constructing an experimental socio-economic index of disadvantage at the household level using General Social Survey (GSS) data. The interest in finer level indexes arises from the need for detailed disadvantage information to complement broad measures such as area based indexes. The GSS covers a broad range of socio-economic variables which enables the incorporation of many dimensions of disadvantage. The 2010 and 2014 GSS datasets were analysed separately to construct a socio-economic index of disadvantage for each period.

The paper discussed the concept of socio-economic disadvantage and how it has evolved over time from a narrow focus on resource and income based indicators to a more broad-based multidimensional concept encompassing both economic and non-economic factors. It also discussed the distinction between area level and individual level measures of disadvantage and provided a rationale as to why the household level was the more appropriate level at which to construct the index of disadvantage compared to an individual or family level.

Both simple and complex measures of disadvantage were explored. The simple measures consisted of counts of indicators and domains of disadvantage while the complex method involved using weights derived from principal component analysis (PCA) to combine the variables to derive a summary or composite measure of disadvantage.

Results from the simple measures showed that a large majority of households experienced few counts of disadvantage and a small proportion experienced severe levels of disadvantage for both 2010 and 2014. The composite method of index construction, which overcomes the limitation of equal weighting of the simple methods, involves the use of an explicit weighting scheme to combine different variables of disadvantage to construct a summary measure of disadvantage. Principal component analysis was used to derive the weights for the compilation of the composite index. The steps used to derive the final set of variables and their corresponding weights in this paper are similar to the approach used for Socio-Economic Indexes for Areas (SEIFA).

An analysis of the results from the composite index showed that the majority of the final set of variables used to construct the index and the distribution of the created index was similar across both periods. The most highly influential variables for both the 2010 and 2014 indexes were from the domains of the health and economics including, financial stress, income and wealth. The results at this stage are experimental and the caveats and limitations discussed in the paper should be kept in mind when interpreting the results. Further work could include additional validation of the methodology and the results and investigation into alternative methods to calculate scores for those records with missing index scores.

CONTENTS

CONSTRUCTING AN EXPERIMENTAL HOUSEHOLD-LEVEL SOCIO-ECONOMIC INDEX OF DISADVANTAGE USING GSS DATA

Courtney Williamson, Summer Wang and Anil Kumar
Analytical Services Branch

ABSTRACT

This paper explored the feasibility of constructing an experimental socio-economic index of disadvantage at the household level using General Social Survey (GSS) data. The interest in finer level indexes arises from the need for detailed disadvantage information to complement broad measures such as area based indexes. The GSS covers a broad range of socio-economic variables which enables the incorporation of many dimensions of disadvantage. The 2010 and 2014 GSS datasets were analysed separately to construct an index of disadvantage for each period. Both simple and complex measures of disadvantage were explored. The simple measures consisted of counts of indicators and domains of disadvantage while the complex method involved using weights derived from principal component analysis to combine the variables to derive a summary or composite measure of disadvantage.

An analysis of the results from the composite index showed that the majority of the final set of variables used to construct the index and the distribution of the created index was similar across both periods. A cross tabulation of the index deciles by selected demographic, geographic and socio-economic characteristics showed that the relationships were in line with expectations. The results at this stage are experimental and the caveats and limitations discussed in the paper should be kept in mind when interpreting the results. Further work could include additional validation of the methodology and the results and investigation into alternative methods to create indexes for those records with missing index values.

1. INTRODUCTION

Addressing the causes of disadvantage remains an important goal of every society. Socio-economic advantage and disadvantage is defined by the Australian Bureau of Statistics (ABS) in terms of people's access to material and social resources, and their ability to participate in society (ABS, 2013). Some measures or indicators that broadly capture these dimensions at individual, group or area level are important in understanding the magnitude, nature and location of disadvantage.

The Socio-Economic Indexes for Areas (SEIFA), produced every five years by the ABS using Census data, provide a measure of socio-economic disadvantage at an area level. While the indexes broadly capture the general level of socio-economic disadvantage of the area they do not necessarily capture finer level disadvantage, such as that of the individuals or households living within the areas. There has been plenty of interest for non-areal measures of disadvantage at finer levels. While the ABS has undertaken exploratory work to create an index at the household level this was confined to using Census data only (Wise and Williamson, 2013). This study extends that work by using General Social Survey (GSS) data which provide a wider selection of variables in a range of social domains. Specifically this paper explores the feasibility of constructing an index of socio-economic disadvantage at the household level using GSS data.

The remainder of the paper is organised as follows. Section 2 provides a brief literature survey of the concept of socio-economic disadvantage and discusses the rationale of and distinction between area level and household-level measures of disadvantage. Section 3 defines disadvantage at the household level distinguishing it from other finer level measures of disadvantage such as the individual level. Section 4 discusses the source of the data and the procedure of variable selection for this study. Section 5 discusses the methodology used for the construction of the household-level index of socio-economic disadvantage. Section 6 presents and discusses the results from the index construction procedure, including a comparison of the results between the two periods considered in this paper. Section 7 presents a brief validation of the created indexes. Section 8 discusses the interpretation and use of the indexes and some of the benefits and challenges associated with the measurement and construction of a household-level index. Section 9 concludes with suggestions for further work.

2. APPROACHES TO MEASURING SOCIO-ECONOMIC DISADVANTAGE

The concept of socio-economic disadvantage is of considerable interest among researchers, practitioners and policy makers given its strong influence on educational, health and labour market outcomes. The concept of disadvantage, which previously was focused narrowly on resource and income based indicators, has broadened to include wider measures of both material deprivation and other types of advantage and disadvantage, including health and development factors (UNICEF, 2013; Abello *et al.*, 2014). Other similar concepts used to describe disadvantage include deprivation, wellbeing and social exclusion (Salmond *et al.*, 2006; McLennan *et al.*, 2011; Michalos *et al.*, 2011; Daly, 2006; Scutella *et al.*, 2009).[1] It is now commonly accepted that disadvantage is multidimensional in nature and affects a person's ability to participate in society in many aspects of life such as, economic, social and political (ABS, 2013; Scutella *et al.*, 2009).

Socio-economic disadvantage can be measured at the area level or at finer levels such as individual, family or household (Bailey *et al.*, 2003; Salmond *et al.*, 2006; Baker and Adhikari, 2007). Area level disadvantage relates to the characteristics of the community or neighbourhood and such an index can be created from the proportions of people in each area with particular characteristics of disadvantage. SEIFA is an example of an area based measure of disadvantage. Finer level disadvantage, on the other hand, is a more personal concept, and it relates to a person or group's ability to access resources and participate in society, based on their material and social circumstances (Baker and Adhikari, 2007). The interest in finer level indexes has arisen from the realisation that while an area based index provides contextual information about the area in which a person lives, within any area there are likely to be groups with characteristics different to the overall population of that area. Inferences made about groups based purely on the characteristics of the area in which they live, can lead to misleading and erroneous conclusions being drawn (Baker and Adhikari, 2007; Wise and Mathews, 2011). The assumption that relationships observed for areas also hold for those living in that area is referred in the literature as ecological fallacy.[2] More in-depth appraisals of available area, household and individual level socio-economic measures and issues associated with moving from one measure to the other can be found in Bailey *et al.* (2003), Morris and Carstairs (1991), Marks *et al.* (2000), Wise and Mathews (2011) and Wise and Williamson (2013).

1 For example the equivalent area-based index in New Zealand is called Deprivation Index (NZDep) (Salmond *et al.*, 2006), in England it is called the Index of Deprivation (McLennan *et al.*, 2011), while in Canada it is called the Canadian Index of Wellbeing (CIW).

2 Ecological fallacy is most likely to be an issue in areas where the characteristics of particular individuals or other population subgroups are too diverse to be meaningfully represented by the average characteristics of people in the area.

Area and individual based measures of disadvantage can be constructed for the population as a whole, or for sub-groups of the population such as children, students, youth, and women (Bradshaw *et al.*, 2009; UNICEF 2013; Ainley and Long, 1995; Abello *et al.*, 2014; Baxter and Taylor, 2014). While finer level and area level measures of disadvantage are based on different concepts there are many commonalities between the two and as such they should be seen as complementary rather than alternatives (Bailey *et al.*, 2003; ABS, 2008).

A domain based approach generally is used in the construction of many socio-economic indexes, whether area based or individual or household based. Domains represent the main themes or broad areas of the concept of disadvantage, such as income, employment, health, education, housing, physical safety and social participation. Indicators underlying each domain are then used to capture a person's or group's ability to participate in society in these specific aspects.

In Australia while national area based measures of disadvantage are well established and accepted, the focus on finer level indexes is of recent origin and still exploratory in nature. Initial work at the ABS led to an experimental index for individuals and families for Western Australia by Baker and Adhikari (2007) using the 2006 Census data. This study was extended by Wise and Mathews (2011) using 2006 Census data to include the whole of Australia which culminated in the construction of the Socio-Economic Indexes for Individuals (SEIFI). Using the 2011 Census data, Wise and Williamson (2013) explored constructing a household-level index by developing a household-level socio-economic index to mitigate issues associated with creating individual and family level indexes compiled previously by the ABS.

This paper extends on the earlier work at the ABS on finer level indexes of disadvantage by using survey data to create an index of disadvantage at the household level. A limitation of using GSS data is that it is sample data and due to surveying fewer households, provides less precision than that of a Census. However, the GSS survey data contains more variables and is able to cover a broader aspect of disadvantage than is possible from Census data. The index developed for this paper provides important additional information about household variations in socio-economic disadvantage and there is added value to be derived from comparing measures developed from different data sources.

3. DEFINING DISADVANTAGE AT HOUSEHOLD LEVEL

A number of ABS research papers have investigated the measurement of disadvantage at different levels, such as individual, family, household, meshblock and area level (Baker and Adhikari, 2007; Wise and Mathews, 2011; Wise and Williamson, 2013). While a finer level index constructed from GSS data could be measured at an individual level, the household level was chosen for several key reasons.

The ABS defines a household as one or more persons, one of whom is an adult, usually resident in the same private dwelling. A household is an understandable and easily quantifiable unit that mitigates several issues that exist with individual level indexes whilst still being a finer level, non-areal based measure. One such issue is the exclusion of the non-working age population due to the lack of available indicators and life-cycle factors. The household is a social unit which tends to share resources and characteristics, which is able to be captured by a household-level index. Income and wealth tends to be shared within households composed of families[3], which make up 71.5% of the households in Australia according to the 2011 Census.[4] Additionally a household-level index provides the benefit of including the whole Australian population in a single index.

Some studies create separate indicators for subpopulations, such as children, youth, working age and older persons due to life-cycle characteristics which may not be equally relevant to all subpopulations (Scutella and Wilkins, 2010; Abello *et al.*, 2014). By considering households as the unit of measurement, some challenges associated with life-cycle variables are able to be mitigated due to households sharing resources and characteristics. The ABS Census-based SEIFI was constructed at the individual level, and due to the lack of relevant indicators for the non-working age population, the index was only calculated for the 15–64 year old population, thus excluding 33% of the total Australian population. Research suggests that family characteristics provide strong indicators for youth and child disadvantage (Lim and Gemici, 2011). Family level indexes have been explored by the ABS previously, however lone person households and group households are not able to be included, resulting in significant population exclusions (Baker and Adhikari, 2007; Wise and Williamson, 2013). A household-level index maximises population inclusion by representing children and youth by their family. The GSS has a wide and varied range of variables relating to aspects of disadvantage which allows the inclusion of a more extensive range of variables relating to disadvantage over the life-cycle. This allows the over 65 year old population to be more accurately included using the GSS dataset rather than the

3 A family is defined by the ABS as two or more persons, one of whom is at least 15 years of age, who are related by blood, marriage (registered or de facto), adoption, step or fostering, and who are usually resident in the same household. Some households contain more than one family.

4 The remaining proportion of households consist of lone person households (24.3%) and group households (4.1%).

Census due to relevant indicators such as dwelling equity, access to services and self-assessed health.

Measures of disadvantage are often constructed at an area level due to confidentiality and data availability reasons (Scutella and Wilkins, 2010). Through investigating disadvantage at a household level, multiple indicators of disadvantage affecting a single household can be investigated. While it is important to identify households with one indicator of disadvantage, it is relevant to investigate households experiencing multiple forms of disadvantage as well as the depth of the disadvantage (Scutella and Wilkins, 2010). The composite index discussed in this report provides a summary of a complex set of information by summarising the common characteristics of disadvantage into a single index. The two simple counts (variable count and the domain count) provide an unweighted count of the variables or domains of disadvantage that a household is experiencing. Implicitly, highly disadvantaged households in the indexes will be experiencing multiple forms of disadvantage. The summary of complex information into an index or count facilitates further investigation into the characteristics of highly disadvantaged households.

