B Data and related issues

The data for the analysis in this paper has been drawn from various Household Expenditure Surveys (HESs) collected by the Australian Bureau of Statistics (ABS). Through this survey, the ABS collects data on household’s spending, in total and by component, forms of income and other household characteristics. The survey was first conducted in 1974-75. For this report, five surveys conducted from 1988-89 to 2009-10 have been used (the 1988-89, 1993-94, 1998-99, 2003-04 and 2009-10 surveys). Over this time sample sizes have varied from below 7000 to over 9000.

The ABS has made numerous changes to the HES since its introduction as it attempts to improve the quality of the survey and its estimates. These changes mean that estimates of many of the variables collected may not be consistent over time. Of importance to this study are changes that may impact on the measurement of household income as these have the potential to influence observed changes and therefore trends in incomes and their distribution from one survey period to the next. This can create uncertainty surrounding the robustness of observed distributional trends and may even misrepresent changes in summary statistics of inequality over time.

Many of the changes to income surveys in the 2000s have impacted on estimates of income — significantly for some groups. Changes have been made to capture newer forms of income such as salary sacrifice benefits and provide for a better accounting of investment and own-business income. Such changes are likely to provide a more complete account of incomes at the top of the distribution, raising measured dispersion of income over time. Thus, it is likely that, in part, some of the upward trend in inequality in the mid to late 2000s reflects better measurement of income rather than underlying changes. This influence has led some researchers to question the reliability of estimates obtained from surveys such as the HES.

While it is likely that changes in survey design and data collection have influenced measured income and trends in summary measures of income dispersion such as the Gini coefficient, it is unlikely that they have had a pervasive impact on the trends observed in the distribution of income across the entire population (as depicted in the Annex to chapter 3). Therefore, given the nature of the analysis presented in this paper, the HES represents a reasonably robust information source from which trends in the distribution of income can be explored. Nevertheless, the major changes in survey design are canvassed in this appendix as they will impact statistics such as the Gini coefficient.

Further, researchers such as Wilkins (2013) have found that different data sources paint a different picture of income inequality and its trends in Australia. These issues are also discussed briefly in this appendix.

## B.1 How has the HES changed?

The ABS’s survey methodology has been evolving since the introduction of the HES in the 1970s. The most significant of the changes to income were introduced in the 2003-04 and 2009-10 surveys. Box B.1 summarises the changes to the ABS’s survey methodology that have influenced income measurement.

### The impact of changes in 2003-04

In 2003-04, the HES was combined with the ABS’s Survey of Income and Housing (SIH), increasing the length of the survey for participants and consequently increasing non-response (and possibly response bias). The integration of the surveys also resulted in an expansion in the information collected on wealth in the HES, from an estimation of owner-occupied dwellings and some household loans to a more comprehensive range of assets and liabilities as per the SIH. The process of providing more detailed information on these items has been assessed by the ABS to have improved the accuracy of reported income from them.

The most significant change occurring in the 2003-04 HES was in the measurement of income from investments and own unincorporated businesses. Previously, these income streams were estimated based on reported income for the previous financial year. From 2003-04, the data have been measured using respondents’ estimates of expected income in the current financial year. According to the ABS, this improves the data significantly, with the year-on-year movements between current and previous year income from these two items now aligning more closely with the related national accounts items. However, the change causes a clear break in the data series, and it is uncertain whether the difference between the two methodologies would remain constant over time. This raises the question of whether this has simply caused a point shift in the data, or whether the impact varies over time, for example in relation to changes in the business cycle (ABS 2006).

|  |
| --- |
| Box B.1 Summary of changes to income measurement in HES |
| 2003-04   * Change in measurement of current income from own unincorporated business and investments from reported income for the previous financial year to respondents' estimates of expected income in the current financial year. * Inclusion in employment income of all salary sacrificed income and non-cash benefits received from employers. * Collection of a more detailed range of income items and information on all assets and liabilities of respondents.   2009-10   * Questions added to the survey on the amount of ‘additional’ overtime respondents expect to earn in the given year (in addition to ‘usual’ overtime). * Netting off of interest paid from interest earned on borrowed funds to purchase shares or units in trusts, previously only gross interest earned was recorded for investments other than rental properties. * Inclusion of termination payments and workers’ compensation lump sums, with an upper boundary of three months wages. * Inclusion of irregular bonuses in employment income (in addition to regular bonuses) * Expansion of family financial support from regular cash payments, mainly child and spousal support, to also include other forms of financial support including goods, services, rent, education (capital transfers, such as cars for example, remain excluded). |
| *Sources*: ABS (2006, 2012a). |
|  |
|  |

