Survey Research Methods (2014)
Vol. 8, No. 3, pp. 195-208

ISSN 1864-3361
http://www.surveymethods.org

Evaluation of weighting methods to integrate a top-up sample with an ongoing longitudinal sample

Nicole Watson
Melbourne Institute of Applied Economic and Social Research
University of Melbourne

Long-running household-based panels tend to top-up (or refresh) their samples on occasions over the life of the panel. The motivation for adding such samples may range from concerns about overall sample size, lack of population coverage or inadequate samples of small target groups. In 2011, a general top-up sample was added to the Household, Income and Labour Dynamics in Australia (HILDA) Survey, a decade after the original sample was selected. This top-up sample added 2150 responding households to the main sample of 7400 responding households, representing a 29 per cent increase in the overall sample size. These top-up samples can be used to improve the cross-sectional and longitudinal weights. Drawing on the experience of other large household-based panels, this paper evaluates six options for integrating the two HILDA samples. The evaluation considers the variability in the weights, bias and the root mean square error of a range of key estimates. The method that performs the best pools the samples together by estimating the sampling and response probabilities for the sample that each unit could have been but was not actually selected into.

Keywords: HILDA; cross-sectional weight; refreshment sample; combining estimates; pooling samples

1 Introduction

Long-running household-based panels tend to top-up, or refresh, their samples from time to time. This may be for several reasons such as to increase sample size, improve the coverage of the population, or to target specific groups in the population. The longest running household-based panel, the US Panel Study of Income Dynamics (PSID) which began in 1968 has added only one top-up sample in 1997 that focused specifically on migrants to the US since the study began (Heeringa, Berglund, Khan, Lee, & Gouskova, 2011). By contrast, the German Socio-Economic Panel Study (SOEP) has had many sample extensions over its 30 year history (Haisken-DeNew & Frick, 2005). Among these are additional samples to expand the coverage of the sample to East Germany in 1990 and recent immigrants in 1994/95, general refresher samples in 1998 and 2000, and an oversample of specific populations including high income households in 2002. The Canadian Survey of Labour and Income Dynamics (SLID) also had a number of new samples as it consisted of two six year rotating panels that overlapped for three years at a time (LaRoche, 2007). The British Household Panel Survey (BHPS), which began in 1991, added a Welsh and Scottish top-up sample in 1999 to boost the overall sample size in these countries and extended the sample to include Northern Ireland in 2001 (Taylor, Brice, Buck, & Prentice-Lane, 2010). In 2009, the BHPS was subsumed into the much larger UK Household Longitudinal Survey (Lynn & Kaminska, 2010). More recently, the Household, Income and Labour Dynamics in Australia (HILDA) Survey added a general top-up sample in 2011 a decade after the study began. The purpose of this top-up sample was to address the under-coverage of recent immigrants whilst also adding to the overall sample size (Watson & Wooden, 2013).

Longitudinal surveys typically provide both cross-sectional and longitudinal weights that facilitate population inference from the sample (Lynn, 2009) and the top-up samples can be used to improve these weights. The focus in this paper is how to best incorporate the top-up sample into the cross-sectional weights. This integrated cross-sectional weight will then form the basis of any longitudinal weight for a panel that begins on or after the wave the top-up sample was introduced. The provision of such weights will help users include in their analysis as much of the sample as possible. The development of these weights is straightforward when the additional sample is from a new part of the population that wasn't included in the original sample, such as the post–1968 immigrant sample for the PSID, the East German sam-

Contact information: Nicole Watson, Melbourne Institute of Applied Economic and Social Research, Faculty of Business and Economics Building, 111 Barry Street, The University of Melbourne (n.watson@unimelb.edu.au)

ple for the SOEP, and the Northern Ireland extension of the BHPS. It is more difficult when the samples are drawn from the same population, as occurred for the Australian, Welsh, Scottish, Canadian and German top-up samples. To calculate these weights accurately, we would need to know the selection and response probabilities for each sample member at each point in time a sample was taken. As these are difficult to know exactly, they need to be estimated in some way.

Somewhat different approaches have been taken across the various surveys for incorporating top-up samples into the ongoing sample. Drawing on the experience of other large household-based panels, this paper evaluates six options for integrating a top-up sample with an ongoing sample using the HILDA Survey as the test bed. There appears to be only one other study that has examined this issue in a longitudinal setting, being O'Muircheartaigh and Pedlow (2002) using a cohort study (the US National Longitudinal Survey of Youth), though both of their samples were selected at the same point in time. This paper adds to this evidence by examining this issue in the context of a household-based longitudinal survey with much larger samples that are selected at different points in time. The impact of each method is also assessed on a wider range of variables.

The paper proceeds as follows. Section 2 describes the design of the HILDA Survey and how the two samples represent overlapping parts of the population. Section 3 sets out the methods tested, the overall weighting process and the evaluation methods. The results are presented in section 4 and section 5 concludes.

2 The HILDA Survey

The HILDA Survey is a household-based longitudinal study that follows individuals over time with annual interviews with all adult members of sampled households. In the vast majority of cases, interviews are conducted face-to-face with some (generally less than 10 per cent) being conducted by telephone. The following sections describe the sample design, the following rules, how the original sample has evolved over time and how the two samples represent overlapping parts of the population.

2.1 Sample design

The original HILDA sample was selected in 2001 via a stratified multi-stage area-based clustered design (Watson & Wooden, 2002). The sample was restricted to households living in private dwellings and excluded those living in very remote parts of Australia.[1] The stratification was by state and within the five most populous states by metropolitan and non-metropolitan areas. In the first stage of selection, 488 Census Collection Districts (CD) were selected with probability proportional to the number of dwellings. Within each CD, approximately 25 dwellings were systematically selected. If a dwelling contained three or fewer households, all were selected. When a dwelling had more than 3 households, 3 were randomly selected. Most dwellings only contained one household. A total of 11,693 in-scope households were included in the 2001 sample and interviews were obtained with 7682 households, resulting in a household response rate of 66 per cent.

The top-up sample was selected in 2011 using a similar design as the 2001 sample with 125 CDs selected (Watson & Wooden, 2013).[2] A total of 2153 households responded from the sample of 3117 in-scope households in the top-up sample, giving a 69 per cent household response rate. As such, the top-up sample provides a population-wide top-up of the sample, representing a 29 per cent increase in the overall sample size. It also helps to rectify the lack of coverage of immigrants arriving in Australia after 2001 and offers the opportunity to examine the effect of a decade of attrition on the ongoing sample.

2.2 Following rules

The following rules adopted in the HILDA Survey are intended to ensure the sample mimics the changes in the population as much as possible and allows for the study of family dissolution. All members of the responding households in 2001 are considered Permanent Sample Members (PSM) and these people are followed over time, even if they move into non-private dwellings or very remote parts of Australia. In addition, others are converted to PSM status if they are:

- born to or adopted by a PSM;
- the other parent of a PSM birth or adoption if they are not already a PSM;
- recent arrivals to Australia since the survey began in 2001.[3]

All other sample members are Temporary Sample Members (TSMs), and are considered part of the sample for as long as they share a household with a PSM. Similar rules apply to the 2011 top-up sample, though recent arrivals in this sample now relate to people arriving in Australia after 2011.