The composite index provides a broad measure of disadvantage which can be analysed to target and investigate groups, which was previously not feasible in a measure tied to geography. This index allows the population to be analysed without the constraint of geography and cross-classified by other demographic characteristics. For example, this index allows household disadvantage to be measured against characteristics such as country of birth, number of children in the household or type of government benefit received.

4. DATA SOURCE AND VARIABLE SELECTION

4.1 General Social Survey data

The GSS is a multi-dimensional social survey that captures a wide range of information on the social dimensions of individuals and households. The topics collected include health, family relationships, education, employment, income, housing, transport, crime and safety, financial stress and community participation. The GSS provides the community and government with detailed information to assist with decision making, including through developing and understanding relationships between social circumstances and outcomes.

The GSS was first conducted in 2002 with a four-yearly cycle. Each iteration collects core information with some changes in each cycle to address emerging or important social topics. The final sample sizes for the surveys were approximately 15,000 private non-remote dwellings for 2010 GSS and approximately 13,000 for 2014 GSS.

In both the 2010 and 2014 GSS, information was collected for each household, usually from a single responsible adult in the household, known as the household reference person (ABS, 2011b). This person responds to the questions about themselves, such as age and country of birth, and questions about the household, such as weekly household income and number of cash flow problems, unless they are not in a position to respond accurately, in which case a more appropriate person in the household is asked. More details about survey design and related survey characteristics of the 2010 and 2014 surveys can be found in Appendix A.

This report analyses the 2010 and 2014 GSS data separately to derive an index of disadvantage for each survey.

4.2 Variable selection

To construct an index of disadvantage, an appropriate set of variables is required that best reflects the concept based on the data available. A domains approach was used to identify the different dimensions of disadvantage. Within each domain, a set of variables that best represent that aspect of disadvantage is selected. The domains commonly considered when measuring disadvantage are education, employment, health, income, wealth, location and consumption (Scutella and Wilkins, 2010; ABS, 2011a). The list of potential variables for index construction was guided by a literature survey, previous studies, subject matter experts and the survey data. The list of GSS variables was reviewed and those associated with the definition of household socio-economic disadvantage were identified.

Table 4.1 presents the 2010 and 2014 GSS list of candidate variables for the construction of the household-level indexes of disadvantage. The first column in each panel shows the relevant domain and the corresponding indicator variables that were chosen to reflect that domain. The second column indicates whether a variable is a household-level (H) or a person-level variable (P). The third and fourth columns indicate whether that variable is available (Y) or not available (N) in GSS 2010 and 2014 respectively. More detailed definitions of the variables are contained in Appendix B.

4.1 List of General Social Survey variables considered for index construction

Domain / Indicator Variables	*H/P*	*GSS 2010*	*GSS 2014*
INCOME			
Household equivalised income	H	Y	Y
Time on government support	P	Y	Y
EDUCATION			
Highest level of education	P	Y	Y
UNEMPLOYMENT			
Unemployed	P	Y	Y
ACCESS TO SOCIETY/SERVICES			
Transport difficulty	P	Y	Y
Difficulty accessing services	P	Y	Y
English poor	P	Y	Y
No social activities	P	Y	N
No social support	P	Y	Y
CRIME AND SAFETY			
Victim of break-in	P	Y	Y
Victim of assault	P	Y	Y
Feeling safe (day)	P	Y	N
Feeling safe (night)	P	Y	Y
Feeling safe walking (night)	P	Y	Y
Neighbourhood problem	P	Y	Y
HOMELESSNESS			
Homeless times	P	Y	Y
Length homeless	P	Y	Y
FINANCIAL STRESS			
Can't raise $2K	H	Y	Y
Cash flow problems	H	Y	Y
Dissaving actions	H	Y	Y
Number of financial stressors	H	Y	Y
Difficulty paying bills	H	Y	Y
Financial exclusions	H	Y	Y
WEALTH			
Dwelling equity	H	Y	Y
Asset value	H	Y	Y
Consumer debt	H	Y	Y
HEALTH			
Self-assessed health	P	Y	Y
Mental health	P	N	Y
Delay doctor	P	Y	N
Delay medication	P	Y	N
Health access	P	N	Y
Disability	P	Y	Y
Employment restriction	P	Y	Y
Education restriction	P	Y	Y
PERSONAL STRESS			
Personal stress	H	Y	Y

As can be seen in table 4.1 the GSS includes a wide range of variables that cover many of the domains commonly associated with socio-economic disadvantage. In addition to income, education, employment, health and housing, the GSS covers dimensions of disadvantage such as crime and safety, financial and social stress, wealth and social participation.

Based on an examination of the quality and correlations between variables, some variables were excluded from the candidate variable list. Variables that had particularly low correlations or high correlations and captured the same concept were excluded from the candidate variable list so that the most appropriate variables would

be selected to successfully utilise the method. The list of candidate variables in table 4.1 formed the starting point for the index construction, with the final list chosen dependent on the assumptions and the method used for index construction as described in the next section.

The variables consist of a combination of categorical and continuous variables. Previous research that has investigated the inclusion of a mix of binary and ordinal variables found that the difference does not greatly affect the results of an index constructed using principal components analysis (Wise and Williamson, 2013; Kolenikov and Angeles, 2009). To utilise the maximum amount of variation within the variables, the variables were left as close to their original structure as possible. Furthermore, the use of ordinal and continuous variables means that the final index has a greater number of possible scores compared to if the variables were all binary. This allows greater distinction between levels of disadvantage in the final index.

4.3 Treatment of missing values

Households with a missing response for a variable included in the output were not included in the analysis. The total percentage of households excluded from the analysis was 16.7% in 2010 and 19.4% in 2014. A large proportion of these households were excluded due to the missing responses for the income variable.

The income variable, household equivalised income in deciles, had a significant level of missing data with 13% in GSS 2010 and 18% in GSS 2014. The missing data were assessed against a number of demographic variables which indicated that the income non-response was fairly equally distributed across education, age, sex and state. The type of household variable had some association with the level of income non-response, with a higher proportion of group and multiple family households not responding. This could be because the household reference person was not aware of the total household income due to the nature of the households, with multiple people having their own private incomes. Overall, despite a larger proportion of group and family households not responding, the largest number of households with income non-response was single family households.

Several alternatives regarding the inclusion of the income variable were compared to assess the best option for the index. Option 1 was to utilise the ordinal income variable represented by deciles, with the lowest decile representing the lowest income and the highest decile representing the highest income, leaving the missing data as untreated. Option 2 involved constructing the variable as a binary variable with Deciles 1 and 2 representing the "low income group" and all others representing the "not low income group". This implicitly places those with missing values in the "not low income group". Option 3 was to drop the income variable from the analysis. The three options produced similar results, with a similar distribution for the index scores

and similar variable loadings, including for the income variable, where applicable. It was decided that the ordinal variable was the most methodologically sound and appropriate option. Utilising the ordinal variable reduced the assumptions needed to be made about the missing data, compared to the binary variable. Due to the missing income values being evenly distributed across a number of key household characteristics it is less likely that this caused bias to the index.

4.4 Individual or household-level index

As can be seen from table 4.1, the GSS consists of a mixture of person-level and household-level variables. Ideally for index construction at the household level, household-level variables should be included which are either directly collected (e.g. household income) or created as a measure from the person level characteristics of all persons living in that household (e.g. proportion of adults unemployed in the household, level of highest education in the household). However, since the GSS collects person level information from the household reference person only and not for or from all persons living in that household it is not possible to construct household measures based on all household members. For the purposes of the construction of the household-level index of disadvantage it has been assumed that the characteristics of the reference person represent or influence the level of disadvantage of the household, in relation to that particular characteristic (e.g. if the reference person is unemployed then this is assumed to have an impact on the whole household). Although this may not hold in all situations, such an assumption is a reasonable one to make in this context, as is done in many other surveys where a person is chosen to represent the household. Furthermore, the household-level weight was used to construct the index. Using a mixture of person and household-level variables and household weights to construct the index means that the constructed index can be considered to be a household-level index rather than an individual level index.

5. METHODOLOGY FOR CONSTRUCTING AN INDEX OF HOUSEHOLD DISADVANTAGE

There are a number of ways to create a summary or composite index for multiple dimensions of socio-economic disadvantage. This consists of simple or complex methods of combining the indicators. Simple methods include a count of the number of indicators of disadvantage or an equal weighted sum of the raw or standardised values of the indicators.[5] Complex methods of index construction consist of the use of some weighting scheme to combine the variables to derive a summary index. These weights could be theoretically based weights[6] or weights derived from multivariate statistical techniques such as principal components or factor analysis (Hagerty and Land, 2007; Lalloué *et al.*, 2013).

In this paper two simple counts and a composite index of disadvantage are presented. The two simple counts involve converting the variables into binary indicators and summing the indicators or summing the number of the domains of disadvantage in which a household has an indicator of disadvantage. This results in either a count of variables of disadvantage or a count of the domains of disadvantage experienced by each household.

The more complex index is based on a weighted sum of the indicators of disadvantage, with weights derived using the principal components analysis (PCA) technique. PCA is a statistical technique that involves summarising a large number of correlated variables into a set of new uncorrelated variables, or principal components, that account for as much of the total variation as possible in the original variables. Each principal component is a linear combination of the original variables. Generally, the first component captures the largest part of the variation in the original set of variables, and can be used to construct the index of interest.[7] Since the objective in this study is to obtain an optimal set of weights that can be used to combine all the relevant variables of disadvantage into a summary index, PCA was considered the

---

5 For example the UNDP Human Development Index (HDI) is calculated using equal weights to combine the (standardised values) of the following three indicators of human development: GDP, life expectancy and education (Hagerty and Land, 2007).

6 Theoretical or empirical weights are explicit weights based on theoretical considerations and responses to the consultation processes. It assigns weights based on the perceived importance of the particular domain or indicator of disadvantage. This approach was used in the construction of the English Indices of Deprivation where of the seven domains of deprivation the income and employment domains were regarded as the most important and hence given higher weights (22.5% each), followed by the health and education domains (13.5% each), followed by crime, housing and environment domains (9.3% each) (Nobel *et al.*, 2004).

7 The alternative approach of using factor analysis (FA) to derive variables weights is more suited in cases where the objective is to identify or decompose the different underlying dimensions or factors of disadvantage. It may be noted that while PCA and FA are similar in many respects their purpose in practice is different. The characteristic that distinguishes between the two techniques is that in PCA all the variability in a variable (total variance) is used in the analysis, while in FA only the variability in a variable that is common with the other variables is used. PCA is the preferred method for data reduction while FA is the preferred method for detecting data structure (Dinov, 2004).

appropriate technique to be used for this purpose. The PCA method used here is similar to the approach used in the construction of SEIFA (ABS, 2013).

The PCA procedure gives an eigenvalue for each component, which indicates the amount of variance in the original data explained by the component. The proportion of variance explained by a principal component is its eigenvalue divided by the sum of all the eigenvalues. In this study the unrotated first principal component was used to derive the weights. Each variable in the analysis is correlated with each component and each correlation is referred to as a loading. Loadings help to interpret which aspects of disadvantage a component may represent. In order to generate the disadvantage scores the loadings are first converted to a weight by dividing it by the square root of the eigenvalue. These weights are then multiplied by the standardised values of the corresponding variables and summed across all the variables to derive the raw scores. The formula for the construction of this index can be expressed as follows.

$$Z_h = \sum_{j=1}^{k} \frac{L_j}{\sqrt{\lambda}} \times X_{j,h}$$

where

Z_h = the raw score for household h;
$X_{j,h}$ = the standardised value of the j-th variable for household h;
L_j = the loading for the j-th variable;
λ = the eigenvalue of the first principal component; and
k = the total number of variables in the index.

There are a number of steps involved in deriving the weights from PCA. The first step involves including all variables in the analysis and examining the variable loadings on the first component. The next step involves removing the variables with loadings below $|0.3|$ which indicates that such variables are not strong indicators of relative disadvantage.[8] Variables are removed one at a time iteratively, starting with the lowest loading variable until no variables with loadings below the $|0.3|$ threshold are left. This is done to ensure that each included variable contributes significantly to the final index. This remaining list of variables and their corresponding weights are then used to compute the disadvantage score.

