THE UNIVERSITY OF MELBOURNE

FACULTY OF BUSINESS & ECONOMICS

Melbourne Institute Working Paper Series

Working Paper No. 19/16

Top Incomes and Inequality in Australia: Reconciling Recent Estimates from Household Survey and Tax Return Data

Richard V. Burkhauser, Markus H. Hahn and Roger Wilkins

MELBOURNE INSTITUTE®
of Applied Economic and Social Research

Top Incomes and Inequality in Australia: Reconciling Recent Estimates from Household Survey and Tax Return Data*

Richard V. Burkhauser[†‡], Markus H. Hahn[‡] and Roger Wilkins[‡]

[†] **Department of Policy Analysis and Management, Cornell University**

[‡] **Melbourne Institute of Applied Economic and Social Research, The University of Melbourne**

Melbourne Institute Working Paper No. 19/16

ISSN 1447-5863 (Online)

ISBN 978-0-73-405216-2

June 2016

* This research is supported by an Australian Research Council Discovery Grant (award DP150102409). We thank the Australian Taxation Office, and in particular Julia Neville, Matt Power, Julia Rymasz and Jamie Fischer, for helpful discussions about the Australian tax records data and for supplying additional tax tables. We also thank Nicolas Hérault and Stephen Jenkins for their comments. This paper uses unit-record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Project was initiated and is funded by the Australian Government Department of Social Services (DSS) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views reported in this paper are those of the authors and should not be attributed to either DSS or the Melbourne Institute.

Melbourne Institute of Applied Economic and Social Research

The University of Melbourne

Victoria 3010 Australia

***Telephone* (03) 8344 2100**

***Fax* (03) 8344 2111**

***Email* melb-inst@unimelb.edu.au**

***WWW Address* http://www.melbourneinstitute.com**

Abstract

We systematically compare differences in recent Australian income inequality estimates derived from tax records and survey data. We use customised tax tables provided by the Australian Taxation Office (ATO) to more precisely measure Australian top incomes as conceptualised in the international top incomes literature. We then document the inability of Household, Income and Labour Dynamics in Australia (HILDA) survey data, as well as the ATO 1% sample tax files, to replicate our top income share values. In contrast, accuracy is substantially improved using the ATO 2% sample introduced in 2011-12. Using HILDA data with enhanced income values from the 2% sample we better track top incomes. More importantly, we are able to use alternative income concepts and income-sharing units to provide more consistent cross-national comparisons than are currently found in the top incomes literature. Furthermore, we improve standard summary measure estimates of inequality found in the traditional survey-based Australian literature. Our findings show great potential value in linking tax records and survey data, but also show that this value is severely limited when using the 1% tax sample available for 2003-04 to 2010-11. Whereas the 2% sample substantially improves top incomes estimates, data without censored and perturbed income variables are needed and could be provided without unduly threatening tax filer confidentiality.

JEL classification: D31, C81

Keywords: Inequality, summary income inequality measures, top income share measures, Australian tax records data, HILDA Survey data

I. Introduction

The success of popular books by research-centred economists on the growth in income inequality, both internationally (e.g., Atkinson, 2015; Piketty, 2014) and in Australia (Leigh, 2013), suggests a heightened interest in this issue by the general public. Yet in Australia, the two main sources of data on the level and growth in income inequality—household surveys and tax return data—appear to find divergent trends since 2000, when consistent yearly household survey data became available.[1]

Empirical researchers are ultimately limited by available data in their ability to provide answers to policy questions. Household survey data and tax record data both have their advantages and disadvantages for those interested in estimating levels and trends in income inequality in Australia and comparing these outcomes with those in other countries. Although household survey data in principal capture a random sample of the entire population together with its income distribution, non-coverage or under-reporting of respondents in top income households are likely to understate levels and miss increases in inequality to the degree they occur at the very top of the income distribution.[2]

In contrast, personal income tax record data in Australia much more fully capture the distribution of income at the top and are available via published tax tables back to the 1920s. However, these advantages come at the cost of being restricted to using the definition of income and income-sharing unit dictated by Australian income taxation law, and to producing summary measures of inequality (the share of *tax-concept gross* income [i.e., taxable income] of top income groups) that do not fully describe the extent and nature of inequality.

In this paper we provide a systematic comparison of recent estimates of income inequality in Australia derived from both tax record data and household survey data. More importantly, we then show that, by combining these two types of data into a single data file, we can measure income and its distribution more accurately than it is measured in both the top incomes tax record-based literature and the more general income inequality literature based on household surveys.

We first use tax tables created for this paper by the Australian Taxation Office (ATO) from their complete Australian unit-record tax data to more precisely measure top incomes in

---

[1] Burkhauser, Hérault, Jenkins, and Wilkins (2016) find this is also the case in the United Kingdom for the years when both household and tax return data are available.

[2] Under-reporting of incomes at the lower end of the income distribution is also an issue for household surveys. We do not attempt to address this issue.

Australia as conceptualised in the international top incomes tax record-based literature.[3] We then document the inability of the HILDA household survey data,[4] as well as the 1% sample of Australian unit-record tax data now made available by the ATO, to replicate our top income share values from the complete unit-record tax data.[5] In contrast, we find accuracy is substantially improved using the 2% sample of these unit-record tax data introduced by the ATO in 2011-12.

When we then use the HILDA Survey unit-record data with adjustments to the top part of the income distribution based on the 2% sample, we are better able to track our top income values from the complete unit-record data for 2011-12 and 2012-13. But, more importantly, we are able to use alternative income concepts and sharing units to undertake more consistent cross-national comparisons than are currently found in the top incomes literature (and we provide an Australia–United States illustration). We are also able to use these enhanced HILDA data to estimate alternative top income share series that consider alternative income concepts (e.g., all sources of income rather than just taxable income, after-tax income) and sharing units (e.g., household unit, equivalised income of persons). Finally, we are able to improve standard summary measure estimates of inequality found in the traditional survey-based Australian literature.

(i). Tax-based vs. Survey-based Income Inequality Findings

The vast majority of researchers use household survey data to monitor income levels and inequality trends in OECD countries. Wilkins (2014a) evaluates the evidence on income inequality in Australia in the 2000s using survey data and standard Gini coefficients and finds only modest changes in inequality.

In their seminal work on top income shares in Australia, Atkinson and Leigh (2007) use published tabulations from tax returns in conjunction with National Accounts data to examine the period from 1921 to 2003. They show substantial increases in the share of

[3] See Atkinson, Piketty and Saez (2011) for a review of this literature. Output from all of the 32 country studies to date is now available in the World Wealth and Income Database (WID), updated each year, with top income share estimates and associated metadata freely downloadable (Alvaredo, Atkinson, Piketty, & Saez, 2015). See: Atkinson and Leigh (2007) and Burkhauser, Hahn, and Wilkins (2015) for Australian examples using previously published ATO tax tables to estimate top incomes.

[4] The Household, Income and Labour Dynamics in Australia (HILDA) Survey is a large household panel survey commonly used to analyse the Australian income distribution. HILDA is administered by the Melbourne Institute, which produces income distribution statistics that are published annually. The most recent survey year available at the time of writing is 2014. See Wilkins (2015).

[5] The Australian Bureau of Statistics also conducts a cross-sectional income distribution study. However, data are only available for every other year and there have been a number of changes to income concept and survey methods over the last decade that make it difficult to determine income distribution trends with these data (Wilkins, 2014a). The HILDA Survey is therefore our preferred household survey data source.

income held by the top 1% in the 2000s.

The need to reconcile and combine data used in these two literatures is illustrated in Figure 1. Burkhauser, Hahn and Wilkins (2015), using the same ATO-published tax table data used by Atkinson and Leigh (2007) and updated by Leigh, reproduce Atkinson and Leigh's estimates of the share of income held by the top 1% (AL series) and then produce alternative estimates for the tax-years 1980-81 to 2013-14 using somewhat different methods (BHW1 series). In Figure 1 we present income inequality trends using these two alternative measures of top income shares, using 2011-12 as the year to which we compare all other years. More traditionally, Wilkins (2014a) uses income data from the HILDA Survey to produce a summary measure—the Gini coefficient—to represent Australian income inequality for the years 2000-01 to 2010-11. We update his series to 2013-14 (the most recent year of HILDA currently available) and translate those findings into our third trend series (Gini series) in Figure 1.

Both measures of top income shares show dramatic growth from 1980-81 to 2000-01, followed by somewhat reduced growth in the years thereafter. But even the reduced post-2000 growth is greater than that found in the household survey-based Gini coefficient trends up to 2011-12. This continues to be the case from 2011-12 to 2013-14, when the Gini values are virtually unchanged while top income shares grow by 8 per cent and 10 per cent. Below we attempt to reconcile these trends.

No study to date has systematically compared Australian inequality estimates derived from survey data and tax record data. To undertake such comparisons properly requires use of the same definitions in both sources. Otherwise, like is not being compared with like. Unfortunately, the HILDA and WID series use different concepts of income and income-sharing unit. The principal income concept in HILDA is household disposable income, which is the sum of the market income (wages, interest, dividends, etc.) and government cash transfers of every individual in the household after the deduction of their personal income taxes and the addition of their tax credits. Importantly, capital gains, either realised or accrued, are not included as part of market income. Most survey-based research uses such an income concept, additionally adjusting income to take into consideration the number of persons in the household (household size-adjusted or *equivalised household disposable* income of persons), which is recommended by the OECD (Förster and d'Ercole, 2012) and the Canberra Group (United Nations, 2011).

In contrast, the concepts of income and sharing unit in the WID series for Australia

reflect the administration of the Australian personal income tax system, which taxes people at the individual level. Hence the Australian income data the WID uses come from the gross income for tax purposes of individuals, where the *tax-concept gross* income is that part of market income and government cash transfers that is subject to taxation, before deductions.[6] In this paper, we draw on Australian tax-records data and HILDA household survey data to investigate how the income concept, the sharing unit, and the inequality measure one uses affect one's view of Australian income inequality and its trends.

The WID data for Australia come from tax tables the ATO creates from the complete population of Australian tax records. The ATO does not make available to external researchers the unit-record data it uses to produce these tables. However, similar to the approach in the United Kingdom, since the 2003-04 tax-year, the ATO has produced a limited yearly cross-sectional sample file that it makes available to external researchers. This was a 1% representative sample of all tax filers up until 2010-11; the sample size was increased to 2% beginning in 2011-12. These unit-record tax files contain relatively few variables. Furthermore, they are substantially altered to protect confidentiality of tax filers, and it is not possible to link individuals belonging to the same family or household.[7]

Consequently, we cannot use the tax data to construct measures based on the definitions of income and income-sharing unit in HILDA or other survey data sets, such as the Australian Bureau of Statistics' two-yearly cross-sectional income distribution surveys. However, with the HILDA data, which contain detailed information on income of each household member by source, we are able to cross-walk from the Australian WID concepts of income and income-sharing unit to those used in the survey-based literature. More importantly, we can use the HILDA unit-record data with adjustments to the top part of the income distribution based on the 1% and 2% samples of the complete Australian unit-record tax data (which we will refer to as AUTAX) to explore the consequences of using different definitions looking at both top income shares and the more general summary inequality measures used in the general survey-based income inequality literature.

Our analysis begins by focusing on fiscal year 2000-01, because this is the first fiscal year of HILDA data and it is the first year we are able to produce improved estimates of top incomes based on the tax tables specially created for this study by the ATO.

---

[6] Atkinson and Leigh (2007) call this *actual* income in their article. More importantly, in their measure of *tax-concept gross* income they include taxable realised capital gains.

[7] In contrast, in the United States, government researchers can gain access to federal unit-record income tax files for approved research purposes and they are able to collaborate with outside researchers on such projects.

In Section II, we discuss the nature of Australian tax records data, the methods top income researchers used to calculate top income shares with these data, and the improvements to tax-based estimates we provide for tax years 2000-01 to 2012-13 with the values derived for us by the ATO.

In Section III, we focus on the period since fiscal year 2003-04, because this is the first year that the ATO began providing its 1% sample of Australian tax records to researchers. We find that neither HILDA nor estimates from the ATO 1% sample accurately replicate the share of *tax-concept gross* income held by the top 1% of tax units based on the tax tables the ATO provided us. But accuracy is substantially improved when using the 2% ATO sample file, which was introduced in fiscal year 2011-12.

In Section IV, we introduce a top income adjustment based on this 2% ATO sample data that substantially improves the HILDA Survey's ability to capture top incomes for 2011-12 and 2012-13. We call this enhanced HILDA data HILDA-AUTAX.

In Sections V and VI we exploit the flexibility of our HILDA-AUTAX data to examine alternative concepts of income and income-sharing unit—including in a cross-national comparison between Australia and the United States. By doing so we show its potential added value in the Australian tax record-based top income and more general survey-based inequality literatures.

In Section VII we summarise our findings and conclude that, while the 2% sample substantially improves top income estimates, data without censored and perturbed income variables are needed and could be provided without unduly threatening tax filer confidentiality.