[1]Very remote areas were defined by the Australian Bureau of Statistics (ABS) as a list of Census Collection Districts considered too remote for cost reasons for the inclusion in certain social surveys (such as their monthly Labour Force Survey supplements).

[2]For the top-up sample, very remote areas were defined as Remoteness Area 4 "Very Remote Australia" in the Australian Standard Geographical Classification (ABS Cat. No. 1216.0, July 2006) as the remoteness definition used in the original HILDA sample was no longer supported by the ABS.

[3]The inclusion of recent arrivals (i.e., immigrants who arrived to Australia after 2001) into the following rules occurred in 2009 and was applied retrospectively. There were some recent arrivals who entered the sample in earlier waves but had moved out by wave 9 so could not be followed.

2.3 Sample changes over time

Over the past decade since the original HILDA sample was selected, the sample has grown, with new births and other new household entrants as shown in Table 1. The sample now includes 23,396 PSMs following the addition of 3482 PSMs (most of whom are babies) to the 19,914 PSMs included in the first wave. There has been 6093 (= 2529 + 3564) TSMs that have joined the sample, 42 per cent of which are still considered "active" sample members as they still live with one or more PSMs. Some sample members have died or moved overseas, whilst others have moved into non-private dwellings and very remote parts of Australia. Only a small number of people arriving in Australia after 2001 (termed "recent arrivals") have joined the sampled households (n = 270), and indeed, these people are not representative of all recent arrivals as they have links to more established households while many others do not.

2.4 Population overlap and classification of the sample

While the sample has changed over time, so has the population. We can think of the original sample augmented by the following rules and the new top-up sample as representing overlapping but not identical populations as shown in Figure 1. The population we are interested in for the purposes of the cross-sectional weights is people living in private dwellings in Australia excluding very remote areas of Australia (i.e., the population defined by Frame B). The original sample followed over time represents the population covered by Frame A and includes some units, identified as a in Figure 1, that are considered out of scope of the frame used to select the top-up sample (Frame B), such as households that have died, moved overseas, moved into non-private dwellings or are living in very remote areas. These units had no chance of selection in the top-up sample. Frame B, on the other hand, includes some units that are not covered by Frame A, identified as b in Figure 1. These are units that had no chance of being selected in 2001 or included via the following rules but now live in private dwellings excluding very remote parts of Australia. Such households contain only people who fall into the following categories:

1. immigrants permanently settling in Australia since 2001;
2. long-term visitors arriving since 2001;
3. Australians not living in Australia in 2001 who have since returned from overseas;
4. people who have moved out of non-private dwellings;
5. people who have moved out of very remote Australia;
6. Australian-born children of these groups.

Based on official migration statistics, it is estimated that these groups form about 7 per cent of the Australian population in 2011, with permanent immigrants being by far the largest missing group (Watson, 2012). The remainder of the

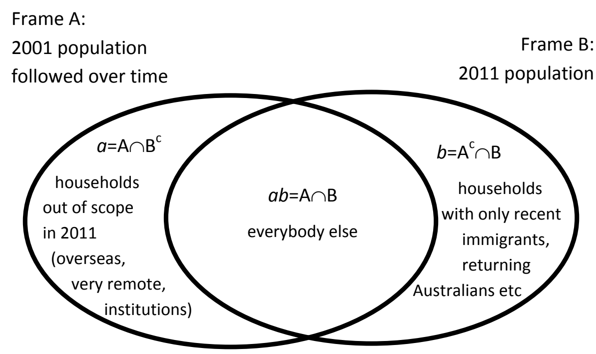

Figure 1. Overlap of frames

units, represented by the intersecting segment ab, are on both frames, and therefore had two chances of being included in the sample, once via the sample in 2001 and again by the sample in 2011.

Using this framework, we can similarly segment the responding households from the two HILDA samples for 2011.

Table 2 shows that there are 90 responding households in the main sample in institutions or very remote areas of Australia (that is, in segment a). These households are given zero cross-sectional weight as they are out of scope of the cross-sectional population of interest.

The vast majority of the households in both samples fall into the overlapping segment of the population, ab, and it is for this part of the sample that we are exploring various means of integrating these two samples together.

The final group identified is those households that contain only recent arrivals. For households where we do not know whether a person arrived in Australia after 2001 or not, their status is imputed from other members of their household. We find that there are 24 households in the main sample and 178 households in the top-up sample that only contain people who arrived in Australia after 2001. The 24 households in the main sample that fall into this category are a result of following recent immigrants after they leave a household that contains a permanent sample member. As these households are quite unrepresentative of all recent immigrant households and it is not possible to identify which of the 178 recent immigrant households they might be similar to, they will be given zero cross-sectional weight. As a result, only those recent immigrant households in the top-up sample will be weighted to represent segment b of the population.

Note also that it was not possible to identify who in the top-up sample had moved out of institutions, moved out of very remote parts of Australia or had returned from living overseas since 2001. Nonetheless, it is expected that the number would be very small and therefore of negligible consequence.

Table 1
Composition of the main sample

	Wave										
	1	2	3	4	5	6	7	8	9	10	11
All sample members	19914	21045	22062	22958	23903	24852	25702	26523	27518	28530	29489
PSM	19914	20146	20410	20682	20989	21307	21678	22096	22507	22941	23396
Active TSM[a]	–	899	1323	1529	1705	1899	2003	1972	2244	2417	2529
Inactive TSM	–	–	329	747	1209	1646	2021	2455	2767	3172	3564
Sample changes[b]											
Deceased	–	68	174	293	397	491	579	683	767	843	937
Moved overseas	–	74	233	374	387	430	483	501	491	513	557
Moved to NPD	–	26	38	39	41	44	48	52	58	73	83
Moved to very remote Australia	75[c]	94	98	107	115	116	114	125	115	110	114
Births	–	219	450	674	920	1157	1413	1661	1926	2184	2457
Recent arrival adults[d]	–	13	26	43	53	85	85	109	151	187	239
Recent arrival children[d]	–	2	5	7	5	10	9	17	22	29	31
Interviewed adult	13969	13041	12728	12408	12759	12905	12789	12785	13301	13526	13603

PSM=Permanent Sample Member; TSM=Temporary Sample Member; NPD=Non-private dwelling.
[a] Active TSMs includes all TSMs not known to have left PSM households. [b] Excludes inactive TSMs.
[c] As the original sample used a slightly different definition of "very remote" some wave 1 sample members are now classified as living in "very remote" Australia.
[d] As recent arrivals were not followed if they left the household of a PSM in waves 1 to 8, the number could go down as inactive TSMs are excluded from this count (this occurs for children in waves 5 and 7).

