

Cross-sectional Representativeness in Longitudinal Surveys: Income Distribution Trends in HILDA*

Bruce Bradbury¹ and Maximilian Longmuir^{1,2}

¹Social Policy Research Centre, University of New South Wales, Australia

²Stone Center on Socio-Economic Inequality / CUNY Graduate Center, USA

August 28, 2026

Abstract

Are household panel surveys suitable for measuring trends in cross-sectional population characteristics such as income inequality and poverty? Can survey weighting effectively compensate for non-random attrition and changing survey composition? Using the Household, Income and Labour Dynamics in Australia (HILDA) Survey, we assess the impact of cross-sectional representativeness on estimates of income inequality trends since 2000-01. We focus on the potential effects of non-random attrition and rule-based sample additions across the income distribution. Taking wave 1 as the baseline, we track later-wave shares across wave-1 equivalised household income deciles, anchor post-wave-1 entrants generated by HILDA's following rules, compare the continuing wave 11 sample with the wave-11 top-up sample, and undertake a reweighting that removes differential attrition by origin income decile. We conclude that while attrition does generate imbalances at both tails of the distribution, much of this drift is attenuated once age composition and later entrants are accounted for. The top-up comparison provides little evidence of a large divergence by wave 11. Our new weights lead to estimates of income trends that are very similar to the standard weights. Subject to limitations concerning baseline coverage and post-baseline migrants, we conclude that HILDA's annual waves remain well suited to cross-sectional trend analysis of income distribution.

Keywords: Representativeness, Survey Data, Cross-sectional analysis, Inequality

JEL Classifications: C81, C83, D31, I32

*This paper uses unit record data from the Household, Income and Labour Dynamics in Australia Survey [HILDA] conducted by the Australian Government Department of Social Services (DSS). The findings and views reported in this paper, however, are those of the authors and should not be attributed to the Australian Government, DSS, or any of DSS' contractors or partners. DOI: 10.26193/6M1BMR. This research is supported by ARC project DP220103811.

1 Introduction

Many large-scale surveys are designed as longitudinal (panel) surveys, tracking a population-representative sample of individuals over time. These surveys are also used cross-sectionally to examine outcomes in particular years, or to estimate time series of population outcomes such as income distribution patterns. This is particularly the case where there is an absence of alternative data sources, or time-series inconsistency in the cross-sectional data collections. In this paper, we examine the suitability of the Household, Income and Labour Dynamics in Australia (HILDA) panel survey for the estimation of population income distribution trends. The issues raised here might also be applicable to the examination of other trends in Australia and for other countries such as the US and Germany where panel datasets such as the PSID or SOEP have often been used in a variety of non-panel applications.

Ensuring cross-sectional representativeness with panel survey data is challenging. Longitudinal data are subject to non-random attrition that accumulates over waves, while following rules, i.e., how to include new partners, births, temporary household members, or immigrants, can change the composition of the survey and affect cross-sectional representativeness. Designing weights to compensate for all these influences is difficult, and even when longitudinal surveys follow best practices, the potential for biased cross-sectional estimates is high.

Interest in the representativeness of cross-sectional estimates derived from longitudinal surveys is longstanding. Panel studies typically begin with a representative baseline sample, but cumulative attrition can erode their suitability for cross-sectional analyses ([Rose, 2000](#)). In addition, well-specified following rules are needed to keep the sample aligned with demographic change as households split, form, and grow ([Lynn et al., 2011](#); [Trivellato, 1999](#); [Watson, 2022](#)). Refreshment samples and top-ups can further reduce compositional drift over time ([Watson and Wooden, 2013](#)). While these design challenges are well understood, existing assessments typically focus on demographic margins or mean outcomes, and rely on weighting and calibration adjustments whose performance for distributional income statistics, especially the tails and poverty measures, has rarely been examined systematically. To our knowledge, this paper provides the first systematic assessment of how cross-sectional representativeness across the income distribution evolves over time. We examine whether differential attrition and sample additions reshape the equivalised household income dis-

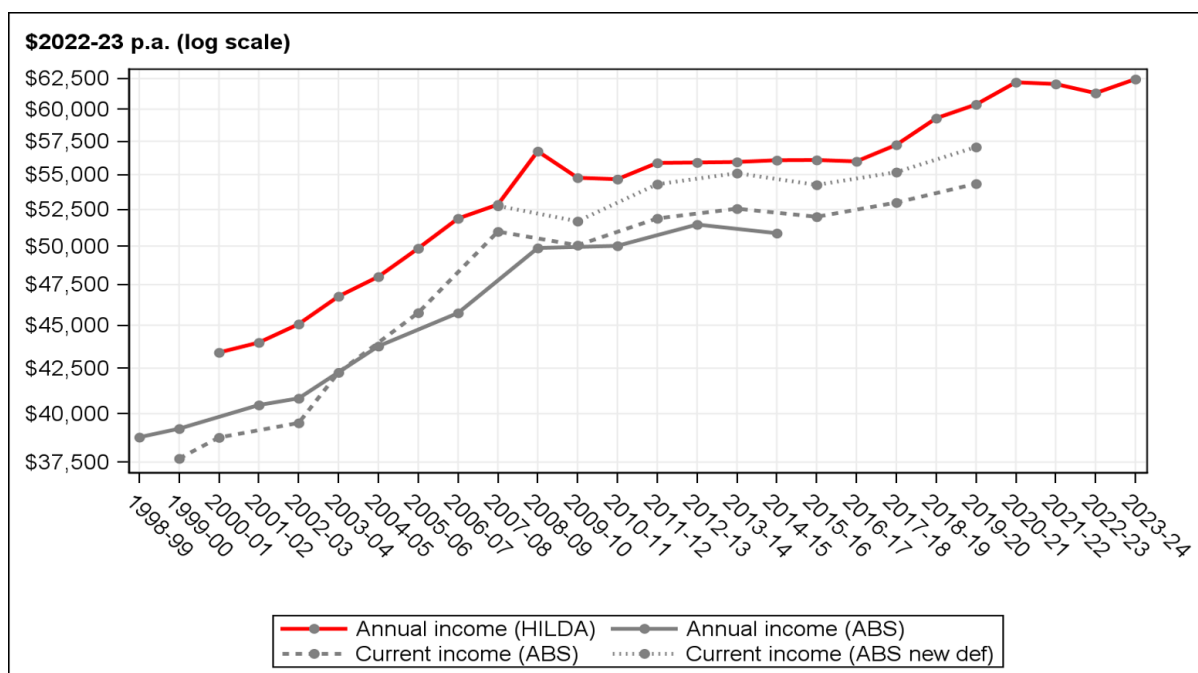
tribution, assess the consequences for estimates of income, inequality, and poverty, and demonstrate a diagnostic approach that can be applied to other longitudinal surveys.

The Household, Income and Labour Dynamics in Australia (HILDA) Survey is the basis of our analysis. HILDA is one of the main data sources for both panel and cross-sectional research in Australia and has been used extensively to study economic inequality, poverty, and policy-relevant outcomes (Davidson et al., 2025; Kaplan et al., 2018; Longmuir, 2023; Wilkins, 2015). At the same time, its design makes it particularly well-suited for assessing more general questions about cross-sectional representativeness in a longitudinal setting.

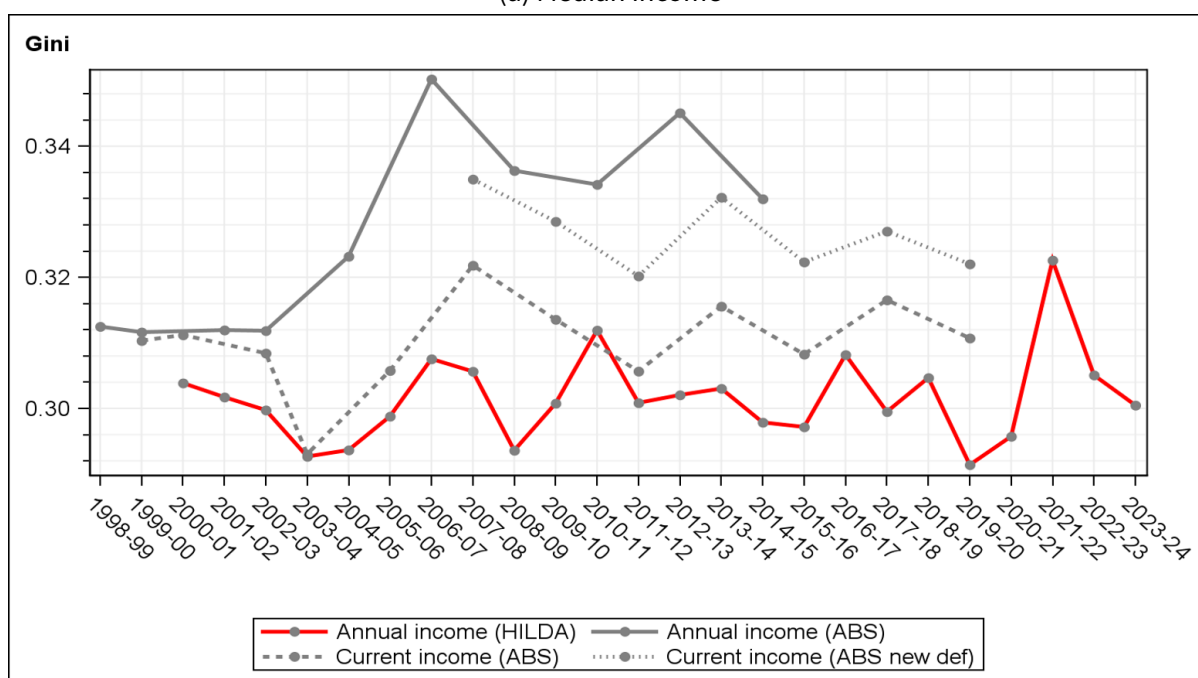
HILDA began with a target of a nationally representative cross-section of households in 2001. In subsequent waves, following rules have been applied, distinguishing Original Sample Members (OSMs), Temporary Sample Members (TSMs, interviewed while co-resident with a continuing member), and Other Permanent Sample Members (OPMs, permanent additions through marriage and birth) (Watson and Wooden, 2012). One well-recognised weakness for cross-sectional research is that migrants to Australia after the initial wave are not included – unless they move into the household of a continuing sample member. This is particularly important in a high-immigration country like Australia, and has been partially addressed via a general top-up sample in 2011 and recent migrant top-ups from 2023 onwards (Summerfield et al., 2025; Watson, 2012).

As its duration has grown, the HILDA survey has been increasingly used to examine trends in the household income distribution. The main alternative to the HILDA has been the series of Surveys of Income and Housing (SIH) conducted by the Australian Bureau of Statistics. These surveys are conducted under the Census and Statistics Act, with very high response rates and large sample sizes, and so in many respects offer an important alternative to HILDA. However, the survey methods have changed substantially over time, leading to concerns about their suitability for measurement of trends (Saunders and Bradbury, 2006; Wilkins, 2014). The irregularity of the SIH (most recent is for 2019-20) has also increased interest in the HILDA for contemporary estimates.

Figure 1a shows median income from these two sources since the start of the century, while Figure 1b shows the corresponding Gini coefficients. The SIH records income both for the previous financial year and on a 'current' (last payment) basis. The HILDA collects full household income for annual income only (some income components are also collected on a current basis). Both surveys use imputation methods to correct for individual (full or partial)



(a) Median income



(b) Gini coefficient

Notes: Compiled by authors using HILDA Release 24 (enumerated person cross-section weights) and ABS SIH surveys (person-weighted), various years. Household (regular) income after income tax, with survey-provider-imputed values for missing individual responses. Modified OECD equivalence scale used, with income deflated by the CPI. Negative income set to zero.

Figure 1. Household equivalent disposable income of persons, 1998-99 to 2023-24, ABS SIH and HILDA surveys

non-response when calculating household income. Over the years where both have been collected, the HILDA median has consistently been between 9 and 14 per cent higher than the SIH annual income median. A similar gap applies to (annualised) SIH current income, though the new version of this with a broader income scope which has been collected since 2007-08 is closer to the HILDA annual estimate.¹ One explanation for the higher incomes in HILDA is the over-representation of people in managerial and professional occupations in the initial wave, which was not corrected by the survey weighting procedure (Watson and Wooden, 2004). Although this suggests that HILDA generally overestimates household income, by itself it is not likely to bias estimates of *trends* in income levels or inequality.

While the gap in median income between the HILDA and SIH surveys has been relatively constant over time, the pattern for income inequality is more complex. In HILDA, the Gini coefficient shows no particular trend since the start of the century.² A similar pattern applies to the SIH current income measure. However, the annual income measure from the SIH, which for many years was the main indicator reported from the surveys,³ shows a significant increase in inequality during the pre-GFC years. There were significant changes in survey methods during this period, inter alia, a move to computer-aided interviewing, but the reasons for the different trends in the annual and current ABS estimates remain unresolved (see Wilkins (2014)). If we could be confident of the time-series comparability of the HILDA data, this would support a conclusion of stability in Australian income inequality over this period.