8 The threshold of $|0.3|$ is an accepted level in the PCA literature (Joliffe, 1986).

Another consideration in PCA is the type of correlation matrix that should be used to derive the components and the variable loadings. As may be noted the starting point in PCA is that there should be a reasonable degree of correlation among the variables. Variables with very low correlations should generally be excluded from the analysis, as should variables with very high correlations if they represent the same concept. Note that if there are two variables with very high correlations but which measure or represent different aspects of disadvantage then there is no reason to exclude one in preference to the other. Polychoric correlation was utilised in this analysis due to the variables predominately being ordinal (e.g. 0–5 scale) and binary (e.g. unemployed, not unemployed).[9] Utilising Pearson correlation for this study, which is appropriate for continuous variables, could have led to biased PCA results (Rigdon and Ferguson, 1991). The method used here to address this bias was to use polychoric correlation for PCA. Polychoric correlation is used when the variables being tested for correlation are binary or ordinal in nature.[10]

9 In this analysis only one variable, dwelling equity, is continuous while the remaining variables are either binary or ordinal.

10 Polychoric correlation is referred to as tetrachoric correlation when the variables are binary in nature.

6. RESULTS

This section presents the results of the simple counts, which is simply a sum of the number of indicators or domains of disadvantage experienced, and the complex method, which is a weighted sum of the variables of disadvantage with weights derived from the PCA procedure.

6.1 Simple count of disadvantage

Table 6.1 shows the percentage distribution of households by the count of indicators of disadvantage, where a bigger number implies more disadvantage and a smaller number means less disadvantage. The number of indicators of disadvantage ranges from 1 to 30 for 2010 and 1 to 26 for 2014. The difference reflects the lesser number of indicators used to construct the 2014 index compared to 2010 due to changes to the survey. Based on this measure all households have at least some degree of socio-economic disadvantage. However, around four-fifths of the households in both periods have ten or fewer indicators of disadvantage with around half having five or fewer indicators. A very small proportion of the households in both periods have more than 20 indicators of disadvantage, with 1.0% in 2010 and 0.5% in 2014.

6.1 Distribution of households, by number of indicators of disadvantage

No. of Indicators	*GSS 2010*	*GSS 2014*
	(%)	
0–5	49.94	52.61
6–10	32.47	32.76
11–15	12.33	11.11
16–20	4.20	3.00
21–25	0.90	0.51
26–30	0.13	0.02
Total	100.0	100.0

The indicator count is significantly influenced by the particular domains of disadvantage that a household may experience. For example if a household experiences disadvantage in the domains that have more indicators, such as safety, then they may end up having a higher count. An alternative way of deriving a simple count of disadvantage that avoids this problem is to do a count of the number of domains of disadvantage rather that the number of indicators of disadvantage. Results from this method are shown in table 6.2. Under this method a household is considered to have a disadvantage under a particular domain if it experiences at least one indicator of disadvantage under that domain. Counting this across all the domains gives a count of the total number of domains of disadvantage (as opposed to number of indicators) for that household. This number can range from zero to ten.

The results show that in both 2010 and 2014 all households have some degree of disadvantage but similar to the results above a majority of the households experience fewer domains of disadvantage. Around two-thirds of households in 2010 and three-quarters in 2014 experience disadvantage across four or less domains and less than 9% in 2010 and 6% in 2014 experience disadvantage across seven or more domains.

However, the limitation with the above measures is that they give equal weighting to all the variables or domains. It is possible that some measures of disadvantage are more important or acute than others. Some implicit or explicit weighting procedure can be used to address this. The above analysis and results, however, makes researchers aware of the different indicators or dimensions of disadvantage, which can be explored further, particularly in cases of households identified as facing disadvantage on several fronts.

6.2 Distribution of households, by number of disadvantage domains

No. of Domains	*GSS 2010*	*GSS 2014*
	(%)	
1	6.15	6.87
2	18.05	21.42
3	23.05	25.56
4	19.52	19.30
5	14.61	13.71
6	9.95	8.05
7	5.82	4.05
8	2.05	0.98
9	0.76	0.07
10	0.04	—
Total	100.0	100.0

6.2 Composite index of disadvantage

This subsection presents the results of the composite index created using PCA. The results are based on the final reduced set of variables determined through the PCA procedure, which had loadings of greater than $|0.3|$ on the first principal component. The first principal component based on this reduced set of variables accounted for 25% of the variance in 2010 and 22% in 2014. The variable loadings for the retained variables used to compile the disadvantage score for 2010 and 2014 are presented in table 6.3. The reduced set consists of 21 variables for 2010 and 20 variables for 2014, with 18 variables common in both periods. Variables that are not common between the two periods simply reflect the fact that these particular variables were not available in both periods. Loadings for the index comprising the full set of variables prior to excluding the variables with lower loadings are presented in Appendix C.

The final variable list shown in table 6.3 comes from the original candidate variables, shown in table 4.1, which were then removed iteratively until all the remaining variable loadings were greater than |0.3|. Under this method, each iteration involves removing the lowest loading variable and re-running PCA. This process is repeated until there are no remaining variables with a loading less than |0.3|. In practice the changes from this are small, however, it can be relevant when a loading is close to the |0.3| cut-off.

6.3 Final PCA variable loadings

Variables	*GSS 2010*	*GSS 2014*
Number of financial stressors	0.75	0.71
Difficulty paying bills	0.63	0.59
Can't raise $2K	0.61	0.59
Delay medication	0.61	—
Employment restriction	0.56	0.58
Disability	0.52	0.55
Mental health	—	0.52
Asset value	0.52	0.48
Homeless times	0.52	0.44
Delay doctor	0.51	—
Household equivalised income	0.50	0.50
Self-assessed health	0.48	0.51
Dissaving actions	0.47	0.47
Time on government support	0.45	0.48
Education restriction	0.45	0.47
Personal stress	0.44	0.43
Length homeless	0.43	0.33
Transport difficulty	0.41	0.39
Difficulty accessing services	0.38	0.39
Health access	—	0.39
Feeling safe (night)	0.38	0.31
Feeling safe (day)	0.36	—
Victim of assault	0.33	0.30

The results from table 6.3 show that variable loadings for the 2010 and 2014 indexes are similar for the common variables in most cases. The loading can be interpreted as the association that the variable has on the index, or its influence on the index. The closer the value is to 1, the more influential a variable is. For example, number of financial stressors (0.75) is more influential than victim of assault (0.33). Comparing the loadings between 2010 and 2014, the loadings for the common variables are not consistently lower, higher or of different sign in one period compared to the other. There are some differences for some variables but they are not vastly different. As such the impact of specific characteristics on the overall household disadvantage appears to be consistent across the two periods.

Table 6.4 presents summary statistics on the composite index of disadvantage for the two periods. For convenience and ease of interpretation, the raw scores were standardised or rescaled to have a mean of 1,000 and standard deviation of 100 to create the index scores of household-level disadvantage. Similar to the approach used for the SEIFA Index of Relative Socio-Economic Disadvantage (IRSD) lower values of this index indicate higher disadvantage while higher values of the index indicate lower disadvantage.

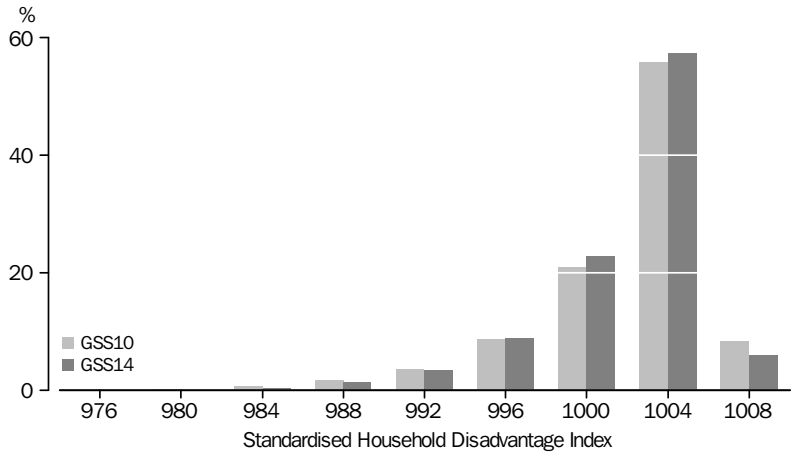
6.4 Composite household index of socio-economic disadvantage – Basic statistics

Summary Statistics	*GSS 2010*	*GSS 2014*
Sample Size	15,028	12,932
Households with Missing Scores (%)	16.7	19.4
Mean Score	1000.0	1000.0
Median Score	1001.3	1001.1
Standard Deviation	100.0	100.0
Minimum Score	971.9	971.8
Maximum Score	1005.2	1005.1
Households below Mean Score (%)	35.8	36.7

As can be seen in table 6.4 the index values for both the periods range roughly from 970 to 1005. Around 17% of households in 2010 and 19% in 2014 have missing scores. These missing scores reflect the fact that these households have missing values for either one or several of the variables that are used to construct the index. The higher median score compared to the mean implies that there is some skewness in the data. Of the households that had a score assigned to them, around 36% in 2010 and 37% in 2014 had scores below the mean (1000). This implies that around two-thirds of the households in both the periods were above mean in terms of the disadvantage measure.

Figure 6.5 below presents the distribution of the index for both the periods. As can be seen from the graph the distributions of the scores across the two periods appear to be similar, with a slightly higher proportion of households in 2010 being in the higher end of the index range compared to 2014, and a slightly higher proportion of households in 2014 being in the 1000–1004 range compared to 2010. The graph also shows there is negative skewness in the scores across both the periods with there being a long left tail, implying that the spread amongst scores is greater for disadvantaged households than for advantaged households.

6.5 Distribution of Household Index of Disadvantage Scores



When interpreting the index it should be noted that the index values are on an arbitrary numerical scale. The values do not represent a constant unit of disadvantage. For example, it cannot be inferred that a household with an index value of 1000 is around 3% more advantaged as a household with an index value of 970. For ease of interpretation the standardised scores are converted into deciles. All households that are assigned with a final index are ordered from the lowest to the highest index, the most disadvantaged 10% of the households are given a decile number of 1, the next most disadvantaged 10% of households are given a decile number of 2 and so on, up to the least disadvantaged 10% of households which are given a decile number of 10. This means that households are divided up into ten equal sized groups, depending on their score.

6.3 Cross-tabulation of selected household characteristics by household index of disadvantage decile

Due to the household-level index not being tied to geography the characteristics of the households can now be assessed against their index decile. Table 6.6 displays cross tabulations between the household socio-economic index of disadvantage deciles and the relevant household characteristics of state of residence, remoteness, tenure type, labour force status, education and SEIFA Index of Relative Socio-Economic Disadvantage for GSS 2010. Appendix D presents comparable results for GSS 2014. These results display the value of the index to identify characteristics of those experiencing extreme disadvantage.

The first panel in table 6.6 shows states and territories by Index of Disadvantage deciles. Most states and territories have a fairly even spread of distribution. The ACT has a higher proportion of the population in more advantaged deciles, with 15.7% of households being in Decile 10.

The second panel in table 6.6 shows remoteness by Index of Disadvantage deciles. This indicates an expected spread of households, with those in major cities tending to be less disadvantaged. Inner regional has the largest proportion of households in Decile 1, with 13.1% of all inner regional households. The largest proportion of all outer regional households also fall in Decile 1, with 11.7% of households. However, there are also more than 10% in Deciles 5 – 8 indicating a moderate lack of disadvantage for these households.

Tenure type appears to have a strong association with disadvantage, as can be seen in panel 4 of table 6.6, with over 50% of renting households falling in Decile 3 or below. For the owner occupiers without a mortgage, 59.8% of households fall within the 4 to 7 decile range. Owners with mortgages appear to experience the least disadvantage with the majority of households being in Decile 6 or above.

Panel 5 and 7 of table 6.6, Labour Force Status and Education, show clearly expected results. There is a strong association between households in which the reference person is employed and a lack of disadvantage, with 14% of these households being in Decile 10, whereas 54.9% of households with an unemployed reference person fall within Deciles 1 and 2. Panel 7 shows that households in which the reference person holds a Bachelor Degree or above are unlikely to be severely disadvantaged with only 15.4% of households falling within the three most disadvantaged deciles, with only 3.7% of those households in Decile 1.