Table 2
Classification of wave 11 responding households representing population overlap groups

	Main sample	Top-up sample	Action in combining sample
Segment a			
Moved into institution (non-private dwellings)	62	–	Excluded from cross-sectional weights
Moved into very remote parts of Australia	28	–	Excluded from cross-sectional weights
Segment ab			
Contains no recent arrivals	7126	1881	Integrate
Contains some recent arrivals	150	94	Integrate
Segment b			
Contains all recent arrivals	24	178	Weight only top-up sample
Total households	7390	2153	

3 Integration methods

The methodology for creating estimates from samples drawn from two or more frames has been developed for sampling problems where there is typically one expensive but complete frame and another cheap and incomplete frame (Hartley, 1962; Cochran, 1977, pgs. 144-146). For example, one might sample from an expensive area-based frame and also a cheaper telephone frame (Sirken & Casady, 1988). In our case, we have two frames from different time points and the coverage of the population is complete with the second frame though a smaller sample has been taken.

Two main ways to integrate independent surveys together have emerged. The first is to combine the estimates from each frame in such a way that minimise the variance of the estimate. Early proponents of this method (e.g. Hartley, 1962) suggested estimators that required a different set of weights for each variable analysed, basically optimising the weight for each variable. More recently, suggestions have been put forward which optimises the weights for a particular variable in order to have one set of weights (Lohr & Rao, 2000; Skinner & Rao, 1996). The second method is to pool the samples using the inclusion probabilities for the two frames (Kalton & Anderson, 1986).

O'Muircheartaigh and Pedlow (2002) adopt the terms "combining estimates" and "pooling samples" to differentiate these two broad methods and the same terminology is used here.

In this paper, five options for combining the estimates and one option for pooling the samples are evaluated together with two options that keep the samples separate. The options that keep the sample separate are used to assess what can be gained from integrating the samples. Option A includes just the main sample (all top-up sample members are given zero weight) and Option B includes just the top-up sample (all main sample members are given zero weight). For the remaining options, we draw on the experience of other large household-based longitudinal panels, namely SLID, SOEP and UKHLS, as well as the US National Longitudinal Survey of Youth (NLSY) which is a cohort study that oversampled certain racial groups. Of the other panels mentioned in Section 1, the BHPS did not provide weights that integrated the Scottish and Welsh boost samples with the ongoing sample and the PSID did not need to integrate their 1997 sample as it was a sample of a separate population of new immigrants.

3.1 Combining estimates

Combining the estimates involves taking a weighted average of the estimates from the two samples. Let's say we have sample A and sample B and we are interested in estimates of a total of a variable of interest Y. The combined estimate would be:

$$\hat{Y}_{\text{combined}} = \theta\hat{Y}_A + (1-\theta)\hat{Y}_B \quad ,$$

where θ is between 0 and 1, $\hat{Y}_A$ is the estimate for Y from sample A, and $\hat{Y}_B$ is the estimate from sample B. When the samples are independent, the optimal choice of θ which minimises the variance of $\hat{Y}_{\text{combined}}$ is:

$$\theta = \frac{\frac{n_A}{\text{deff}_A}}{\frac{n_A}{\text{deff}_A} + \frac{n_B}{\text{deff}_B}}$$

where n_A and n_B are the number of elements in each sample, and deff_A and deff_B are the design effects for the estimates $\hat{Y}_A$ and $\hat{Y}_B$ (O'Muircheartaigh & Pedlow, 2002). The design effect is the ratio of the actual variance of a sample to the variance of a simple random sample (without replacement) of the same number of elements. Where the total is a weighted estimate of observed elements in the sample, the weights for each element will therefore be:

$$w_{\text{combined},i} = \begin{cases} \theta w_{Ai} & \text{if } i \in S_A \\ (1-\theta) w_{Bi} & \text{if } i \in S_B \end{cases}$$

where w_{Ai} is the original weight for element i in sample A and w_{Bi} is the similar weight for sample B. That is, for all elements in sample A we take a fraction θ of the original weight and $(1-\theta)$ of the weight in sample B. θ can therefore be thought of as a panel allocation factor. As the design effects can be different depending on the variable of interest, we evaluate a number of choices for θ.

A range of estimates were considered for determining the optimal value of θ, some at the household level and others at the person level, as shown in Table 3. These estimates are restricted to the part of the sample representing the overlapping population ab (being households with some or no recent arrivals that are not living in institutions or very remote parts of Australia).[4]

The first option that combines the estimates from the two samples (Option C) assigns the panel allocation factor based on the proportion of households in the main sample, being θ=0.787. This option assumes the design effects for the two samples are the same. In reality, there is no reason to expect them to be the same: the main sample has been subject to 10 years of attrition which would be detrimental to the design effect but it has also unclustered over time which would benefit the design effect.

The second option to combine estimates (Option D) sets the panel allocation factor as the optimal θ for one particular estimate. Skinner and Rao (1996) suggest choosing θ to minimise the variance of the number of units in ab and this is the approach taken by SLID in combining two rotating panels (LaRoche, 2007). As SLID appears to undertake the integration at the person-level rather than the household-level, we adopt a related measure here being the average number of adults per household. This also aligns well with the fact that most of the HILDA data users tend to produce person level estimates rather than household level estimates, so optimising the weight for the number of persons (i.e., average household size times the total number of households in ab) would be of greater benefit than just the number of households. The panel allocation factor for Option D is 0.800.

In the third option to combine estimates, we take the average of the optimal θ across a range of variables presented in Table 3. Incidentally this panel allocation factor (0.787) turns out to be identical to the one generated from the sample sizes (and therefore is already tested via Option C). Reassuringly, the range of optimal θ for the variables presented in Table 3 is not particularly large, so we can be reasonably confident that the choice we make for θ within this range will not be greatly detrimental to other estimates.

The next option we test (Option E) is based on the method adopted by NLSY (O'Muircheartaigh & Pedlow, 2002). They propose that, as it is inconvenient to use the different design effects based on specific variables, a general factor be used to capture the impact of unequal weighting on sample efficiency. Under this method, the design effects are approximated using the coefficient of variation of the weights in the following way:

$$\widehat{\text{deff}}_A \cong 1 + \text{CV}(w_A)^2$$
$$\widehat{\text{deff}}_B \cong 1 + \text{CV}(w_B)^2$$

where $\text{CV}(w_A)$ is the coefficient of variation of the weights in sample A and $\text{CV}(w_B)$ is the same for the weights in sample B. This construction of the design effect relies on a property of weights initially observed by Kish (1992) that arbitrary weights increase variances by a factor of $1 + L$ where L can be adequately estimated by the squared coefficient of variation of the weights. O'Muircheartaigh and Pedlow (2002) also suggest using different factors for a small number of groups that may have similar weights due to design or non-response reasons (in their case, race and sex). For HILDA, we have differential non-response across geographic regions and we have chosen three broad groups to express this. As a result, the optimal θ developed in this way are 0.745 in Sydney, 0.799 in the other major cities (Melbourne, Brisbane, Adelaide and Perth) and 0.809 for the rest of Australia.