HILDA's survey methodology and questionnaire have remained stable, but its sample composition may nevertheless have changed after wave 1 because of attrition and its survey following rules. Although attrition between adjacent waves is low by panel-survey standards, the losses accumulated over more than two decades are substantial and non-random. HILDA's weights seek to compensate for this selection, but income is not included directly among the calibration benchmarks. In addition, the HILDA following rules do not automatically lead to a sample that is representative of the Australian population over time. Before the recent migrant top-ups, post-2001 migrants could enter HILDA only through the 2011

¹A new annual income measure was also introduced at this time (not shown), but this had little impact on the median.

²We have also calculated the trend for a balanced-panel sample, which is thus not subject to attrition bias. This shows a steady upward trend in the Gini. However, this sample is necessarily older in each year, which could explain the trend.

³The original current income definition in the ABS survey excluded some income sources such as overtime bonuses which were included in the new current income measure collected from 2007-08 onwards.

top-up or by joining a household containing a continuing sample member. As well, the recruitment of new members to the sample follows design rules that primarily satisfy the requirements of longitudinal rather than cross-sectional research. The weighting system has to compensate for all these factors. Is it thus suitable for the measurement of trends in income inequality in Australia? If not, can alternative weights be created to achieve this?

We address these questions with three tests. First, we look at the survival of people in the survey who start from the different parts of the income distribution in wave 1. Second, the HILDA survey had a substantial top-up to its sample in wave 11. This top-up is compared with the main sample to test for any biases over the first 11 waves. Third, we calculate a new set of weights accounting for baseline income position and test whether these make a substantive difference to conclusions about trends in income distribution.

These three diagnostics assess potential biases arising from differential response and the recruitment of people linked to different parts of the wave-1 income distribution – but it is important to note that they do not test all the potential sources of bias in trend estimates of incomes. Conceivably, differential response based on other characteristics could also introduce bias. The exclusion of migrants is one example, which is partly tested by our wave 11 comparison, but there could be others.

Assessing cross-sectional representativeness in a longitudinal survey is inherently difficult. A standard approach is to benchmark distributional estimates against administrative sources or other sources of repeated cross-sections. As noted above, there are concerns about the over-time comparability of the primary survey source of distributional data, the ABS SIH surveys. Administrative data sources for household income are currently under development, but have not yet been evaluated for time-series comparability.

Instead, we take HILDA's wave 1 as the reference point and study how the income distribution evolves thereafter. We first construct equivalised household income deciles for each enumerated person in wave 1 and then track, in each subsequent wave, the distribution of observed sample members across these fixed origin-deciles. In a perfectly stable panel with no differential attrition, each origin-decile would continue to account for 10% of the wave- t population. Deviations from 10%, therefore, provide a transparent diagnostic of distributional selection in subsequent waves. To take account of the associations between income and age, our main results calculate these origin-deciles within 10-year age groups.

We then extend the origin-decile framework to individuals who enter the observed sample after wave 1 through HILDA's following rules, excluding the top-up samples from this calculation. Specifically, we anchor each TSM and OPM to a wave-1 person and assign each entrant the anchor's wave-1 income decile. Using the partner, parent, and household rules described below, we obtain an anchor for all eligible later entrants in the main sample.

In wave 11, HILDA introduced a substantial top-up sample. Drawn using a design closely aligned with that of the original survey, the top-up added 5,451 enumerated persons, i.e., around 27 per cent of the original sample and 30 per cent of the wave-11 main sample. This provides a complementary diagnostic of whether compositional biases had accumulated over the survey's first 11 waves. In Section 4.2, we compare the distributional characteristics of the top-up and wave-11 main samples, both as observed and after excluding migrants who arrived in Australia after wave 1.

To assess the practical implications for distributional analysis, we estimate a new set of weights that ensure that the weighted shares across origin deciles defined within age groups are exactly 10% in every wave. While this calibration does not guarantee representativeness for other socio-economic dimensions, it provides a direct way to gauge how sensitive median income and inequality estimates are to attrition and sample evolution.

Focusing on attrition from the original wave-1 cohort, we document systematic imbalance across baseline equivalised household income deciles. Relative to the 10% benchmark, the bottom of the distribution becomes under-represented over time: the 1st and 2nd deciles decline to 7–8% each by the end of the panel. In contrast, the 9th and 10th deciles are over-represented, reaching roughly 11.5% and 12%, respectively. Note that these results are based on unweighted estimates.

However, when we construct deciles within baseline age brackets and then aggregate, the decile imbalance is substantially attenuated, indicating that much of the differential attrition is consistent with age and life-course composition effects rather than income rank per se. Finally, when we anchor temporary and additional sample members to wave 1, the decile shares remain close to the 10% benchmark in every wave, approximately 9.1–10.7% in the unweighted series (and similarly tight in the weighted series), with no systematic gradient at either the bottom or the top of the distribution.

The wave-11 top-up provides broadly consistent evidence. The top-up sample has moderately lower mean and median income than the continuing sample, while most inequal-

ity measures and the poverty rate are broadly similar. Estimates obtained by restricting the official weights to top-up members show somewhat larger differences, particularly for the P90/P10 ratio and poverty rate. Because these weights were not designed to represent the top-up independently, the weighted comparisons should be interpreted cautiously. Excluding recent migrants makes the target populations more comparable and generally narrows the differences between the two samples, although the changes are small and not uniform across measures. Overall, the diagnostic provides little evidence of a large divergence in the distributional statistics examined.

Finally, we calculate a new set of weights which ensure that later waves include precisely a weighted 10 per cent of people linked to each within-age group decile in wave 1. We do this for the subsample that excludes the wave 11 top-up sample, starting from the survey weight provided for use with this sub-sample. Consistent with our earlier analysis, we find that weights need to be adjusted downwards for people linked to the top deciles (by around 5%) and upwards for those linked to lower income deciles in wave 1.

However, the impacts on median income and the Gini coefficient are very small. Interestingly, they suggest that the main underestimation of inequality because of differential attrition was in the first decade of the survey, with the effect diminishing thereafter. We interpret this as reflecting income mobility over time attenuating the attrition effect. This mobility includes both personal mobility and the following rules introducing new sample members who have income outcomes that differ from the wave 1 outcomes of the people they are linked to.

We also observe patterns consistent with the absence of new migrants biasing inequality measures downwards. However, none of these effects are large. We conclude that the finding of a generally stable level of Australian income inequality over the first two decades of this century is robust to the challenges examined here.

2 HILDA Design Features

This section summarises the HILDA Survey features most relevant for cross-sectional distributional analysis: (i) the construction of cross-sectional weights used to recover population-representative estimates in each wave, and (ii) the following rules that determine which

individuals remain in-scope as households split, form, and change membership over time (Wooden et al., 2002; Watson and Wooden, 2012; Watson, 2022).

The HILDA Survey began in 2001 with a stratified, multi-stage clustered area sample of households in private dwellings, excluding very remote parts of Australia. Primary sampling units were 488 Census Collection Districts (CCDs), selected within strata defined by state with probability proportional to the number of dwellings (Watson, 2012; Watson and Wooden, 2012). The estimated survey response rate in wave 1 was 66 per cent (Wooden et al., 2002). Subsequent waves have maintained wave-on-wave response rates above 93.5 per cent – a very high response rate for national panel surveys (Summerfield et al., 2025). Because the original sample represents the 2001 private-dwelling population living in Australia in 2001, later entrants to the population, especially post-2001 immigrants and returnees from overseas, are only partially captured through joiners and births, motivating the 2011 top-up sample (Watson and Wooden, 2013).

In this study, we treat the quality of the initial cross-section as given, and focus our assessment on the design of cross-sectional weights and the choice of following rules. These are the two central survey design levers that govern cross-sectional representativeness in later waves of a household panel and estimates of change over time (Lynn et al., 2011; Watson, 2022).

2.1 Cross-sectional Weights for Representativeness

Cross-sectional weights are intended to make each wave of HILDA representative of the contemporaneous in-scope population, despite unequal selection probabilities and selective non-response (Watson, 2012; Summerfield et al., 2025). HILDA provides cross-sectional weights at multiple levels, namely household, enumerated persons, and responding persons. Throughout this paper, our cross-sectional assessment is conducted at the enumerated-person level, because enumerated persons reflect the full set of household members in each wave (including those who are not interviewed, e.g. because they are too young or not available), and this is the conventional basis for distributional analysis of equivalised household income. These weights are 'integrated' in that all household members receive the same weights, and the household and enumerated person weights are constructed jointly.

According to Watson (2012), wave-1 cross-sectional weights are constructed in four main steps:

First, *domain identification* determines which sampled addresses, households, and persons belong to the in-scope population. HILDA begins with the issued sample addresses, removes addresses found to be out of scope,⁴ and adds additional households where multiple households occupy a sampled address. In wave 1, this yielded 11,693 in-scope households, of which 7,682 responded. Within responding households, enumerated-person weights are then defined for all usual residents, including interviewed adults, non-responding adults, and children. This yielded 19,914 enumerated persons in responding households in wave 1 (Watson, 2012; Summerfield et al., 2025).

Second, *initial (design) weights* are calculated as the inverse of each household's probability of selection under HILDA's multi-stage area-based sample design. In wave 1, selection occurred through successive stages: CCDs were selected with probability proportional to size, dwellings were then sampled within listed areas. Where more than three households occupied a sampled dwelling, a random sample of three households was taken. Design weights therefore reflect unequal probabilities of inclusion across geographic areas, dwellings, and, in a limited number of cases, households within dwellings.

Third, the design weights are *adjusted for differential non-response* using auxiliary information available for all selected households, including both responding and non-responding cases. In wave 1, HILDA estimates a logistic regression model of household response. The model includes interviewer-observed dwelling characteristics, such as dwelling type, external condition, and security-related features. It also includes neighbourhood characteristics of the selected CCD, such as geographic location, population density, language composition, labour-force composition, and ABS Socio-Economic Indexes for Areas (SEIFA) (Australian Bureau Of Statistics, 2011). The design weight is multiplied by the inverse of the predicted response probability, so that households resembling non-respondents receive a larger weight adjustment.

Fourth, *calibration to external population benchmarks* aligns the household non-response-adjusted weights with known external totals. HILDA uses the ABS-developed GREGWT procedure (see Bell (2000)) to calibrate the weights simultaneously to multiple household- and person-level benchmarks. The person benchmarks include State, part of State, sex, and age, based on ABS Estimated Resident Population figures, while further benchmarks capture household composition, labour-force status, and marital status. This final step anchors the

⁴This includes vacant or non-residential dwellings, or dwellings where all household members were outside Australia for six months or more.

weighted HILDA sample to the external demographic, household, and labour-market structure of the Australian population. The calibration benchmarks also define the population for which the weights are valid: persons living in very remote areas and, in later waves, persons living in non-private dwellings are excluded from the relevant benchmarks and therefore receive zero cross-sectional weights ([Watson, 2012](#)).

For our income-focused analysis, it is important to underline that HILDA's weights do not benchmark the income distribution directly. Instead, income-related selection is addressed indirectly through adjustment and calibration variables that are correlated with economic resources, including SEIFA, geography, dwelling characteristics, household composition, labour-force status, marital status, age, and sex. The weights therefore improve income representativeness to the extent that these variables capture income-related differences in selection and response, but they do not guarantee representativeness of the income distribution itself.

From wave 2 onward, cross-sectional weights are no longer based on a fresh draw from the population. Each wave's observed cross-section is generated by the original wave 1 sample, subsequent (non-)response outcomes, and household dynamics over time ([Watson, 2012](#)). Prior to wave 11, new individuals enter the cross-section through changes in household composition, such as births, household splits, household formation, and new co-residents of original sample members. Later-wave weights therefore adjust for these observation pathways, then for household non-response, before being calibrated to the same types of external benchmark margins used in wave 1 ([Watson, 2012](#); [Summerfield et al., 2025](#)).

From wave 11, HILDA integrates the original sample with a general top-up sample in the construction of cross-sectional weights. The top-up sample was introduced to refresh the representativeness of the survey, especially for population groups that could not be fully captured through the original wave 1 panel and its following rules, such as post-2001 immigrants. The wave 11 top-up sample is first weighted in a way that is closely aligned, but not identical, to the original wave 1 sample.⁵ The main and top-up samples are then combined by accounting for overlap in the populations, before the resulting cross-sectional weights are calibrated to the same types of external household and person benchmarks used in earlier

⁵Specifically, CCD selection was based on occupied private dwellings, the CCDs were selected from a geographically ordered list rather than from separately defined geographic groups, and the wave-11 top-up used a somewhat extended non-response model.

waves ([Watson, 2012](#); [Watson and Wooden, 2013](#); [Summerfield et al., 2025](#)). From wave 11 onward, weighted cross-sectional estimates therefore refer to the integrated main and top-up sample rather than to the original panel alone, although HILDA also provides weights for the original sample only (which we use for some of our calculations). HILDA subsequently introduced migrant-focused refreshment samples in waves 23 and 24. These samples are incorporated into the released cross-sectional weighting system. Because their members cannot be linked to a wave-1 anchor, they are included in our full-sample estimates but excluded from the origin-decile and recalibration analyses. We therefore interpret results for waves 23 and 24 with caution ([Summerfield et al., 2025](#)).