Table 6.6, panel 6 shows that multiple family households tend to be more disadvantaged, with 48.2% of households falling within Deciles 1 to 3. One family households are fairly evenly distributed across the Deciles, however there is a slight tendency towards the more advantaged deciles.

The last panel in table 6.6, SEIFA deciles by GSS Index of Disadvantage deciles, shows the value of this index as a complement to an area based index. Due to the broad nature of area based measure, finer level disadvantage can be missed. As expected there is correlation between the two measures, however, there are many households that have a differing SEIFA and household-level decile of disadvantage.

6.6 Percentage of households in GSS Index of Disadvantage Deciles by selected household characteristics

	Decile 1	*Decile 2*	*Decile 3*	*Decile 4*	*Decile 5*	*Decile 6*	*Decile 7*	*Decile 8*	*Decile 9*	*Decile 10*	*Total*
						(%)					
State and Territory											
New South Wales	9.3	9.9	9.5	10.5	10.1	10.6	10.6	9.3	10.1	10.2	100.0
Victoria	10.4	10.2	10.2	9.3	9.7	8.5	9.6	10.2	10.6	11.3	100.0
Queensland	11.8	11.0	8.7	10.0	9.7	10.9	9.1	10.6	9.8	8.4	100.0
South Australia	9.3	8.6	11.9	10.4	10.1	9.6	10.1	11.1	8.8	10.1	100.0
Western Australia	9.0	9.6	11.8	10.4	10.2	11.3	10.7	10.2	8.8	8.0	100.0
Tasmania	10.6	9.2	12.1	8.8	10.7	9.5	8.2	8.9	11.2	10.8	100.0
Northern Territory	9.7	9.2	11.1	7.9	8.3	10.3	10.5	9.0	14.1	9.9	100.0
Aust. Capital Territory	7.6	6.6	9.5	7.8	9.0	9.1	11.5	10.9	12.3	15.7	100.0
Remoteness											
Major Cities	8.8	9.9	9.7	10.2	9.8	9.9	10.8	10.2	10.1	10.7	100.0
Inner Regional	13.1	10.8	11.4	10.5	10.1	10.2	7.3	9.0	9.1	8.7	100.0
Outer Regional	11.7	8.9	9.2	7.4	10.2	10.8	10.7	11.2	11.5	8.4	100.0
Remote Area	11.5	8.6	12.8	10.8	11.4	13.0	8.2	8.5	10.1	5.0	100.0
Tenure Type											
Owner w/o mortgage	3.6	6.4	9.4	13.0	11.9	13.1	12.4	8.8	10.4	11.0	100.0
Owner with mortgage	7.3	8.0	8.3	8.0	9.0	10.3	10.4	13.6	12.4	12.7	100.0
Renter	21.1	16.5	12.7	9.5	8.3	6.3	7.1	6.9	6.9	4.9	100.0
Labour Force Status											
Employed	6.2	8.3	8.0	8.3	8.7	9.0	11.0	13.1	13.4	14.0	100.0
Unemployed	36.7	18.2	8.7	7.4	5.9	5.3	3.6	2.2	7.7	4.5	100.0
Not in Labour Force	15.0	12.5	13.7	13.3	12.4	12.4	8.7	5.0	4.0	3.0	100.0
Household Type											
One family household	8.4	9.4	9.0	10.0	9.8	9.9	9.8	10.8	11.0	11.8	100.0
Multiple family h'hold	13.8	15.8	18.6	6.2	7.9	6.9	7.5	6.3	3.7	13.2	100.0
Lone person	13.9	11.4	11.9	10.2	10.5	10.8	10.8	8.1	7.3	5.0	100.0
Group household	15.5	8.1	15.8	8.7	7.4	8.7	9.1	9.1	10.7	7.0	100.0
Education											
Year 12 and below	11.8	11.2	13.8	11.2	12.1	10.7	8.5	8.2	6.6	6.0	100.0
Diploma / Certificate	12.3	10.8	7.9	9.9	7.2	10.1	10.8	10.2	10.7	10.0	100.0
Bachelor & above	3.7	6.1	5.6	7.9	9.3	8.9	11.9	13.5	15.4	17.7	100.0
SEIFA – IRSD deciles											
SEIFA 1	28.2	14.7	14.1	11.0	10.0	6.3	6.1	5.3	2.8	1.5	100.0
SEIFA 2	14.7	15.2	12.7	11.5	9.5	10.1	9.0	6.3	6.8	4.4	100.0
SEIFA 3	13.8	8.7	12.8	9.0	12.7	10.0	7.9	11.8	7.9	5.5	100.0
SEIFA 4	9.4	12.0	12.1	8.4	12.3	11.1	10.5	8.5	7.4	8.4	100.0
SEIFA 5	10.2	10.1	12.7	12.0	9.5	11.7	9.4	7.2	9.0	8.1	100.0
SEIFA 6	8.8	8.7	9.3	9.7	8.3	10.2	11.1	12.2	12.4	9.3	100.0
SEIFA 7	5.6	8.4	7.2	10.2	10.2	11.2	11.1	10.5	13.9	11.8	100.0
SEIFA 8	6.2	10.8	8.8	10.9	8.5	8.1	7.9	11.8	12.6	14.6	100.0
SEIFA 9	2.8	5.9	6.1	8.8	8.1	12.1	14.2	15.4	13.1	13.6	100.0
SEIFA 10	1.7	5.3	4.0	8.9	10.4	9.7	12.0	11.6	14.6	21.9	100.0

7. VALIDATION OF THE INDEX

To ensure that the constructed index is measuring the desired concept and to confirm the results, validation of the index was performed. This validation is important to establish the credibility of the index and identify any issues that may have been missed in the construction of the index. This section briefly discusses results from the validation exercise.

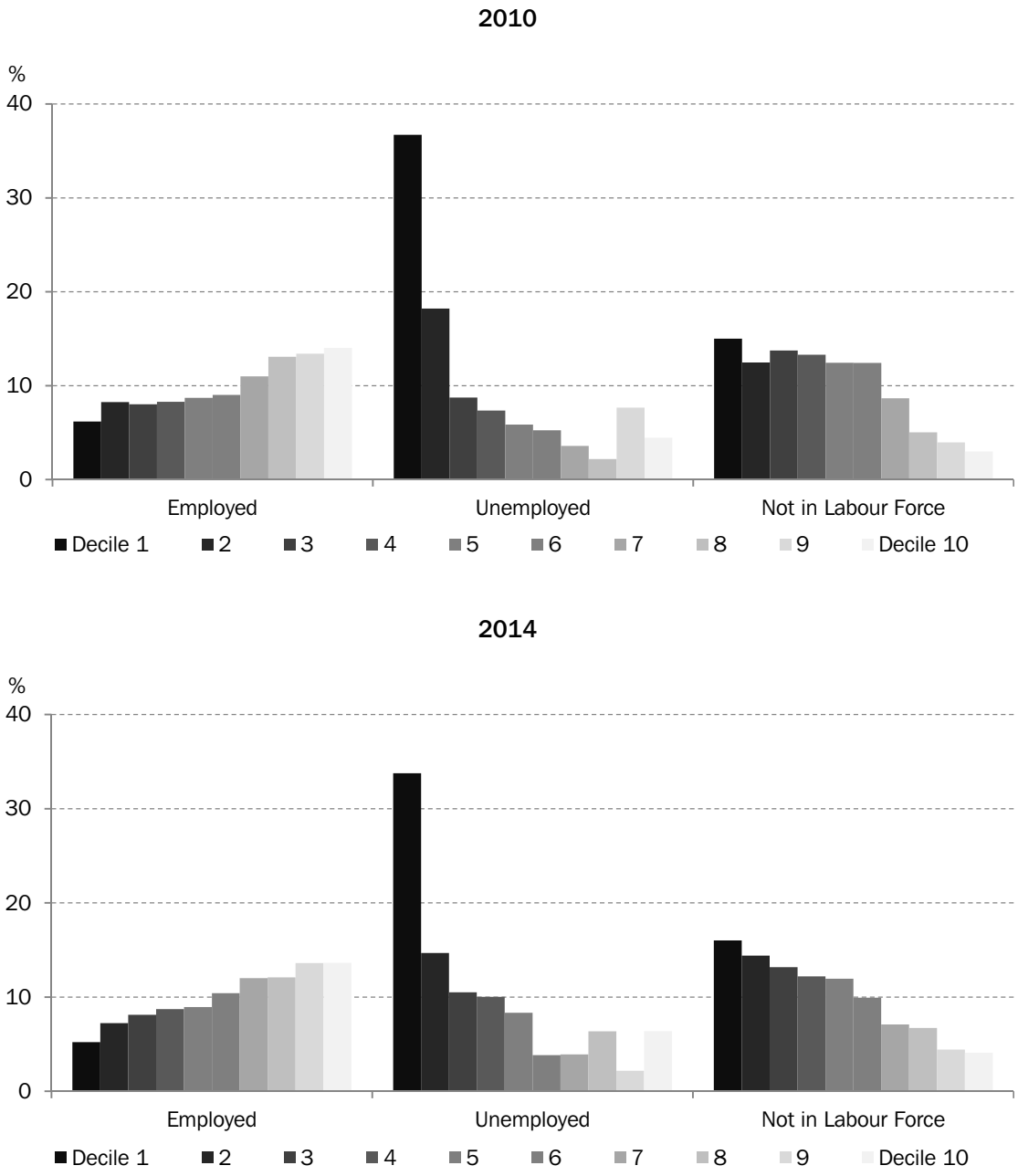
In the previous section it was shown that a large majority of the final set of variables used to compile the index were common across both periods and their loadings were quite similar. This consistency between the loadings across the two periods provides an indication of the robustness of the index results constructed based on the PCA method.

As an initial validation to check the robustness of the created index, the characteristics of the most disadvantaged households were compared with the least disadvantaged households. With a robust model, it is expected that the household with a lower index score should have more disadvantage characteristics, represented by lower values across the variables selected, and vice versa. As expected the validation showed that increases in the index score are associated with increases in the values of each variable.

As further validation of the index, a cross tabulation of the created index against a number of selected variables or household characteristics not included in the index construction was undertaken to establish the robustness of both the 2010 and 2014 results. Consistent with the index construction process, the weighted population was used for the cross tabulation. The household characteristics examined include state, SEIFA, remoteness, tenure type, and family composition, with graphs created based on percentages calculated within each category of each variable on the x axis.

Figure 7.1 shows labour force status by household disadvantage deciles for 2010 and 2014. From the graphs it can be seen that for both periods, households with employed reference persons are generally found in the less disadvantaged deciles, while a larger proportion of households with unemployed reference persons are generally found in the more disadvantaged deciles, particularly Deciles 1 and 2. Similar pattern could be found with the households whose reference persons are not in the labour force (NILF). While the overall results are consistent with expectations, there are some anomalies in this pattern, particularly for the unemployed group which has an unexpectedly large proportion in advantaged Deciles 9 and 10 for the GSS 2010 and Decile 8 and 10 for GSS 2014. This could be because there are employed persons in the households who are not the reference person.

7.1 Percentage of households residing in deciles of GSS Index of Disadvantage by Labour Force Status



This limited validation of the index generally confirms that the produced index appears to be robust and credible. There is scope for further validation of the methodology and the results. This could include consultation with external experts, inspecting the validity of the rankings of households in more depth and testing the selection of variables and the sensitivity of the derived weights by taking multiple random samples of households and re-deriving the significant variables and their weights.

8. DISCUSSION

The construction of a household-level index of disadvantage using GSS data has provided insight into the disadvantage experienced by Australian households. The variables included in the index provide an overview of the combination of factors associated with disadvantage. By utilising variables which have a loading of |0.3| or above, the index highlights the variables most correlated with the concept of disadvantage.

Five out of the ten highest loading variables for both the 2010 and 2014 indexes were from the domains of financial stress, income and wealth. Four out of the ten highest loading variables were from the health domain. This is consistent with other socio-economic indexes created by the ABS, although the GSS index provides a more detailed view due to the wider range of variables available (ABS, 2013; Wise and Williamson, 2013). A limitation of the GSS dataset, however, is that it is based on a sample of the Australian population. The GSS provides depth of topics, but a smaller sample of households surveyed compared to the Census, that provides less depth but more coverage.

The results are data driven, which reduces the subjectivity of the variable selections and weightings. The variables are selected based on the correlations within the dataset, rather than the hypothesised importance of the variables. This can be viewed as a strength of the index, although it has the implication of reducing the choice and control that the researcher has over the final variables in the index.