The ABS provides estimates of income based on both the new and old collection methods. Summary estimates of gross income for all households and household earnings from an own unincorporated business, hereafter termed self-employed household income, using both the old and new 2003-04 basis indicate that the impact of the changes on gross household income are small (gross household income is made up of employee income, income from own unincorporated businesses, direct government payments and other sources) (table B.1). However, the impact of the changes on self-employed household income are larger.

The changes reduce both the mean and median self-employed earnings and its standard deviation. One possible explanation for the reduction in the standard deviation is the reduction in the number of households reporting negative self-employment incomes. This effect can be seen more clearly in figure B.1 (left panel). Some negative self-employment income under the old definition (the dotted line), appears to have been shifted to slightly positive income under the new definition (the solid line), as reflected in the higher peak of the solid line next to the y-axis.

Table B.1 Impact of survey changes on gross and self-employed household income in 2003-04

2003-04 dollars

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Gross household income | |  | Self-employed earnings | |
| Summary statistic | Previous HES basis | 2003-04 HES basis |  | Previous HES basis | 2003-04 HES basis |
| Mean | 1 124 | 1 126 | 648 | | 609 |
| Median | 921 | 928 | 464 | | 400 |
| Standard deviation | 954 | 919 | 1150 | | 975 |
| Gini coefficient | 0.41 | 0.40 | 0.72 | | 0.68 |

a Self-employed is defined as those with non-zero earnings from own unincorporated businesses.

*Source*: Author estimates based on ABS (Household Expenditure Survey, cat. no. 6503.0, confidentialised unit record files).

The reason for little change in gross household income stemming from the revised measurement of own unincorporated business income is that only around 8 per cent[[1]](#footnote-1) of households earn income from this source. In terms of the distribution, it is difficult to observe the impact (figure B.1, right hand panel).

Figure B.1 Distribution of gross and self-employed household income, old and new basis in 2003-04

Proportion of households

|  |  |
| --- | --- |
| Self-employed earnings | Gross income |

*Data source*: Author estimates based on ABS (Household Expenditure Survey, cat. no. 6503.0, confidentialised unit record files).

### The impact of changes in 2009-10

In the 2009-10 HES, the concept of household income was altered to comprise all current receipts, whether monetary or in kind, that are intended to support current consumption. In previous surveys, income was limited to regular cash income only. Employment income was expanded to include all payments received by individuals as a result of their current or former involvement in paid employment. These include non-cash benefits, bonuses, termination payments and payments for irregular overtime, and the inclusion of lump sum workers’ compensation receipts.

A wider range of data on financial support received from family members resident outside the household was also included. In addition to regular payments previously collected, financial support was extended to include other forms of financial support, including goods and services received which were purchased by others such as rent, education, food, clothing, car registration and utilities. Capital transfers, such as the purchase of property or cars, remain excluded.

Some changes were also made to investment income; netting out of interest paid from income earned on money borrowed to purchase shares or units in trusts, and the reclassification of income earned as a silent partner in a partnership and some private trust income as investment income rather than unincorporated business income.

The impact of these changes can been seen by examining the differences between the new income estimates compared to those obtained using the previous method. While the 2009-10 HES does not have comparable estimates on a pre-2003-04 nor a 2003-04 basis, differences between the 2005-06 and 2009-10 methods can be explored, although there is a small difference in the measurement of wages and salaries between the 2003-04 HES/SIH and the 2005-06 SIH.[[2]](#footnote-2) Summary statistics for gross household income (table B.2) indicate that the changes made in the 2009-10 HES have slightly increased income on average, raising both the mean and median incomes by less than five per cent, and also increasing the standard deviation and Gini coefficient a little.