2.2 Following Rules

Following rules specify who remains eligible to be followed and interviewed as households move, split, and form, and therefore shape how a household panel evolves over time ([Watson, 2022](#)). They matter for cross-sectional analysis because the composition of the observed cross-section in each wave reflects not only the initial sample design and subsequent (non-)response, but also the cumulative implications of how the survey treats household dissolution (splits) and household formation (new co-residents and joiners) ([Watson, 2022](#); [Lynn et al., 2011](#)).

The simplest following rule is to continue following only people in the households interviewed in the initial wave. However, this means that the sample will grow older over time as new children are not added and other individuals who enter the target population after the initial wave, such as new immigrants, are not represented. Expanding the rules to include children and new partners of survey members allows greater scope for longitudinal analysis, but it also requires weights to be inferred for people who were not directly sampled in the first wave. At the broadest end, the survey may follow every new co-resident who enters a household containing an initial sample member. This approach can continuously refresh the sample, especially when attrition is high, but it relies on stronger weighting assumptions and may generate disproportionate long-run sample growth relative to more restrictive following rules ([Watson, 2022](#)). Overall, these design choices involve trade-offs between maintaining cross-sectional representativeness, preserving longitudinal interpretability, and containing fieldwork costs, with long-run effects on sample size and composition.

In HILDA, the wave 1 responding households form the basis of an indefinite-life panel, and the sample is extended over time through rule-based additions ([Summerfield et al., 2025](#)). The set of *continuing sample members* (CSMs) comprises all members of the original wave 1 households (Original Sample Members, OSMs) and *other permanent sample members* (OPMs), who become permanent through specified conversion criteria. Most prominently, these include children born to or adopted by continuing sample members, as well as certain new entrants who qualify for conversion, for example by having a child with a continuing sample member or by meeting the rules for recent arrivals to Australia. All other people who share a household with a continuing sample member in wave 2 or later are classified as *temporary sample members* (TSMs): they are enumerated and, where eligible, interviewed while they co-reside with a continuing sample member, but they do not become permanent panel members by default ([Summerfield et al., 2025](#)). We provide an overview of the number and shares of OSMs, OPMs, and TSMs in Appendix A.

These rules also imply clear exit mechanisms from the survey. Continuing sample members can be absent from a given wave because of non-response, temporary ineligibility (for example, being overseas for an extended period), or death.⁶ Importantly, eligibility to be followed can persist even if an individual is not observed in a particular wave, so some exits are only temporary from the perspective of the cross-section ([Watson, 2022](#)). By contrast, TSMs exit mechanically when they no longer live with a continuing sample member, because they are not followed to a new address once they split off, unless they are converted to permanent sample-member status under the survey's rules ([Summerfield et al., 2025](#)).

The choice of cross-sectional weighting method in a household panel is tightly linked to its following rules, since the rules determine which selection pathways into a wave-*t* household exist and therefore which probabilities the weights need to recover. [Kalton and Brick \(1995\)](#) set out several options, and [Lavalley \(1995\)](#) formalised the *weight share method* (WSM), which has since become the international standard: each wave-*t* household receives a weight derived from the wave 1 design weights of the OSMs, with TSMs and OPMs absorbing a share through co-residence. [Schonlau et al. \(2013\)](#) show that SOEP, BHPS, HILDA and PSID all use WSM, but that its inputs – which wave 1 design weights propagate forward, and to whom – depend on each survey's following rules. Wider rules, such as in the SOEP, expand the set of people who become permanent sample members, at the cost of greater weight vari-

⁶People moving out of the survey scope, such as to remote areas or to nursing homes, can be kept in the survey but with zero weight.

ance and faster sample growth (Schonlau et al., 2011; Watson, 2022). Narrower rules, such as in the PSID, keep weights tighter but leave subpopulations the rules cannot reach, most notably post-baseline immigrants, to be restored through refreshment or top-up samples, whose integration introduces a separate weighting problem of combining or pooling samples (O’Muircheartaigh and Pedlow, 2002; Watson and Wooden, 2013). HILDA sits between these poles, and its weighting procedure reflects this directly: wave 1 design weights are adjusted for the multiple pathways through which wave-t household members could have been observed (a WSM step), then for non-response, then calibrated to external benchmarks, with the wave 11 top-up integrated through a combined-samples approach (Watson, 2012).

For distributional studies, following rules may matter for income representativeness because the entry and exit processes they generate are unlikely to be income-neutral. Household splits, new partnerships, and other changes in co-residence can introduce or remove individuals with systematically different income profiles, and the mix of continuing members and temporary co-residents may shift over time in ways that alter the income distribution observed in each wave.

Since cross-sectional weighting is calibrated to demographic and household benchmarks rather than to the income distribution itself, the extent to which later-wave distributional estimates remain representative depends on how following rules shape the observed cross-section and on how effectively weighting adjustments correct for the resulting selection and compositional change. This motivates the empirical assessment and methodology described in the next section.

3 Method

Our empirical strategy proceeds in three parts. Part 1 quantifies the extent to which cross-sectional income distributions in later waves remain representative of the wave 1 population once we account for (a) differential (non-)response over time and (b) the survey’s following rules that generate later entrants through household splits and formations. Part 2 uses the wave-11 top-up sample as an external diagnostic of whether the continuing panel had drifted away from the broader cross-sectional income distribution by 2011. Part 3 recalibrates cross-sectional enumerated person weights using wave 1 income-decile cells, and estimates the impact of this recalibration on median income and the Gini coefficient over time.

In the absence of suitable calibration data to directly test the income representativeness of the HILDA cross-sectional data, we begin by examining changes in sample characteristics over time, and in particular, how they are correlated with household income. Are people who start in different parts of the income distribution in wave 1 (while controlling for age) more likely to stay in the survey and/or have other new entrants attached to them? Does the existing weighting model sufficiently control the loss of sample balance arising from this? For example, unless adjusted for, the representation of high income groups will increase over time if people in higher income households are less likely to die, more likely to have children, less likely to drop out of the survey or more likely to separate and re-partner.

We begin with the wave 1 enumerated-person cohort and follow these individuals through waves 2–24 without adding later entrants (e.g., temporary co-residents or new permanent members). Using equivalised regular disposable household income⁷ in wave 1, we assign each wave 1 individual to baseline income deciles under two baseline concepts: (i) unweighted equal-count deciles, and (ii) weighted deciles based on wave 1 cross-sectional enumerated person weights. In addition, we construct age-adjusted deciles within wave 1 age groups.⁸ In each subsequent wave, we then compute the within-wave shares of the wave 1 cohort across these baseline deciles, both weighted by the wave-specific cross-sectional enumerated-person weights excluding the wave 11 top-up sample and unweighted by counts. Deviations from 10% within-wave shares are consistent with differential attrition or differential observability along the baseline income distribution.⁹ The weighted within-age decile classification is also used in the subsequent recalibration exercise.

Next, we extend the analysis to include later-wave enumerated persons who can be linked back to a wave 1 “anchor” person and thereby inherit wave 1 baseline deciles.¹⁰ The linkage is implemented as a forward pass through waves 2 – 24.¹¹ In each wave $t > 1$, among individuals who are not yet anchored, we apply the following priority order (1 to 3) to assign an anchor:

⁷This excludes *irregular* income, like redundancy payments, and other unusual lump-sum payments. Income is adjusted by the modified OECD equivalence scale.

⁸Baseline age groups are defined using wave 1 age: 0–14, 15–24, 25–34, 35–44, 45–54, 55–64, 65–74, and 75+ years.

⁹Note that we only use income from wave 1 in this calculation. This means that our calculation is not influenced by any potential regression to the mean of extreme incomes in subsequent years – though this mobility does influence our final reweighting outcomes.

¹⁰It is necessary to link individuals to an anchor person, rather than a household, as the within-age income deciles may differ for enumerated persons in the same household.

¹¹Once an individual is assigned a wave 1 anchor, the anchor is fixed in all subsequent waves.

1. *Partner rule*: If the individual's partner is already anchored, the individual inherits the partner's wave 1 anchor and baseline deciles.
2. *Parent rule*: If an anchored parent is observed, the individual inherits that parent's wave 1 anchor and baseline deciles. If only one parent was surveyed in prior waves, we anchor the child to that parent. If both parents were observed, we prioritise the mother's anchor over the father's.
3. *Household rule*: If at least one anchored person is present in the individual's current household, the individual inherits the anchor from a single deterministic "household anchor" person selected within that household.

The deterministic household anchor is defined as follows: within each household, among persons who are already anchored, we select exactly one anchor source by

- (a) preferring adults (age ≥ 18),
- (b) preferring the oldest individual,
- (c) breaking any remaining ties by a stable ordering on the personal identifier.

In this way, we are able to link all enumerated persons to a wave 1 anchor, and they are assigned the decile value of their anchor. We then recompute the within-wave decile shares analogously to Step (a), now also including all anchored later entrants.

Top-up sample members in wave 11 as well as migrant sample members in waves 23 and 24 are not included as they cannot be explicitly anchored to wave 1. For this reason, the top-up sample is not assessed in this step, and we use the enumerated-person weights that exclude the wave 11 top-up sample.¹²

To assess whether the continuing sample had diverged from a newly drawn cross-sectional sample by wave 11, we compare distributional statistics for equivalised regular disposable household income between the continuing and top-up samples. By design, the income distribution of the top-up sample should be similar to the continuing sample in wave 11, once recent migrants are excluded. Comparing the distributional statistics of the adjusted top-up with those of the continuing sample therefore provides a check for cross-sectional differences at that point in the panel's life.

¹²This introduces a small bias in the last two waves, particularly wave 24, as the enumerated-person weights include the wave 23 and 24 top-ups, but our linking to wave 1 does not.

HILDA does not provide separate weights for the top-up sample. We therefore use the official combined-sample weights for this comparison, but interpret the resulting estimates cautiously. Another caveat is that HILDA does not identify exactly whether migrants arrived after 2001. The closest available measure identifies migrants who arrived after 1999, which we use as an imperfect proxy. This is stricter than the ideal restriction, because it also excludes migrants who arrived in 2000 or 2001 and are part of the original wave 1 target population. Nonetheless, after applying this restriction, the top-up sample should approximate the same underlying population that the continuing sample was designed to represent in 2011. The comparison therefore remains informative about whether the continuing sample had drifted from the cross-sectional income distribution by wave 11.

Finally, we construct alternative enumerated-person weights for the wave-1-linked sample. We begin with HILDA's published main-sample weight: the standard enumerated-person weight in waves 1–10 and the weight excluding the wave-11 top-up from wave 11 onward. We reproduce two vectors of person-level calibration controls.

We preserve two sets of person-level calibration controls. The first comprises 250 cells defined by geography, sex, and geography-specific age groups. The second comprises 30 cells defined by state and labour-force status, with unemployment and non-participation combined for the Northern Territory and the Australian Capital Territory. For each wave, the corresponding control totals are obtained by summing the published main-sample weights within these cells.¹³

We then replicate these calibration vectors by summing the weights across these characteristics, and supplement them with a vector with 10 per cent of the population in each of 10 cells.¹⁴ We then generate new weights using the ABS GREGWT macro,¹⁵ which fits each vector in turn until convergence is reached. This ensures that the weighted data reflect the existing calibration variables as well as ensuring that 10 per cent of the weighted sample in each wave is in each wave 1 within-age decile category.

Finally, we compare income medians and the Gini coefficient using the existing enumerated-person weight, and when using this new weight.

¹³Where current-wave benchmark inputs are missing, we use information from earlier waves only. Cases that remain unclassified are retained in a separate residual category.

¹⁴Not precisely 10 per cent because discrete weights mean that the percentage in each decile is slightly different.

¹⁵We use the exponential distance function, but otherwise the default convergence parameters. Convergence required fewer than three iterations in each wave.

4 Results

This section presents results from the three tests: the evolution of the main sample across baseline income deciles, the comparison of the continuing and wave-11 top-up samples, and the recalibration of the main-sample weights by baseline income position.

