FACULTY OF
BUSINESS &
ECONOMICS

THE UNIVERSITY OF
MELBOURNE

HILDA PROJECT DISCUSSION PAPER SERIES
No. 1/16, December 2016

Evaluating potential improvements to the income imputation methods for the HILDA Survey

Nicole Watson and Ning Li

The HILDA Project was initiated, and is funded, by the Australian Government Department of Social Services

MELBOURNE INSTITUTE®
of Applied Economic and Social Research

Acknowledgements

Thanks to John Henstridge, Robert Breunig, Markus Grabka and Mark O'Shea for their helpful suggestions and comments on this research.

Contents

Introduction

An inevitable problem faced by surveys is missing data. In longitudinal surveys, there are three types of non-response: i) *item* non-response, where an individual provides an interview but does not provide a response to a particular item; ii) *wave* non-response, where an individual is interviewed in some waves but not others; and iii) *unit* non-response, where an individual is not interviewed at any wave. When interviews are sought with multiple people in a household, household-level variables that combine individual responses may be missing due to any of these forms of non-response.

Usually a combination of weighting and imputation is adopted to counter the effects of missing data. Consider, for example, non-response in the income data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Survey is a longitudinal household-based survey that began in 2001. Interviews are conducted with all adult members (aged 15 or over) in the household where possible and respondents are asked for details of their income from various sources. Missing income data is imputed if an individual is interviewed but does not provide an amount for a particular income source or if they are aged 15 and over and have not been interviewed but are part of a responding household (where at least one other person is interviewed). As a result, total person income and total household income can be calculated for all responding households once missing items are imputed at the person level (children aged 14 or under are assumed to have zero income). Of the 14,000 adults in responding households in wave 14, for instance, the proportion missing total person income due to item, wave, and unit non-response is 10%, 2.5% and 2.5% respectively.[1] Various longitudinal weights are used in concert with imputation to handle wave and unit non-response. Weights are provided to address wave and unit non-response at the individual level (for details of the weighting methodology see Watson, 2012).

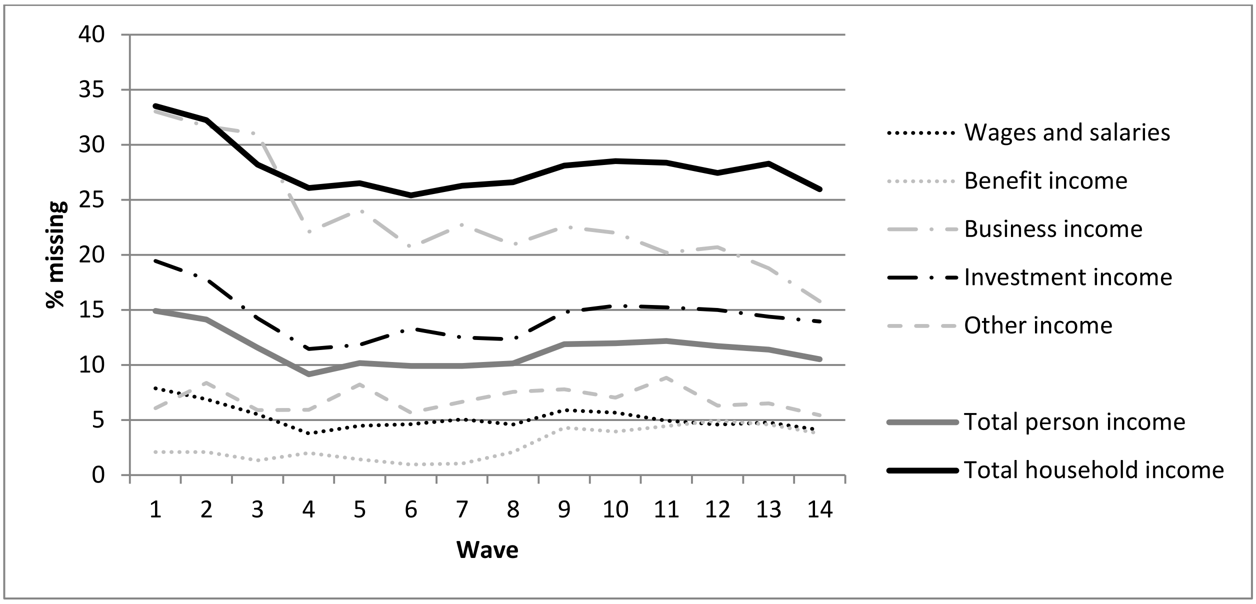
Missing values often concentrate on questions about finances, such as income, wealth and expenditure. This is because respondents either don't know or are not willing to report the values. Using HILDA Survey data, Figure 1 shows that the percentage of missing values varies substantially according to income component. For the five income components presented, the proportion of missing values relates to those respondents who report receiving income from the particular source (virtually every respondent answers the screener question as to whether they receive income from a particular source or not). The missing rate for total individual income or total household income relates to all respondents (irrespective of whether they receive any income or not). Most people with wages and salaries or benefit income report an amount, resulting in missing rates that are generally below 5%. Business income is more problematic, with around 30% of those with business income are unable or unwilling to provide an amount in the early waves and around 20% in the more recent waves. When all components of income are aggregated at the person level, 10 to 15% of respondents are missing total income. As expected, the proportion of respondents without total household income is much higher (25-33%).