The final combining estimates option we test (Option F) mimics the integration adopted by the German Socio-Economic Panel (as outlined by Spiess and Rendtel, 2000). They use the equation above for θ to define a range within which the optimal θ will lie for any given estimate. They propose that the ratio of the design effect for the newly selected to the ongoing sample ($\frac{\text{deff}_B}{\text{deff}_A}$) would lie in the range 0.5 to 0.8 on the grounds that the new sample has been selected on the same basis as the main sample and the main sample deteriorates over time due to attrition. They then choose a particular convenient value for θ within this range. Applying their method to the HILDA data suggests the optimal range for θ is between 0.648 and 0.747, and we might choose a convenient value of say 0.7. The reason that the SOEP adopt a single convenient value for θ is that they provide the weights to their users in such a way that the user can deconstruct the weights and create weights specific to their purposes should they choose to do so. In the case of the HILDA Survey, applying a convenient value does not carry the same benefit for users as we apply a benchmarking step following the integration step in the HILDA weighting process that is absent in the SOEP weighting process. Nevertheless, we still test this approach as Option F.

3.2 Pooling samples

When pooling the samples, we need to know (or estimate) the probability of selection and response that each element had in either sample A or B. The pooled weight for household i is given by:

$$w_{\text{pooled},i} = \frac{1}{p_{iA} + p_{iB}} \quad ,$$

[4]Both PSMs and active TSMs are included in the person-level estimates for the main sample.

Table 3
Optimal theta for key variables

	Main (Sample A)		Top-up (Sample B)		
	Deff	Sample	Deff	Sample	θ
HH level variables					
Number adults in HH	2.02	7,276	2.19	1,975	0.800
Total gross FY household income	2.03	7,276	3.06	1,975	0.847
Own dwelling	2.14	7,254	2.72	1,971	0.824
Income support reliant	2.78	7,276	3.50	1,975	0.823
Lone parent HH with dependants	1.15	7,276	1.26	1,975	0.802
Person-level variables					
Wages and salaries FY	1.93	8,655	1.72	2,271	0.773
Usual hours worked	1.52	8,631	1.31	2,265	0.767
Has permanent job	1.52	7,325	1.09	1,900	0.735
Supervisor	1.54	8,655	1.31	2,269	0.765
Job satisfaction	1.88	8,651	1.21	2,270	0.710
Married/defacto	2.49	13,432	1.91	3,658	0.739
Number of children had	1.82	13,444	2.18	3,659	0.815
Life satisifaction	2.41	13,444	2.23	3,654	0.773
Has university degree	2.70	13,442	3.68	3,657	0.833
Long term health condition	2.95	13,448	3.07	3,660	0.793
Average optimal θ					0.787
Proportion of HHs in main sample					0.787

where p_{iA} and p_{iB} is the probability of selection and response for household *in* sample A and B respectively.[5] As we do not observe these probabilities in the sample the element was not selected into, we need to estimate it in some way.[6] Kaminska and Lynn (2012), when considering how to integrate the BHPS and UKHLS samples, suggest modelling these probabilities based on the characteristics of the sample members. This is not unlike how some longitudinal surveys (namely HILDA and SOEP) assign weights to household members that join the households over time (Schonlau, Kroh, & Watson, 2013).

The modelling process for this option (Option G) proceeds as follows. Using a logit transformation, the probability of selection and response is modified so that it is an unbounded continuous variable. A linear regression model is fitted with relatively simple covariates, being 12 geographic location indicators and six age categories and the sex of the household reference person. The adjusted-R^2 for the model of probabilities in the main sample is 0.210 and for the top-up sample it is 0.176. The model for each sample is then used to make out of sample predictions ($\hat{p}_{iB}$ for households included in the ongoing sample and $\hat{p}_{iA}$ for households included in the top-up sample) for each household for the sample they were not included in. The pooled weights are calculated as:

$$w_{\text{pooled},i} = \begin{cases} \frac{1}{p_{iA}+\hat{p}_{iB}} & \text{if } i \in S_A \\ \frac{1}{\hat{p}_{iA}+p_{iB}} & \text{if } i \in S_B \end{cases}$$

We can reformulate the pooled weight to identify the adjustment factor that the original weight for each household i is multiplied by in each sample. These adjustment factors are shown in Figure 2. The factors on the left are for the top-up sample and those on the right are for the main sample. The average adjustment factor for the main sample is very close to those calculated under Options C to E to combine the samples.

[5]This assumes that the probability of a household being selected in both samples ($p_{iA}p_{iB}$) is effectively zero given small sampling rates (otherwise it would need to be taken into account in the denominator).

[6]In the NLSY example examined by O'Muircheartaigh and Pedlow (2002), they were able to obtain the exact selection probabilities and they do not make an adjustment for response propensities when pooling the samples, presumably because the two samples were selected at the same time and any non-response adjustment could be undertaken once the samples had been integrated.

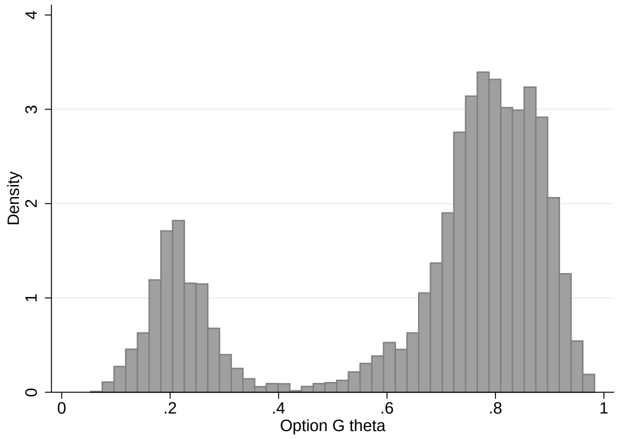

Figure 2. Adjustment factor for weights when pooling samples (Option G)

3.3 Incorporating the sample integration into the weighting process

The weighting process that produces the cross-sectional weights has a number of different steps, of which the integration of the two samples is one part. These steps are detailed in Watson (2012) and a summary is provided here. Prior to the integration step, the design weights for the two samples are calculated and the main sample weights are adjusted for TSMs joining the household after the initial sample was selected.[7] Next both samples are adjusted for household-level non-response. Following this, the two samples are split into the population overlap groups described in section 2.4 and the portion of the sample representing the overlapping population (*ab*) is integrated according to the different methods. The portion of the top-up sample representing recent immigrant households (*b*) retains the non-response adjusted design weight. The equivalent portion of the main sample receives zero weight. That is, for the methods that combine the samples (options C-F), the integrated weight is given by:

$$w_{\text{combined},i} = \begin{cases} 0 & \text{if } i \in S_{A,a} \\ \theta w_{Ai} & \text{if } i \in S_{A,ab} \\ (1-\theta) w_{BI} & \text{if } i \in S_{B,ab} \\ 0 & \text{if } i \in S_{A,b} \\ w_{Bi} & \text{if } i \in S_{B,b} \end{cases},$$