II. Australian tax records data and the calculation of top income shares

Conceptually, it is straightforward to use tax records data to measure the share of income going to top income groups such as the top 1%. One simply tallies up the incomes of individuals in the top income group and divides this by the total income of all individuals. However, operationally the available tax data in Australia do not allow such a simple calculation. There are three issues that must be considered. First, income that is not taxable does not appear in the tax data; rather, what is available is a *tax-concept gross* income measure. This measure comprises taxable market income and government cash transfers, and excludes non-taxable income such as family cash transfers, disability benefits and private transfers. Second, not all people receiving income are required to lodge a tax return, even if that income is taxable. As a result, the sum of *tax-concept gross* income (the income control

total used in the denominator) constructed from tax records is less than all *tax-concept gross* income received by Australian residents. Third, prior to 2003-04, researchers only have access to published tax tables that report the number of people and the sum of *tax-concept gross* income in each of a number of income ranges, so the *tax-concept gross* income of a top income group cannot be accurately identified (barring an unlikely coincidence between the number of people in the top income group and the boundary of two income ranges).

Nonetheless, following approaches adopted by other country studies, Atkinson and Leigh (2007) produce estimates of Australian top *tax-concept gross* income shares over the period 1921 to 2003. Their study, and subsequent updates, formed the Australian contribution to the WID up to 2010-11. (See Figure 1, which includes a slightly corrected version of the Atkinson and Leigh series from 1980-81 onward.) The nature of the data they use varies somewhat over the period examined, but from 1959-60 onward it is relatively consistent. For this period, Atkinson and Leigh use the published taxation statistics in conjunction with population data to calculate *tax-concept gross* income of top income groups (the numerator) and they use the National Accounts to calculate the sum of *tax-concept gross* income (the denominator).

To identify the number of individuals or 'tax units' in the top income groups for the numerator Atkinson and Leigh use ABS population data. Australia has always had individual (as opposed to family) taxation, so the total number of tax units in the population is assumed by Atkinson and Leigh to equal the number of individuals in the population aged 15 or over. Thus, for example, the number of tax units in the top 1% is simply 1% of the population aged 15 and over.[8]

However, because the tax tables report tax income data by category bin (income range), an assumption is required about the distribution of income within the income bin containing the bottom value of the top income group. That is, for the income bin that contains both members of the top income group and individuals just below the top income group, Atkinson and Leigh must estimate how much of the income in that bin goes to members of the top income group. To do so, they adopt the 'mean split histogram' method (Atkinson,

[8] In principle, all persons with taxable income in excess of a minimum threshold must file a tax return, including children of any age. Thus, strictly speaking, an analysis of the distribution of tax unit income across tax units is simply an analysis of the distribution of personal (tax) income across the entire population, inclusive of children. In practice, very few individuals under 15 years of age lodge tax returns. For example, data for 2012-13 show 0.1% of all tax filers are under 15 years of age. Atkinson and Leigh (2007) therefore define the population of tax units to be all individuals aged 15 and over, an approach that we also adopt.

2005).

Recognising that the tax data do not capture all taxable income needed for the denominator, to calculate the sum of *tax-concept gross* income of all Australian residents (known as the income control total), Atkinson and Leigh use the household income account of the National Accounts. (See Appendix Table A1 for itemisation of the National Accounts components included by Atkinson and Leigh in their control total for income.) To be consistent with the *tax-concept gross* income numerator, the income control total that forms the denominator should only include that portion of household income from the National Accounts that is in the tax data. However, because these accounts include all cash income received by persons living in Australia, some of this income, such as family cash transfers, is included even though such income is not taxed. Moreover, as explained further below, the National Accounts exclude some components, including realised taxable capital gains and dividend imputation credits that are included in Atkinson and Leigh's *tax-concept gross* income numerator.

Atkinson and Leigh acknowledge several of the limitations of their approach to estimating top income shares discussed above; Burkhauser et al. (2015) address two of the more important ones.[9] Major tax reforms in Australia in the mid-1980s substantially impacted the personal income tax base. Two reforms were particularly important. From 1 July 1987, a system of full dividend imputation credits was introduced, allowing dividend recipients to claim tax credits for the imputed company tax paid on those dividends. These credits entered the personal income tax base, resulting in a substantial increase in measured incomes in the tax tables from the 1987-88 tax year, which disproportionately went to top income groups. However, while these credits are included in the tax tables, they are not part of a *tax-concept gross* income concept and should be excluded from this definition of income.

The second major change, of more serious consequence for ascertaining trends in top income shares found by Atkinson and Leigh (2007), is that most realised capital gains became taxable from 1 July 1986, but only on assets acquired after 19 September 1985. This resulted in a steady rise in the share of realised capital gains entering the tax base from 1986, leading to spurious measured increases in top income shares from 1986 to at least the mid-2000s. In addition to these issues, realised capital gains and dividend imputation credits do not enter the

[9] In Burkhauser et al. (2015), we also address a problem with the annual updates undertaken by Leigh in each year after 2003, whereby changes to income concepts and definitions in the National Accounts, as well as revisions to estimates subsequent to first release, led to spurious measured changes in Leigh's top income shares series.

National Accounts data that is used in Atkinson and Leigh's control total for income, leading to a further overestimation of top income shares. To capture a more consistent measure of *tax-concept gross* income, therefore, it is important to exclude all capital gains, including realised taxable capital gains, from the definition of income. That can best be done by producing a *tax-concept gross* income series for Australia that explicitly states that it does not include realised capital gains in any years.[10]

In Burkhauser et al. (2015), we address these problems by producing top income share estimates for the period 1970-71 to 2010-11 excluding dividend imputation credits and taxable realised capital gains. By doing so, we produce a *tax-concept gross* income series that is more consistently measured over time as well as more internally consistent (given that realised capital gains and dividend imputation credits do not enter the income control total).

A problem for the construction of this series (which we here label BHW1) is that the tax tables do not contain tabulations of income excluding dividend imputation credits and realised capital gains. Rather, in addition to *tax-concept gross* income, the tables report total imputation credits and total realised capital gains for each taxable income bin range. We therefore simply subtract total imputation credits and total capital gains from total income for each taxable income bin range, effectively assuming that these two income components are distributed across individuals in the same way as other income components. In practice, this is not the case. We draw on unit-record tax data and show that this leads to a small degree of underestimation of top income shares.[11]

Motivated by the limitations of the published tax tables, in 2015 we requested and received customised tables from the ATO based on the full population of tax records which provided total incomes of various top income groups excluding all imputation credits and realised capital gains. The ATO was only able to produce these data back to 2000-01. (In Appendix Table A2, we more fully define our improved numerator.) With it we are able to produce more precise estimates of top *tax-concept gross* income shares for the period 2000-01 to 2012-13. This new series (which we here label BHW2) addresses all limitations of the published tax tables to produce a *tax-concept gross* income measure—the classification of *tax-concept gross* income by taxable income ranges, the need for distributional assumptions

[10] In the WID series, most countries provide only a total gross income series without realised capital gains. A few additionally provide a total gross income series with capital gains.
[11] BHW1 has since 2011-12 been the Australian contribution to the WID, but at the time of this writing the WID website had not yet updated these estimates for 2011-12, 2012-13 and 2013-14, or made revisions for earlier years.

within income ranges, and the inability to accurately remove dividend imputation credits and realised capital gains.

As part of our reconsideration of our top income share values, we also reviewed the calculation of the control total for income. Beginning in tax year 1989-90 it is possible to exclude employer social contributions from the National Accounts data used in the denominator. These comprise contributions to employee retirement savings accounts and to workers' compensation insurance premiums, which are not taxable and so are not included as part of *tax-concept gross* income in the numerator. Hence their inclusion in the control total artificially lowers top *tax-concept gross* income share estimates. In our new preferred top income share estimates we subtract these employer social contributions from the control total for income. For all these reasons our *tax-concept gross* income series now provides the most consistent and accurate measure of the top income shares for 2000-01 to 2012-13.[12]

Figure 2 shows the income share of the top 1% over the 2000-01 to 2012-13 period using the methods developed for the Atkinson and Leigh (AL) series, our original BHW1 series using the standard ATO tax tables and our BHW2 series that uses our customised tax tables. (Appendix Table B1 repeats these values for the top 1% series and provides the income share estimates for the top 0.1%, 0.5%, 5% and 10%). The AL series includes taxable realised capital gains and dividend imputation credits, and so, all else equal, should be higher than the other two series. Indeed, most of the difference between AL and BHW1 is due to these two income components. However, moving from BHW1 to BHW2 results in a substantial shift up in the level of top income shares. The main source of the level difference between BHW1 and BHW2 is the exclusion of employer social contributions from the BHW2 income control total, although the mean split histogram method of interpolating incomes used in BHW1 also generates substantial downward bias in estimates. We report the relative importance of our various adjustments in Appendix Tables A3 to A7, which decompose the sources of the differences between BHW1 and BHW2.

As BHW2 is the most accurate tax-based series for measuring *tax-concept gross* income (excluding taxable realised capital gains) and is the most comparable to survey data (which in particular do not capture capital gains), and because it is available for all but the last year of the HILDA Survey data (2013-14), this is the series against which we benchmark the

[12] Nonetheless, our revised income control total contains income components that are not part of *tax-concept gross* income, including non-taxable personal income and employee non-cash benefits. We have not developed methods for removing these components, which would require data not recorded in the National Accounts. Consequently, our BHW2 series will still somewhat underestimate top *tax-concept gross* income shares.

HILDA Survey for purposes of integration with the WID literature. In 2011-12, the top 1% of Australian tax units received 8.71 per cent of *tax-concept gross* income in Australia, based on our BHW2 measure.

III. Using the HILDA household survey data and the ATO sample data to measure top *tax-concept gross* income shares

Our household survey data come from Waves 1 to 14 of the HILDA Survey, conducted from 2001 to 2014. The HILDA Survey is a nationally representative household panel study conducted by annual in-person interview with all household members aged 15 years and over. The study is designed to be infinitely lived, following all original sample members and their descendants. It commenced in 2001 with 13,969 respondents in 7,682 households, and in 2011 a general sample top-up added 4,008 individuals living in 2,000 households. For more information about the study, including rates of attrition, see Summerfield et al. (2015).

Each year, the HILDA Survey collects detailed information on annual income for the period July 1 to June 30, which coincides with the tax year and hence the tax records-based income data. The unit-record data contain values for each of 29 income components for every sample member aged 15 years and over. Because all household members aged 15 and over are interviewed, the data allow us to replicate the tax-based measures of top income shares, and to examine how estimates are affected by changes in income concept and sharing unit. Moreover, we can examine the full distribution of income, whereas we cannot do so with the tax data, which do not capture incomes of individuals who do not file tax returns. Published tables by the ATO show that, over the 2001 to 2013 period, between 25 and 30 per cent of tax units did not file tax returns in any given year. Although some of these tax units will have no income, many will have income that is either non-taxable or below the tax-free threshold.

We construct a *tax-concept gross* income variable for individual tax units[13] from the HILDA Survey data by first subtracting non-taxable income components from the personal gross income variable provided in the HILDA unit-record data. The main non-taxable components comprise the Disability Support Pension, Carer Allowance, Family Tax Benefit (Parts A and B), regular private transfers, compensation payments and the 'salary sacrificed' component of wages and salaries. We then add in taxable irregular income components—

[13] Every person aged 15 and older is treated as a separate tax unit. All income components in HILDA are asked and reported at the person level; household income is an aggregate of these components.

redundancy and severance payments, and lump sum workers' compensation—that are not included in the HILDA personal gross income variable. (The HILDA income variable is constructed to be consistent with 'Canberra Group' standards for measurement of total income in household surveys. These guidelines exclude large irregular cash receipts as well as capital gains, whether realised or not. See United Nations [2011] for details.)

Sample sizes are a constraint in the HILDA data. For example, depending on the year, there are between 141 and 195 observations in the top 1% of individual tax units. We therefore do not attempt to examine top income groups smaller than the top 0.5%.

In Figure 3 we report top 1% income share series based on three alternative sources of data for the years 2003-04 to 2012-13—the only years for which we have information from each. (The values for all series in Figure 3 can be found in Appendix Table B2.) The first is our preferred BHW2 top 1% income series introduced in Figure 2, whose numerator comes from the customised ATO tax table of *tax-concept gross* income for the top 1% of tax units. The control totals are the WID's National Accounts control totals minus employer social contributions.

The second series comes from our HILDA Survey estimates of the share of income held by the top 1% of tax units. In this series the numerator contains *tax-concept gross* income data from HILDA, but we use the same National Accounts income control totals used in the BHW2 series. Because of its use of survey data we expect this *tax-concept gross* income HILDA series to underestimate the BHW2 series.

The third series comes from a unit-record sample file of tax filers that the ATO has produced for each tax year over the period 2003-04 to 2013-14. We call this sample file the AUTAX file, and will use it later together with the HILDA Survey data to enhance the value of the HILDA data in capturing the top part of the income distribution. But before doing so, it is important to test the ability of this sample to capture top incomes. As noted, this was a 1% sample of all tax filers up until 2010-11 and a 2% sample since 2011-12. However, in all tax years, these samples were subject to top-coding and bottom-coding, as well as perturbations to protect confidentiality of tax filers. Hence it is not certain that the AUTAX files can accurately replicate the BHW2 top income shares, which are derived from the unaltered income data of the full population of tax filers.