The index is dependent on the initial variable list chosen for the analysis, as the method simply calculates correlations within the dataset to determine the weights. The GSS provides a broad and extensive range of variables for the index, such as in the domains of safety, crime and wealth, which facilitates more granular analysis of disadvantage compared to indexes constructed from Census data. The detailed nature of the dataset provides a broad base to measure disadvantage and has a wide range of appropriate variables which is likely to improve the quality of the index.

Missing income responses led to a large proportion of households being excluded from the analysis. The spread of these missing responses was investigated and the impact that this could be having on the index was evaluated. The missing records did not seem to be affecting the results to a large degree, however, some way of calculating values for the missing responses could be examined in future research.

A significant challenge of constructing finer level indexes, such as individual level indexes, is accurately accommodating the whole population due to the changing relevance of factors in life-cycle stages and a limitation of available data items. The choice of a household-level index mitigates some of these challenges by considering the household as a unit, which will have less extreme life-cycle effects. A limitation of

the GSS dataset is that the person level data relates to a single person in the household. The assumption that this information is relevant for and representative of the household may not hold in all cases. Conversely, person variables do remain relevant for whole households, which is a limitation of individual level indexes. For example, consider a married couple with differing incomes. In an individual level index, they would be considered to have different levels of disadvantage, although this might not be an appropriate reflection. Additionally, characteristics such as unemployment or whether a person has a disability are likely to impact upon the rest of the household.

9. CONCLUSION AND FURTHER RESEARCH

This paper explored the feasibility of constructing an index of socio-economic disadvantage at the household level using GSS data. It extends on earlier ABS work on finer level indexes of disadvantage at individual, family and household level using Census data. The interest in finer level indexes emerges from the fact that area based measures of disadvantage may not necessarily capture the disadvantage of those living within the areas.

The GSS captures a broader range of socio-economic variables than the Census, which allows a detailed investigation of the relationship between these variables and the concept of disadvantage. Data from both GSS 2010 and GSS 2014 were examined separately and variables relevant to the concept of socio-economic disadvantage were selected from each dataset. The GSS variables selected from each survey were very similar, indicating the stability of the measure. A mixture of person and household-level variables was used to construct an index of household-level disadvantage using household-level weights.

This paper examined both simple and complex measures of disadvantage. The simple measures involved counts of the indicators and domains of disadvantage. Results from the simple measures showed that a large majority of the households experienced few counts of disadvantage and a very small proportion experienced severe levels of disadvantage for both 2010 and 2014. A limitation with such simple measures is that they give equal weighting to all the indicators or domains of disadvantage used to compile the count measure. This ignores the possibility that some indicators or domains of disadvantage may be more important or acute than some others. The results from this simple analysis, however, helps shed light on the different dimensions of disadvantage, which can be explored further, particularly in the cases of households identified as facing disadvantage on several fronts.

The composite method of index construction, which overcomes the limitation of equal weighting of the simple methods, involves the use of an explicit weighting scheme to combine the different variables of disadvantage to construct a summary measure of disadvantage. The PCA technique was used to derive the weights for the compilation of the composite index. The steps used to derive the final set of variables and their corresponding weights in this paper are similar to the approach used for SEIFA.

An analysis of the results from the composite index showed that a large majority of the final set of variables used to compile the index were common and with similar variable loadings across both periods. The distribution of the index was also very similar across the two periods. Five out of the ten highest loading indicators for both the 2010 and 2014 indexes were from the domains of financial stress, income and wealth. Four out of the ten highest loading indicators were from the health domain. Factors

such as education, unemployment and English proficiency level were excluded from the index because they did not meet the cut-off threshold for variable selection.

A cross tabulation of the household-level index decile by selected demographic, geographic and socio-economic characteristics showed that the relationships between the household-level disadvantage measure and these variables were in line with expectations. Limited validation of the index that was undertaken also generally confirmed that the produced index was robust.

This research has shown that it is possible to construct an index of household-level disadvantage using GSS data. Such an index can potentially be used to categorise households into appropriate groupings at area level and more specifically target disadvantaged groups which is not feasible with an area based measure.

However, the results at this stage should be treated as experimental. There are a number of limitations worth considering regarding to the results. The results produced here are data driven and the final index is dependent on the set of variables selected for analysis. A different set of underlying variables could produce different results. While the GSS contains a wider range of socio-economic variables than the Census, the variables in the survey appear to be more skewed to towards certain areas (such as financial stress, health, crime, personal stress). The mixture of person and household-level variables and the assumptions that have been made to create a household index should also be born in mind when interpreting this household based index.

Further research as part of this study could include further validation of the methodology and the results as identified in this paper. For those records with missing scores alternative methods to deal with the missing data could also be investigated such as creating a new survey weighting or imputing income values.

ACKNOWLEDGEMENTS

The authors would like to thank Rosalynn Mathews for her initial work on this project. We would also like to thank the following members of ABS staff who provided comments and feedback: Ruel Abello, Fiona Shalley, Stephen Becker, Tom Sullivan, Tracey Brandrup, Caroline Daley, Heather Burgess, Angie McCosker and Allan McLean. Responsibility for any errors or omissions remains solely with the authors.

REFERENCES

Abello, A.; Cassells, R.; Daly, A.; D'Souza, G. and Miranti, R. (2014) *Youth Social Exclusion in Australian Communities: A New Index*, NATSEM Working Paper 14/25, University of Canberra.
<http://www.natsem.canberra.edu.au/storage/1-WP_25_Youth_Social_Exclusion_in_Australian.pdf>

Ainley, J. and Long, M. (1995) "Measuring Student Socioeconomic Status", in J. Ainley, B. Graetz, M. Long and M. Batten, (eds) *Socioeconomic Status and School Education*, Australian Government Publishing Service, Canberra, pp. 53–76.

Australian Bureau of Statistics (2008) *Information Paper: An Introduction to Socio-Economic Indexes for Areas (SEIFA), 2006*, cat. no. 2039.0, ABS, Canberra.
< http://www.abs.gov.au/ausstats/abs@.nsf/mf/2039.0 >

—— (2011a) *Information Paper: Measures of Socio-economic Status, 2011*, cat. no. 1244.0.55.001, ABS, Canberra.
< http://www.abs.gov.au/ausstats/abs@.nsf/mf/1244.0.55.001 >

—— (2011b) *General Social Survey: Summary Results, Australia, 2010*, cat. no. 4159.0, ABS, Canberra.
< http://www.abs.gov.au/ausstats/abs@.nsf/mf/4159.0 >

—— (2013) *Socio-Economic Indexes for Areas (SEIFA) – Technical Paper, 2011*, cat. no. 2039.0.55.001, ABS, Canberra.
< http://www.abs.gov.au/ausstats/abs@.nsf/mf/2033.0.55.001 >

Bailey, N.; Flint, J.; Goodlad, R.; Shucksmith, M.; Fitzpatrick, S. and Pryce, G. (2003) *Measuring Deprivation in Scotland: Developing a Long-Term Strategy*, Scottish Executive Central Statistics Unit.
< http://www.scotland.gov.uk/publications/2003/09/18197/26538 >

Baker, J. and Adhikari, P. (2007) "Socio-Economic Indexes for Individuals and Families", *Methodology Advisory Committee Papers*, cat. no. 1352.0.55.086, Australian Bureau of Statistics, Canberra.
< http://www.abs.gov.au/ausstats/abs@.nsf/mf/1352.0.55.086 >

Baxter, J. and Taylor, M. (2014) "Measuring the Socio-economic Status of Women Across the Life Course", *Family Matters*, 95, Australian Institute of Family Studies.
< https://aifs.gov.au/sites/default/files/fm95g.pdf >

Bradshaw, J.; Noble, M.; Bloor, K.; Huby, M.; McLennan, D.; Rhodes, D.; Sinclair, I. and Wilkinson, K. (2009) "A Child Well-Being Index at Small Area Level in England", *Child Indicators Research*, 2(2), pp. 201–219.

Dinov, I. (2004) *Principal Component Analysis*, Course Notes on Statistical Methods in Biomedical Imaging (UCLA Stat 233), University of California, Los Angeles. <http://www.stat.ucla.edu/~dinov/courses_students.dir/04/Spring/Stat233.dir/STAT233_notes.dir/PCA.pdf>

Hagerty, M. and Land, K. (2007) "Constructing Summary Indexes of Quality of Life: A Model for the Effect of Heterogeneous Importance Weights", *Sociological Methods and Research*, 35(4), pp. 455–496.

Joliffe, I.T. (1986) *Principal Components Analysis*, Springer Series in Statistics.

Kolenikov, S. and Angeles, G. (2009) "Socio-economic Status Measurement with Discrete Proxy Variables: Is Principal Component Analysis a Reliable Answer?", *Review of Income and Wealth*, 55(1), pp. 128–165.

Lalloué, B.; Monnez, J.-M.; Padilla, C.; Kihal, W.; Le Meur, N.; Zmirou-Navier, D. and Deguen, S. (2013) "A Statistical Procedure to Create a Neighborhood Socio-economic Index for Health Inequalities Analysis", *International Journal for Equity in Health*, 12(21). < http://www.equityhealthj.com/content/12/1/21 >

Lim, P. and Gemici, S. (2011) *Measuring the Socioeconomic Status of Australian Youth*, National Centre for Vocational Education Research, Adelaide.

Marks, G.N.; McMillan, J.; Jones, F.L. and Ainley, J. (2000) *The Measurement of Socio-economic Status for the Reporting of Nationally Comparable Outcomes of Schooling*, Australian Council for Educational Research, Melbourne and Australian National University, Canberra.

McLennan, D.; Barnes, H.; Noble, M.; Davies, J.; Garrat, E. and Dibben, C. (2011) *The English Indices of Deprivation 2010*, Department for Communities and Local Government, London.

Michalos, A.C.; Smale, B.; Labonté, R.; Muharjarine, N.; Scott, K.; Moore, K.; Swystun, L.; Holden, B.; Bernardin, H.; Dunning, B.; Graham, P.; Guhn, M.; Gadermann, A.M.; Zumbo, B.D.; Morgan, A.; Brooker, A.-S. and Hyman, I. (2011), *The Canadian Index of Wellbeing, Technical Report 1.0. Waterloo*, ON: Canadian Index of Wellbeing and University of Waterloo.

Morris, R. and Carstairs, V. (1991) "Which Deprivation? A Comparison of Selected Deprivation Indexes", *Journal of Public Health*, 13(4), pp. 307–311.

Rigdon, E.E. and Ferguson, C.E., Jr. (1991) "The Performance of the Polychoric Correlation Coefficient and Selected Fitting Functions in Confirmatory Factor Analysis with Ordinal Data", *Journal of Marketing Research*, 28(4), pp. 491–497.

Salmond, C.; Crampton, P.; King, P. and Waldegrave, C. (2006) "NZiDep: A New Zealand Index of Socio-economic Deprivation for Individuals", *Journal of Social Science and Medicine*, 62(6), pp. 1474–1485.

Scutella, R.; Wilkins, R. and Horn, M. (2009) *Measuring Poverty and Social Exclusion in Australia: A Proposed Multidimensional Framework for Identifying Socio-Economic Disadvantage*, Working Paper No. 4/09, Melbourne Institute of Applied Economic and Social Research.
< http://melbourneinstitute.com/downloads/working_paper_series/wp2009n04.pdf >

Scutella, R. and Wilkins, R. (2010) "Measuring Social Exclusion in Australia: Assessing Existing Data Sources", *Australian Economic Review*, 43(4), pp. 449–463.