where w_{Ai} is the design weight for element i in the main sample (sample A) that has been adjusted for non-response and TSM joiners, w_{Bi} is the non-response adjusted design weight for the top-up sample (sample B), and a, ab, and b identify the population segments that the sample elements represent (as identified in Figure 1). For the pooling method (Option G), the integrated weight is given by:

$$w_{pooled,i} = \begin{cases} 0 & \text{if } i \in S_{A,a} \\ \frac{1}{p_{iA}+\hat{p}_{iB}} & \text{if } i \in S_{A,ab} \\ \frac{1}{\hat{p}_{iA}+p_{iB}} & \text{if } i \in S_{B,ab} \\ 0 & \text{if } i \in S_{A,b} \\ w_{Bi} & \text{if } i \in S_{B,b} \end{cases},$$

where $p_{iA} = \frac{1}{w_{Ai}}$ and $p_{iB} = \frac{1}{w_{Bi}}$ are the selection and response probabilities of being included in the main sample (sample A) and top-up sample (sample B) respectively, and $\hat{p}_{iA}$ and $\hat{p}_{iB}$ are the estimated equivalent probabilities of being in each sample.

After this integration step, the resulting weights are then simultaneously calibrated to various known household and person benchmarks. This produces the household weight. The responding person weight, which is applicable to all adults interviewed within the responding households, is then derived from the household weight via a person-level non-response adjustment and calibration to additional person-level benchmarks.

In this paper, we examine the properties of the final household- and person-level weights as these are the ones provided to users of the dataset. Seven different sets of household and person level weights are produced corresponding to the seven integration options examined:

A set $\theta = 1$ to give the main sample only (i.e. $w_{\text{opt}A,i} = w_{Ai}$ if $i \in S_A$, 0 otherwise);

B set $\theta = 0$ to give the top-up sample only (i.e. $w_{\text{opt}B,i} = w_{Bi}$ if $i \in S_B$, 0 otherwise);

C assign θ based on the relative sample size of the two samples (as it turns out in our example this is equivalent to using the average of θs that are optimal for a range of estimates);

D assign θ based on optimising one particular estimate, chosen to be the average number of adults per HH;

E assign θ based on approximate design effects calculated from the coefficient of variation of the weights;

F assign θ to a convenient value within a plausible range; and

G pool the samples and estimate the probability of selection and response in each sample.

[7]The household weight is reduced to allow for the multiple pathways into the household that we observe (i.e. through the PSMs we did follow or through the TSMs we could have followed had they been selected). All household members (PSMs and TSMs) are given the adjusted household weight in this step and are included in the weighting process from this point onwards.

Table 4
Distribution of the weights, Options A-G compared

Variable	N	Mean	Std. Dev.	Min	Max	Quartile range	Coeff. of variation
Household weights							
A: Main sample	7300	1177	850	53	14330	655	0.72
B: Top-up sample	2153	3989	1605	842	15821	1886	0.40
C: Combine on sample size	9429	911	681	36	11881	478	0.75
D: Combine on optimal theta	9429	911	687	37	12011	483	0.75
E: Combine on coeff. of variation	9429	911	674	37	11344	477	0.74
F: Combine on convenient theta	9429	911	671	32	11001	547	0.74
G: Pool samples	9429	911	578	33	8765	394	0.63
Responding person weights							
A: Main sample	13491	1328	1109	51	20000	774	0.84
B: Top-up sample	4009	4468	2094	864	26613	2393	0.47
C: Combine on sample size	17460	1026	882	35	17652	571	0.86
D: Combine on optimal theta	17460	1026	889	35	17933	576	0.87
E: Combine on coeff. of variation	17460	1026	872	36	17417	574	0.85
F: Combine on convenient theta	17460	1026	873	31	15832	645	0.85
G: Pool samples	17460	1026	758	32	10363	488	0.74

Cases with zero weight have been excluded. This includes respondents living in non-private dwellings or very remote parts of Australia that are considered out of scope for the cross-sectional population.

3.4 Evaluation methods

The seven options are evaluated in three ways. First, the variability in the weights is examined. A weighting strategy that has low variability in the resultant weights yet still achieves accurate estimates will be preferred for efficiency reasons over one with high variability.

Second, the bias is calculated for a wide range of cross-sectional estimates. The bias is taken as the difference between the relevant HILDA estimate ($\hat{Y}$) and an equivalent estimate from the Australian Bureau of Statistics ($\hat{Y}_{\mathrm{ABS}}$):

$$\mathrm{Bias}(\hat{Y}) = \hat{Y} - \hat{Y}_{\mathrm{ABS}}$$

The ABS estimates come from the monthly Labour Force Survey (generally for September 2011) and one of its supplementary surveys, the Survey of Education and Work.

Third, the root mean square error (RMSE) is assessed for these estimates. The RMSE gives a measure of the quality of an estimate ($\hat{Y}$) that considers both the bias in the estimate and the variability in the estimate. It is calculated as:

$$\mathrm{RMSE}(\hat{Y}) = \sqrt{\mathrm{SE}(\hat{Y})^2 + \mathrm{Bias}(\hat{Y})^2}$$

where $\mathrm{SE}(\hat{Y})$ is the standard error of the estimate $\hat{Y}$ and bias is defined above. A lower RMSE is better than a higher one.

4 Results

A summary of the distribution of the weights under the seven options is provided in Table 4. There is very little difference in the distribution of the weights between the four different options to combine the samples (Options C-F). There is sizeable reduction in the variability in the weights when we pool the samples (Option G). This finding is consistent with O'Muircheartaigh and Pedlow (2002) who show that pooling the samples produces much less variability in the weights when the selection probabilities are quite different between the two samples.

Next, we consider what impact these seven options have on the bias. Table 5 provides the estimates of bias from the various options together with the ABS estimate. The estimates of bias that are significantly different from zero are highlighted and the option with the lowest bias is indicated in bold. The estimates include relationship in household, highest level of education, country of birth, year of arrival, indigenous status, and for those employed we consider usual hours worked and employment status. Note that the estimates for the five options to combine the estimates (Options C-F) do not always fall between the estimates for Options A and B due to the calibration step in the weighting process which occurs after the samples are combined.