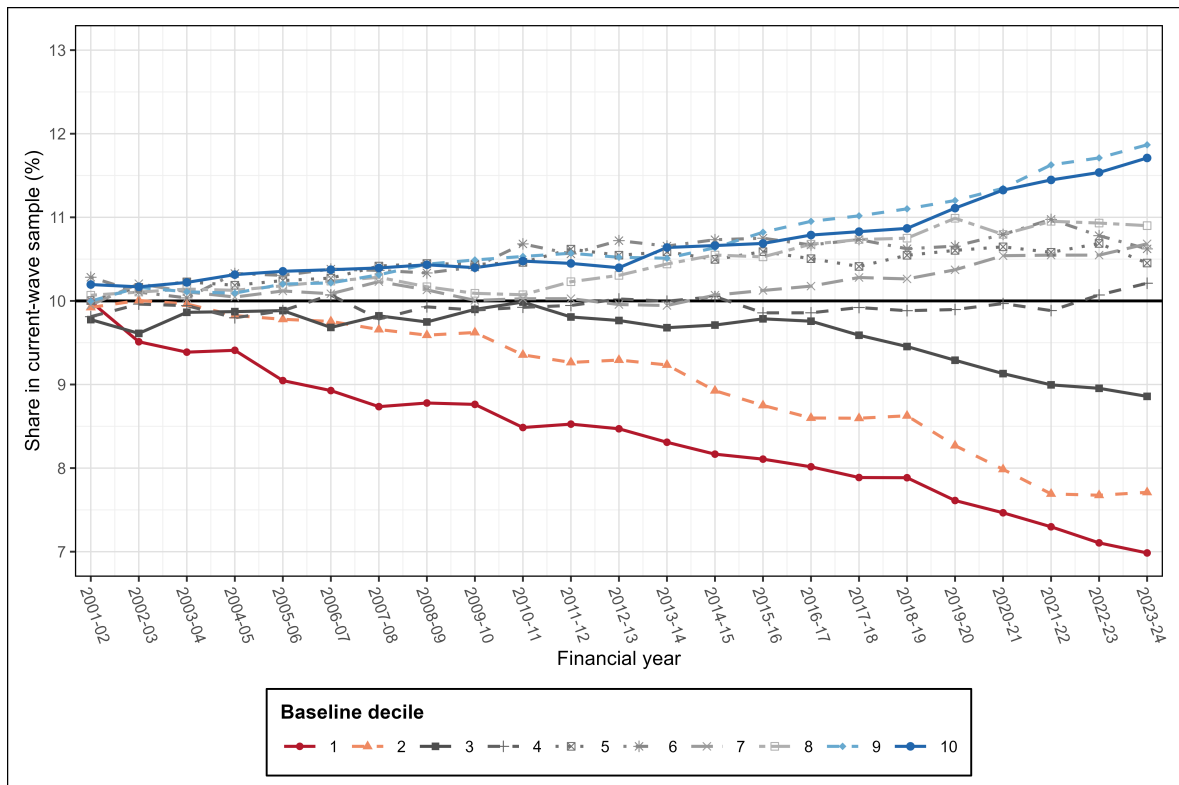
4.1 Stability of the Main Sample Across Baseline Income Deciles

The first test treats the wave-1 income distribution as an origin distribution and asks how evenly that origin distribution is represented in later waves. The purpose is not to estimate income inequality directly, but to assess whether panel evolution is selective with respect to baseline income position. We track later-wave composition across fixed wave-1 income deciles, first for the wave-1 cohort and then after accounting for baseline age composition and linked later entrants. If later-wave observation and rule-based entry were unrelated to baseline income rank, each decile would continue to account for approximately 10% of the observed sample. Although this is a stylised benchmark, it provides an interpretable reference for the magnitude and direction of distributional drift over the life of the panel.

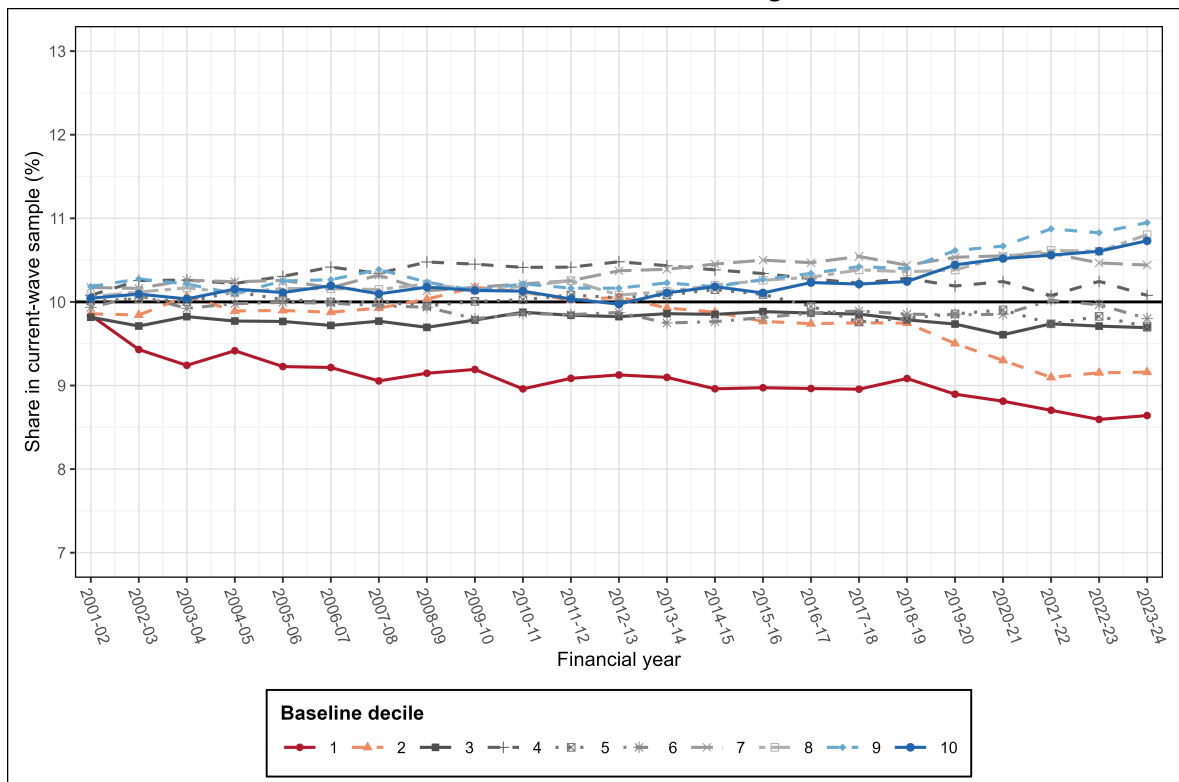
We use a common visualization across specifications. The x-axis plots the financial year based on HILDA waves b to x (2002–2024), while the y-axis reports the within-wave shares of individuals across the income deciles defined in wave 1. Figure 2 presents the simplest specification: unweighted baseline deciles are defined using equivalised disposable household income at the enumerated-person level in wave 1, and the analysis follows only the wave-1 cohort, excluding later entrants. Panel 2a reveals a differential gradient across income deciles over time. The lower tail becomes progressively under-represented, most notably deciles 1 and 2. In contrast, the upper deciles rise gradually. By wave 24, the shares of deciles 1 and 2 have fallen to approximately 7.0% and 7.7%, respectively, while those of deciles 9 and 10 have risen to approximately 11.9% and 11.7%.¹⁶

A key interpretation issue is that baseline income deciles contain strong life-cycle differences. Lower equivalised household income in wave 1 is mechanically more common among younger individuals (e.g., students, early-career workers, and young families) and among the oldest old, while middle-to-upper deciles are more concentrated in prime work-

¹⁶Note that we do not apply weights at this stage, as following wave 1 only would require longitudinal weights. As this study is interested in the cross-sectional representativeness, we focus simply on the degree of differential attrition in the data.



(a) Shares of income deciles (unweighted)



(b) Shares of within-age group deciles (unweighted)

Notes: Based on HILDA Release 24; calculations by the authors. Each panel shows the unweighted distribution of observed members of the wave-1 cohort across (unweighted) income deciles defined in wave 1. Panel (a) uses deciles defined for the full wave-1 cohort, while Panel (b) uses deciles defined separately within age groups. Baseline deciles are based on equivalised regular disposable household income. Later entrants are excluded.

Figure 2. OSM shares by equivalised household income deciles

ing ages. Because mortality, moves out of scope, and survey nonresponse vary with age and household stage, differential observation across age groups can generate changes in baseline-decile shares even if attrition is income-neutral conditional on age. We therefore also define income deciles within baseline age groups.

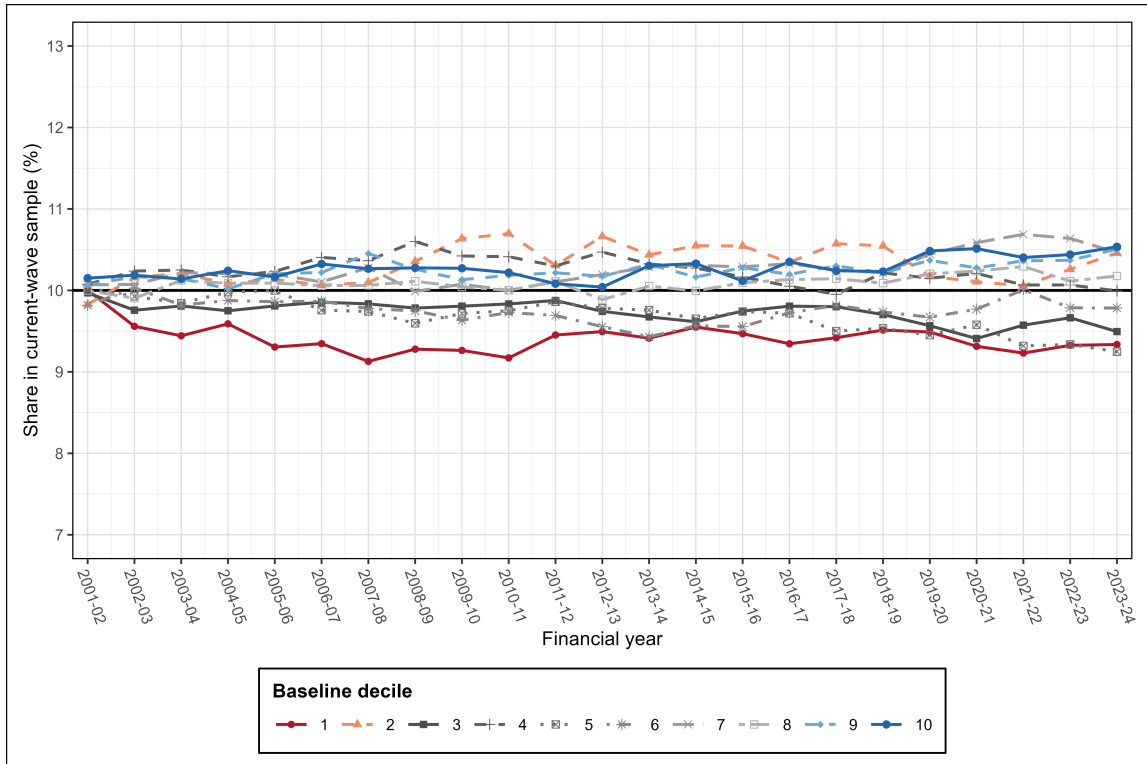
Panel 2b shows that defining deciles within age groups substantially reduces the imbalance. The upper deciles remain below 11% throughout, while the second decile remains above 9%. Decile 1 nevertheless declines gradually to approximately 8.6% by the final wave.

We interpret these results as evidence that accounting for baseline age profiles is crucial, and that a substantial share of the discrepancies observed in Figure 2 reflects life-course composition effects rather than income-related attrition alone. The next question is whether later entrants generated by HILDA's following rules reduce or amplify the remaining imbalance across deciles when they are incorporated into the observed cross-section. To examine this, we anchor all later entrants to wave 1, as described in Section 3. The resulting estimates are presented in Figure 3.