Not surprisingly, in Figure 3, the HILDA-based series in all years but one captures a smaller share of the top 1% than is found in the BHW2, and the HILDA series is more volatile over time because its sample size is smaller. Nonetheless, the HILDA values are

within range of the BHW2 values. More surprising is that the 1% AUTAX series does a poorer job of replicating the BHW2 values than does the HILDA series in several years. In contrast, the 2% AUTAX sample does a much better job than HILDA and comes very close to replicating the BHW2 series for the two years that all three series are available.

In Figure 4 we compare BHW2, HILDA and AUTAX results for other top income share groups. (The values for all series in Figure 4 can be found in Appendix Table B2.) For the top 0.5% to 1%, 1% to 5% and 5% to 10% of tax filers, both the 1% and 2% AUTAX samples come close to replicating the BHW2 values. It is only within the top 0.5% income group that the 1% AUTAX fails to do so. Hence it is the inclusion of the top 0.5% within the top 1.0% bin that is driving the inaccuracy of 1% AUTAX top 1% values.

Undoubtedly part of the reason for the failure of the AUTAX sample to replicate values based on the full tax record file is related to sample size, and the increase in sample size explains a large share of the improvement in the ability of the AUTAX data to replicate the results from the complete tax record file. But topcoding is also likely to have caused some of the undercount of top shares. Figure A1 in the appendix shows, for various top income groups, the proportion of persons in the AUTAX file with at least one topcoded income source. Because the unit-record data do not directly provide information on who is topcoded, the estimates of the share of those in each top income group who are likely to have been topcoded is a lower bound. (See the discussion in Appendix A for details of how this figure was constructed.) But not surprisingly, it is among tax units in the top 0.5% income group that we find the highest percentage topcoded. While this has fallen over time it is still substantial. Topcoding in the 2% to 10% top income groups is very low.

Based on this evidence, we concluded that the AUTAX 1% sample file is unsuitable for examining top income shares. In contrast, the AUTAX 2% file is much closer to the BHW2 series. We therefore focus our subsequent efforts on improving the ability of the HILDA data to better capture the top part of the income distribution beginning in 2011-12 when the 2% sample becomes available. Doing so will provide the reader with a sense of the added value that access to the full ATO data file would provide to both the top incomes literature and the survey-based income inequality literature.

IV. Adjusting the household survey data using unit-record tax data

To improve the capture of top incomes in HILDA for 2011-12 through to 2013-14, we adjust the top 1% of incomes using the 2% AUTAX unit-record tax data. To do this, we simply replace the value for *tax-concept gross* income for each person in the top 1% of

HILDA with the value for the person at the same rank in the distribution of *tax-concept gross* income in the AUTAX unit-record tax data.[14]

We show the added value of our enhanced HILDA-AUTAX data in capturing the share of *tax-concept gross* income held by the top 1% of tax units in Figure 5. (The values for all series in Figure 5 can be found in Appendix Table B3.) Here we focus on income years 2011-12 and 2012-13 and repeat our BHW2, AUTAX and HILDA top 1% income share series from Figure 4 using the same (BHW2) National Accounts income control.

To these series we now add the top 1% values based on our enhanced HILDA-AUTAX values using the BHW2 National Accounts income control. By construction, we dramatically improve our capture rate compared to the HILDA values and now almost exactly match the BHW2 and AUTAX values. Our final series HILDA-AUTAX (HILDA-AUTAX income controls) uses the same income values used in HILDA-AUTAX but we now exclusively use HILDA-AUTAX-based controls for our income control total denominator. That is, the denominator is equal to the weighted sum of the *tax-concept gross* income in HILDA-AUTAX, where the weights are the population weights from the HILDA data. Using these weights we overestimate the share of *tax-concept gross* income held by the top 1% because our HILDA-AUTAX-based income controls understate the BHW2 National Accounts-based income controls. (See Appendix Figure B1 for a comparison of BHW2 National Accounts income controls and the HILDA-AUTAX income controls.) Using the HILDA-AUTAX income control denominator values has the disadvantage of adding error to top income shares estimates, but this disadvantage is offset by the ability to examine top income shares for alternative income concepts and sharing units. By contrast, it is very difficult, and often impossible, to produce income controls for alternative income concepts and sharing units using National Accounts data because of the aggregate nature of the data.

V. Using our adjusted household survey data for cross-national comparisons of top incomes in Australia and the United States

Panel A of Figure 6 reports the World Income and Wealth Database (WID) top 1%

[14] Specifically, we use the population weights provided with the HILDA and AUTAX unit-record data to restrict each sample to the top 1% of persons in the distribution of (*tax-concept gross*) income. Each of these sub-samples is expanded by replicating each observation according to its population weight, such that the two expanded sub-samples have the same number of observations. We then replace the income value of each HILDA observation with the income value of the observation in AUTAX at the same rank in the distribution. The HILDA expanded sub-sample is then collapsed to its original size, with each observation in the collapsed sample assigned the mean income of its collapsed observations.

tax-concept gross income share levels for the United States together with our various measures of the top 1% *tax-concept gross* income share for Australia in 2011-12 and 2012-13 (from Figures 2 and 5). It is this U.S. top income data series that is used to compare levels and trends in the share of income held by the top 1% of tax units in the United States to the share of income held by the top 1% of tax units in Australia, as discussed in Atkinson, Piketty and Saez (2011). Table 1 describes the income concept, sharing unit, and data sources for each of the 12 top income share measures used in Figure 6.

The share of *tax-concept gross* income controlled by the top 1% in the United States (labelled Series 1 in Table 1) is substantially higher (17.47 per cent in 2011-2012) than the four Australian estimates of *tax-concept gross* income share. BHW1 (Series 2 in Table 1) produces an estimated top 1% income share of 7.68 per cent in 2011-12. Our new BHW2 series (Series 3) produces an estimated top 1% income share of 8.71 per cent in 2011-12. Our HILDA-AUTAX series with the BHW2 National Accounts-based control total (Series 4: *Individual tax-concept gross income (NA)*) by construction is approximately equal to BHW2. Moving to the HILDA-AUTAX income controls for the HILDA-AUTAX series (Series 5: *Individual tax-concept gross income*) increases the top 1% share—to 9.43 per cent in 2011-12—but it is still well below that of the United States.

As discussed above, using the HILDA-AUTAX income controls introduces some error into our estimates of top income shares, but it allows us to adjust our income concept and sharing unit to make like-on-like comparisons. Below we use the HILDA-AUTAX data with HILDA-AUTAX income controls to make comparisons with the United States.

The U.S. estimates of top income shares are constructed quite differently from the Australian estimates, even when taxable realised capital gains are excluded. First, in Australia, *tax-concept gross* income includes not only market income but also government transfers because most government transfers are taxable in Australia, whereas in the United States *tax-concept gross* income includes only market income because very few government transfers are taxed. Moreover, to maintain consistency in their measure over time, Piketty and Saez do not include social security benefits, which were not taxed until 1984 and then only on the part of this benefit above a constant (nominal) threshold.

Second, the Australian tax unit (and hence assumed sharing unit) is the individual, with all individuals aged 15 and over treated as their own tax unit. In the United States, the tax unit is the family, whereby married couples and their co-resident children under the age of 20 are treated as part of the same tax unit. Thus, in Australia, top income shares refer to the

personal income share of the top *x*% of individuals aged 15 and over, whereas in the United States top income shares refer to the family income of the top *x*% of families.

Third, the control total for income is derived differently in the two countries. In Australia, the control total is derived from the household account in the National Accounts. (See Appendix Table A1 for itemisation of the components). In the United States, the control total for income is equal to total market income as captured by tax records, plus 20 per cent of the mean income of tax filers multiplied by the number of non-filing tax units (Piketty and Saez, 2003). In the United States, approximately 7 per cent of tax units are non-filers. In Australia, by contrast, the figure is 25 to 30 per cent, although if a U.S. tax unit were to be adopted, this figure would likely be considerably lower.

To crosswalk from Australian to U.S. definitions, we construct a U.S.-style income concept and tax unit measure using our augmented HILDA-AUTAX Survey data with HILDA-AUTAX income controls. The income variable is straightforward to replicate, being simply market income as captured by the HILDA-AUTAX data. (Here we define market income as all income other than government transfers. In order of importance, it includes wages and salaries, investment income, unincorporated business income, private pensions, and miscellaneous other payments.)

Replicating the U.S. tax unit is less straightforward, but here we take the approach of Piketty and Saez (2003). Following this approach, tax units comprise singles aged 20 and over together with any co-resident children under the age of 20, and married couples together with any co-resident children aged under age 20. If an individual has never been married, is under age 20 and is not living with a parent, that individual is assigned to the tax unit of the household's primary family or, if there is no primary family, to the oldest adult in the household. If the household contains no adults aged 20 and over, the individual is treated as his own tax unit. Finally, if an unmarried co-resident couple has children, the children are assumed to be part of the father's tax unit.

We do not attempt to replicate the U.S. income control total using Australian data, other than to adjust for the restriction to market income (by excluding government benefits). Although it is possible to apply the same criteria—total market income captured by tax records plus 20 per cent of the average income of filers multiplied by the number of non-filers—this is unlikely to be appropriate in the Australian context given differences in rates of filing and likely differences in the relationship between incomes of filers and non-filers.

Panel B of Figure 6 shows the potential added value of using the flexibility of

household survey data to create like-on-like comparisons for cross national comparative purposes. It shows the importance of alternative choices of income and sharing unit, as well as the importance of the introduction of equivalence scales that take into consideration the number of persons in a sharing unit. We make like-on-like comparisons using the three years for which we have AUTAX values from the 2% sample: 2011-12, 2012-13 and 2013-14. We use these same data for the other panels in Figure 6 that track the share of income of other top income groups. In all cases, the estimate of the control total for income (the denominator) comes from HILDA-AUTAX, but with the value of the control total depending on the income concept and sharing unit.

The results show the sensitivity of top income shares to the components of income measured and the sharing unit used. We begin by once more looking at the U.S. and Australian *tax-concept gross* income series used for cross-national comparisons in the tax-based top income literature. In 2011-12 the top 1% of tax units (the family) in the United States held 17.47 per cent of U.S. *tax-concept gross* income (which only includes market income) while the top 1% of tax units (single persons) in Australia (Series 5) holds 9.43 per cent of *tax-concept gross* income—a measure that includes many government transfers. (The values for all series in Panels B to E of Figure 6 can be found in Appendix Table B4.) But when we use the same market income measure for Australia (Series 6: *Individual tax-concept market* income*)*, rather than the measure that includes government transfers (Series 5), we find that the share of income held by the top 1% of single persons rises to 9.97 per cent of *tax-concept market* income. Clearly the inclusion of government transfers in Series 5 overstates the difference in top income share between these two countries. In contrast, if we keep the same Australian *tax-concept gross* income measure but use the U.S. family tax unit (Series 7: *Family tax-concept gross* income), the top 1% of family tax units in Australia only holds 8.41 per cent. Using a family tax unit clearly lowers the share of income held by top income tax units. When we use a U.S. market income of family tax unit measure (Series 8: *Family tax-concept market* income) to get our like-to-like comparison, these two countervailing forces offset each other and the top 1% of families now are found to hold 8.88 per cent of the market income—a value slightly lower than the one used in cross-national top income comparisons.

The remaining panels in Figure 6 repeat these same comparisons but do so for the top 0.5%, the top 1% to 5%, and the top 5% to 10%. In almost all cases, the further up the distribution the shares refer to, the more important is the shift from an individual to a family definition of tax unit and the less important is the shift from a *tax-concept gross* income

definition that includes a substantial amount of government transfers to one like that in the U.S., which includes only market income. This is not surprising, because at the 0.5% and above level almost all income is market income. But when we move down to the 1% to 5% top income group the shift to market income is more important than the shift to a family tax unit. Panel E of Figure 6 shows that, in the top 5% to 10% income group, not only are the estimates from the Australian *family tax-concept market* income series greater than the Australian *individual tax-concept gross* income series, but they are greater than the *U.S. family market* series.

We can now turn to issues related to income inequality comparisons that go far beyond those that can be considered with tax data alone. It is common practice in income distribution analyses using household survey data to examine the distribution of size-adjusted post-tax post-transfer household income (often referred to as *equivalised household disposable* income) across all individuals in the population, including children. We will now begin to cross-walk toward this concept.

Above in our discussion of Panels A and B of Figure 6, we showed that in 2011-12, 8.41 per cent of the *tax-concept gross* income in Australia was held by the top 1% of families. However, this excludes non-taxable income from consideration—a restriction caused by the inability of tax data to provide such information. But such income is received by families and should be included in any measure of their income, including the share of income held by the top 1%. As can be seen in Panel B of Figure 6, when non-taxable income is included, the share of *survey-concept gross* income (taxable and non-taxable income based on Canberra Group rules) held by the top 1% of families (Series 9: *Family survey-concept gross* income) drops to 7.76 per cent. This drop compared to Series 8 is in large part because not all government transfers are taxed in Australia, and a disproportionate amount of these transfers go to the bottom 99% of families.[15,16]

Whereas the use of a family sharing unit was important to control for the differences in top incomes between the United States and Australia, it is standard practice in the broader income distribution literature to use the household rather than the family as the sharing unit.

---

[15] In Australia, transfer payments consist of welfare benefits (income support payments), supplements paid to carers of people with severe disabilities, family benefits, foreign government pensions and other *ad hoc* government transfer payments.