The main sample (Option A) has the greatest number of estimates that are biased (with 25 of 37 estimates being significantly different from zero). This falls to 11 biased estimates in the top-up sample, though remember the top-up sample is a much smaller sample so fewer significant differences are expected apart from any improvements from including an appropriate sample of recent immigrants. The integrated samples (Options C-G) are fairly indistinguishable from each other in terms of bias reduction with 15 or 16 of

Table 5
Bias estimates, options A-G compared

Characteristic	ABS	Bias A	Bias B	Bias C	Bias D	Bias E	Bias F	Bias G
Relationship in household								
Couple with dependents	26.4	0.62	**0.29**	0.63	0.64	0.62	0.58	0.57
Couple without dependents	32.8	−0.71	**−0.39**	−0.73	−0.74	−0.72	−0.68	−0.67
Lone parent with dependents	3.8	−0.25	0.18	−0.16	−0.16	−0.14	−0.12	**−0.09**
Lone parent without dependents	1.7	0.79*	**0.08***	0.49*	0.50*	0.49*	0.45*	0.42*
Dependent student	7.1	1.19*	**0.68**	1.07*	1.08*	1.07*	1.03*	1.09*
Nondependent child	8.5	1.44*	−0.48	0.35	0.36	0.34	0.25	**0.23**
Other family member	2.6	**0.20**	1.31*	0.80	0.79	0.79	0.85*	0.77*
Nonfamily member	17.0	−3.27*	**−1.67**	−2.46*	−2.47*	−2.45*	−2.37*	−2.32*
Highest level of education (15-64 year olds)								
Postgraduate (masters or doctorate)	4.6	**−0.55**	1.17	0.60	0.59	0.62	0.67	0.71
Grad diploma or grad certificate	2.1	2.85*	**2.77***	3.01*	3.02*	3.02*	2.99*	3.15*
Bachelor or honours	17.0	−2.52*	**−0.22**	−1.24*	−1.26*	−1.25*	−1.11	−1.21*
Advanced diploma or diploma	9.1	−0.44	0.36	−0.16	−0.17	−0.16	**−0.10**	−0.15
Cert IV or III	17.4	4.05*	5.21*	3.45*	3.42*	3.40*	3.64*	**3.29***
Year 12	20.6	**−1.64***	−4.42*	−2.72*	−2.68*	−2.72*	−2.93*	−2.84*
Year 11 or below (incl. Cert I, II, nfd)	29.1	**−1.80***	−4.83*	−3.01*	−2.98*	−2.96*	−3.21*	−2.95*
Country of birth								
Australia	70.1	4.59*	−2.44	−1.49	−1.47	−1.47	−1.63	**−0.95**
Main English speaking country	10.8	−1.92*	1.43	0.08	**0.06**	0.09	0.24	0.15
Other country	19.1	−2.67*	1.01	1.41	1.41	1.38	1.39	**0.80**
Year of arrival (if born overseas)								
Before 1971	24.0	3.04*	**−0.99**	−2.22	−2.24	−2.26	−2.09	−2.24
1971–1980	11.6	2.84*	1.27	−0.13	−0.16	−0.12	**0.04**	−0.39
1981–1990	16.8	8.08*	−4.84*	0.22	0.30	**0.12**	−0.31	−0.31
1991–2000	16.4	7.86*	−0.33	0.73	0.74	0.69	0.65	**−0.12**
2001–2010	28.5	−19.63*	4.57	1.21	**1.16**	1.35	1.51	2.71
2011	2.7	−2.19*	0.31	0.20	0.20	0.23	**0.20**	0.36
Indigenous	2.1	0.40	0.50	0.14	0.14	0.14	0.16	**0.13**
Employed persons: Usual hours worked								
1–15	11.6	**0.57**	1.13	0.70	0.69	0.72	0.74	0.81
16–29	13.0	**0.48**	1.19	1.06	1.06	1.05	1.07	1.01
30–34	5.7	0.34	0.34	0.29	0.29	0.30	**0.28**	0.38
35–39	23.4	−3.99*	−4.73*	−3.65*	−3.63*	−3.64*	−3.79*	**−3.52***
40	19.7	**−2.94***	−4.59*	−3.38*	−3.37*	−3.41*	−3.48*	−3.45*
41–44	3.2	**1.02***	1.19*	1.15*	1.15*	1.15*	1.16*	1.11*
45–49	7.1	2.15*	2.40*	2.12*	2.12*	2.11*	2.17*	**2.05***
50–59	9.3	2.22*	2.16*	1.74*	1.73*	1.72*	1.78*	**1.60***
60 or more	6.7	0.32	1.15	0.17	0.15	0.20	0.27	0.21
Employed persons: Employment status								
Employee	89.2	1.22*	**0.47**	1.24*	1.25*	1.22*	1.19*	1.09*
Employer	2.9	**−0.76***	−0.78	−0.83*	−0.83*	−0.80*	−0.83*	−0.80*
Own account worker (incl. contrib. fam. worker)	7.9	−0.46	0.31	−0.42	−0.43	−0.42	−0.36	**−0.29**

General remarks: The option with the lowest bias is typesetted bold faced. ABS estimates for relationship in household, country of birth, year of arrival and indigenous status exclude institutionalised population, otherwise the estimates apply to all civilians aged 15 and over. HILDA estimates are also for aged 15 and over including the defence force but excluding institutionalised population and very remote parts of Australia.

Abbreviations: A=main sample. B=top-up sample. C=combine on sample size. D=combine on optimal theta. E=combine on CV. F=combine on convenient theta. G=pooled samples.

ABS sources: Relationship in household, country of birth, year of arrival and usual hours worked: ABS Cat. No. 6291.0.55.001 (Labour Force Australia, Detailed Electronic Delivery, September 2011). Highest level of education is from ABS Cat.No. 62270DO001-201105 (Education and Work, Australia, May 2011); Indigenous status: ABS Cat. No. 62870DO001-2011 (Labour Force Characteristics of Aboriginal and Torres Strait Islander Australians, 2011); Employment status: ABS Cat. No. 6291.0.55.003 (Labour Force, Australia, Detailed, Quarterly, August, 2011.)

* $p < 0.05$ (only shown for bias estimates)