UNICEF Office of Research (2013) *Child Well-being in Rich Countries: A Comparative Overview*, Innocenti Report Card 11, UNICEF Office of Research, Florence.
< http://www.unicef-irc.org/publications/pdf/rc11_eng.pdf >

Wise, P. and Mathews, R. (2011) "Socio-Economic Indexes For Areas: Getting a Handle on Individual Diversity Within Areas", *Methodology Research Papers*, cat. no. 1351.0.55.036, Australian Bureau of Statistics, Canberra.
< http://www.abs.gov.au/ausstats/abs@.nsf/mf/1351.0.55.036 >

Wise, P. and Williamson, C. (2013) "Building on SEIFA: Finer Levels of Socio-Economic Summary Measures", *Methodology Advisory Committee Papers*, cat. no 1352.0.55.135, Australian Bureau of Statistics, Canberra.
< http://www.abs.gov.au/ausstats/abs@.nsf/mf/1352.0.55.135 >

All URLs last viewed on 16 July 2015

APPENDIXES

A. COMPARISON OF GSS 2010 AND GSS 2014 ACROSS SELECTED SURVEY CHARACTERISTICS

Survey information	*GSS 2010*	*GSS 2014*
Sample size	15,028	12,932
Sampling unit	Household	Household
Response rate	87.6%	80.2%
Reference period	4 months – Aug to Nov 2010.	4 months – Mar to Jun 2014.
Survey design	Sample targeted to capture socio-economic disadvantage.	Sample moderately targeted to capture socio-economic disadvantage.
Scope	Covers population in private dwellings.	Covers population in private dwellings.
Coverage	Covers urban and rural areas across all states and territories but excludes very remote areas.	Covers urban and rural areas across all states and territories but excludes very remote areas, and Discrete Indigenous Communities.
Sample selection	Any responsible 18+ person randomly selected from household as Household Reference Person who answers person-level and household-level questions.	Any responsible 15+ person randomly selected from household as Household Reference Person who answers person-level and household-level questions.
Interview method	Face-to-face personal interview.	Face-to-face personal interview.
Mode of survey	Personal interview (CAI).	Personal interview (CAI).
Weighting method	Initial selection probability weights benchmarked to: by age, sex, state, part of state (POS), household composition, SEIFA and LFS to derive final person and household-level weights.	Initial selection probability weights benchmarked to: by age, sex, state, part of state (POS), household composition, and SEIFA to derive final person and household-level weights.

B. DESCRIPTION OF GSS VARIABLES USED IN INDEX CONSTRUCTION

B.1 Description of GSS variables used in index construction

Variable Name	*Variable Description*	*Variable Type*	*Scales**
INCOME			
Household equivalised income	Equivalised household gross weekly income	Categorical	1–10
Time on government support	Time spent on government support as main source of income in the last two years	Categorical	1–25
EDUCATION			
Highest level of education	Highest educational attainment	Categorical	1–11
UNEMPLOYMENT			
Unemployed	Labour force status is unemployed	Binary	0=Unemployed; 1=Employed or Not in the labour force
ACCESS TO SOCIETY/SERVICES			
Transport difficulty	Perceived level of difficulty with transport	Categorical	1–5
Difficulty accessing services	Whether could not obtain health care when it was needed	Binary	0=Could not obtain health care; 1=Could obtain health care;
English poor	Proficiency in spoken English	Categorical	1–5
No social activities	Whether has had social activities in last three months	Binary	0=No social activities; 1=Has had social activities
No social support	Ability to get support in times of crisis from persons living outside the household	Binary	0=Not able to get support; 1=Able to get support
CRIME AND SAFETY			
Victim of break-in	Victim of actual or attempted break-in in the last 12 months	Binary	0=Victim; 1=Not a victim
Victim of assault	Victim of assault or break-in in the last 12 months	Binary	0=Victim; 1=Not a victim
Feeling safe (day)	Feelings of safety at home alone during day	Categorical	1–6
Feeling safe (night)	Feelings of safety at home alone after dark	Categorical	1–6
Feeling safe walking (night)	Feelings of safety walking alone in local area after dark	Categorical	1–6
Neighbourhood problem	Degree of severity of main type of problem in local area	Categorical	1–4
WEALTH			
Dwelling equity	Equity in dwelling	Continuous	
Asset value	Value of investment	Categorical	0–4
Consumer debt	Value of consumer debt	Categorical	1–5
FINANCIAL STRESS			
Can't raise $2K	Could not raise $2,000 within a week	Binary	0=Could not raise $2K; 1=Could raise $2K
Cash flow problems	Number of different types of cash flow problems in last 12 months	Categorical	1–10
Dissaving actions	Number of different types of dissaving actions taken in last 12 months	Categorical	1–10
Number of financial stressors	Number of financial stress indicators experienced in the last 12 months	Categorical	1–21
Difficulty paying bills	Frequency in experiencing difficulty in paying bills in last 12 months	Categorical	1–7
Financial exclusions	Number of different types of financial exclusions experienced in last 12 months	Categorical	1–5

B.1 Description of GSS variables used in index construction – continued

Variable Name	*Variable Description*	*Variable Type*	*Scales**
HEALTH			
Self-assessed health	Self-assessed health status	Categorical	1–5
Mental health	Whether has a mental health condition	Binary	0=Has mental condition; 1=No mental condition
Delay doctor	Delayed medical consultation from a doctor because could not afford it	Binary	0=Has delayed medical consultation; 1=Has not delayed medical consultation
Delay medication	Delayed purchasing prescribed medication because could not afford it	Binary	0=Has delayed medication; 1=Has not delayed medication
Health access	Whether could not obtain health care when it was needed	Binary	0=Could not obtain health care; 1=Could obtain health care
Disability	Has profound or severe core activity restriction	Categorical	1–7
Employment restriction	Has employment restriction due to disability (under 65 years old)	Binary	0=Has restriction on employment; 1=Has no restriction on employment
Education restriction	Has education restriction due to disability (under 65 years old)	Binary	0=Has restriction on education; 1=Has no restriction on education
HOMELESSNESS			
Homeless times	Whether has been in the situation without a permanent place to live	Binary	0=Has experienced homelessness; 1=Haven't experienced homelessness
Length homeless	Length of time of most recent experience without a permanent place to live	Categorical	1–10
PERSONAL STRESS			
Personal Stress	How many types of personal stressors has experienced in the last 12 months	Categorical	1–15

* The scales of all variables are ordered from most disadvantaged to least disadvantaged

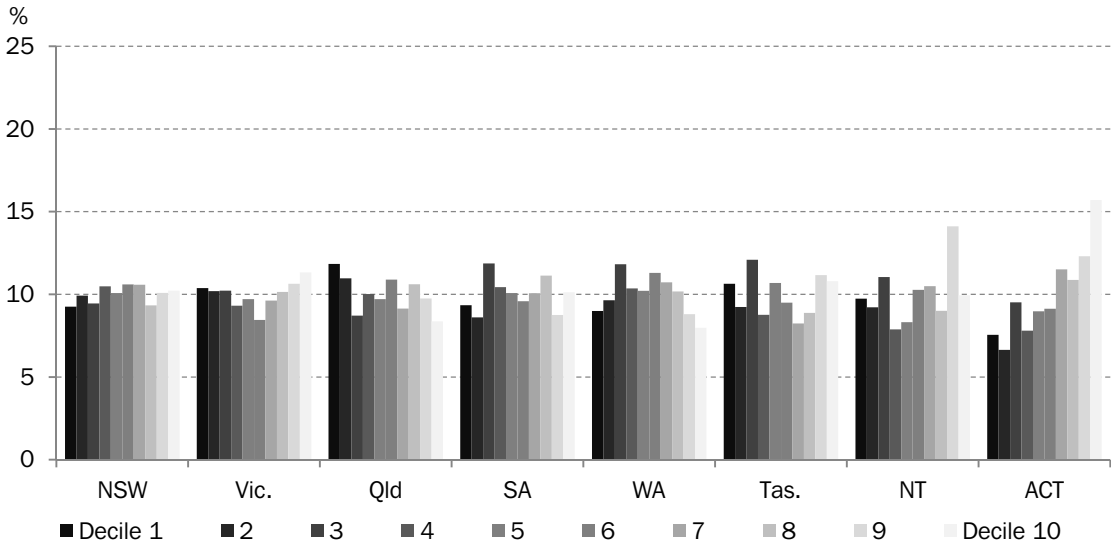
C. VARIABLE LOADINGS WITH ALL VARIABLES INCLUDED

Variables	*2010 loadings*	*2014 loadings*
Number of financial stressors	0.73	0.70
Can't raise $2K	0.61	0.60
Difficulty paying bills	0.61	0.59
Employment restriction	0.55	0.57
Asset value	0.53	0.50
Household equivalised income	0.53	0.52
Disability	0.52	0.54
Homeless times	0.50	0.42
Self-assessed health	0.50	0.52
Time on government support	0.48	0.50
Dissaving actions	0.44	0.46
Education restriction	0.44	0.46
Personal stress	0.43	0.43
Length homeless	0.42	0.33
Feeling safe (night)	0.42	0.35
Transport difficulty	0.41	0.40
Difficulty accessing services	0.37	0.39
Victim of assault	0.33	0.30
Victim of break-in	0.29	0.24
Neighbourhood problem	0.27	0.25
Highest level of education	0.27	0.26
Dwelling equity	0.26	0.22
Feeling safe walking (night)	0.26	0.23
Unemployed	0.19	0.21
No social support	0.16	0.20
Financial exclusions	0.06	0.06
Consumer debt	0.04	0.07
English proficiency	0.02	0.01
Delay medication	0.59	—
Delay doctor	0.49	—
Feeling safe (day)	0.39	—
No social activities	0.17	—
Mental health	—	0.50
Health access	—	0.38

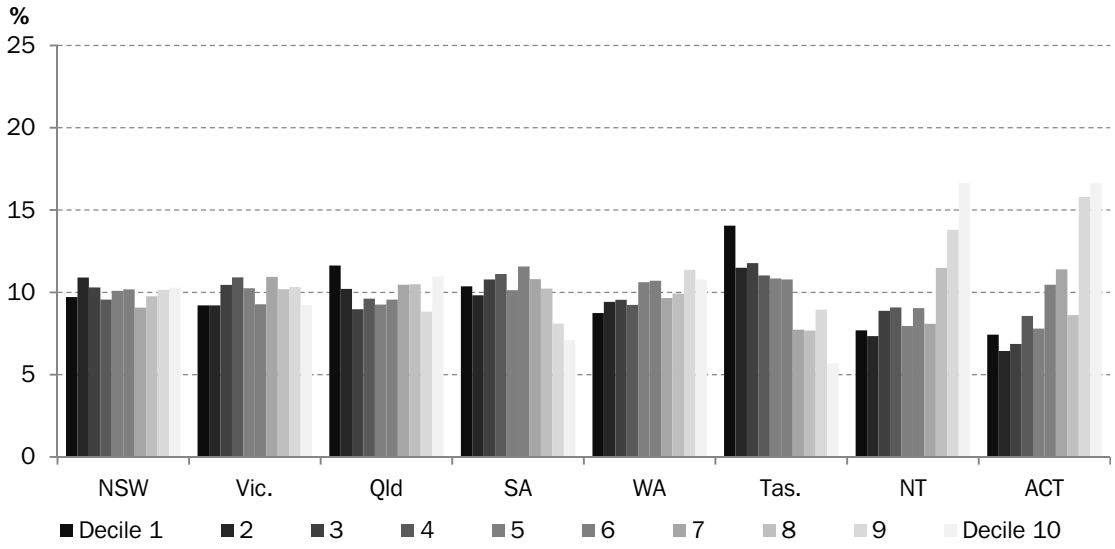
D. CROSS-TABULATION GRAPHS

D.1 Percentage of households residing in deciles of GSS Index of Disadvantage, by State and Territory