[16] Because HILDA follows the Canberra Group (United Nations, 2011) conventions with respect to its definition of income, it does not collect information on capital gains, either taxed or non-taxed. HILDA does however collect information on other sources of irregular income. We included this income as taxable income in our total gross income measure, but subtract it here.

The HILDA data allow us to make this adjustment. We do this in Panel B of Figure 6—Series 10: *Household survey-concept gross* income—and find that the share of income held by the top 1% of households drops to 6.51 per cent in 2011-2012.

Another important adjustment in the more general income inequality literature is to control for the number of persons in the household. This effectively uses the individual as the unit of interest but adjusts each individual's household income for the number of people in the household. A commonly-used adjustment is to divide household income by the square root of the number of people in the household. Hence, each household member—those aged 15 and over as well as those under the age of 15—contributes one observation to the analysis of the distribution of size-adjusted household income of persons in a given year. We do this in Panel B of Figure 6—Series 11: *Equivalised household survey-concept gross income*—and find that the share of equivalised household income (taxable and non-taxable) held by the top 1% of all persons drops to 5.96 per cent in 2011-12.

Finally, in the more general income inequality literature, disposable or after-tax equivalised income is considered the more appropriate measure of income. We are able to estimate household after-tax income by applying the HILDA tax and benefit model to each individual's *survey-concept gross* income. See Wilkins (2014b) for details on the tax and benefit model.[17] We do this in Panel B of Figure 6—Series 12: *Equivalised household disposable* income—and find that the share of equivalised household disposable income (taxable and non-taxable) held by the top 1% of all persons drops to 4.80 per cent in 2011-12. The remaining panels in Figure 6 repeat these same comparisons, but do so for the top 0.5%, the top 1% to 5%, and the top 5% to 10%. In all cases, the same ordering occurs.

In such a short series it is difficult to say much about trends. However, it is clear that there has been an increase in the share of income held by the top 1% between 2011-12 and 2013-14 but no increases—and even decreases—for the top 1% to 5% and top 5% to 10% income groups. What these data do show is that this share is greatest when we focus on the before-tax market income of individuals but it declines as we broaden our measure of income to include government transfers, expand the sharing unit to the family or household, take into consideration the number of persons in the household and subtract income taxes.

[17] Note that although the HILDA-AUTAX data contain post-tax post-transfer income variables, for the top 1% who have their incomes adjusted, the post-tax post-transfer income needs to be recalculated. This is done by applying the HILDA tax and benefit model to these adjusted (gross) incomes.

VI. Using our adjusted household survey data to examine inequality measures used in the survey-based income inequality literature

A weakness of survey-based data in general and the HILDA data in particular, is that they understate the incomes of those in the top part of the income distribution. As discussed above, the 2% AUTAX data allow us to better capture the top part of the income distribution. Using these data we assigned values to the *tax-concept gross* income of the top 1% of tax units in HILDA to create our HILDA-AUTAX (HILDA-AUTAX income controls) data. We then showed (in Panel B of Figure 6) that doing so enabled us to create plausible measures of the share of *equivalised household disposable* income held by the top 1% of all Australians.

In Table 2 we examine how inequality measures for *equivalised household disposable* income are affected in 2011-12, 2012-13 and 2013-14 by the adjustment of the HILDA data using the 2% AUTAX sample. The first column examines the share of income held by the top 1% of all Australian residents, showing that our enhanced data increases the share of income held by the top 1% by 9.3 per cent, 7.6 per cent and 12.0 per cent in each of the respective years.

In the rest of the columns in Table 2 we examine overall inequality measures used in the survey-based income literature. While the Gini coefficient, mean log deviation and Theil (GE0 and GE1) indexes all increase using our HILDA-AUTAX (HILDA-AUTAX income controls) data, the increases are much smaller. This reflects the fact that, while the unadjusted HILDA data produce lower income estimates for the top 1%, this does not greatly affect estimates that focus on differences in either the lower or middle parts of the distribution. Consistent with this, we see little or no change in median income, or indeed in the P90/P50 ratio. The HILDA-AUTAX (HILDA-AUTAX income controls) data does, however, markedly increase the Theil GE 2 (half squared coefficient of variation), the most top-sensitive of our summary measures. This measure increases from 0.198 to 0.235 (18.6 per cent) in 2011-12, from 0.219 to 0.281 (29.6 per cent) in 2012-13 and from .229 to .317 (38.2 per cent) in 2013-2014. The AUTAX adjustment also increases the mean value of *household equivalised disposable* income in each year, since the increased *tax-concept gross* income values of the top 1% also increased the *equivalised household disposable* incomes of their households, but only slightly.

VII. Summary and Conclusion

No previous study of Australia has systematically compared inequality estimates derived from tax records and household survey data. Exploiting the flexibility provided by

access to unit-record household survey data, notably the ability to cross-walk between different income concepts and sharing units so that like is compared with like, we have provided such a comparison.

We have shown, using our HILDA-AUTAX data (with National Accounts control totals), that we can produce *tax-concept gross* income top income shares that, by construction, match those we produced using the customised tax tables provided to us by the ATO for this paper (BHW2). We then showed that our HILDA-AUTAX series (with HILDA-AUTAX control totals), although adding some error to these top income share values, allows us to employ different income concepts and sharing units. Doing so, we have shown how we can improve the like-on-like comparisons in cross-national studies in the top incomes literature.

When we adjust for the fact that *tax-concept gross* income in the United States is effectively market income whereas *tax-concept gross* income in Australia also includes most government transfers, we find that Australian top income shares rise using a like-on-like market income measure, but fall when we use a family tax unit measure that is comparable to the U.S. measure. These offsetting factors vary in their overall effect on comparisons of the share of income held by the top income groups. They make top income shares relatively smaller in Australia relative to the United States for those in the top 0.5% and 1% groups, but considerably larger in the top 1% to 5% and 5% to 10% groups. But more importantly, the HILDA-AUTAX series allows researchers to examine top incomes using a measure of income consistent with the established survey-based literature—namely, *equivalised household disposable* income. This cannot be done with tax records data alone.

We have also demonstrated the value of using HILDA-AUTAX data in the more general survey-based literature. We show that inequality levels using summary inequality measures, especially those that take into account income differences in the top part of the income distribution—the top 1% share and the GE2—are substantially increased when using our enhanced HILDA data set that better captures income at the top of the income distribution.[18]

Overall, we find that there is great potential value in linking administrative tax record data to HILDA data to both improve the comparability of Australian inequality measures to those of other countries in the top incomes literature, and to better capture the top part of the

[18] If the goal is simply to obtain improved estimates of overall inequality, an alternative approach is to calculate summary inequality measures separately for the very rich using tax data and for the rest of the population using survey data and then combine the summary measures. See, e.g., Atkinson (2007), Alvaredo (2011), and Jenkins (2015). Of course, this approach also requires like-for-like definitions in both sources.

income distribution in the traditional Australian income inequality literature based on household surveys.

But we also find that this added value from tax records data is limited by our inability to capture the very top part of the income distribution in the 1% samples the ATO provides to researchers for tax-years 2003-04 to 2010-11. Although we find that the 2% sample made available since then provides much more accurate estimates of the *tax-concept gross* income held by top income groups, it still contains censored and perturbed income variables that affect its ability to provide a complete picture of the top part of the income distribution.

Consequently, better tax data is required to fully and accurately adjust top incomes in Australian household survey data like HILDA. This could be achieved by the ATO partnering with the Australian Bureau of Statistics (ABS) to retrospectively produce a matched ABS-ATO Survey of Income and Housing (SIH), using methods similar to the ones we have used with the 2% ATO sample to produce tax records-based values for the top 1% of tax filers. Or the ATO could do so with HILDA. Something similar is already done in the United Kingdom for the Households Below Average Income (HBAI) Survey.[19]

Alternatively, the ATO could produce a public-release unit-record sample file with income data that is neither censored (top-coded or bottom-coded) nor perturbed. Even with a 2% (or larger) sample, this could be implemented without unduly threatening the confidentiality of tax filers, although it may require suppression of some demographic data. Indeed, simply removing regional information from the sample file would probably be sufficient to ensure confidentiality. This would greatly increase the ability of researchers to capture top income shares in Australia, as well as better estimate overall levels and trends in Australian income inequality.

[19] Burkhauser et al. (2016) show, however, that the method used by the UK Department of Work and Pensions to link tax record data to the HBAI does not result in the HBAI fully capturing the incomes of the top 1%. Burkhauser et al. then propose an alternative method similar in design to the method presented in this paper.

VIII. References

Alvaredo, F. (2011), 'A Note on the Relationship between Top Income Shares and the Gini Coefficient', *Economic Letters*, **110** (3), 274–277.

Alvaredo, F., Atkinson, A.B., Piketty, T., and Saez, E. (2015), The World Top Incomes Database. Available at: http://g-mond.parisschoolofeconomics.eu/topincomes/. Accessed 1 October 2015.

Atkinson, A.B. (2005), 'Top Incomes in the UK over the Twentieth Century', *Journal of the Royal Statistical Society*, *Series A*, **168** (2), 325–343.

Atkinson, A.B. (2007), 'Measuring Top Incomes: Methodological Issues', in Atkinson, A.B. and Piketty, T. (eds.), *Top Incomes over the Twentieth Century: A Contrast between Continental European and English-speaking Countries*. Oxford University Press, Oxford.

Atkinson, A.B. (2015), *Inequality: What Can Be Done?* Harvard University Press, Cambridge, MA.

Atkinson, A.B., and Leigh, A. (2007), 'The Distribution of Top Incomes in Australia', *Economic Record*, **83** (262), 247–261.

Atkinson, A.B., Piketty, T., and Saez, E. (2011), 'Top Incomes in the Long Run of History', *Journal of Economic Literature*, **49** (1), 3–71.

Burkhauser, R.V., Hahn, M., and Wilkins, R. (2015), 'Measuring Top Incomes Using Tax Record Data: A Cautionary Tale from Australia', *Journal of Economic Inequality*, **13** (2), 181–205, doi: 10.1007/s10888-014-9281-z.

Burkhauser, R.V., Hérault, N., Jenkins, S.P., and Wilkins, R. (2016), 'What Has Been Happening to UK Income Inequality since the Mid-1990s? Answers from Reconciled and Combined Household Survey and Tax Return Data', NBER Working Paper w21991.

Förster, M.F., and d'Ercole, M.M. (2012), 'The OECD Approach to Measuring Income Distribution and Poverty', in Besharov, D. and Couch, K. (eds.), *Counting the Poor – New Thinking About European Poverty Measures and Lessons for the United States.* Oxford University Press, Oxford; 27–58.

Jenkins, S.P. (2015), 'To What Extent Is Inequality Increasing? Some Data Issues', LIS Lecture presented at the 6th ECINEQ conference, Luxembourg, 13 July 2015. http://www.lisdatacenter.org/news-and-events/2015-summer-lecture/

Leigh, A. (2013), *Battlers and Billionaires: The Story of Inequality in Australia*. Redback Press, Collingwood VIC.

Piketty, T. (2014), *Capital in the Twenty-first Century*. Harvard University Press, Cambridge, MA.

Piketty, T., and Saez, E. (2003), 'Income Inequality in the United States, 1913–1998', *Quarterly Journal of Economics,* **118** (1), 1–39.

Summerfield, M., Freidin, S., Hahn, M., Li, N., Macalalad, N., Mundy, L., Watson, N., Wilkins, R., and Wooden, M. (2015), *HILDA User Manual – Release 14.* Melbourne Institute of Applied Economic and Social Research, Melbourne.

United Nations (2011), *Canberra Group Handbook on Household Income Statistics*, *Second Edition.* United Nations, New York and Geneva.

Wilkins, R. (2014a), 'Evaluating the Evidence on Income Inequality in Australia in the 2000s', *The Economic Record*, **90** (288), 63–89.

Wilkins, R. (2014b), 'Derived Income Variables in the HILDA Survey Data: The HILDA Survey 'Income Model'', HILDA Project Technical Paper No. 1/14. Melbourne Institute of Applied Economic and Social Research, Melbourne.

Wilkins, R. (2015), *The Household, Income and Labour Dynamics in Australia Survey: Selected Findings from Waves 1 to 12*. Report commissioned by the Australian Government Department of Social Services. Melbourne Institute of Applied Economic and Social Research, Melbourne.