Table 6
Bias estimates, options A-G compared

Characteristic	RMSE						
	A	B	C	D	E	F	G
Relationship in household							
Couple with dependents	0.89	1.18	0.88	0.89	0.87	0.86	**0.83**
Couple without dependents	0.99	1.27	0.98	0.99	0.97	0.95	**0.92**
Lone parent with dependents	0.32	0.43	0.24	0.25	0.24	0.23	**0.21**
Lone parent without dependents	0.83	**0.32**	0.53	0.54	0.53	0.49	0.45
Dependent student	1.24	**0.86**	1.11	1.12	1.11	1.07	1.12
Nondependent child	1.52	0.81	0.52	0.53	0.51	0.45	**0.41**
Other family member	**0.44**	1.42	0.90	0.89	0.88	0.94	0.85
Nonfamily member	3.29	**2.01**	2.50	2.52	2.50	2.42	2.37
Highest level of education (15-64 year olds)							
Postgraduate (masters or doctorate)	**0.61**	1.32	0.71	0.70	0.72	0.77	0.81
Grad diploma or grad certificate	2.87	**2.81**	3.03	3.03	3.03	3.01	3.16
Bachelor or honours	2.56	**1.02**	1.35	1.37	1.36	1.25	1.33
Advanced diploma or diploma	0.55	0.73	0.36	0.37	0.36	**0.35**	0.35
Cert IV or III	4.09	5.33	3.49	3.46	3.44	3.68	**3.32**
Year 12	**1.71**	4.49	2.75	2.72	2.76	2.96	2.87
Year 11 or below (incl. Cert I, II, nfd)	**1.92**	4.96	3.07	3.04	3.01	3.27	3.00
Country of birth							
Australia	4.68	2.90	1.84	1.82	1.83	1.95	**1.44**
Main English speaking country	1.96	1.66	0.58	**0.57**	0.58	0.63	0.60
Other country	2.83	1.99	1.79	1.79	1.78	1.79	**1.37**
Year of arrival (if born overseas)							
Before 1971	3.36	**2.27**	2.54	2.56	2.57	2.43	2.56
1971–1980	3.12	1.82	0.96	0.97	0.95	0.92	**0.91**
1981–1990	8.23	4.97	1.17	1.20	1.14	1.12	**1.11**
1991–2000	8.15	1.69	1.54	1.55	1.49	1.43	**1.18**
2001–2010	19.64	5.13	2.63	**2.61**	2.69	2.75	3.64
2011	2.20	0.71	0.60	0.61	0.62	**0.60**	0.70
Indigenous	0.51	0.68	0.26	0.26	0.26	0.27	**0.24**
Employed persons: Usual hours worked							
1–15	**0.71**	1.40	0.80	0.79	0.82	0.84	0.89
16–29	**0.69**	1.62	1.24	1.24	1.22	1.24	1.18
30–34	0.46	0.60	0.39	0.40	0.40	**0.39**	0.46
35–39	4.04	4.85	3.71	3.69	3.69	3.84	**3.57**
40	**2.98**	4.66	3.41	3.40	3.44	3.52	3.48
41–44	**1.06**	1.30	1.19	1.19	1.19	1.20	1.14
45–49	2.19	2.50	2.17	2.16	2.16	2.22	**2.09**
50–59	2.28	2.27	1.79	1.79	1.77	1.83	**1.65**
60 or more	0.50	1.39	0.38	**0.38**	0.40	0.45	0.38
Employed persons: Employment status							
Employee	1.31	**1.00**	1.32	1.32	1.30	1.27	1.17
Employer	**0.78**	0.88	0.84	0.84	0.82	0.85	0.82
Own account worker (incl. contrib. fam. worker)	0.64	0.85	0.57	0.58	0.57	0.53	**0.48**

General remarks: The option with the lowest RMSE is typesetted bold faced.
Abbreviations: A=main sample. B=top-up sample. C=combine on sample size. D=combine on optimal theta. E=combine on CV. F=combine on convenient theta. G=pooled samples.

these bias estimates being significantly different from zero. The two groups of estimates that show the greatest improvement in bias (in Options B-G) are, not surprisingly, year of arrival and country of birth. As an example, the proportion of adults born abroad who arrived in Australia between 2001 to 2010 has risen from 9 per cent in the main sample only (Option A) to around 30 per cent in the integrated sample (regardless of the integration method), which is now in line with the ABS estimate. Country of birth is similarly affected by the integration of the top-up sample. The proportion of adults born in Australia based on the main sample only is 75 per cent and this is pulled more into line with the ABS figure of 70 per cent via the various integration options. There also appears to be less bias in the top-up sample (Option B) for the variables relating to relationship in household than in the main sample (Option A) and integrating these two samples help alleviate some of these biases.

While Option G generally brings the HILDA estimates closer to the ABS estimates, there are two variables that are further away from the ABS estimate than those from the main HILDA sample alone (Option A). These variables are hours worked and highest level of education. These differences may stem from differences in the collection methodology or questions asked which will in turn limit the validity of these comparisons. The Labour Force Survey obtains information about all adults in the household from any responsible adult whereas the HILDA Survey interviews each adult in the household. Wooden, Wilkins and McGuinness (2007) shows that probably for this reason the HILDA estimates on hours worked align more closely with the ABS Survey of Employment Arrangements and Superannuation, where all adults are interviewed, than the Labour Force Survey. To some extent this collection methodology will also impact on the highest level of education information collected with qualifications not being known to the responsible adult in the household. Further the questions asked about education are quite different between the HILDA Survey and the Labour Force Survey. Respondents to the HILDA Survey are asked to recount all of their education qualifications in their first interview and this is updated over time with subsequent education activity reported in later interviews. The ABS question in the Labour Force Survey asks for the highest level of education. It is possible that the respondent filters out some less important or less relevant qualifications when answering the more aggregated question used by the ABS. There is also some suggestion of this in the HILDA Survey, with wave 1 respondents and wave 11 top-up respondents aged 15–64 showing fewer Certificate III or IV and fewer graduate diplomas or certificates than respondents aged 15–64 in other waves. Nevertheless, the differences between the estimates from the main sample and the combined sample for these two variables are generally less than 1 percentage point.

Finally, we examine what impact these seven options have on the root mean square error as shown in Table 6. The option with the lowest RMSE is indicated in bold and is the best estimate. The method that provides the lowest RMSE on the majority of occasions is Option G where we have pooled the estimates. These improvements in RMSE under Option G are fairly consistently reproduced for most of the estimates though the differences to other integration options are not statistically significant.[8] Most of the time, the improved RMSE comes about via reduced variability in the estimates due to the lower coefficient of variation of the Option G weights. Figure 3 shows the percentage change in the standard errors of the estimates under Option G (pooling the samples) and Option C (combining the estimates based on sample size). For almost all estimates, there is a reduction in the standard error with Option G, with an average reduction of about 4 per cent. While not shown here, there is almost no difference in the standard errors of the estimates between the five options to combine estimates (Options C-F).

5 Conclusion

In this paper, we have examined the properties of six alternative ways to integrate an ongoing longitudinal sample with a top-up sample. By assessing the variability of the weights, the bias in the estimates and the overall root mean square error, we found that the method that performed the best was the one that estimates the sampling and response probabilities for the sample that each unit could have been but was not actually selected into, thus pooling the samples together. The pooling method often results in modest reductions in bias and RMSE. It has the lowest coefficient of variation of the weights which generally results in smaller estimated standard errors. This approach is also consistent with the modelling approach taken more broadly within the HILDA weighting process for new entrants who join existing households.

The essential difference between pooling samples (Option G) and combining estimates (Options C–F) is the relative importance given to each member of the sample. When pooling samples, each sample member is given a weight that reflects its likely sampling and response probability, so the households in the top-up sample who are least likely to be sampled or least likely to respond had they been in the main sample are given a higher weight than others. This permits the top-up sample to counteract the effect of attrition in the ongoing sample in the construction of cross-sectional weights. This will also have a bearing on the longitudinal weights that are created for balanced panels beginning with this combined sample (though it will have no effect on the balanced panels that start from earlier waves). In contrast, when combining

[8] The size of the standard error of the bias (given it is based on estimates from two independent surveys) is alone sufficient for the RMSEs to not be statistically different across the integration options.