2010

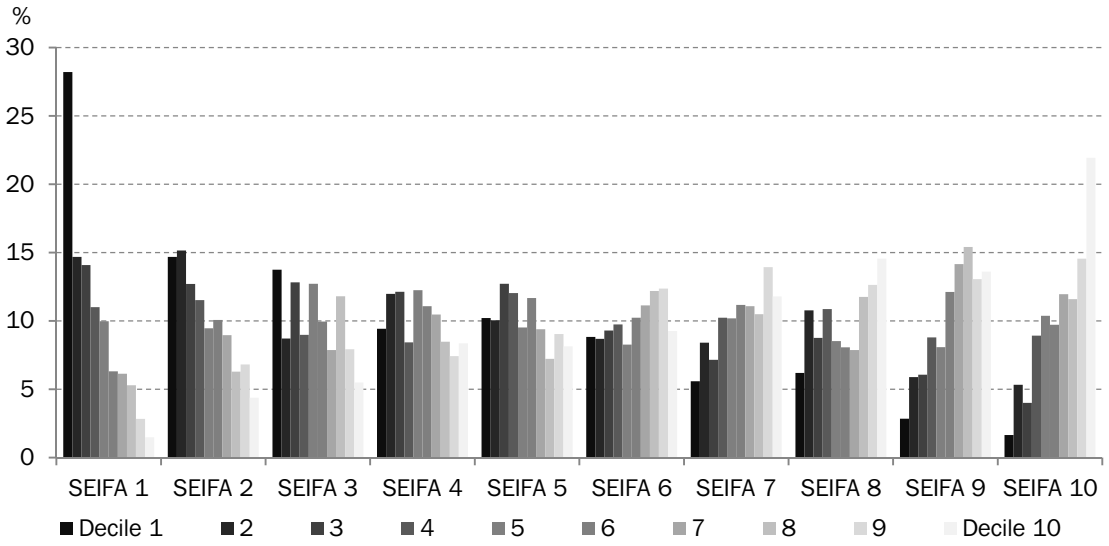


2014

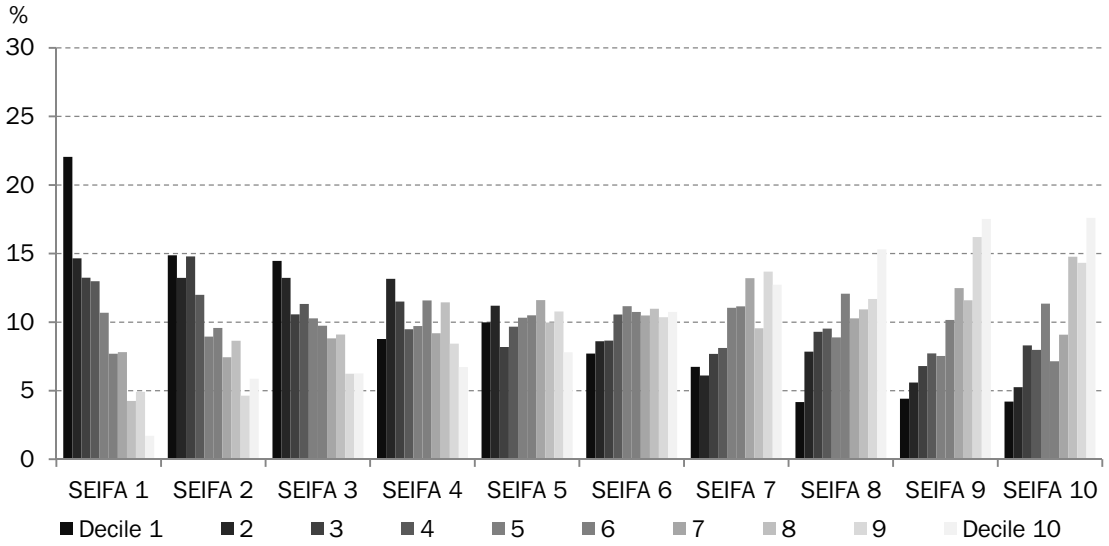


D.2 Percentage of households residing in deciles of GSS Index of Disadvantage by SEIFA IRSAD Deciles

2010

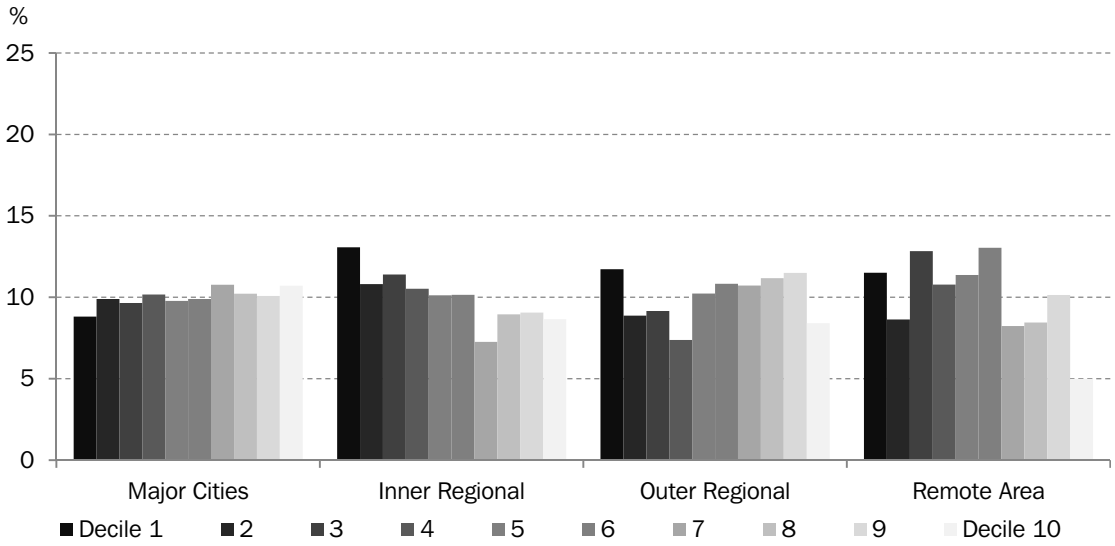


2014

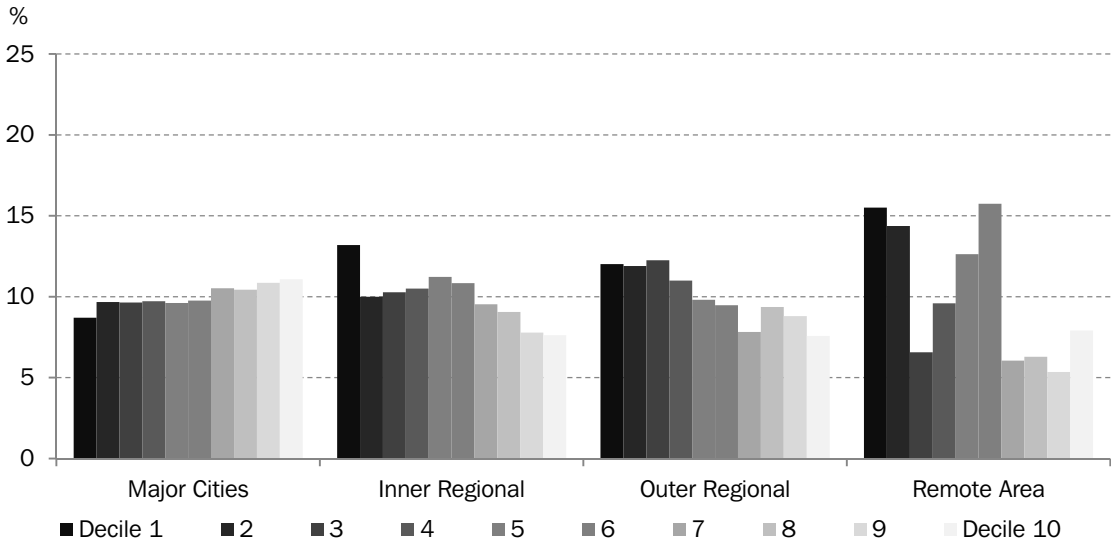


D.3 Percentage of households residing in deciles of GSS Index of Disadvantage by Remoteness

2010

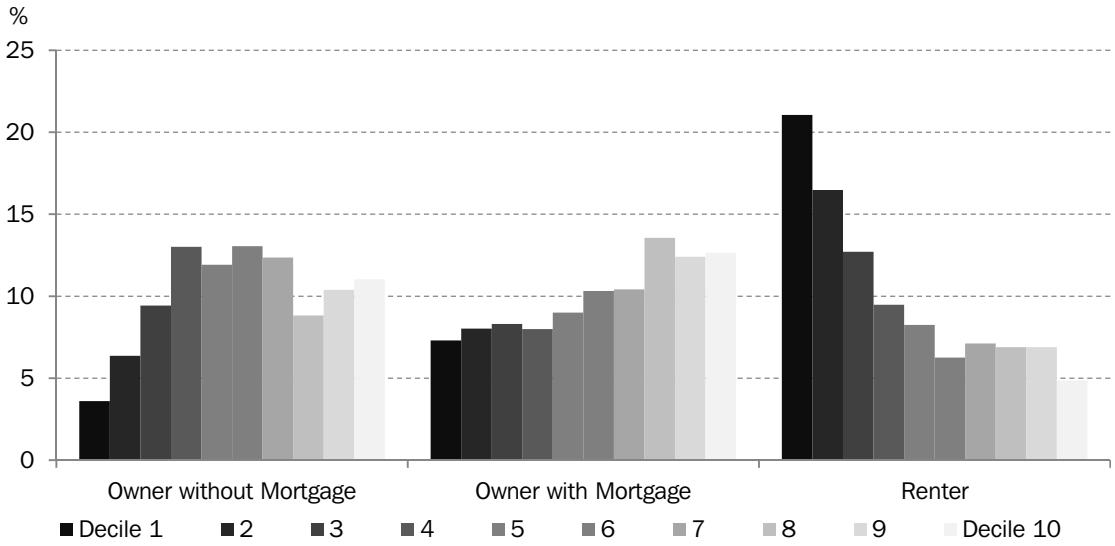


2014

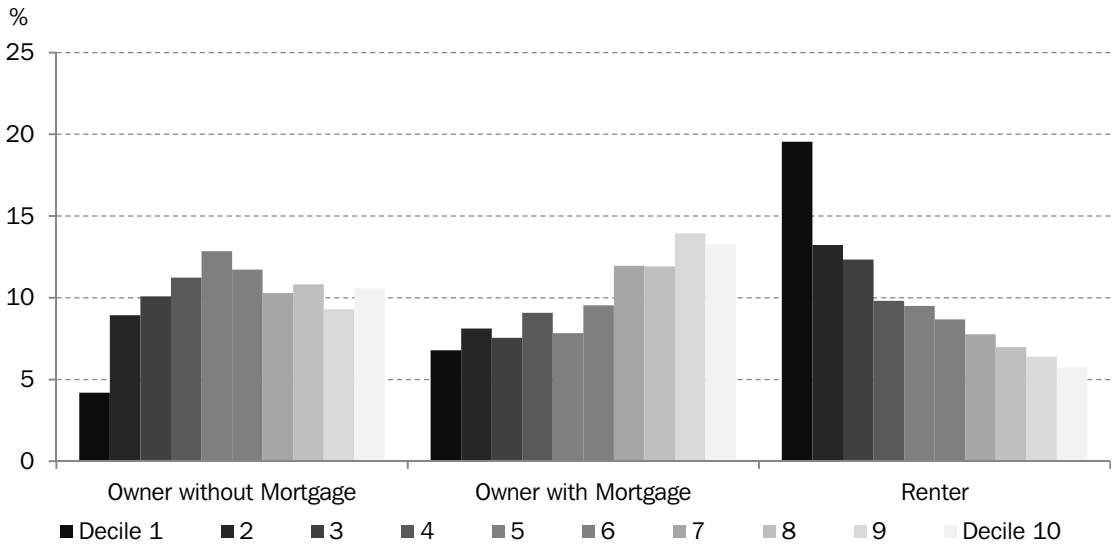


D.4 Percentage of households residing in deciles of GSS Index of Disadvantage by Tenure Type