IX. Tables and figures

Table 1: Descriptions of top income share measures used in Figure 6

Name of top income share measure	Income concept	Unit of analysis	Data source of numerator	Data source of denominator
1. US family market	Market[1]	Family / US tax unit	US tax data[8]	US tax data[12]
2. BHW1	Tax-concept gross[2]	Person, 15 years+	Public tax tables[9]	National Accounts (V1)[13]
3. BHW2	Tax-concept gross[3]	Person, 15 years+	Customised tax tables [10]	National Accounts (V2)[14]
4. Individual tax-concept gross (NA)	Tax-concept gross[4]	Person, 15 years+	HILDA-AUTAX[11]	National Accounts (V2)[14]
5. Individual tax-concept gross	Tax-concept gross[4]	Person, 15 years+	HILDA-AUTAX[11]	HILDA-AUTAX[11]
6. Individual market	Market[5]	Person, 15 years+	HILDA-AUTAX[11]	HILDA-AUTAX[11]
7. Family tax-concept gross	Tax-concept gross[4]	Family / US tax unit	HILDA-AUTAX[11]	HILDA-AUTAX[11]
8. Family market	Market[5]	Family / US tax unit	HILDA-AUTAX[11]	HILDA-AUTAX[11]
9. Family survey-concept gross	Survey-concept gross[6]	Family / US tax unit	HILDA-AUTAX[11]	HILDA-AUTAX[11]
10. Household survey-concept gross	Survey-concept gross[6]	Household	HILDA-AUTAX[11]	HILDA-AUTAX[11]
11. Equivalised household survey-concept gross	Equivalised survey-concept gross[6]	Person, any age	HILDA-AUTAX[11]	HILDA-AUTAX[11]
12. Equivalised household disposable	Equivalised disposable[7]	Person, any age	HILDA-AUTAX[11]	HILDA-AUTAX[11]

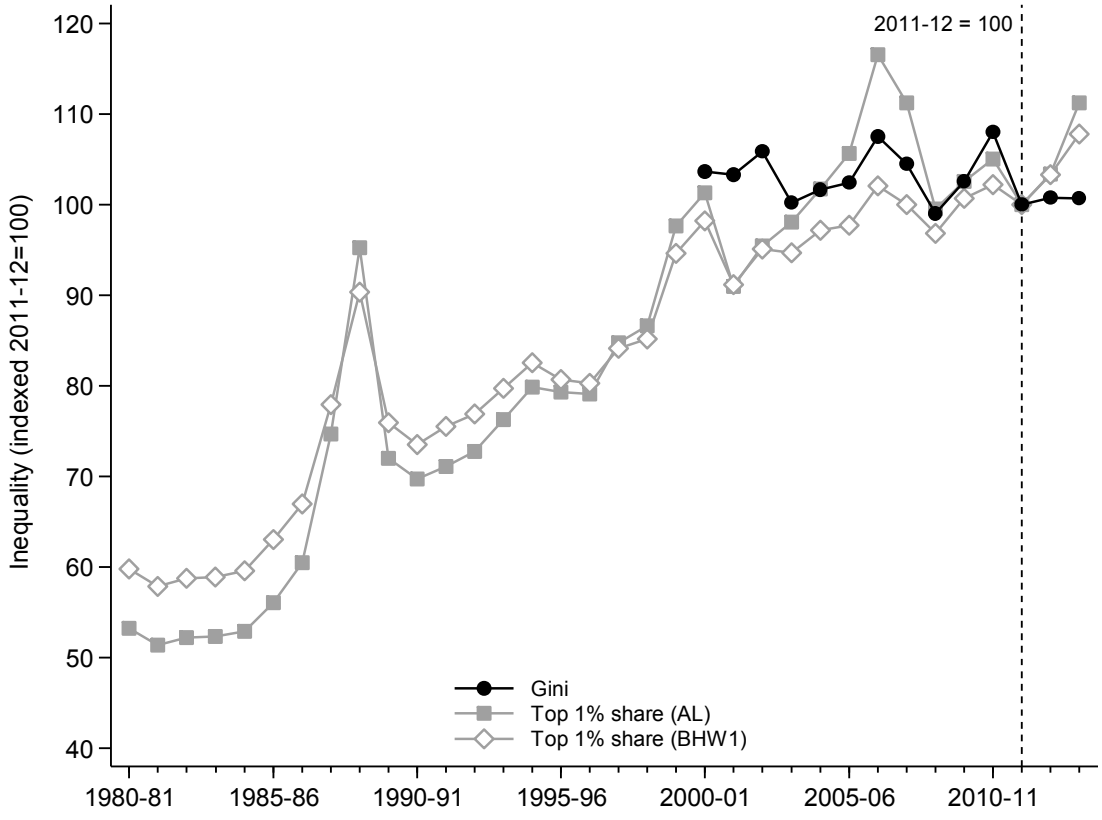
Notes:
[1] Piketty and Saez (2003) use a market income concept (even though they refer to it as 'gross income') that comprises 'salaries and wages, small business and farm income, partnership and fiduciary dividends, interest, rents, royalties, and other small items' reported on tax returns. The series we report here is their series that excludes taxable realised capital gains.
[2] This income concept is based on 'Total Income' as defined by the ATO, which is the sum of all taxable income reported on Australian tax returns. Total Income includes wages and salaries, investment income, business income, (taxable) government transfers, other reported income and also dividend imputation credits and taxable realised capital gains. For our *tax-concept gross* income measure we subtract dividend imputation credits and taxable realised capital gains from Total Income.
[3] We use the same income concept as described in Note 2, but with minor adjustments as described in Appendix Table A2.
[4] To replicate *tax-concept gross* income using *HILDA-AUTAX* data, we take HILDA personal gross income and subtract Disability Support Pension, Carer Allowance, Family Tax Benefit (Parts A and B), regular private transfers, compensation payments and 'salary sacrificed' wages and salaries. We then add in taxable irregular income components—redundancy and severance payments, and lump sum workers' compensation.
[5] Obtained by subtracting taxable government transfers and irregular income from *tax-concept gross* income.
[6] Obtained by taking *tax-concept gross* income and adding non-taxable government transfers and regular private transfers, and subtracting all irregular income.
[7] Obtained by taking *survey-concept gross* income and subtracting income tax calculated using the HILDA tax model. See Wilkins (2014b).
[8] Piketty and Saez (2003) took their data from US Internal Revenue Service (IRS) tax records.
[9] Data taken from the detailed tax tables that are published each year in the Australian Taxation Statistics.
[10] Data taken from the customised tables that were provided to the authors by the ATO.
[11] HILDA-AUTAX is the result of using HILDA data and replacing its top 1% *tax-concept gross* income with the top 1% *tax-concept gross* income from the 2% sample of the ATO unit-record data.
[12] Based on US IRS tax data. Piketty and Saez (2003) derive their denominator as market income of tax filers plus income of non-filers (calculated as 20 per cent of average market income of filers).
[13] Household income components from the National Accounts (based on Items 1-8 in Appendix Table A1).
[14] Household income components from the National Accounts (based on Items 1-9 in Appendix Table A1).

Table 2: Comparison of inequality measures based on Unadjusted HILDA and Adjusted HILDA (HILDA-AUTAX) data, 2011-12, 2012-13 and 2013-14

	Top 1% share	Gini	Mean log deviation (GE0)	Theil index[1] (GE1)	Half squared coefficient of variation[1] (GE2)	Mean	Median	90-50 ratio
2011-12								
Unadjusted equivalised household disposable income	4.39	0.3026	0.1637	0.1561	0.1980	57,044	50,494	1.91
Adjusted equivalised household disposable income	4.81	0.3062	0.1680	0.1654	0.2351	57,351	50,508	1.92
2012-13								
Unadjusted equivalised household disposable income	4.72	0.3050	0.1649	0.1632	0.2194	57,489	50,493	1.90
Adjusted equivalised household disposable income	5.00	0.3077	0.1682	0.1722	0.2808	57,725	50,493	1.90
2013-14								
Unadjusted equivalised household disposable income	4.85	0.3046	0.1602	0.1638	0.2293	57,852	50,390	1.94
Adjusted equivalised household disposable income	5.37	0.3089	0.1656	0.1778	0.3123	58,224	50,390	1.94

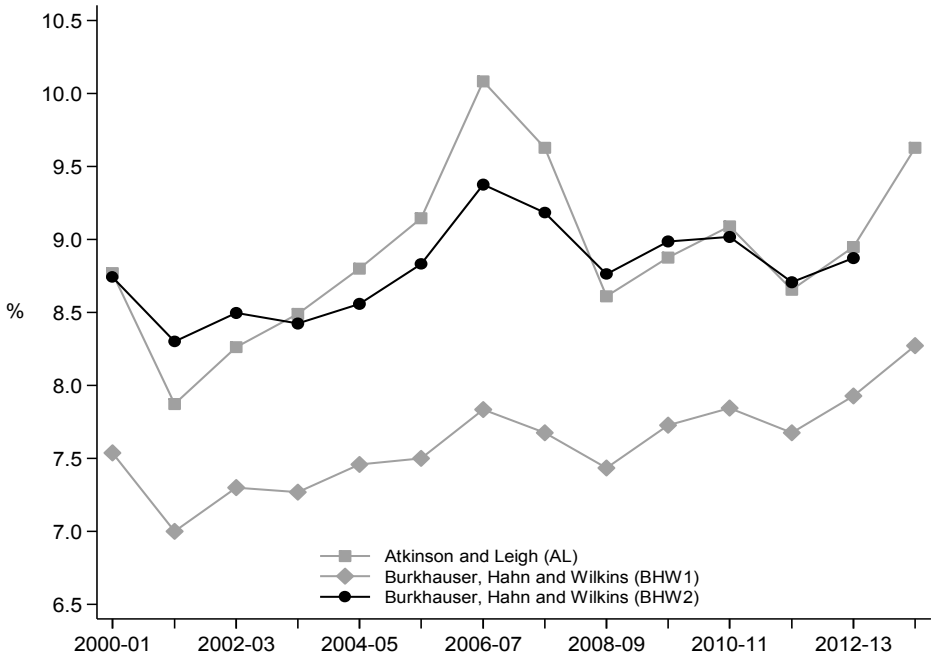
Note: [1] Negative and zero incomes are excluded.

Figure 1: Comparison of inequality trends in Australia using alternative measures: Top 1% share of income and Gini income inequality coefficients



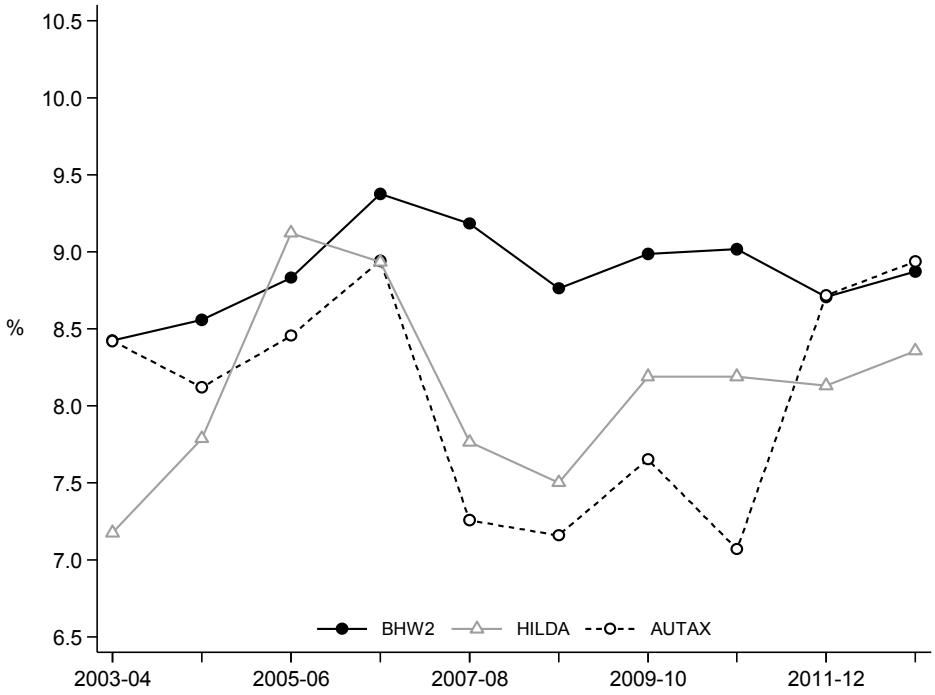
Sources: Burkhauser, Hahn and Wilkins (2015, updated) and authors' own calculations.