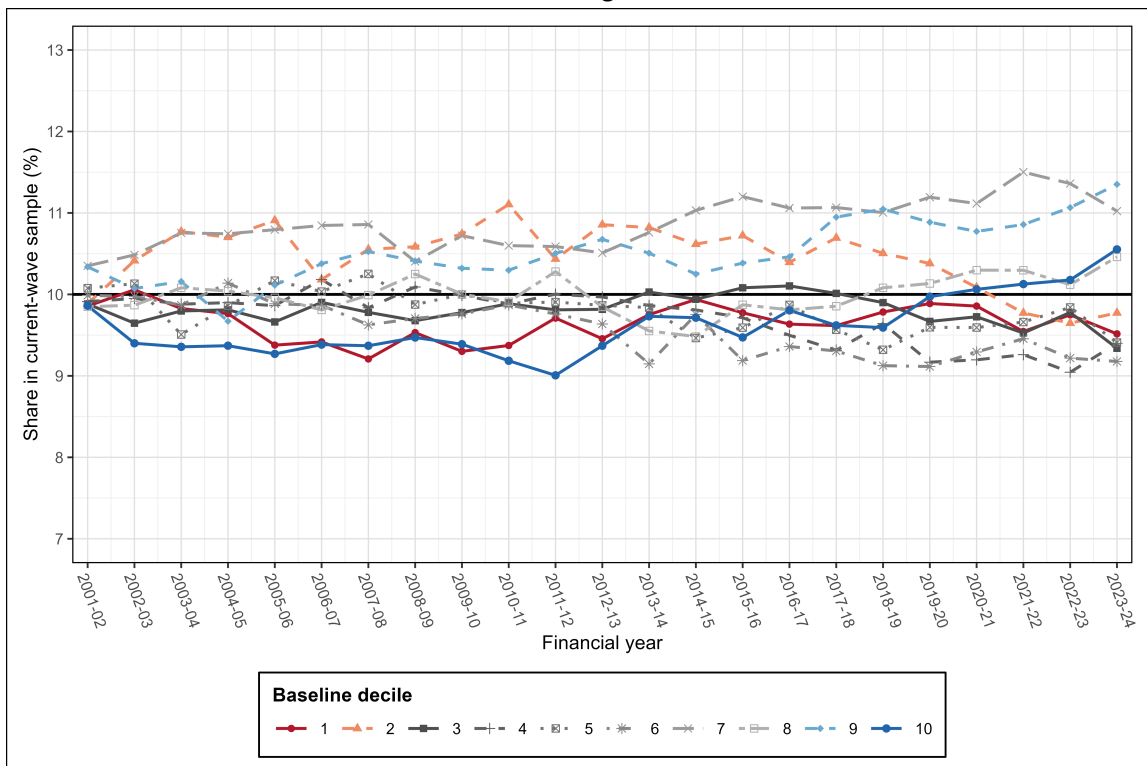
Including linked later entrants further reduces the imbalance. In the unweighted specification in Panel 3a, decile shares remain between approximately 9.1% and 10.7%, with no systematic gradient across the income distribution. In particular, the under-representation of decile 1 and the more pronounced discrepancies in the latest waves largely disappear. Under the weighted specification in Panel 3b, the shares range from approximately 9.0% to 11.5%. Although the weighted series displays somewhat greater variation, the lower two deciles remain close to the 10% benchmark.

The findings are robust to alternative age-bin definitions and anchoring rules. Appendix C replicates the main figures using five broader baseline age groups to define wave-1 income deciles, with results that are quantitatively very similar. The conclusions are also unchanged when we vary the anchoring procedure in Appendix D, either by anchoring children before partners or by restricting parental links to biological parents only. We therefore do not view the results as sensitive to these methodological choices.

Taken together, these findings suggest that there is no substantial income-related drift in the cross-section once later entrants generated by HILDA's following rules are incorporated appropriately. Section 4.3 examines whether the residual imbalance materially affects estimates of median income and inequality.



(a) Unweighted



(b) Weighted at enumerated-person level

Notes: Based on HILDA Release 24; calculations by the authors. Each panel shows the within-wave distribution of the wave-1-linked sample across income deciles defined within wave-1 age groups. Later entrants inherit the wave-1 decile of their assigned anchor. Panel (a) is unweighted, while Panel (b) uses wave-specific enumerated-person weights excluding the wave-11 top-up sample. Top-up and migrant-refreshment sample members are excluded.

Figure 3. Shares across wave-1 within-age income deciles, including TSMs and OPMs

4.2 Top-Up Sample Comparisons

The preceding analysis focuses on the evolving main sample and excludes the general population top-up introduced in wave 11, as well as the migrant refreshment samples added in waves 23 and 24. We now use the wave-11 top-up as a complementary diagnostic of cross-sectional representativeness. Since the top-up was newly drawn using a design closely aligned with that of the original sample, large differences between the top-up and the continuing non-top-up population would be consistent with compositional drift in the continuing panel by wave 11.¹⁷

We present both unweighted and weighted comparisons. The unweighted results compare the observed income distributions of the two samples. For the weighted results, estimates for the non-top-up population use HILDA's enumerated-person weights excluding the top-up sample. HILDA does not, however, provide weights designed to represent the top-up independently. We therefore calculate weighted top-up estimates by restricting the official integrated-sample weights to top-up members. These estimates provide a useful diagnostic but should be interpreted cautiously because the restricted weights do not constitute a separately calibrated weighting system for the top-up.

The target populations also differ because the top-up includes migrants who arrived after wave 1 and who could not generally enter the continuing sample. We therefore repeat the comparison after excluding recent migrants from the top-up. HILDA does not identify precisely whether migrants arrived after 2001, so we use the available `anaf99` indicator as a proxy. This restriction also excludes migrants who arrived in 2000 or 2001 and is therefore somewhat broader than the ideal adjustment.

Table 1 reports the unweighted and weighted results in Panels A and B, respectively. The first row reports distributional statistics for the non-top-up population, while the remaining rows express the corresponding top-up estimates as percentages of the non-top-up estimates. A ratio of 100 therefore indicates equality between the two groups.

The top-up sample has moderately lower mean and median income than the non-top-up population, but the differences are moderate in size. In the unweighted comparison, the top-up mean and median are 96.2% and 95.8% of the corresponding non-top-up estimates. The weighted comparison points in the same direction, with ratios of 93.7% for the mean

¹⁷Such differences could also reflect differences in sample design, coverage, or nonresponse between the original and wave 11 top-up sample.

Group	Mean	P50	P90/P10	Gini	Top 10	Top 1	Pov50%
<i>Panel A: Unweighted</i>							
Non-top-up population	46,887 (721)	40,768 (475)	4.190 (0.055)	0.311 (0.004)	23.71 (0.36)	4.77 (0.28)	12.40 (0.65)
Top-up / Non-top-up ratio	96.2 (2.7)	95.8 (2.5)	101.2 (5.1)	102.3 (3.5)	102.3 (3.9)	96.2 (16.1)	98.5 (10.5)
Top-Up excl. recent mig / Non-top-up ratio	96.5 (2.8)	95.8 (2.6)	100.8 (4.6)	102.2 (3.5)	102.4 (3.9)	99.4 (24.0)	98.6 (9.9)
<i>Panel B: Weighted</i>							
Non-top-up population	47,535 (2,402)	41,221 (2,262)	4.000 (0.604)	0.305 (0.019)	23.39 (1.27)	4.81 (0.80)	11.24 (3.41)
Top-up / Non-top-up ratio	93.7 (5.8)	94.5 (6.4)	110.9 (17.3)	105.3 (7.5)	104.0 (8.0)	91.5 (16.9)	111.7 (39.1)
Top-up excl. recent mig / Non-top-up ratio	94.5 (5.9)	94.9 (5.5)	108.6 (15.4)	104.4 (7.5)	103.4 (6.1)	93.0 (17.5)	110.8 (27.7)

Notes: Based on HILDA Release 24; calculations by the authors. The analysis is conducted at the enumerated-person level using equivalised household disposable regular income. Non-top-up rows report levels; top-up rows report ratios relative to the non-top-up population, expressed in percent. Standard errors are reported in parentheses. Weighted standard errors use the HILDA grouped JK1 jackknife with the 45 official replicate weights. For the non-top-up row, point estimates use the non-top-up weight and replicate estimates are constructed from the corresponding official replicate weights restricted to the same subset. Weighted top-up estimates use the full-sample enumerated-person weights restricted to top-up members. Recent migrants are excluded using the `anaf99` migration proxy.

Table 1. Wave 11 distributional statistics: non-top-up population and top-up ratios

and 94.5% for the median. This suggests that the wave-11 top-up is somewhat less affluent than the continuing non-top-up sample, but not sharply different in its overall income level.

Most other distributional statistics are similar in the unweighted comparison. The P90/P10 ratio is only 1.2% higher in the top-up sample, while the Gini coefficient and top-10% share are both approximately 2.3% higher. The top-1% share and poverty rate are slightly lower. The weighted estimates show somewhat larger differences, particularly for the P90/P10 ratio and poverty rate, which are 10.9% and 11.7% higher in the top-up sample. However, the associated standard errors are large, and these differences are not estimated precisely. We therefore interpret these differences cautiously and do not view them as evidence of a clear distributional break between the top-up and non-top-up populations.

Excluding recent migrants generally brings the top-up estimates slightly closer to the non-top-up estimates, particularly in the weighted comparison. The changes are nevertheless small and not uniform across measures. This suggests that differences in recent-migrant composition account for only a limited part of the observed differences between the two groups, subject to the limitations of the available migration proxy.

Overall, Table 1 provides little evidence that the continuing non-top-up population had drifted strongly away from a newly drawn cross-sectional sample by wave 11. The top-up sample appears slightly lower-income and, in the weighted results, somewhat more unequal at the lower tail. However, the broad pattern is one of similarity rather than discontinuity. This supports the interpretation that the wave-11 top-up refreshed the sample without revealing a major break in income representativeness.

4.3 Reweighting with Wave 1 Income as a Calibration Benchmark

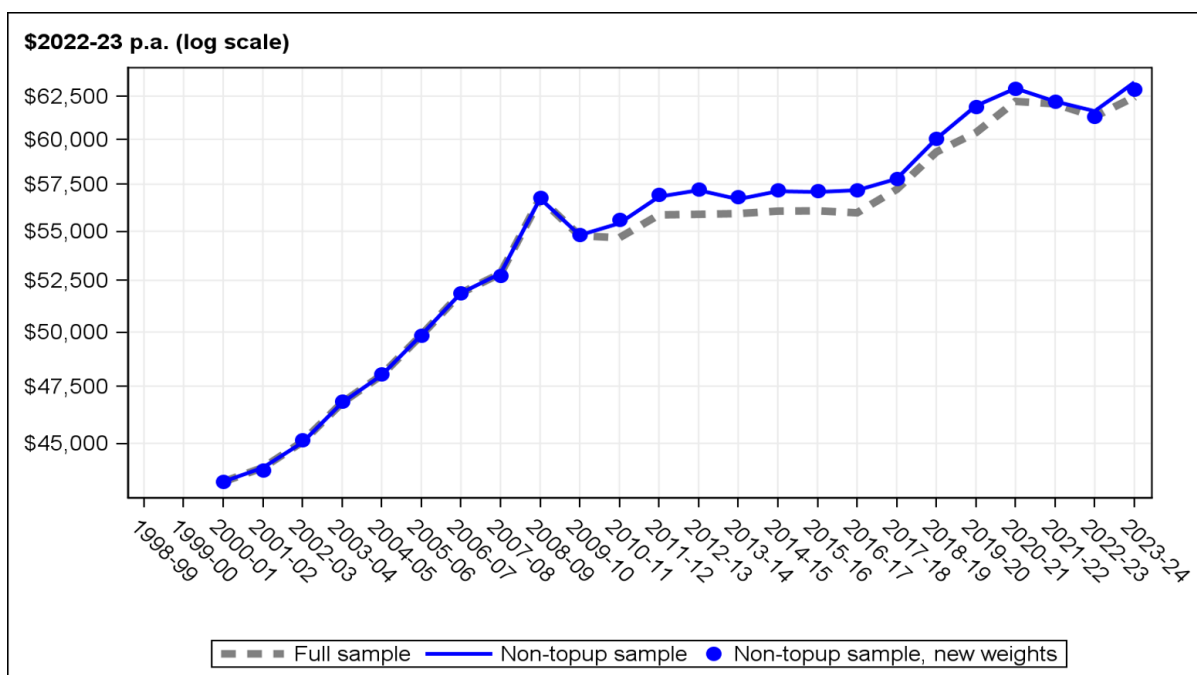
How much does the imbalance observed in Section 4.1 matter for substantive distributional estimates? In this section, we evaluate this by calculating new weights which ensure that a weighted 10 per cent of the population in each wave is linked to a person in each within-age decile in wave 1. Since this linkage only applies to the main, or non-top-up sample, we focus our attention on this sample and compare our results to those obtained when using the existing enumerated person weight for the main sample.

These new weights tend to be lower for people linked to wave 1 individuals from the top half of the income distribution, reflecting the differential attrition described above. To take the most extreme example in Figure 3b, in wave 22, the survey weights implied that 11.5 per cent of people were linked to decile 7 in wave 1. Following joint recalibration to the demographic, labour-force, and origin-decile controls, the new weight for this group is approximately 85 per cent of the published weight.