Imputing cases with missing values avoids biases that may be introduced if researchers restrict their datasets to complete cases only (Little and Rubin 2002, p41). Further, having this imputation undertaken by the data producer ensures the majority of researchers will use the same imputed values. Of course, this should not preclude the researcher being able to undertake their own imputation should they wish to do so. This is facilitated by making available post-imputed data along with either imputation flags or pre-imputed data.

---

[1] The top-up sample, introduced in wave 11, has been excluded from the analyses in this paper.

Figure 1: Percentage of respondents with missing data, by income component



Note: For each income component, the percentage is calculated of those respondents receiving non-zero income for that component.

Once it is decided that variables need to be imputed, the question is then how best to do so. Watson and Starick (2011) undertook an evaluation of eight longitudinal imputation methods in a simulation study using income data from the first five waves of the HILDA Survey. The quality of the imputed data was assessed in terms of its predictive, distributional and estimation accuracy for cross-section estimates as well as estimates of change over time. They found that the method proposed by Little and Su (1989), when combined with a population carryover method (Williams and Bailey 1996) of zeros for non-respondents, performed the best overall. Also, using broad age imputation classes improved the imputation of some income components, such as wages and salaries and benefit income. For cases where the longitudinal imputation method cannot be used, Watson and Starick recommend a cross-sectional nearest neighbour method be used as the fall-back method.

The Little and Su method is currently used by a number of longitudinal studies, including the HILDA Survey (Hayes and Watson 2009), the German Socio-Economic Panel (Haisken-DeNew and Frick 2005), the UK Household Longitudinal Study (Knies 2015), the Swiss Household Panel (Lipps 2010), and the Longitudinal Study of Australian Children (Mullan et al 2015). Improvements to this method may, therefore, be of interest to the data producers and data users of these and other panel studies. Further, harmonising imputation methods across longitudinal studies may also be important for cross-country comparisons. For example, Frick and Grabka (2010) showed that some of the cross-national variation in an analysis of wages was simply due to the different imputation methods used.

The Little and Su method uses a multiplicative model to incorporate the trend across waves (referred to as the column effect), the recipients departure from the trend in waves where the income component has been reported (referred to as the row effect), and a residual donated from another respondent with complete income information (the residual effect). These are described below.

The column (wave) effects are essentially an adjustment for inflation and real income growth over time. It is calculated as:

$$c_j = \frac{\overline{Y}_j}{\overline{\overline{Y}}}$$

where $\overline{Y}_j$ is the sample mean of variable Y for wave j for complete cases

$\overline{\overline{Y}}$ is the global mean of variable Y based on complete cases, calculated as

$$\overline{\overline{Y}} = \frac{1}{m}\sum_{j=1}^{m} \overline{Y}_j$$ where m is the total number of waves

The row (person) effects are calculated for all cases:

$$\overline{Y}^{(i)} = \frac{1}{m_i}\sum_j \frac{Y_{ij}}{c_j}$$

where m_i is the number of recorded waves for case i

Y_{ij} is the variable of interest for case i, wave j

Cases are ordered by the row effect, $\overline{Y}^{(i)}$, and incomplete case i is matched to the closest complete case d.