Figure 2: Comparison of alternative estimates of top 1% share of *tax-concept gross* income based on tax records data



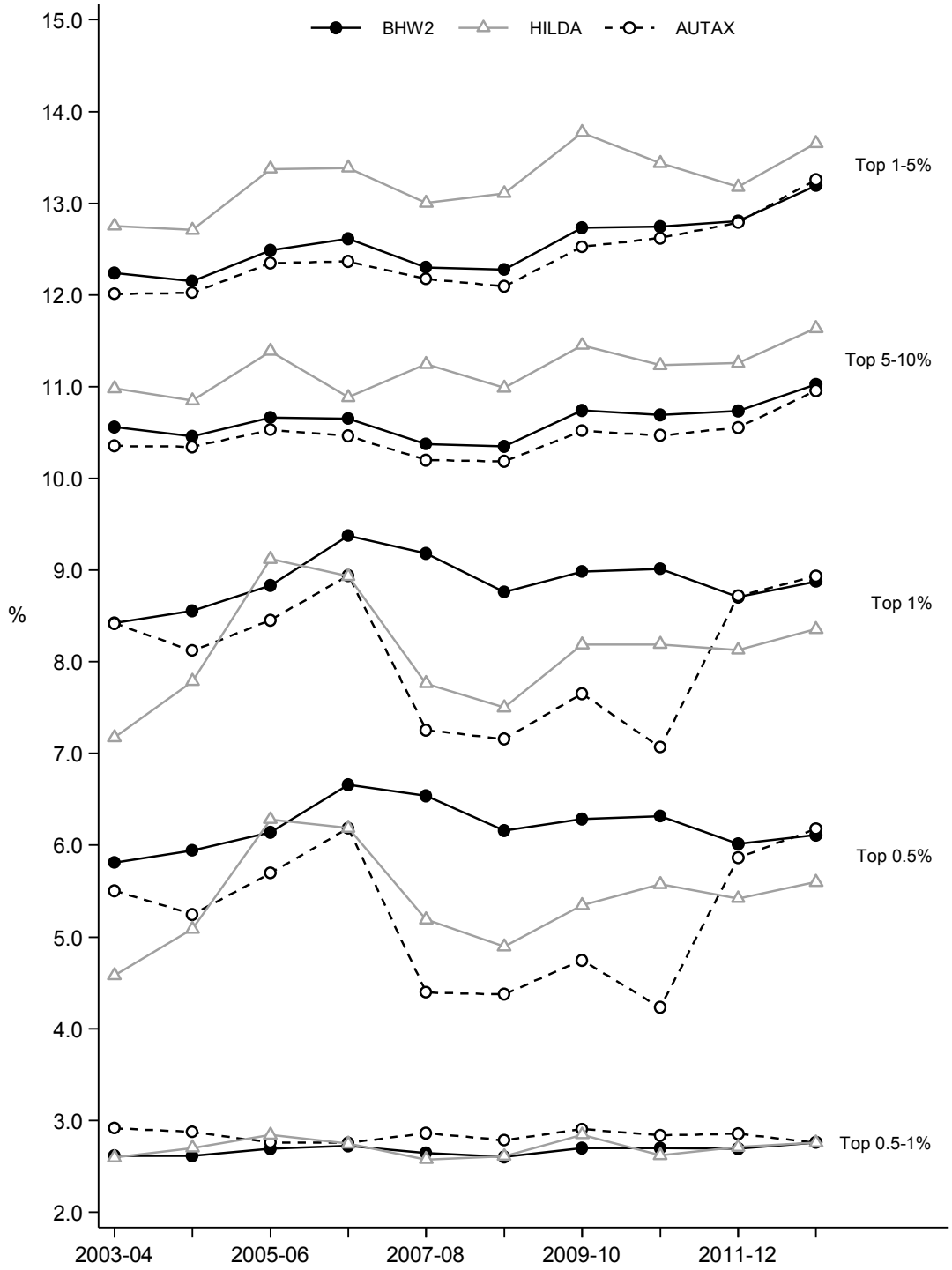
Sources: AL: Atkinson and Leigh (2007, updated) until 2010-11 and Burkhauser, Hahn and Wilkins (2015, updated) since 2011-12 using the same methods; BHW1: Burkhauser, Hahn and Wilkins (2015, updated); and BHW2: authors' own calculations.
Notes: The income control totals are calculated from the National Accounts. See Appendix Table B1 for the exact values of the top 1% shares.

Figure 3: Top 1% share of *tax-concept gross* income: HILDA and tax sample data (AUTAX) compared with tax records data



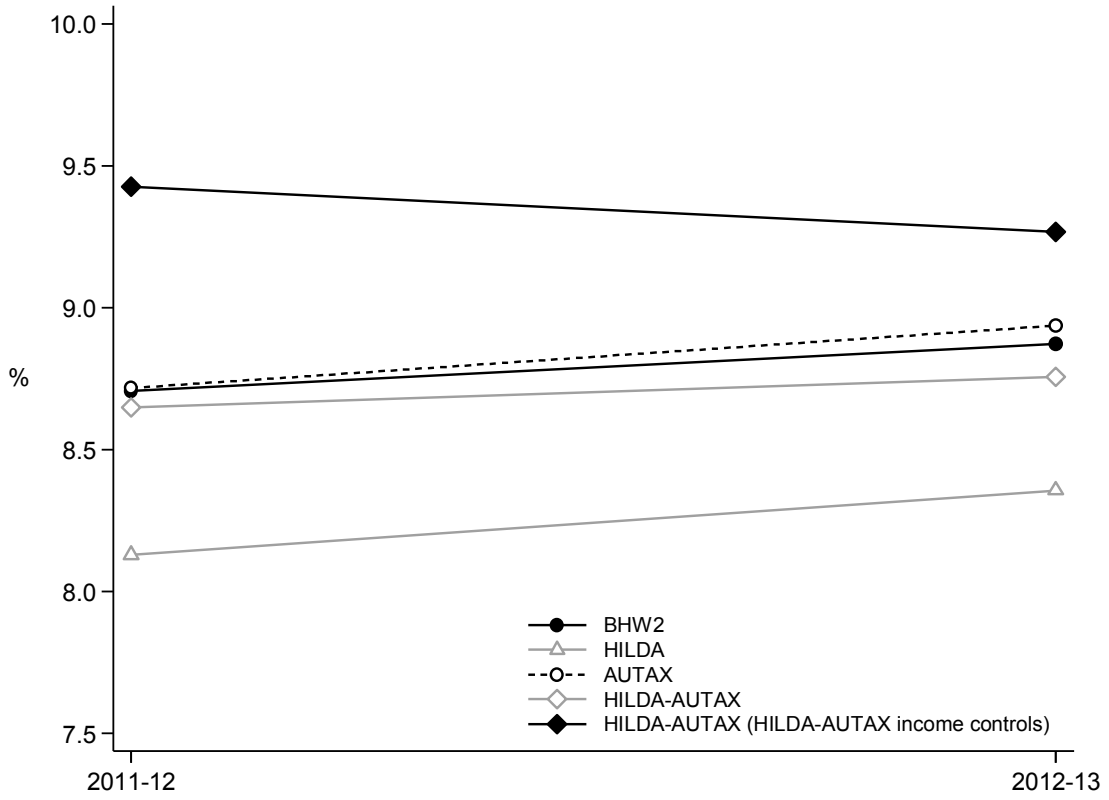
Notes: The income control totals are calculated from the National Accounts. See Appendix Table B2 for the exact values of the top 1% shares.

Figure 4: Top income shares of *tax-concept gross* income: HILDA and tax sample data (AUTAX) compared with tax records data



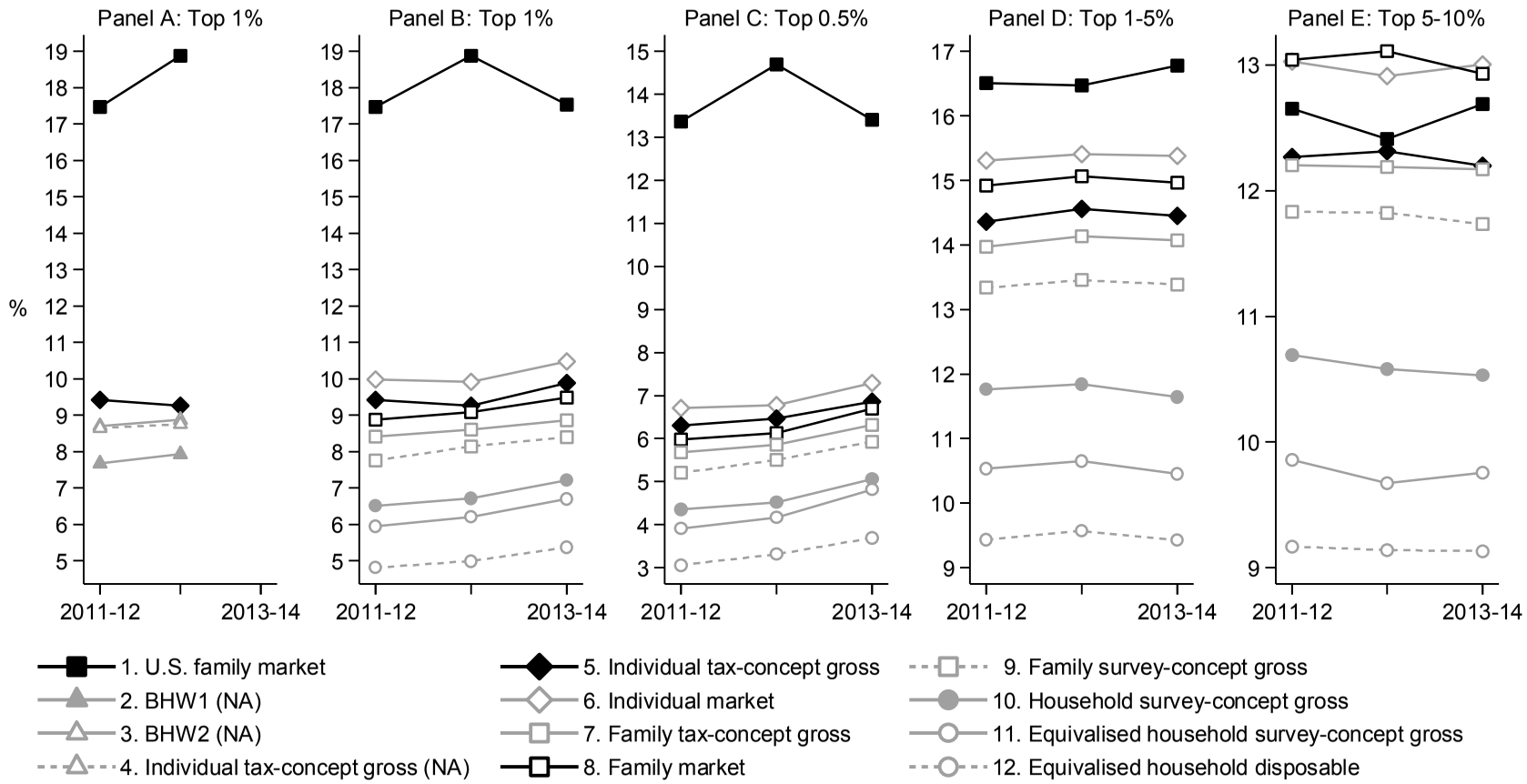
Notes: The income control totals are calculated from the National Accounts. See Appendix Table B2 for the exact values of the top income shares.

Figure 5: Top 1% share of *tax-concept gross* income: HILDA, tax sample data (AUTAX) and improved HILDA-AUTAX data compared with tax records data



Notes: For all series other than HILDA-AUTAX (HILDA-AUTAX income control totals), the income control totals are calculated from the National Accounts. See Appendix Table B3 for the exact values of the top 1% shares.

Figure 6: Crosswalk from top income shares of U.S. family market income to equivalised household disposable income of all Australian residents



Sources: 'US family market': Piketty and Saez (2003, updated); 'BHW1 (NA)': Burkhauser, Hahn and Wilkins (2015, updated); all other income values: authors' own calculations.
Notes: See Table 1 for a description of the 12 top income share measures. See Appendix Table B1 and B3 for the exact values of the top 1% shares in Panel A. See Appendix Table B4 for the exact values of the top income shares in Panels B–E.

Appendix A

1. Income definitions

Table A1: Components of control total for income (our National Accounts denominator)

	ABS description
Positive components:	
1. Gross mixed income	The surplus or deficit accruing from production by unincorporated enterprises. It includes elements of both compensation of employees (returns on labour inputs) and operating surplus (returns on capital inputs).
2. Compensation of employees	The total remuneration, in cash or in kind, payable by an enterprise to an employee in return for work done by the employee during the accounting period. It is further classified into two sub-components: wages and salaries; and employers' social contributions. Compensation of employees is not payable in respect of unpaid work undertaken voluntarily, including the work done by members of a household within an unincorporated enterprise owned by the same household. Compensation of employees excludes any taxes payable by the employer on the wage and salary bill (e.g., payroll tax).
3. Property income receivable: Interest	
4. Property income receivable: Dividends	
5. Secondary income receivable: Social benefits receivable: Workers' compensation	
6. Secondary income receivable: Social benefits receivable: Social assistance benefits	Includes current transfers to persons from general government in return for which no services are rendered or goods supplied. Principal components include: scholarships; maternity, sickness and unemployment benefits; family allowances; and widows', age, invalid and repatriation pensions.
Negative components:	
7. Property income payable: Interest payable: Unincorporated enterprises	
8. Household sector consumption of fixed capital	The reduction in the value of fixed assets used in production during the accounting period resulting from physical deterioration, normal obsolescence or normal accidental damage. Unforeseen obsolescence, major catastrophes and the depletion of natural resources are not taken into account.
9. Employers' social contributions (separately available from 1989-90 onwards)	Includes contributions to superannuation made by employers and payments of workers' compensation premiums.

Note: The control total for income is the sum of the positive components minus the sum of the negative components. Atkinson and Leigh (2007) use components 1-8 in their control total for income; for BHW2, we use components 1-9—that is we additionally subtract employers' social contributions.