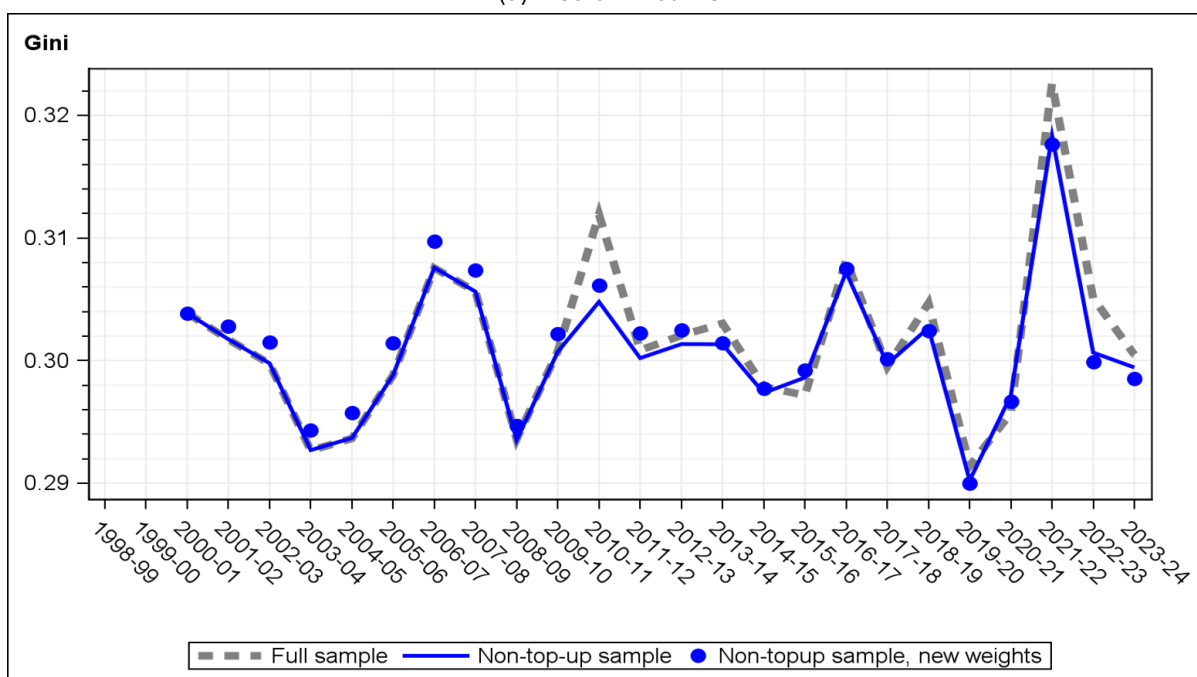
In Figure 4a we show the median equivalised household income of people, for the full sample (dashed), the non-top-up sample (blue line) and the new weights for the non-top-up sample (blue dots). Figure 4b presents the corresponding Gini coefficients.

For median income, there is negligible difference between the estimates based on the new and old weights. There are, however, some small differences in the Gini coefficients. In most years, moving to the new weights would increase measured inequality, but only by a small amount – much smaller than the year-to-year fluctuation. (Note that the estimates for the last two years should be treated with caution, as the starting weights for our analysis are calibrated against the sample including the (small) wave 23 and 24 top-ups, but these top-ups are not included in these results as they do not have a linked first wave person).

Interestingly, the impact of the reweighting on the Gini is greater in the first decade of the HILDA survey, peaking at wave 7 and then declining. This is different to the pattern shown in



(a) Median income



(b) Gini coefficient

Notes: Compiled by authors using HILDA Release 24. 'Full sample' is the full HILDA cross-sectional sample, enumerated person weights. 'Non-top-up sample' is the sample excluding the wave 11 top-up sample, enumerated person weight (includes wave 23 and 24 top-ups). 'Non-top-up sample, new weights' is this sample excluding the wave 23 and 24 top-ups, with weights adjusted to ensure equal fractions of the sample linked to each within-age group decile in wave 1. Household (regular) equivalised income after income tax (modified OECD equivalence scale), with income deflated by the CPI. Negative income rounded to zero.

Figure 4. Median and Gini coefficient of equivalised household disposable income under alternative samples and weights

3b, which shows a slightly increasing dispersion over time. This reflects the varying impacts of attrition, recruitment and income mobility over time. While attrition might cumulate its impact over time, mobility and new sample entrants will work in the opposite direction. If people change their place in the income distribution as time goes on then weights based on wave 1 income position will have less impact on the later wave income distribution. Similarly for new sample members. For example, new spouses might have very different incomes, and children might leave home and move into poorer households, even compared to other people the same age.

The figures also show the effect of excluding the wave-11 top-up and subsequent migrant refreshment samples, which is more substantial than the effect of correcting for the differential attrition on which this paper focuses. From wave 11, the full sample includes the top-up sample, including migrants who arrived in Australia after 2001 and who are not generally represented in the continuing sample. Median income in the full sample is slightly lower than in the continuing non-top-up sample, although the difference attenuates over time. The Gini coefficient is also noticeably higher in wave 11, but this difference attenuates quickly. These patterns are consistent with the weighted top-up comparison in Table 1: in wave 11, the top-up median is 94.5% of the non-top-up estimate, while its Gini coefficient is 5.3% higher. Excluding recent migrants from the top-up changes these ratios only slightly, to 94.9% and 4.4%, respectively, suggesting that the initial differences reflect the broader composition of the top-up sample rather than its migrant component alone.

Overall, the conclusion from these two figures is that differential sample attrition and recruitment associated with initial starting position in the income distribution does not bias estimates of trends in median income and inequality of income across Australian households since 2001. Similarly, the impact of not including migrants does not appear to have a substantial impact on trend estimates.

5 Conclusion

This paper examined whether a leading household panel survey can maintain cross-sectional representativeness across the income distribution over time. Using the HILDA Survey, we assessed how cumulative attrition, survey following rules, and top-up samples shape the distribution of equivalised regular disposable household income when annual panel waves

are used for cross-sectional analysis. Our focus was not on whether HILDA remains representative in every dimension, but on a narrower and policy-relevant question: whether the income distribution observed in later waves drifts systematically away from the wave-1 cross-section.

Tracking the wave-1 cohort alone reveals a clear imbalance across baseline income deciles. The lowest deciles, particularly the first and second, become increasingly under-represented over time, while the upper deciles become over-represented. Accounting for baseline age profiles, however, substantially reduces this imbalance, indicating that much of the apparent drift reflects differences in life-cycle composition rather than income-related attrition alone.

A second main result is that HILDA's following rules appear to mitigate rather than amplify the remaining imbalance. When later entrants were linked to wave-1 individuals, the decile shares in subsequent waves remained close to the 10% benchmark in both the weighted and unweighted specifications, with no strong systematic gradient at either tail of the distribution. This suggests that the additions to the sample generated by household formation, co-residence, and permanent changes in household membership help preserve the income distribution more effectively than an analysis restricted to the wave-1 cohort would imply. The subsequent reweighting exercise provides a substantive check on this interpretation: imposing exact balance across the anchored origin-decile groups has negligible effects on median income and only small effects on the Gini coefficient. More generally, however, the distributional consequences of following rules for cross-sectional income estimates remain underexplored. Our findings are specific to HILDA: surveys with broader following rules, such as the SOEP, or with more restrictive rules, such as the PSID, may generate different patterns of entry and therefore different effects on the observed income distribution. The diagnostic framework developed here can be used to examine these differences, but the substantive findings should not be assumed to transfer directly to other panels.

The wave-11 top-up comparison provides supporting evidence. The top-up sample has moderately lower mean and median income than the non-top-up population, while most inequality and poverty measures are broadly similar in the unweighted comparison. The weighted estimates show somewhat larger differences, particularly for the P90/P10 ratio and poverty rate, but are imprecisely estimated and rely on integrated-sample weights that were not calibrated to represent the top-up independently. Excluding recent migrants generally

narrows the differences only slightly. Overall, this diagnostic provides little evidence of a large divergence by wave 11, although it is not a definitive test of population representativeness.

Finally, by recalibrating the main-sample weights to restore balance across wave-1 within-age income deciles, we find little evidence that differential representation by baseline income position materially biases the estimated trends in median income or the Gini coefficient. The effect on inequality is greatest during HILDA's first decade and subsequently diminishes. Nevertheless, our analysis takes the quality of the initial HILDA cross-section as given. [Watson and Wooden \(2004\)](#) report that managerial and professional groups were over-represented in wave 1, which may help explain why HILDA records higher household incomes than the SIH. A persistent difference in income levels does not by itself imply biased trends. Trend estimates could, however, be affected if groups omitted or under-represented in the initial sample, and not adequately captured by subsequent top-ups, experienced systematically different income growth or distributional changes. Because these counterfactual trajectories of groups absent from the baseline sample cannot be observed in HILDA, our diagnostics cannot test this possibility. Subject to this qualification, our findings show that HILDA remains a strong basis for examining trends in household income and inequality, provided that researchers attend to sample definition, age composition, later entrants, and the treatment of top-up samples.

References

- Australian Bureau Of Statistics (2011). Socio-economic indexes for areas (SEIFA). *Canberra: Australian Bureau of Statistics*.
- Bell, P. (2000). Weighting and standard error estimation for ABS household surveys. *ABS Methodology Advisory Committee: Australian Bureau of Statistics*.
- Davidson, P., Naidoo, Y., and Bradbury, B. (2025). Measuring Poverty in Australia — The Role of Income. *Australian Economic Review*, 58:S45–S57.
- Kalton, G. and Brick, J. M. (1995). Weighting schemes for household panel surveys. *Survey Methodology*, 21(2):33–34.

- Kaplan, G., La Cava, G., and Stone, T. (2018). Household economic inequality in Australia. *Economic Record*, 94(305):117–134.
- Lavalley, P. (1995). Cross-sectional weighting of longitudinal surveys of individuals and households using the weight share method. *Survey Methodology*, 21(1):25–32.
- Longmuir, M. (2023). Fair crack of the whip? The distribution of augmented wealth in Australia from 2002 to 2018. *The Journal of Economic Inequality*, 21(4):835–866.
- Lynn, P. et al. (2011). *Maintaining cross-sectional representativeness in a longitudinal general population survey*. Institute for Social and Economic Research, University of Essex Colchester, UK.
- O’Muircheartaigh, C. and Pedlow, S. (2002). Combining samples vs. cumulating cases: a comparison of two weighting strategies in NLSY97. In *American Statistical Association Proceedings of the Joint Statistical Meetings*, pages 2557–2562. American Statistical Association.
- Rose, D. (2000). Researching social and economic change. *The Uses of Household Panel Studies*.
- Saunders, P. and Bradbury, B. (2006). Monitoring trends in poverty and income distribution: Data, methodology and measurement. *The Economic Record*, 82(258):341–364.
- Schonlau, M., Kroh, M., and Watson, N. (2013). The implementation of cross-sectional weights in household panel surveys. *Statistics Surveys*, 7:37–57.
- Schonlau, M., Watson, N., and Kroh, M. (2011). Household survey panels: how much do following rules affect sample size? In *Survey Research Methods*, volume 5, pages 53–61.
- Summerfield, M., Garrard, B., Kamath, R., Nesa, M. K., Macalalad, N., Watson, N., and Wilkins, R. (2025). *HILDA User Manual – Release 24*. Melbourne Institute: Applied Economic and Social Research, University of Melbourne, Melbourne, Australia.
- Trivellato, U. (1999). Issues in the design and analysis of panel studies: A cursory review. *Quality and Quantity*, 33(3):339–351.
- Watson, N. (2012). Longitudinal and cross-sectional weighting methodology for the HILDA Survey.

- Watson, N. (2022). Choosing who to follow: The long-run impact of following rules on the sample size and composition of household panel surveys. In *Survey Research Methods*, volume 16, pages 301–328.
- Watson, N. and Wooden, M. (2004). Assessing the Quality of the HILDA Survey Wave 2 Data. Technical Report 5/04, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.
- Watson, N. and Wooden, M. (2013). Adding a Top-Up Sample to the Household, Income and Labour Dynamics in Australia Survey. *Australian Economic Review*, 46(4):489–498.
- Watson, N. and Wooden, M. P. (2012). The HILDA survey: a case study in the design and development of a successful household panel survey. *Longitudinal and life course studies*, 3(3):369–381.
- Wilkins, R. (2014). Evaluating the Evidence on Income Inequality in Australia in the 2000s. *Economic Record*, 90(288):63–89.
- Wilkins, R. (2015). Measuring income inequality in Australia. *Australian Economic Review*, 48(1):93–102.
- Wooden, M., Freidin, S., and Watson, N. (2002). The household, income and labour dynamics in Australia (HILDA) survey: Wave 1. *Australian Economic Review*, 35(3).