The imputed value is then calculated as:

$$\hat{Y}_{ij} = \left(c_j\right)\left(\overline{Y}^{(i)}\right)\left(\frac{Y_{dj}}{c_j \overline{Y}^{(d)}}\right) = Y_{dj}\frac{\overline{Y}^{(i)}}{\overline{Y}^{(d)}}$$

The Little and Su method has a number of weaknesses as the length of the panel increases:

(i) the donor pool of complete cases grows smaller;

(ii) the imputation in early waves is always under review;

(iii) the information from each wave is not ordered in any way (i.e., reported values in the waves closest to the missing wave have the same influence on the individual's row effect as those further away);

(iv) the last reported value could be quite far away from the missing wave and the contemporaneous information from the wave with missing data is unused; and

(v) income patterns change over time so a donor chosen based on information in all available waves may not be the most suitable for the early waves or the later waves.

This paper examines several variations to the Little and Su method that restrict the calculation of the row effect to a subset (or window) of waves and weights the contribution of each wave. Different window sizes and different weights are tested. Further, the potential benefits of using the cross-sectional nearest neighbour imputes in the calculation of the row effects are also explored. A series of simulated datasets with missing values, based on HILDA Survey data, are used to assess the accuracy of the imputation. Estimates of both the level of income and the change in income are assessed for their predictive, distributional and estimation accuracy.

Variations to the Little and Su method

Several alternatives for how the Little and Su method could be applied in a longitudinal setting are considered in this paper. We restrict the calculation of the row effect to a limited set (i.e. a window) of waves rather than to all waves, which effectively gives zero weight to the waves outside of that window. The size of the window may be important so we assess windows of size 3, 5, 7, and 9 waves. The weight given to each wave within the window may also be important. In the Little and

Su method described above, equal weight is given to the information from each wave. This means, for example, that if a missing value for wages and salaries in wave 3 needs to be imputed then the (column adjusted) amounts the respondent received in wave 2 and wave 14 contribute equally to that individual's row effect. As a result, we consider two ways of weighting the waves within the window: i) applying a rectangle window (i.e. a uniform kernel) where each wave is given equal weight in the calculation of the row effect; and ii) applying a smooth kernel where the waves closer to the middle of the window have a higher weight than those further away. We also assess whether it is beneficial to use the cross-sectional nearest neighbour imputes in the calculation of the row effect.[2] These cross-sectional imputes incorporate information about the respondent's current circumstances (such as hours worked, occupation, industry, job tenure, education, and partner income if available) in a regression equation to predict each income component. This predicted value is used to find a donor with a similar predicted value and the donor's reported value is taken as the recipient's imputed value for the missing income component. The use of these cross-sectional imputes are assessed with windows of various sizes where each wave is given equal weight.

As a result, 14 imputation methods are assessed altogether:

Method 1 – Rectangle window of 14 waves (i.e., the current Little and Su method);

Method 2 – Rectangle window of 14 waves using cross-sectional nearest neighbour imputes;

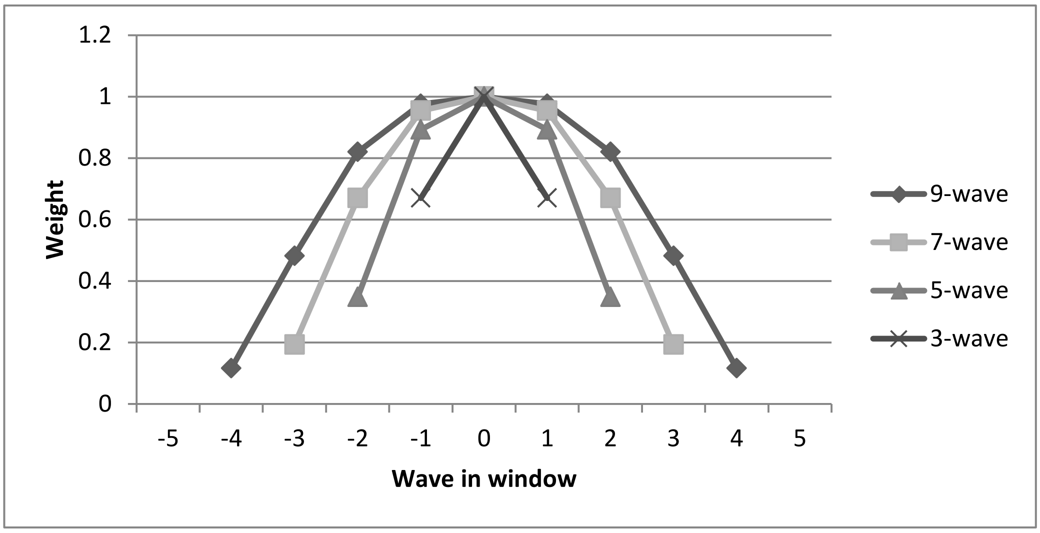
Methods 3 to 6 – Rectangle window of 3, 5, 7, and 9 waves respectively;