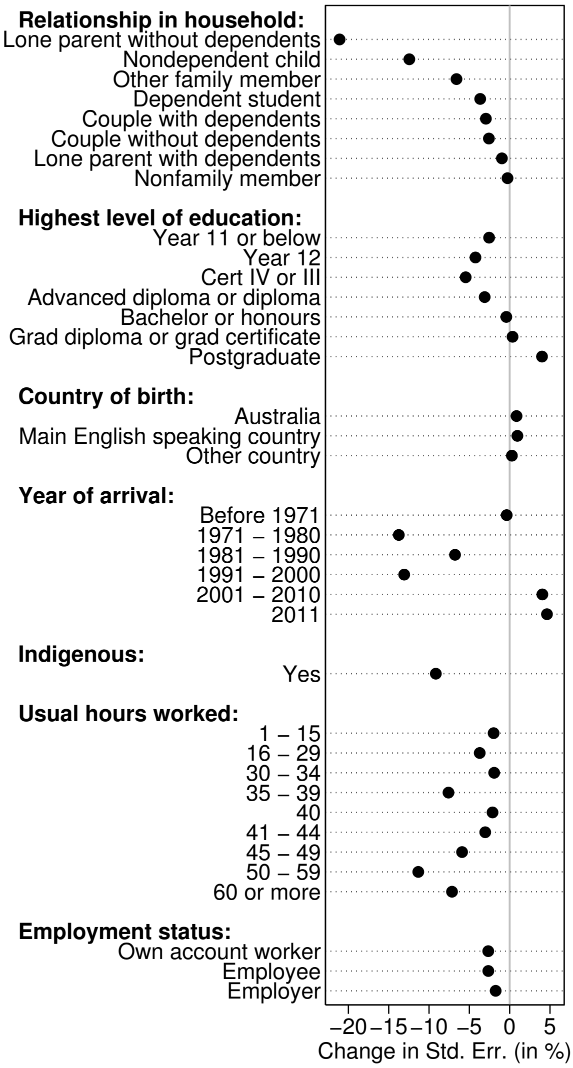

Figure 3. Percentage change in standard error for pooled sample (Option G) compared to combined sample based on relative sample size (Option C).

estimates, each person in the top-up sample is given the same level of importance when integrating the samples. The precise level of importance is defined in different ways giving rise to the four different options examined. It is not, therefore, surprising that by pooling the estimate would produce the best results as long as the sampling and response propensity can be approximated.

The finding that pooling samples is more efficient than combining estimates (due to a lower coefficient of variation of the weights) is expected to be broadly applicable to other longitudinal surveys with a top-up sample. It is expected that the benefits of pooling the samples would be even greater than what we have found in the case of the HILDA Survey for those surveys that oversample certain parts of the population, have highly differential non-response, or have a greater difference in the size of the main sample compared to the top-up sample. This is because when pooling samples, the sample members with similar characteristics (in so far as they can be modelled) would have similar weights which would reduce the overall coefficient of variation of the weights. Panel studies with a main sample and top-up sample of similar sizes (such as occurred in the SLID) have the least to gain from pooling samples as both samples have relatively similar weights prior to the integration.

6 Acknowledgments

This research has been supported, in part, by an Australian Research Council Discovery project grant (#DP1095497). It makes use of unit record data from the HILDA Survey (Release 12). 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, however, are those of the author and should not be attributed to either DSS or the Melbourne Institute. The author thanks John Henstridge, Robert Breunig, Stephen Horn, Frank Yu, the journal's associate editor and two referees for their helpful comments on an earlier version of this paper.

References

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). Singapore: John Wiley and Sons.

Haisken-DeNew, J. P. & Frick, J. R. (2005). *Desktop companion to the German Socio-Economic Panel (SOEP)*. Berlin: DIW Berlin.

Hartley, H. O. (1962). Multiple frame surveys. In *Proceedings of the social statistics section* (Vol. 19, *6*, pp. 203–206). American Statistical Association.

Heeringa, S. G., Berglund, P. A., Khan, A., Lee, S., & Gouskova, E. (2011). PSID cross-sectional individual weights, 1997-2009. In *PSID technical report*. Ann Arbor: Survey Research Centre.

Kalton, G. & Anderson, D. W. (1986). Sampling rare populations. *Journal of the Royal Statistical Society, Series A*, *149*(1), 65–82.

Kaminska, O. & Lynn, P. (2012). Combining refreshment or boost samples with an existing panel sample: challenges and solutions. *Paper presented at the 2012 Panel Survey Methods Workshop*.

Kish, L. (1992). Weighting for unequal pi. *Journal of Official Statistics*, *8*(2), 183–200.

LaRoche, S. (2007). Longitudinal and cross-sectional weighting of the Survey of Labour and Income Dynamics. *Income Research Paper Series*, *75F0002MIE*(007).

Lohr, S. L. & Rao, J. N. K. (2000). Inference from dual frame surveys. *Journal of the American Statistical Association*, *95*(449), 271–280.

Lynn, P. (2009). Methods for longitudinal surveys. In P. Lynn (Ed.), *Methodology of longitudinal surveys*. Chichester: John Wiley and Sons.

Lynn, P. & Kaminska, O. (2010). Weighting strategy for Understanding Society. *Understanding Society Working Paper Series, 2010-05*.

O'Muircheartaigh, C. & Pedlow, S. (2002). Combining samples vs. cumulating cases: a comparison of two weighting strategies in NLSY97. *American Statistical Association Proceedings of the Joint Statistical Meetings*, 2557–2562.

Schonlau, M., Kroh, M., & Watson, N. (2013). The implementation of cross-sectional weights in household panel surveys. *Statistics Surveys*, *7*, 37–57.

Sirken, M. G. & Casady, R. J. (1988). Sampling variance and nonresponse rates in dual frame, mixed mode surveys. In R. M. Groves, P. P. Biemer, L. Lyberg, J. T. Massey, W. Nicholls, & J. Waksberg (Eds.), *Telephone survey methodology*. New York: John Wiley and Sons.

Skinner, C. J. & Rao, J. N. K. (1996). Estimation in dual frame surveys with complex designs. *Journal of the American Statistical Association*, *91*(433), 349–356.

Spiess, M. & Rendtel, U. (2000). Combining an ongoing panel with a new cross-sectional sample. *DIW Discussion Paper Series*, (198).

Taylor, M. F., Brice, J., Buck, N., & Prentice-Lane, E. (2010). *British Household Panel Survey user manual volume A: introduction, technical report and appendices* (M. F. Taylor, Ed.). Colchester: Institute for Social and Economic Research.

Watson, N. (2012). Longitudinal and cross-sectional weighting methodology for the HILDA Survey. *HILDA Technical Paper Series*, *2*(12).

Watson, N. & Wooden, M. (2002). The household, income and labour dynamics in australia (HILDA) Survey: wave 1 survey methodology. *HILDA Technical Paper Series*, *1*(2).

Watson, N. & Wooden, M. (2013). Adding a top-up sample to the Household, Income and Labour Dynamics in Australia survey. *The Australian Economic Review*, *46*(4), 489–98.

Wooden, M., Wilkins, R., & McGuinness, S. (2007). Minimum wages and the working poor. *Economic Papers*, *26*(4), 295–307.