Table B.2 Impact of survey changes on gross and self-employed household income in 2009-10

2009-10 dollars

|  |  |  |  |
| --- | --- | --- | --- |
|  | Gross household income | |  |
| Summary statistic | 2005-06 HES basis | 2009-10 HES basis |  |
| Mean | 1 619 | 1 684 |
| Median | 1 287 | 1 316 |
| Standard deviation | 1 607 | 1 689 |
| Gini coefficient | 0.42 | 0.43 |

a Self-employed is defined as those with non-zero earnings from own unincorporated businesses.

*Source*: Author estimates based on ABS (Household Expenditure Survey, cat. no. 6503.0, confidentialised unit record files).

The difference between the 2005-06 definitions and 2009-10 appear to be negligible for lower income earners (those earning less than $1000 per week) (figure B.2). The 2009-10 changes create some change in the distribution for higher earners.

Figure B.2 Distribution of gross household income, old and new basis in 2009-10

Proportion of households

|  |
| --- |
|  |

*Data source*: Author estimates based on ABS (Household Expenditure Survey, cat. no. 6503.0, confidentialised unit record files).

## B.2 Differences between data sources

Differences in survey methodology and definitions of income result in inconsistencies not only within the HES series over time but also when comparing measures of income distribution across different data series.

As noted in this paper, information on income was obtained from the ABS HES surveys, however, other sources of income information exist. The HILDA survey, for example, has been used by other researchers to explore trends in the distribution of income in Australia over time. Estimates using the HILDA survey data are likely to differ from those obtained using ABS data due to differences in survey design. Further, some researchers have found inconsistencies in observed trends amongst data collected by the ABS depending on whether it is reported on an annual or weekly basis. These differences are outlined below.

### Weekly, annual and HILDA based estimates of inequality

In the SIH (and HES) income is measured in two ways — weekly income is calculated from reported current income (that is the income received at the time the data was collected) and annual income (calculated from previous financial year data). In the SIH, the two series show different changes in income inequality between 2002-03 and 2005-06. Overall, the weekly series suggests very little overall change in household equivalised income inequality whereas the annual series suggests a large increase in income inequality (figure B.3).[[3]](#footnote-3)

When income data from the HILDA survey is included in the comparison, the general trend in estimated income inequality appears consistent with results from the weekly SIH data series, although with less variation from year to year. The HILDA survey methodology has remained consistent over time avoiding the sampling problems with ABS data discussed above. However, HILDA is a longitudinal study, surveying the same households every year, which exposes it to different biases in regard to assessing changes in the distribution of income relative to the SIH and HES. HILDA may not take into account some demographic changes in the population which could change the structure of the labour market and ultimately the overall distribution of household earnings. For example, immigrants entering Australia over the survey period could not enter the sample between years 2001 and 2010 (unless they joined a household with original sample members) as only households selected in wave 1 could be reinterviewed in this period. Further, the SIH and HILDA surveys have different designs. One focuses on income while the other covers a wider range of topics. Given this, it is likely that both question design, coverage of topics and benchmark data against which household weights are generated will differ. These factors will influence the results of any summary statistics such as the Gini coefficient.

Figure B.3 Gini coefficient for equivalised disposable household income

|  |
| --- |
|  |

aPre-2007 basis, estimates were adjusted for some changes in methodology over time, see Wilkins (2013) for more detail.

*Data source*: Wilkins (2013).

The key lesson from these data comparisons is that differences in survey methodology and definitions of income can result in different estimates of income inequality. Estimates of year-on-year changes in the Gini coefficient are not consistent over different data series making it difficult to determine what exactly has happened to income inequality in Australia using this statistic alone. According to Wilkins (2013) close analysis of changes in distribution of employment, wages and investment income and other decomposition analysis may yield more reliable information on the changes in income distribution in Australia over this period compared to simple headline measures such as the Gini coefficient.

1. Share differs between 2003-04 and previous measurement basis. [↑](#footnote-ref-1)
2. Previously estimates included only some salary sacrificed amounts, this was changed to all amounts being included. [↑](#footnote-ref-2)
3. Summary estimates from the SIH used to provide a more comparable time series to HILDA data than what could be obtained from HES data. [↑](#footnote-ref-3)