Methods 7 to 10 – Rectangle window of 3, 5, 7 and 9 waves respectively using cross-sectional nearest neighbour imputes; and

Methods 11 to 14 – Smooth kernel window of 3, 5, 7 and 9 waves respectively.

The smooth kernel adopted in this evaluation is the tri-cube kernel, $K(z) = (1 - |z|^3)^3$, as illustrated in Figure 2 for the four different window sizes. Analysts using smooth kernel functions tend to find that the choice of kernel is not as important as the bandwidth over which it is applied (Wand and Jones, 1995, p.13). Consistent with this, some initial comparisons of the imputation results using the Tri-cube kernel and those using the Epanecnikov quadratic kernel indicated they were very similar.

Figure 2: Weight from the tri-cube kernel function with windows of various sizes



[2] This is in addition to the situation in which the cross-sectional nearest neighbour imputes are used to calculate the row effect where the respondent has not reported income in any wave.

Having multiple windows across the span of the panel raises the question of how to position those windows. We opted to centre the window across a run of missingness (i.e. where consecutive waves have missing values) and also, where possible, the window makes maximum use of the reported information in surrounding waves.

Evaluation of alternative imputation methods

Ideally we would like to compare the imputed value with the value that the respondent would have reported if they had known the value or had not refused. As this is not possible, we simulate missing data in the HILDA Survey by deliberately setting some cases to missing. The 'true' values are those in the dataset before they are set to missing. The values imputed via the different methods are then evaluated against these 'true' values.

Dataset

Data from the first 14 waves of the HILDA Survey are used to generate a series of datasets with simulated missing values. The simulated missingness is added to the missingness already in the HILDA data and the various imputation methods impute both the actual and simulated missing values. There are between 13,300 and 15,100 adults in responding households each wave.

An alternative way to simulate missingness is to restrict the HILDA datasets to 'complete' cases only, as was done by Watson and Starick (2011). To be considered 'complete' a respondent would need to provide valid values to the three income components examined here in all waves they were eligible for. With a 14-wave panel, this approach greatly alters the shape of the panel, reducing the wave 1 sample by 63% and the wave 14 sample by 43% (this was far less of a problem with the 5-wave panel that Watson and Starick analysed). The reason the early waves are more affected than the later waves is that it is easier for a person who becomes eligible to be interviewed in later waves (i.e., a child turning 15 during the panel or a new entrant to the household) to provide complete responses for the small number of waves they are eligible for. Nonetheless, this evaluation was replicated with the simulated datasets based on complete cases and the overall conclusions are similar.

Variables

We focus on the three largest income components: wages and salaries, benefit income, and business income. Together these three components contribute an average of 86% to total person level income. Also, for the purposes of this evaluation, a measure of 'total' income is constructed as the sum of these three components.

Simulation

We assume that the probability an income amount is missing depends on a range of characteristics of the respondent but not on the income amount itself, that is, we adopt the missing at random assumption (Rubin 1976). A sample of cases with reported values in the HILDA data are set to missing, approximately doubling wave non-response and item non-response for wages and salaries and benefits. Item non-response for business income is increased by 30%, as doubling would have left a substantially smaller donor pool available for imputing the missing cases.