Online Appendix

A Sample Composition by Wave

Wave	Total obs.	Enum.persons	OSM	OPM	TSM
a	19,914	19,914	19,914	0	0
b	18,295	18,295	17,164	339	792
c	17,690	17,690	15,981	644	1,065
d	17,209	17,209	15,106	956	1,147
e	17,467	17,467	14,911	1,313	1,243
f	17,453	17,453	14,458	1,646	1,349
g	17,280	17,280	13,901	2,004	1,375
h	17,144	17,144	13,440	2,357	1,347
i	17,632	17,632	13,305	2,764	1,563
j	17,855	17,855	13,022	3,142	1,691
k	23,415	23,415	18,119	3,502	1,794
l	23,182	23,182	17,279	3,989	1,914
m	23,299	23,299	16,678	4,496	2,125
n	23,114	23,114	16,156	4,849	2,109
o	23,305	23,305	15,762	5,349	2,194
p	23,508	23,508	15,405	5,846	2,257
q	23,443	23,443	14,999	6,249	2,195
r	23,269	23,269	14,414	6,578	2,277
s	23,265	23,265	13,975	6,966	2,324
t	22,957	22,957	13,434	7,204	2,319
u	22,451	22,451	12,815	7,267	2,369
v	21,751	21,751	12,100	7,329	2,322
w	21,833	21,833	11,891	7,500	2,442
x	23,759	23,759	13,556	7,569	2,634
Wave	OSM (%)	OPM (%)	TSM (%)	Obs. income	Pos. hhwt alt
a	100.0	0.0	0.0	19,914	19,839
b	93.8	1.9	4.3	18,295	18,198
c	90.3	3.6	6.0	17,690	17,578
d	87.8	5.6	6.7	17,209	17,095
e	85.4	7.5	7.1	17,466	17,349
f	82.8	9.4	7.7	17,453	17,329
g	80.4	11.6	8.0	17,280	17,173
h	78.4	13.7	7.9	17,144	17,019
i	75.5	15.7	8.9	17,632	17,513
j	72.9	17.6	9.5	17,855	17,731
k	77.4	15.0	7.7	23,415	23,282
l	74.5	17.2	8.3	23,182	23,002
m	71.6	19.3	9.1	23,299	23,135
n	69.9	21.0	9.1	23,114	22,942
o	67.6	23.0	9.4	23,305	23,138
p	65.5	24.9	9.6	23,507	23,310
q	64.0	26.7	9.4	23,443	23,275
r	61.9	28.3	9.8	23,269	23,095
s	60.1	29.9	10.0	23,265	23,087
t	58.5	31.4	10.1	22,957	22,779
u	57.1	32.4	10.6	22,451	22,277
v	55.6	33.7	10.7	21,750	21,598
w	54.5	34.4	11.2	21,833	21,691
x	57.1	31.9	11.1	23,759	23,399

Notes: Based on HILDA Release 24; calculations by the authors. OSM, OPM, and TSM are defined using `hhsm`. Counts include Top-up sample. Obs. income counts rows with non-missing equivalised household disposable income. Pos. hhwt alt counts rows with a positive enumerated-person weight excluding the top-up sample.

Table A1. Sample composition by wave

B Retention of Wave-1 Cohort

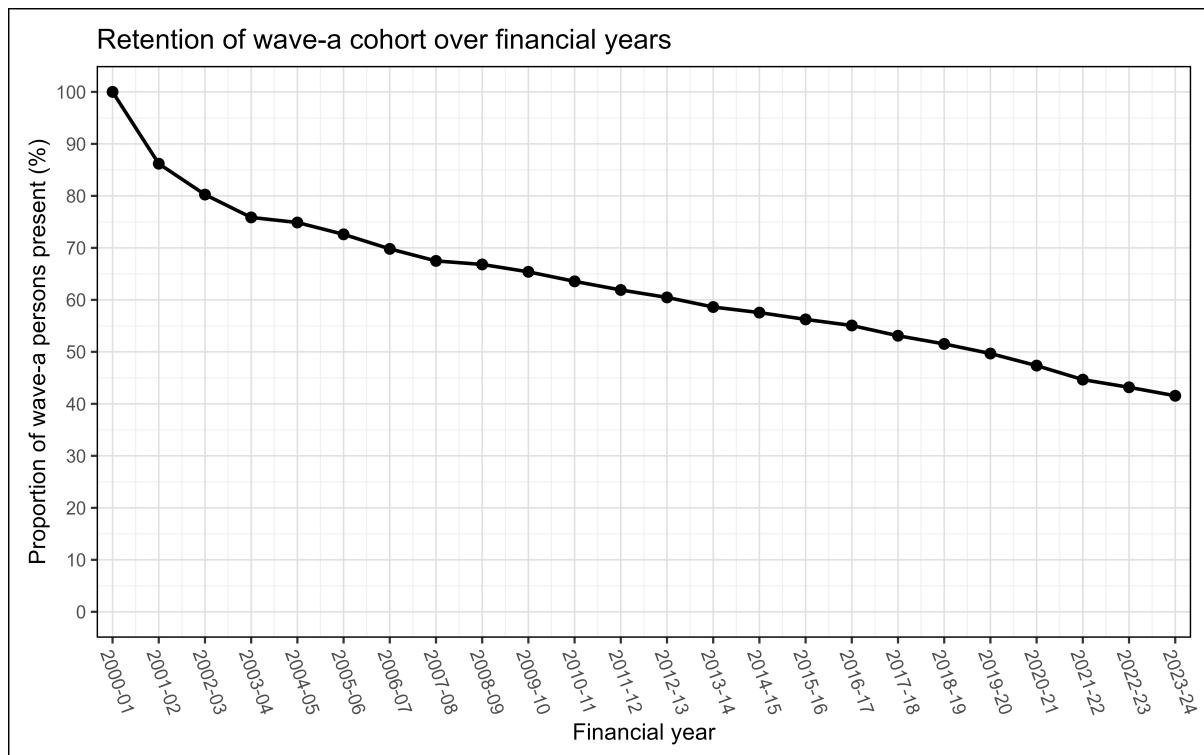


Figure A1. Retention rates of the wave-1 cohort in following waves (unweighted)

C Broader Age-Bin Robustness

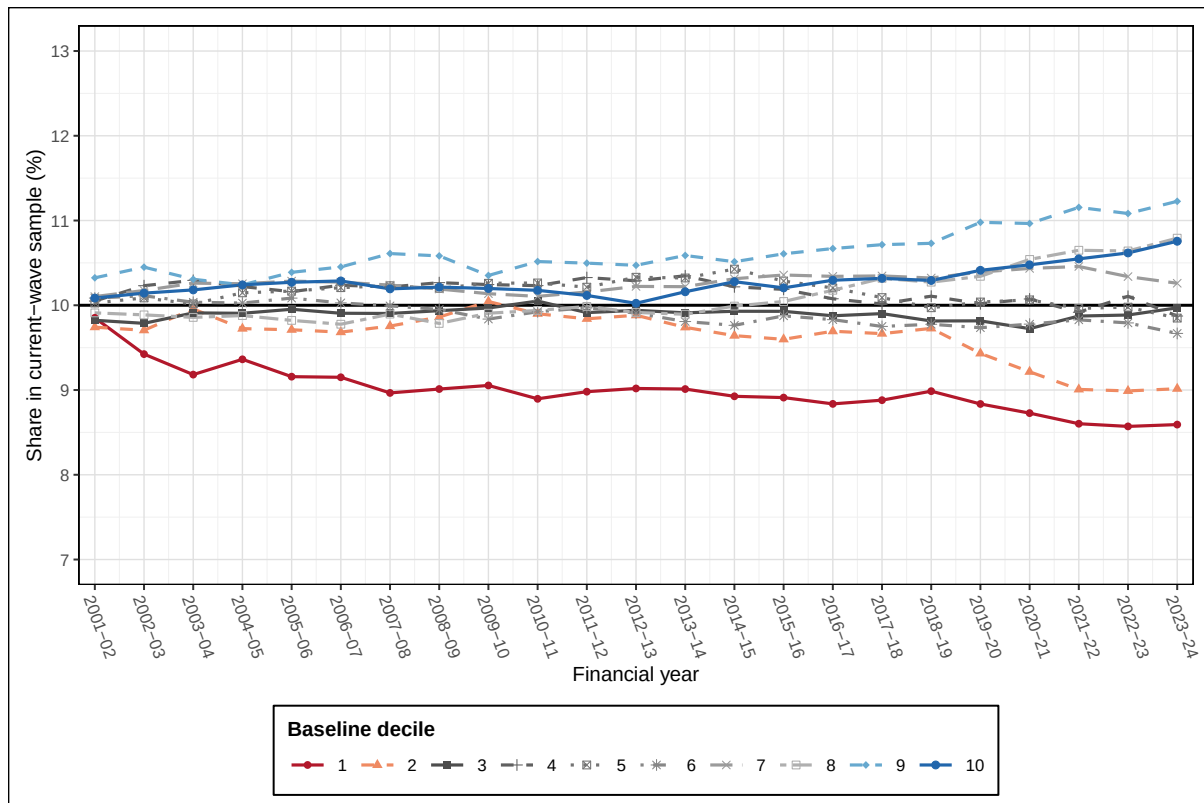
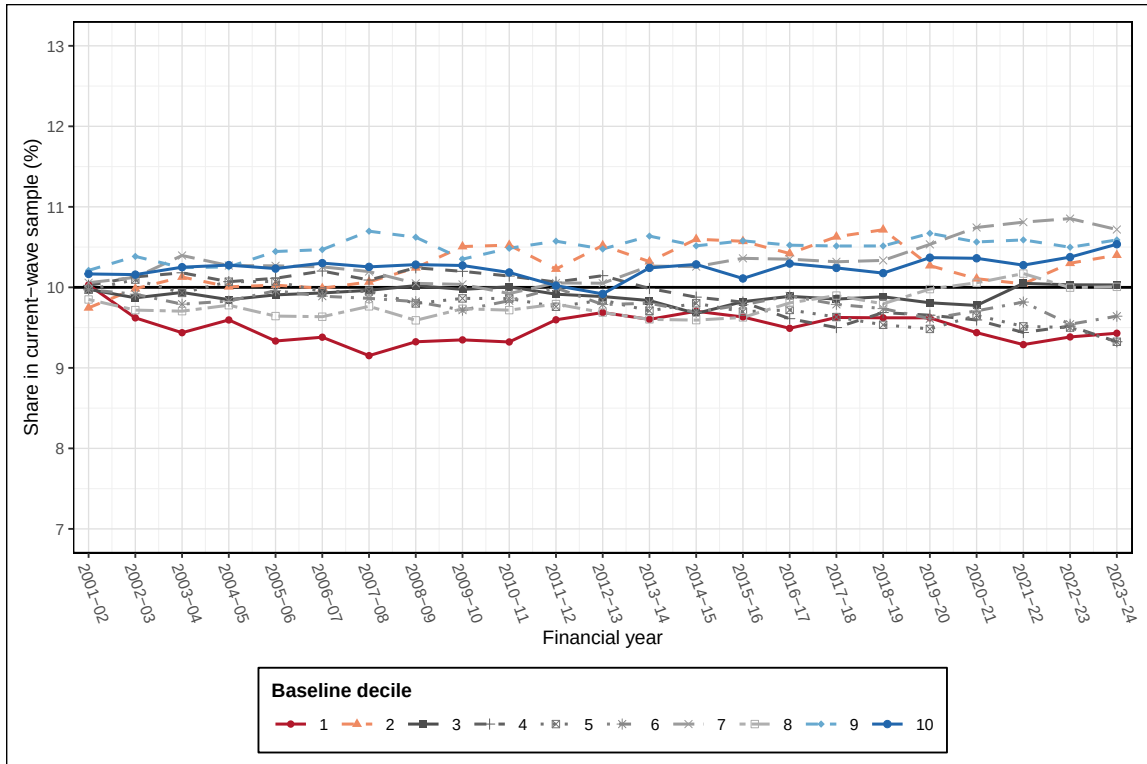
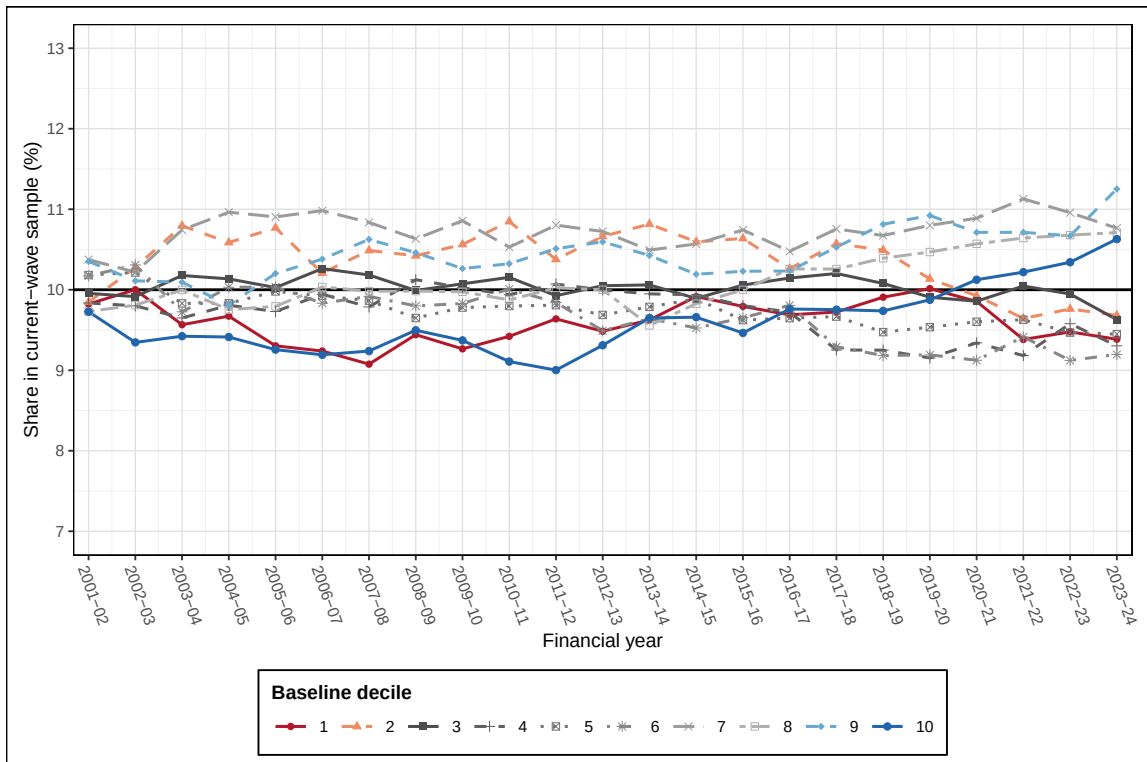


Figure A2. OSM age-adjusted decile shares using broader baseline age bins (0-18, 19-34, 35-54, 55-64, 65+, unweighted).