Table A2: Numerator for our top tax-concept gross income share series

	Item name in Tax Statistics 2004	Description
	Total income or loss	
plus	Exempt foreign employment income	Foreign income that is not taxed in Australia and is therefore not included in 'Total income or loss'.
	Minor adjustments correcting for income deferral to later years	
less	Net farm management withdrawals	Previous years' income withdrawn from the Farm Management Deposit Scheme.
plus	Net farm management deposits	Current year's income deposited into the Farm Management Deposit Scheme.
less	Total deferred losses	Non-commercial business losses that have been deferred to later years. First year of this item is 2000-01.
plus	Deferred non-commercial business losses PY – PP & NPP	Previous years' deferred non-commercial business losses. Adjustment made from 2001-02 onwards.
	Imputation credits	
less	Imputation credit primary	Imputation or franking credits directly received.
less	Imputation credit subsidiary	Imputation or franking credits indirectly received through partnerships or trusts.
less	Australian franking credits from a New Zealand company	Imputation or franking credits received through a New Zealand company that participates in the Australian dividend imputation system. Adjustment made from 2003-04 onwards.
	Capital gains	
less	Net capital gain	Taxable realised capital gains.
=	**Our measure of *tax-concept gross* income**	

2. Decomposition of sources of differences between BHW1 and BHW2

With the available data, it is possible to decompose the sources of the differences in top income shares estimates into four components. These comprise differences due to: (1) Sorting of individuals in BHW1 by our *tax-concept gross* income ranges rather than 'Total' income ranges found in the standard tax tables; (2) 're-ordering' of individuals in tax tables in BHW1 due to the exclusion of dividend imputation credits and capital gains; (3) the combined effects of the (slight) changes in income definition for the numerator (see Table A2) plus errors in BHW1 in the imputation of income of individuals within the income range in which the bottom of the top income group falls[20]; and (4) inclusion of employer social contributions in the control total for income in BHW1 (see Table A1).

Using the BHW1 series, BHW2 series, and other series based on the published tax tables and customised tax tables obtained from the ATO, the contributions of the four

[20] The additional income components included in BHW2 are relatively small. For example, for the top 1%, they amount to between -0.4 and +0.1 per cent of their total income (or approximately -0.04 to +0.01 per cent of the total income of all tax filers). Consequently, component (3) mostly comprises the effects of imputation.

components can be identified as follows:

$$
\begin{aligned}
S_{BHW2} - S_{BHW1} &= S_{TOT\_CG}^{Ex\_CG} - S_{BHW1} && (1)\\
&+\left(S_{CUST}^{Ex\_CG} - S_{CUST}^{Inc\_CG}\right) - \left(S_{TOT}^{Ex\_CG} - S_{TOT}^{Inc\_CG}\right) && (2)\\
&+S_{CUST}^{Inc\_CG} - S_{TOT}^{Inc\_CG} && (3)\\
&+S_{BHW2} - S_{CUST}^{Ex\_CG} && (4)
\end{aligned}
\qquad (0)
$$

where:

- S_{BHW1} is the BHW1 estimated income share (the income share excluding dividend imputation credits and capital gains estimated from the published tax tables sorted by taxable income);
- $S_{TOT\_CG}^{Inc\_CG}$ is the income share including dividend imputation credits and capital gains estimated from the published tax tables sorted by Total income;
- $S_{TOT\_CG}^{Ex\_CG}$ is the income share excluding dividend imputation credits and capital gains estimated from the published tax tables sorted by Total income;
- $S_{CUST}^{Inc\_CG}$ is the income share including dividend imputation credits and capital gains estimated from the customised tax tables obtained from the ATO;
- $S_{CUST}^{Ex\_CG}$ is the income share excluding dividend imputation credits and capital gains estimated from the customised tax tables obtained from the ATO using *tax-concept of gross* income; and
- S_{BHW2} is the BHW2 estimated income share (the income share excluding dividend imputation credits and capital gains estimated from the customised tax tables obtained from the ATO (*tax-concept gross income*), and using the revised control total for income that excludes employer social contributions which now more closely matches *tax-concept gross income*).

This decomposition is only possible from 2006-07, when the tax tables with persons sorted by Total income became more detailed—specifically, the number of Total income bins more than doubled. Tables A3 to A7 respectively undertake this decomposition exercise for the top 0.1%, 0.5%, 1%, 5% and 10%.

Table A3: Decomposition of differences between BHW1 and BHW2—Top 0.1% income share

	2006-07	2007-08	2008-09	2009-10	2010-11	2011-12	2012-13	Mean
Alternative series of top income shares								
S_{BHW1}	2.37	2.31	2.24	2.29	2.38	2.20	2.28	2.30
$S_{TOT}^{Inc\_CG}$	3.74	3.58	2.98	2.98	3.17	2.84	2.95	3.18
$S_{TOT}^{Ex\_CG}$	2.36	2.32	2.26	2.32	2.39	2.22	2.29	2.31
$S_{CUST}^{Inc\_CG}$	4.01	4.02	3.21	3.13	3.28	2.90	2.99	3.36
$S_{CUST}^{Ex\_CG}$	2.81	2.79	2.59	2.56	2.59	2.37	2.42	2.59
S_{BHW2}	3.05	3.03	2.80	2.77	2.81	2.57	2.62	2.81
Decomposition of difference between BHW2 and BHW1								
Total difference $\left(S_{BHW2} - S_{BHW1}\right)$	0.68	0.72	0.56	0.48	0.43	0.37	0.34	0.51
Contribution of each component								
(1) Sorting variable $\left(S_{TOT}^{Ex\_CG} - S_{BHW1}\right)$	-0.01	0.01	0.03	0.02	0.02	0.01	0.01	0.01
(2) Re-ordering $\left(\left(S_{CUST}^{Ex\_CG} - S_{CUST}^{Inc\_CG}\right) - \left(S_{TOT}^{Ex\_CG} - S_{TOT}^{Inc\_CG}\right)\right)$	0.17	0.03	0.09	0.09	0.09	0.09	0.08	0.09
(3) Imputation & income definition $\left(S_{CUST}^{Inc\_CG} - S_{TOT}^{Inc\_CG}\right)$	0.28	0.44	0.23	0.15	0.11	0.06	0.04	0.19
(4) Income control total $\left(S_{BHW2} - S_{CUST}^{Ex\_CG}\right)$	0.24	0.24	0.21	0.22	0.22	0.20	0.20	0.22

Table A4: Decomposition of differences between BHW1 and BHW2—Top 0.5% income share

	2006-07	2007-08	2008-09	2009-10	2010-11	2011-12	2012-13	Mean
Alternative series of top income shares								
S_{BHW1}	5.52	5.37	5.18	5.38	5.47	5.26	5.42	5.37
$S_{TOT}^{Inc\_CG}$	7.57	7.19	6.26	6.39	6.60	6.17	6.37	6.65
$S_{TOT}^{Ex\_CG}$	5.47	5.36	5.21	5.40	5.48	5.27	5.42	5.37
$S_{CUST}^{Inc\_CG}$	7.99	7.79	6.63	6.67	6.83	6.34	6.47	6.96
$S_{CUST}^{Ex\_CG}$	6.13	6.02	5.69	5.80	5.83	5.55	5.64	5.81
S_{BHW2}	6.66	6.54	6.16	6.28	6.32	6.01	6.11	6.30
Decomposition of difference between BHW2 and BHW1								
Total difference $\left(S_{BHW2} - S_{BHW1}\right)$	1.14	1.16	0.98	0.91	0.85	0.75	0.70	0.93
Contribution of each component								
(1) Sorting variable $\left(S_{TOT}^{Ex\_CG} - S_{BHW1}\right)$	-0.05	-0.01	0.02	0.02	0.01	0.01	0.01	0.00
(2) Re-ordering $\left(\left(S_{CUST}^{Ex\_CG} - S_{CUST}^{Inc\_CG}\right) - \left(S_{TOT}^{Ex\_CG} - S_{TOT}^{Inc\_CG}\right)\right)$	0.23	0.06	0.12	0.12	0.12	0.12	0.11	0.13
(3) Imputation & income definition $\left(S_{CUST}^{Inc\_CG} - S_{TOT}^{Inc\_CG}\right)$	0.43	0.60	0.37	0.28	0.23	0.16	0.11	0.31
(4) Income control total $\left(S_{BHW2} - S_{CUST}^{Ex\_CG}\right)$	0.53	0.51	0.47	0.49	0.49	0.46	0.47	0.49

Table A5: Decomposition of differences between BHW1 and BHW2—Top 1% income share

	2006-07	2007-08	2008-09	2009-10	2010-11	2011-12	2012-13	Mean
Alternative series of top income shares								
S_{BHW1}	7.83	7.67	7.43	7.73	7.84	7.68	7.93	7.73
$S_{TOT}^{Inc\_CG}$	10.32	9.81	8.74	8.97	9.18	8.76	9.02	9.26
$S_{TOT}^{Ex\_CG}$	7.84	7.73	7.53	7.80	7.89	7.75	7.97	7.79
$S_{CUST}^{Inc\_CG}$	10.85	10.51	9.21	9.33	9.49	8.96	9.16	9.64
$S_{CUST}^{Ex\_CG}$	8.63	8.46	8.10	8.29	8.32	8.04	8.19	8.29
S_{BHW2}	9.38	9.18	8.76	8.98	9.02	8.71	8.87	8.99
Decomposition of difference between BHW2 and BHW1								
Total difference $\left(S_{BHW2} - S_{BHW1}\right)$	1.54	1.51	1.33	1.26	1.17	1.03	0.95	1.26
Contribution of each component								
(1) Sorting variable $\left(S_{TOT}^{Ex\_CG} - S_{BHW1}\right)$	0.00	0.06	0.10	0.07	0.04	0.08	0.05	0.06
(2) Re-ordering $\left(\left(S_{CUST}^{Ex\_CG} - S_{CUST}^{Inc\_CG}\right) - \left(S_{TOT}^{Ex\_CG} - S_{TOT}^{Inc\_CG}\right)\right)$	0.27	0.03	0.10	0.13	0.13	0.08	0.08	0.12
(3) Imputation & income definition $\left(S_{CUST}^{Inc\_CG} - S_{TOT}^{Inc\_CG}\right)$	0.52	0.69	0.47	0.36	0.30	0.21	0.14	0.38
(4) Income control total $\left(S_{BHW2} - S_{CUST}^{Ex\_CG}\right)$	0.75	0.72	0.67	0.70	0.69	0.67	0.68	0.70

Table A6: Decomposition of differences between BHW1 and BHW2—Top 5% income share

	2006-07	2007-08	2008-09	2009-10	2010-11	2011-12	2012-13	Mean
Alternative series of top income shares								
S_{BHW1}	18.82	18.44	18.22	19.01	19.21	19.16	19.85	18.96
$S_{TOT}^{Inc\_CG}$	22.54	21.60	20.21	20.86	21.11	20.75	21.49	21.22
$S_{TOT}^{Ex\_CG}$	18.91	18.56	18.39	19.08	19.26	19.22	19.95	19.05
$S_{CUST}^{Inc\_CG}$	23.48	22.65	21.08	21.60	21.77	21.24	21.79	21.94
$S_{CUST}^{Ex\_CG}$	20.24	19.79	19.44	20.03	20.08	19.86	20.37	19.97
S_{BHW2}	21.99	21.48	21.04	21.72	21.76	21.51	22.07	21.65
Decomposition of difference between BHW2 and BHW1								
Total difference $\left(S_{BHW2} - S_{BHW1}\right)$	3.16	3.04	2.82	2.71	2.55	2.36	2.21	2.69
Contribution of each component								
(1) Sorting variable $\left(S_{TOT}^{Ex\_CG} - S_{BHW1}\right)$	0.09	0.12	0.16	0.07	0.05	0.06	0.09	0.09
(2) Re-ordering $\left(\left(S_{CUST}^{Ex\_CG} - S_{CUST}^{Inc\_CG}\right) - \left(S_{TOT}^{Ex\_CG} - S_{TOT}^{Inc\_CG}\right)\right)$	0.38	0.18	0.18	0.20	0.17	0.15	0.12	0.20
(3) Imputation & income definition $\left(S_{CUST}^{Inc\_CG} - S_{TOT}^{Inc\_CG}\right)$	0.94	1.05	0.87	0.75	0.66	0.49	0.30	0.72
(4) Income control total $\left(S_{BHW2} - S_{CUST}^{Ex\_CG}\right)$	1.75	1.69	1.60	1.69	1.68	1.66	1.70	1.68

Table A7: Decomposition of differences between BHW1 and BHW2—Top 10% income share

	2006-07	2007-08	2008-09	2009-10	2010-11	2011-12	2012-13	Mean
Alternative series of top income shares								
S_{BHW1}	28.27	27.74	27.44	28.57	28.76	28.76	29.80	28.48
$S_{TOT}^{Inc\_CG}$	32.53	31.33	29.75	30.74	30.98	30.66	31.71	31.10
$S_{TOT}^{Ex\_CG}$	28.43	27.92	27.70	28.74	28.92	28.94	29.94	28.65
$S_{CUST}^{Inc\_CG}$	33.76	32.59	30.89	31.76	31.87	31.35	32.18	32.06
$S_{CUST}^{Ex\_CG}$	30.04	29.35	29.00	29.93	29.95	29.76	30.54	29.80
S_{BHW2}	32.64	31.86	31.39	32.46	32.45	32.25	33.09	32.30
Decomposition of difference between BHW2 and BHW1								
Total difference $\left(S_{BHW2} - S_{BHW1}\right)$	4.37	4.12	3.95	3.89	3.69	3.49	3.28	3.83
Contribution of each component								
(1) Sorting variable $\left(S_{TOT}^{Ex\_CG} - S_{BHW1}\right)$	0.16	0.18	0.26	0.17	0.15	0.19	0.14	0.18
(2) Re-ordering $\left(\left(S_{CUST}^{Ex\_CG} - S_{CUST}^{Inc\_CG}\right) - \left(S_{TOT}^{Ex\_CG} - S_{TOT}^{Inc\_CG}\right)\right)$	0.39	0.18	0.17	0.18	0.14	0.13	0.14	0.19
(3) Imputation & income definition $\left(S_{CUST}^{Inc\_CG} - S_{TOT}^{Inc\_CG}\right)$	1.23	1.26	1.14	1.01	0.89	0.70	0.46	0.96
(4) Income control total $\left(S_{BHW2} - S_{CUST}^{Ex\_CG}\right)$	2.60	2.50	2.39	2.53	2.50	2.48	2.55	2.51