2010

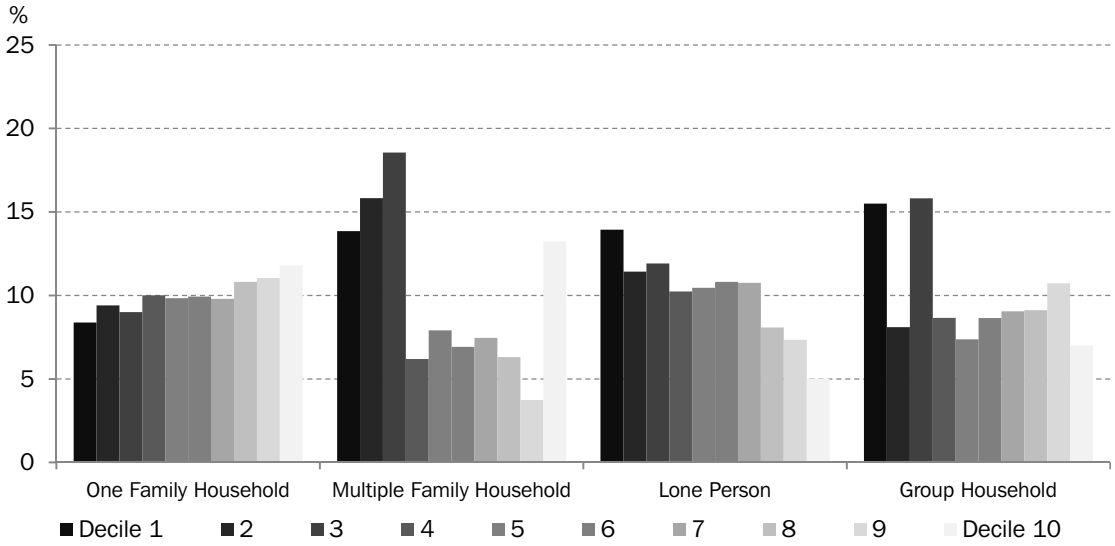


2014

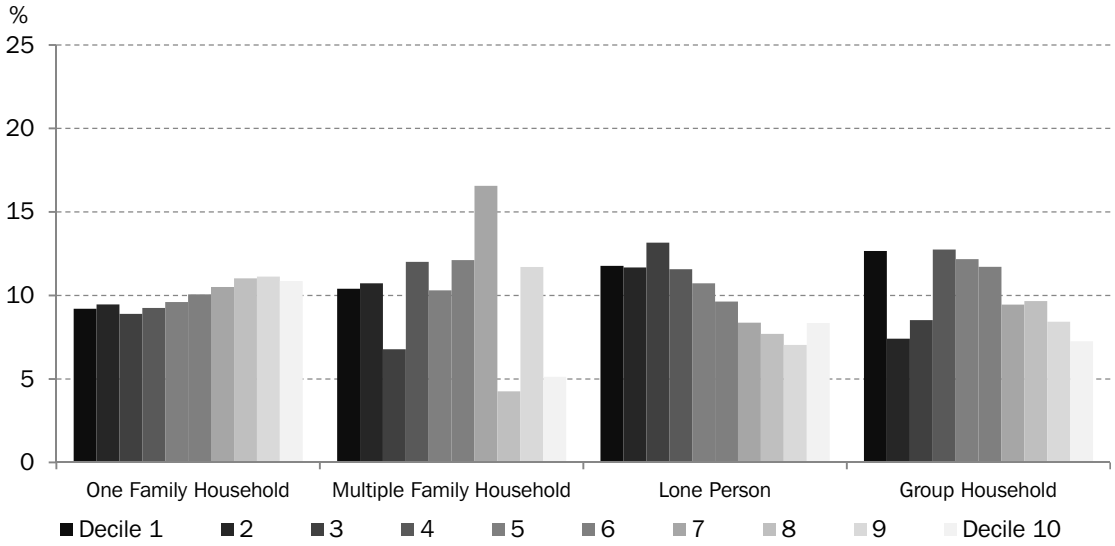


D.5 Percentage of households residing in deciles of GSS Index of Disadvantage by Family Composition

2010

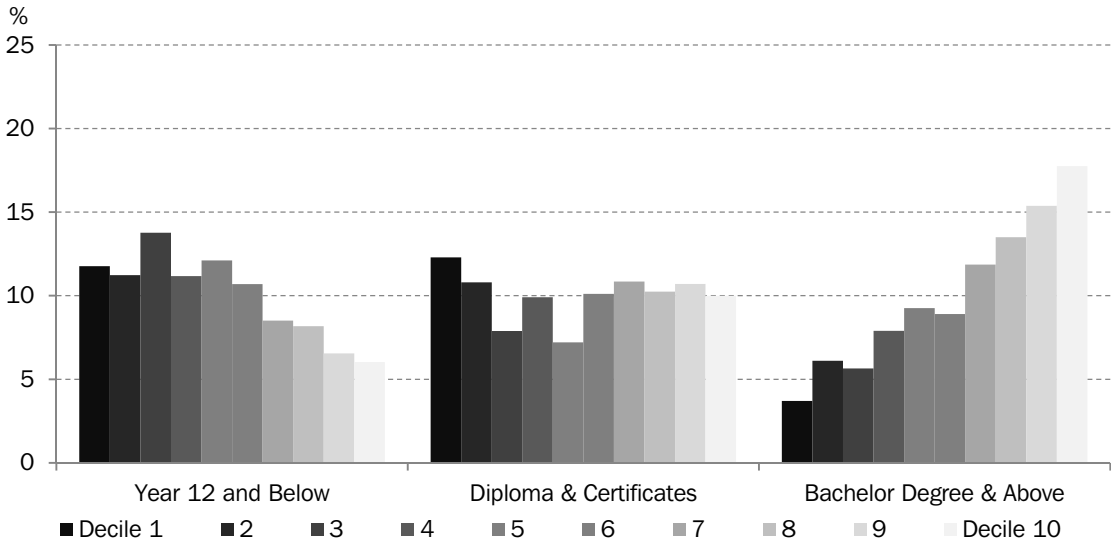


2014

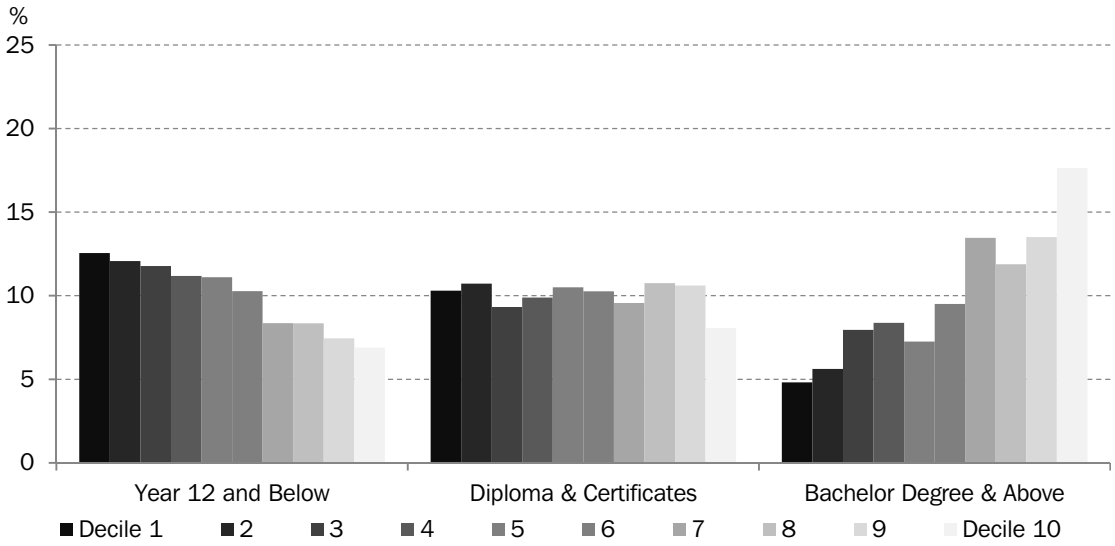


D.6 Percentage of households residing in deciles of GSS Index of Disadvantage by Education

2010



2014



E. CROSS-TABULATION OF HOUSEHOLD INDEX DECILES AGAINST SELECTED HOUSEHOLD CHARACTERISTICS – GSS 2014

	Decile 1	*Decile 2*	*Decile 3*	*Decile 4*	*Decile 5*	*Decile 6*	*Decile 7*	*Decile 8*	*Decile 9*	*Decile 10*	*Total*
						(%)					
State and Territory											
New South Wales	9.7	10.9	10.3	9.6	10.1	10.2	9.1	9.8	10.2	10.3	100.0
Victoria	9.2	9.2	10.5	10.9	10.3	9.3	10.9	10.2	10.3	9.2	100.0
Queensland	11.6	10.2	9.0	9.6	9.3	9.6	10.5	10.5	8.8	11.0	100.0
South Australia	10.4	9.8	10.8	11.1	10.1	11.6	10.8	10.2	8.1	7.1	100.0
Western Australia	8.7	9.4	9.6	9.2	10.6	10.7	9.7	9.9	11.4	10.8	100.0
Tasmania	14.1	11.5	11.8	11.0	10.8	10.8	7.7	7.7	8.9	5.7	100.0
Northern Territory	7.7	7.3	8.9	9.1	8.0	9.0	8.1	11.5	13.8	16.6	100.0
Aust. Capital Territory	7.4	6.4	6.9	8.6	7.8	10.5	11.4	8.6	15.8	16.6	100.0
Remoteness											
Major Cities	8.7	9.7	9.6	9.7	9.6	9.8	10.5	10.4	10.9	11.1	100.0
Inner Regional	13.2	10.0	10.3	10.5	11.2	10.8	9.5	9.1	7.8	7.6	100.0
Outer Regional	12.0	11.9	12.3	11.0	9.8	9.5	7.8	9.4	8.8	7.6	100.0
Remote Area	15.5	14.4	6.6	9.6	12.6	15.7	6.1	6.3	5.4	7.9	100.0
Tenure Type											
Owner w/o mortgage	4.2	8.9	10.1	11.2	12.9	11.7	10.3	10.8	9.3	10.6	100.0
Owner with mortgage	6.8	8.1	7.6	9.1	7.8	9.5	12.0	11.9	13.9	13.3	100.0
Renter	19.5	13.2	12.3	9.8	9.5	8.7	7.8	7.0	6.4	5.8	100.0
Labour Force Status											
Employed	5.2	7.2	8.1	8.7	8.9	10.4	12.0	12.1	13.6	13.6	100.0
Unemployed	33.8	14.7	10.5	10.0	8.3	3.8	3.9	6.4	2.2	6.4	100.0
Not in Labour Force	16.0	14.4	13.2	12.2	12.0	9.9	7.1	6.7	4.4	4.1	100.0
Household Type											
One family h'hold	9.2	9.5	8.9	9.3	9.6	10.1	10.5	11.0	11.1	10.9	100.0
Multiple family h'hold	10.4	10.7	6.8	12.0	10.3	12.1	16.6	4.3	11.7	5.1	100.0
Lone person	11.8	11.7	13.2	11.6	10.7	9.6	8.4	7.7	7.0	8.4	100.0
Group household	12.7	7.4	8.5	12.8	12.2	11.7	9.5	9.7	8.4	7.3	100.0
Education											
Year 12 and below	9.8	9.4	9.2	8.7	8.6	8.0	6.5	6.5	5.8	5.4	100.0
Diploma / Certificate	8.0	8.4	7.3	7.7	8.2	8.0	7.5	8.4	8.3	6.3	100.0
Bachelor & above	3.8	4.4	6.3	6.6	5.7	7.5	10.6	9.3	10.6	13.9	100.0
SEIFA – IRSD deciles											
SEIFA 1	22.1	14.7	13.2	13.0	10.7	7.7	7.8	4.3	4.9	1.7	100.0
SEIFA 2	14.9	13.2	14.8	12.0	8.9	9.6	7.4	8.6	4.6	5.9	100.0
SEIFA 3	14.5	13.2	10.6	11.3	10.3	9.7	8.8	9.1	6.2	6.3	100.0
SEIFA 4	8.8	13.2	11.5	9.5	9.7	11.6	9.2	11.4	8.4	6.7	100.0
SEIFA 5	10.0	11.2	8.2	9.7	10.3	10.5	11.6	10.0	10.8	7.8	100.0
SEIFA 6	7.7	8.6	8.7	10.6	11.2	10.7	10.5	11.0	10.4	10.7	100.0
SEIFA 7	6.7	6.1	7.7	8.1	11.0	11.2	13.2	9.6	13.7	12.7	100.0
SEIFA 8	4.2	7.9	9.3	9.5	8.9	12.1	10.3	10.9	11.7	15.3	100.0
SEIFA 9	4.4	5.6	6.8	7.7	7.5	10.2	12.5	11.6	16.2	17.5	100.0
SEIFA 10	4.2	5.3	8.3	8.0	11.4	7.2	9.1	14.8	14.3	17.6	100.0

FOR MORE INFORMATION . . .

INTERNET	**www.abs.gov.au** The ABS website is the best place for data from our publications and information about the ABS.
LIBRARY	A range of ABS publications are available from public and tertiary libraries Australia wide. Contact your nearest library to determine whether it has the ABS statistics you require, or visit our website for a list of libraries.

INFORMATION AND REFERRAL SERVICE

Our consultants can help you access the full range of information published by the ABS that is available free of charge from our website, or purchase a hard copy publication. Information tailored to your needs can also be requested as a 'user pays' service. Specialists are on hand to help you with analytical or methodological advice.

PHONE	1300 135 070
EMAIL	client.services@abs.gov.au
FAX	1300 135 211
POST	Client Services, ABS, GPO Box 796, Sydney NSW 2001

FREE ACCESS TO STATISTICS

All statistics on the ABS website can be downloaded free of charge.

WEB ADDRESS	www.abs.gov.au