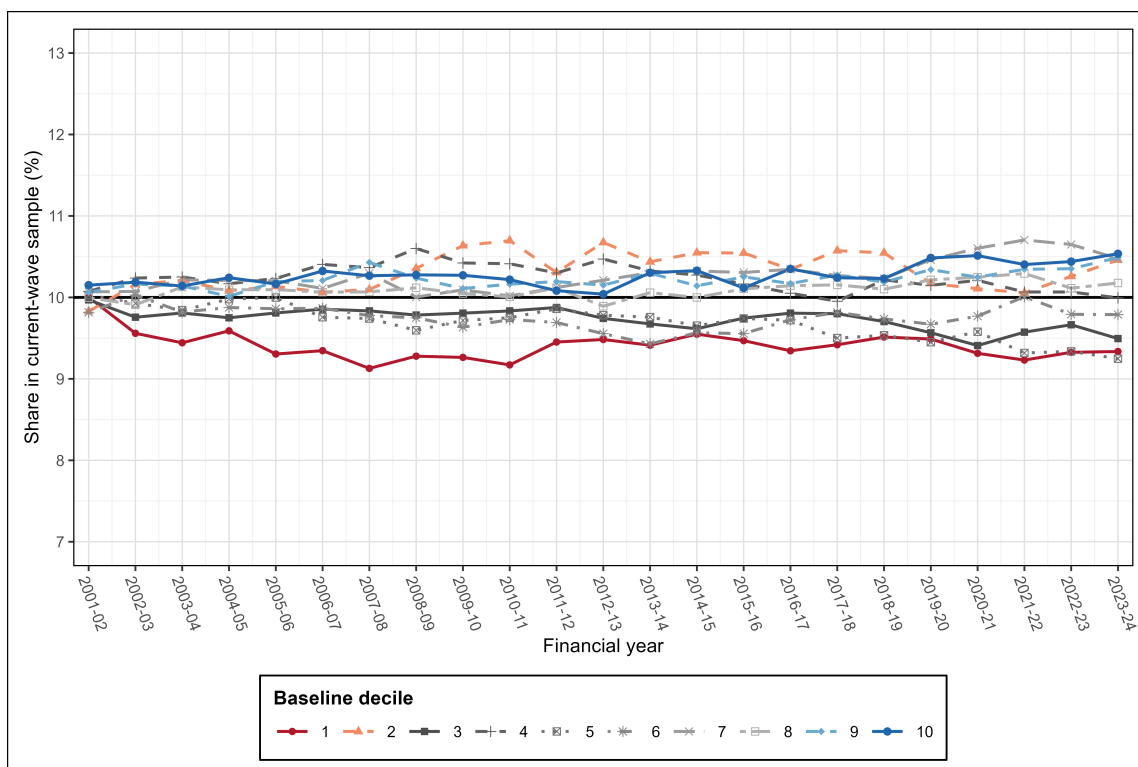
(a) Linked age-adjusted decile shares using broader baseline age bins (0-18, 19-34, 35-54, 55-64, 65+, unweighted).



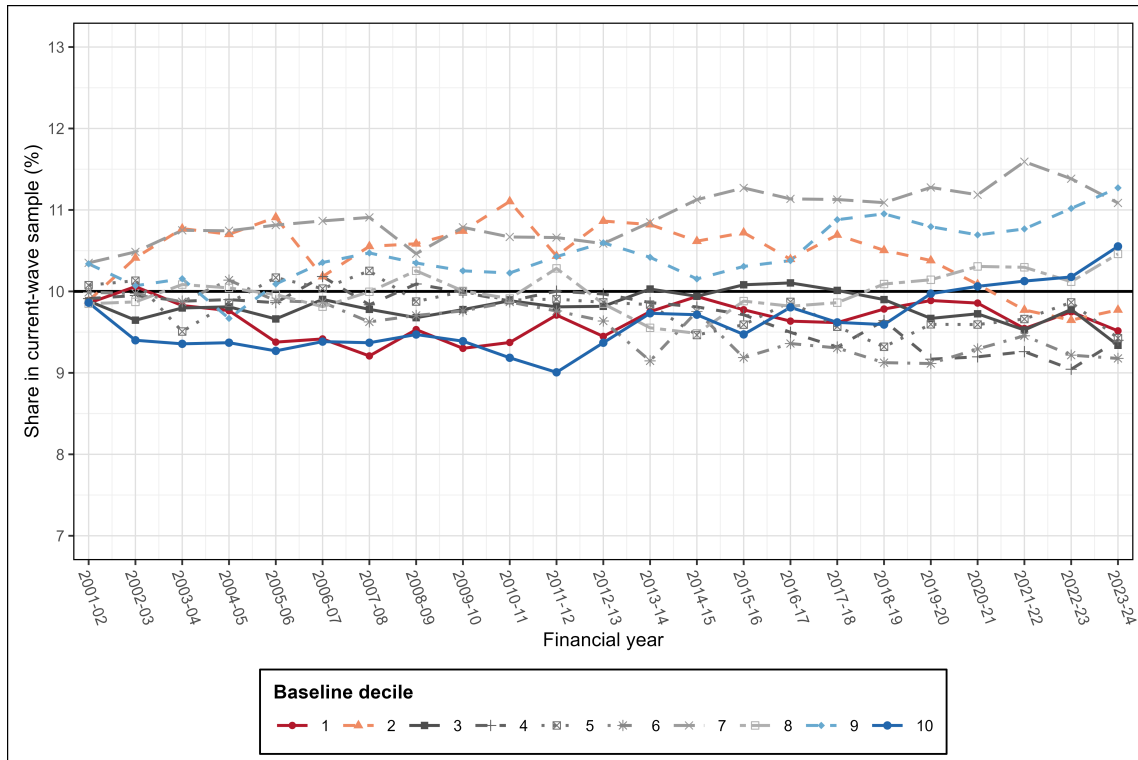
(b) Linked age-adjusted decile shares using broader baseline age bins (0-18, 19-34, 35-54, 55-64, 65+, weighted).

Figure A3. Robustness of the linked age-adjusted figures to broader baseline age bins.

D Anchoring Rule Robustness

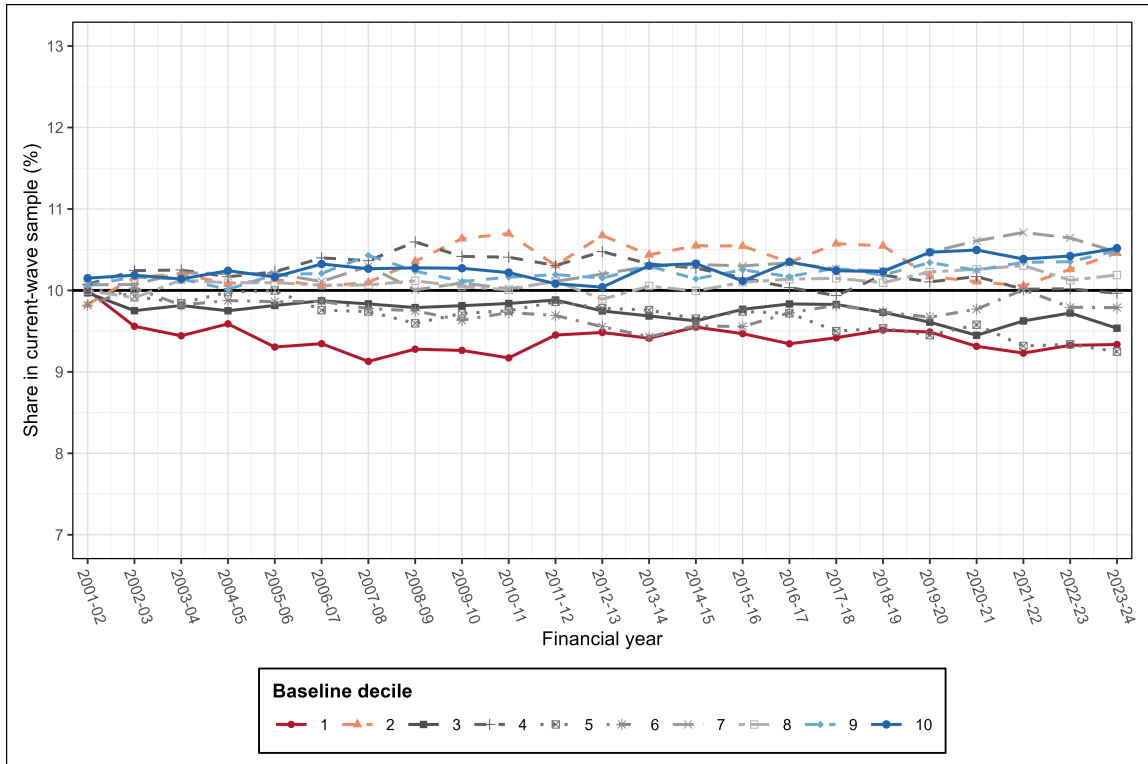


(a) Parent-first rule order (unweighted).

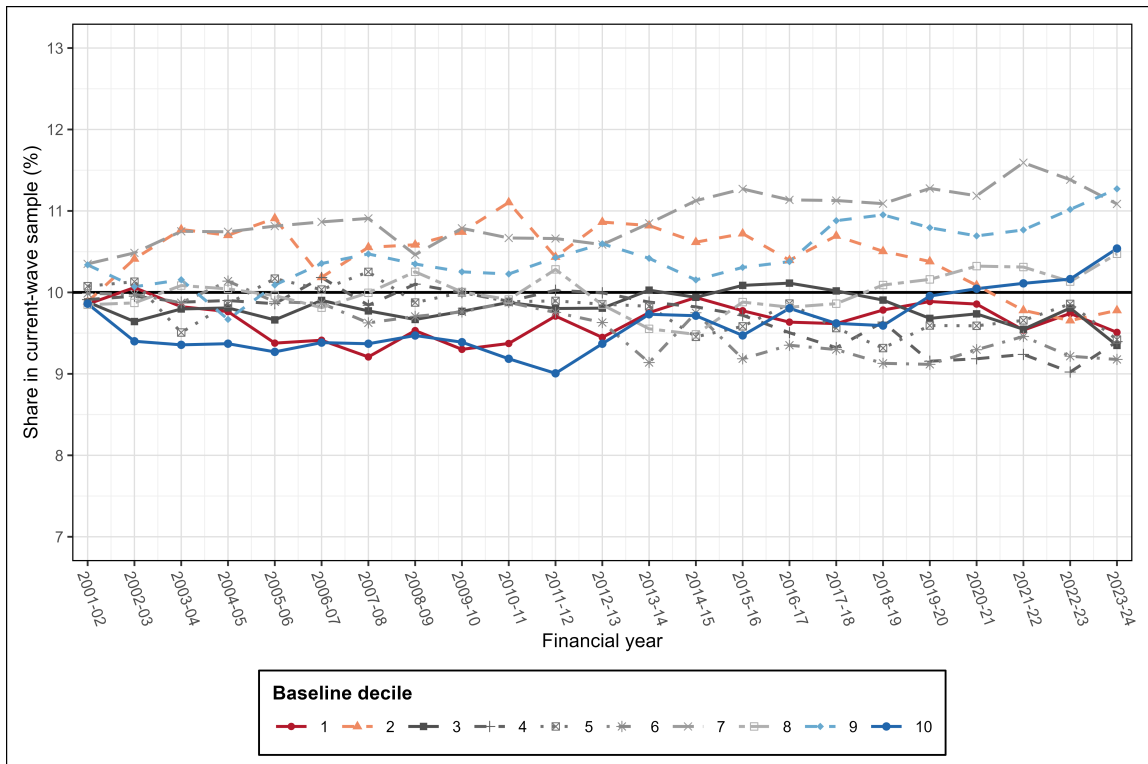


(b) Parent-first rule order (weighted).

Figure A4. Robustness of linked age-adjusted decile shares to using parent-first rather than partner linkage.



(a) Biological-parent-only rule (unweighted).



(b) Biological-parent-only rule (weighted).

Figure A5. Robustness of linked age-adjusted decile shares to restricting parent linkage to biological-parent identifiers only.