3. Topcoding in the Australian tax sample file

Figure A1: Proportion of persons with at least one top-coded income source, by top Total income group

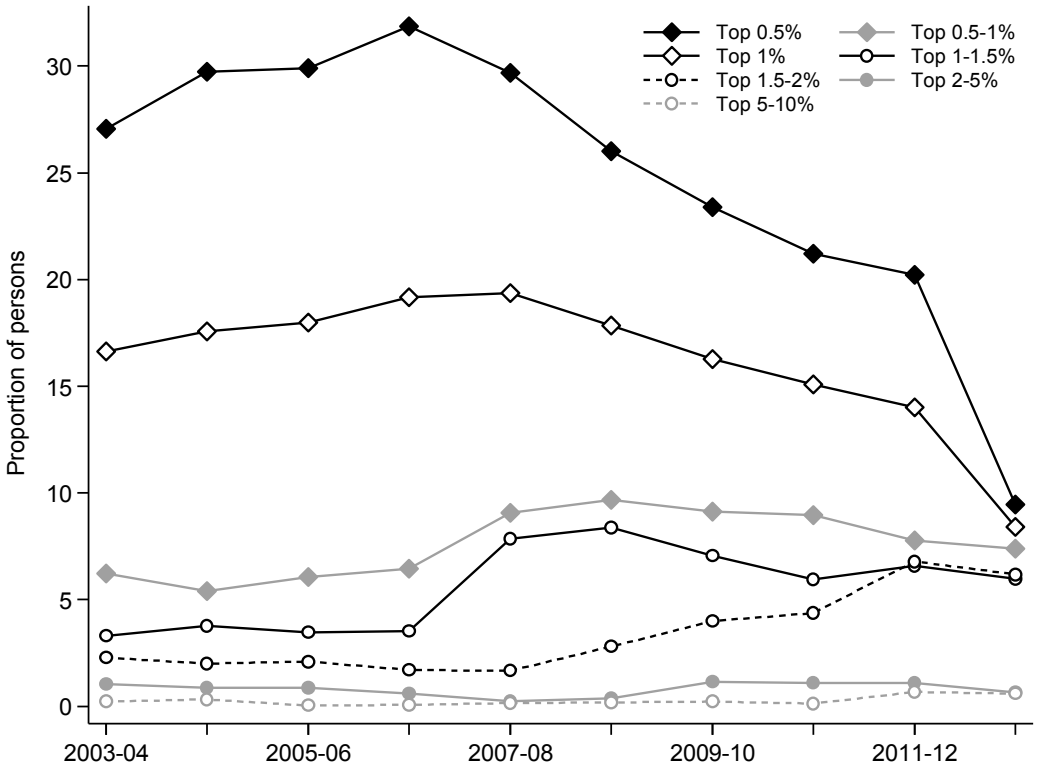


Figure A1 shows the proportion of persons within a top Total income group—such as the top 1%—that have a least one topcoded income source. But, because the unit-record data makes it impossible to identify all topcodes, the estimates shown should be regarded as lower bounds. For each income source of 'Total income or loss', we attempt to identify the topcodes as follows:

- Restrict search to the top 10% of values for the income source.
- Find all values with a frequency of 10 or higher:
  - If three or fewer values are found, use these values as the topcodes.
  - If more than three values are found, use the three most frequent values as the topcodes.

The above procedure does not perform well for aggregate income sources that are made up of less-aggregated income sources that are topcoded. For example, the unit-record data only contains a 'Wages and salaries' income item that may consist of up to five different wages or salary sources. Each of the five wages or salary components have been topcoded. This means that, for persons with more than two jobs during the financial year, we cannot identify any topcodes.

Appendix B: Additional Tables and Figures

Table B1: Alternative estimates of Australian top *tax-concept gross* income shares based on tax records data

Year	Top 0.1%	Top 0.5%	Top 1%	Top 5%	Top 10%
			Atkinson and Leigh (AL, updated)		
2000-01	2.97	6.25	8.77	20.39	30.45
2001-02	2.38	5.45	7.87	19.17	28.95
2002-03	2.52	5.74	8.26	19.73	29.62
2003-04	2.67	5.97	8.49	19.81	29.54
2004-05	2.91	6.29	8.80	20.11	29.83
2005-06	3.06	6.54	9.14	20.78	30.60
2006-07	3.68	7.44	10.08	22.06	32.01
2007-08	3.52	7.10	9.63	21.20	30.85
2008-09	2.93	6.18	8.61	19.88	29.32
2009-10	2.93	6.33	8.88	20.69	30.46
2010-11	3.13	6.55	9.09	20.99	30.73
2011-12	2.80	6.13	8.65	20.62	30.39
2012-13	2.92	6.34	8.95	21.36	31.53
2013-14	3.36	6.94	9.63	22.30	32.69
			Burkhauser, Hahn and Wilkins (BHW1, updated)		
2000-01	2.27	5.24	7.54	18.79	28.65
2001-02	1.95	4.77	7.00	17.95	27.54
2002-03	2.07	5.00	7.30	18.38	28.05
2003-04	2.05	4.96	7.27	18.06	27.49
2004-05	2.11	5.07	7.46	18.14	27.52
2005-06	2.12	5.16	7.50	18.46	27.95
2006-07	2.37	5.52	7.83	18.82	28.27
2007-08	2.31	5.37	7.67	18.44	27.74
2008-09	2.24	5.18	7.43	18.22	27.44
2009-10	2.29	5.38	7.73	19.01	28.57
2010-11	2.38	5.47	7.84	19.21	28.76
2011-12	2.20	5.26	7.68	19.16	28.76
2012-13	2.28	5.42	7.93	19.85	29.80
2013-14	2.50	5.72	8.27	20.41	30.52
			Burkhauser, Hahn and Wilkins (BHW2)		
2000-01	2.65	6.06	8.74	21.40	32.38
2001-02	2.34	5.66	8.30	20.79	31.55
2002-03	2.42	5.82	8.49	21.06	31.91
2003-04	2.44	5.81	8.42	20.66	31.22
2004-05	2.56	5.94	8.56	20.71	31.16
2005-06	2.66	6.14	8.83	21.32	31.98
2006-07	3.05	6.66	9.38	21.99	32.64
2007-08	3.03	6.54	9.18	21.48	31.86
2008-09	2.80	6.16	8.76	21.04	31.39
2009-10	2.77	6.28	8.98	21.72	32.46
2010-11	2.81	6.32	9.02	21.76	32.45
2011-12	2.57	6.01	8.71	21.51	32.25
2012-13	2.62	6.11	8.87	22.07	33.09
2013-14	-	-	-	-	-

Sources: AL: Atkinson and Leigh (2007, updated) until 2010-11 and Burkhauser, Hahn and Wilkins (2015, updated) since 2011-12 using the same methods; BHW1: Burkhauser, Hahn and Wilkins (2015, updated); and BHW2: authors' own calculations.

Notes: The income control totals are calculated from the National Accounts. The top 1% estimates (column 3) were used to create Figure 2.

Table B2: Comparing BHW2, HILDA and AUTAX based estimates of top *tax-concept gross* income shares

Year	Top 1%	Top 0.5%	Top 0.5-1%	Top 1-5%	Top 5-10%
			BHW2		
2003-04	8.42	5.81	2.61	12.24	10.56
2004-05	8.56	5.94	2.61	12.15	10.46
2005-06	8.83	6.14	2.69	12.49	10.66
2006-07	9.38	6.66	2.72	12.61	10.65
2007-08	9.18	6.54	2.64	12.30	10.38
2008-09	8.76	6.16	2.60	12.28	10.35
2009-10	8.98	6.28	2.70	12.73	10.74
2010-11	9.02	6.32	2.70	12.74	10.69
2011-12	8.71	6.01	2.69	12.81	10.73
2012-13	8.87	6.11	2.76	13.19	11.02
			HILDA		
2003-04	7.18	4.58	2.60	12.75	10.98
2004-05	7.79	5.09	2.70	12.71	10.85
2005-06	9.12	6.28	2.84	13.37	11.39
2006-07	8.93	6.19	2.74	13.39	10.88
2007-08	7.76	5.19	2.57	13.00	11.25
2008-09	7.50	4.89	2.61	13.11	10.99
2009-10	8.19	5.35	2.84	13.77	11.45
2010-11	8.19	5.57	2.62	13.44	11.23
2011-12	8.13	5.42	2.71	13.18	11.26
2012-13	8.36	5.60	2.76	13.65	11.64
			AUTAX		
2003-04	8.42	5.50	2.92	12.01	10.35
2004-05	8.12	5.25	2.87	12.02	10.34
2005-06	8.45	5.70	2.76	12.35	10.53
2006-07	8.94	6.19	2.75	12.37	10.46
2007-08	7.26	4.39	2.86	12.17	10.20
2008-09	7.16	4.38	2.78	12.09	10.18
2009-10	7.65	4.75	2.91	12.53	10.52
2010-11	7.07	4.23	2.84	12.62	10.47
2011-12	8.72	5.86	2.86	12.79	10.55
2012-13	8.94	6.18	2.76	13.26	10.96

Notes: The income control totals are calculated from the National Accounts. This table's values were used to create Figures 3 and 4.

Table B3: Comparing BHW2, HILDA, AUTAX and HILDA-AUTAX based estimates of top *tax-concept gross* income shares

	Top 1%	Top 0.5%	Top 0.5-1%	Top 1-5%	Top 5-10%
2011-12					
BHW2	8.71	6.01	2.69	12.81	10.73
HILDA	8.13	5.42	2.71	13.18	11.26
AUTAX	8.72	5.86	2.86	12.79	10.55
HILDA-AUTAX	8.65	5.79	2.86	13.18	11.26
HILDA-AUTAX (HILDA-AUTAX income controls)	9.43	6.31	3.12	14.36	12.27
2012-13					
BHW2	8.87	6.11	2.76	13.19	11.02
HILDA	8.36	5.60	2.76	13.65	11.64
AUTAX	8.94	6.18	2.76	13.26	10.96
HILDA-AUTAX	8.76	6.11	2.65	13.76	11.64
HILDA-AUTAX (HILDA-AUTAX income controls)	9.27	6.46	2.81	14.56	12.31

Notes: For all series other than HILDA-AUTAX (HILDA-AUTAX income controls), the income control totals are calculated from the National Accounts. The top 1% estimates (first column) were used to create Figure 5.

Table B4: Crosswalk from U.S. family market income to Australian equivalised household disposable income

	Top 1%	Top 0.5%	Top 0.5-1%	Top 1-5%	Top 5-10%
2011-12					
U.S. family market	17.47	13.37	4.10	16.51	12.65
Individual market	9.97	6.72	3.26	15.31	13.03
Individual tax-concept gross	9.43	6.31	3.12	14.36	12.27
Family market	8.88	5.98	2.90	14.92	13.00
Family tax-concept gross	8.41	5.69	2.73	13.98	12.20
Family survey-concept gross	7.76	5.20	2.55	13.34	11.84
Household survey-concept gross	6.51	4.35	2.15	11.76	10.69
Equivalised household survey-concept gross	5.96	3.91	2.05	10.54	9.86
Equivalised household disposable	4.81	3.06	1.75	9.43	9.17
2012-13					
U.S. family market	18.88	14.70	4.17	16.47	12.41
Individual market	9.91	6.78	3.13	15.41	12.91
Individual tax-concept gross	9.27	6.46	2.81	14.56	12.31
Family market	9.09	6.13	2.96	15.06	13.11
Family tax-concept gross	8.61	5.86	2.75	14.12	12.20
Family survey-concept gross	8.14	5.51	2.64	13.45	11.82
Household survey-concept gross	6.72	4.52	2.20	11.84	10.58
Equivalised household survey-concept gross	6.21	4.17	2.04	10.66	9.67
Equivalised household disposable	5.00	3.31	1.68	9.57	9.14
2013-14					
U.S. family market	17.54	13.41	4.13	16.78	12.69
Individual market	10.48	7.29	3.18	15.38	13.00
Individual tax-concept gross	9.89	6.86	3.03	14.45	12.20
Family market	9.48	6.69	2.79	14.99	12.96
Family tax-concept gross	8.86	6.31	2.55	14.08	12.16
Family survey-concept gross	8.41	5.93	2.47	13.39	11.74
Household survey-concept gross	7.22	5.06	2.16	11.64	10.53
Equivalised household survey-concept gross	6.71	4.82	1.89	10.46	9.76
Equivalised household disposable	5.37	3.69	1.68	9.42	9.13

Sources: 'US family market income': Piketty and Saez (2003, updated); all other income values: authors' own calculations.

Notes: The income control totals are calculated from the HILDA data. This table's values were used to create Panels B to E in Figure 6. 'Individual tax-concept gross' is the same income definition as 'HILDA-AUTAX (HILDA-AUTAX income control)' in Appendix Table B3.

Figure B1: Comparison of different *tax-concept gross* income control totals