The assignment of missingness is achieved by modelling the response mechanism in the HILDA dataset and applying this to the cases with reported values. To account for the dependence of wave non-response across waves, the model for each wave includes an indicator of response in prior waves and some basic characteristics of respondents (age, sex, labour force status, relationship in the household, place of residence, value of the house, usual rent/mortgage repayments, and presence of a long-term health condition). The dependence of missingness between income components within a wave is accounted for by sequentially modelling the income components and including a response indicator for each income component in prior waves and previous income components of that wave together with a range of respondent characteristics. The respondent characteristics include marital status, highest level of education, occupation status, whether a multiple job holder, usual hours worked, whether the respondent speaks a language other than English, time since school spent working, time since school spent unemployed, along with the basic respondent characteristics included in the earlier model. As the missingness for each item is simulated, the probability that subsequent items are missing are recalculated.

A total of 45 simulated datasets were created.

A comparison of the characteristics of the simulated datasets to the original datasets is given in Table 1. Apart from the proportion of cases with missing the income components, the characteristics of the respondents in the original and simulated datasets are very similar.

Table 1: Characteristics of simulated datasets

	Original datasets	Simulated datasets
% respondents missing income (of those receiving component)		
Wages and salaries	5.2	10.3
Government benefits	2.9	5.5
Business income	20.7	27.2
Sum of components	5.7	10.1
% non-respondents (and therefore missing all income components)	5.9	11.8
Average number of adults in household	1.9	1.9
Characteristics of respondents		
Average age	44.0	44.1
% female	52.6	53.6
% employed	63.5	63.4
% married or defacto	62.0	61.8
% university degree or higher	21.2	20.8

Evaluation criteria

The quality of the imputed data is assessed via three criteria which measure the predictive, distributional and estimation accuracy. Watson and Starick (2011) used 11 criteria to assess these three dimensions (5 for predictive accuracy, 2 for distributional accuracy, and 4 for estimation accuracy). Most of these criteria were based on those proposed by Chambers (2001) for the Euredit Project which evaluated various imputation methods for mainly cross-sectional European surveys. We use one criteria for each dimension to simplify the reporting of the results. The particular criteria chosen in each dimension was the one that closely reflected the overall results in the earlier evaluation. All three of our measures of accuracy are defined on the *n* imputed values for a wave (or pair of waves) where the 'true' value is known.

The predictive accuracy is assessed via a regression approach useful for highly skewed data. The imputed and true values are transformed by taking the natural logarithm ($\ln(y+1)$) and then the transformed imputed values, $\hat{y}_t$, are regressed against the transformed true values, y_t^*, in the model $y_t^* = \beta\hat{y}_t$. The t-test statistic for $\beta = 1$ is calculated. The smaller the test statistic, the better the imputation method.

The distributional accuracy of the imputation is assessed by the maximum distance between the empirical distribution function of the true values and that of the imputed values (as measured by the Kolmogorov-Smirnov distance). The smaller the difference, the better the imputation method.

The estimation accuracy of the imputed data is measured as the absolute difference in the means of the true and imputed values.

Results

In order to summarise the results across the different criteria and to test for significant differences in the performance of the methods, the scores from each criteria are standardised and transformed to a log scale following the approach adopted in the Euredit Project (Chambers and Zhao 2003). The log standardised score is calculated as follows:

$$\log c_s = log\left(\frac{r}{|\widetilde{r_{nz}}|} + k\right)$$

where r is the residual after subtracting the median score for criteria c (i.e., $r = c - \tilde{c}$) where the median score $\tilde{c}$ is calculated across all methods, variables, waves and simulations, $\widetilde{r_{nz}}$ is the median of the non-zero residuals, and k is a small constant to ensure the resulting log score is valid. The standardisation is undertaken separately for respondents and non-respondents.

The log standardised scores are presented in Table 2 for the three dimensions of accuracy – predictive, distributional and estimation – for both cross-sectional and longitudinal estimates for each of the three components of income and the sum of these components. The average score across the three accuracy dimensions is also provided. The results are separately reported for respondents and non-respondents as the imputation values required for respondents are all non-zero and those for non-respondents can be zero or non-zero leading to much more variability in the imputed values. Methods with low average log standardised scores are better than those with high scores. The lowest score is highlighted in bold and statistically significant differences in the performance of the methods from the method with the lowest score are identified by an asterisk. These significant differences are found using Tukey's test, which reduces the probability of falsely identifying a significant result when multiple comparisons are being made.

We find that for both wages and salaries and benefit income of respondents, the imputed values from the Little and Su method with a 3-wave window where equal weight is given to the contribution from each wave performs significantly better than the full window (of 14 waves). There is no significant advantage in applying a smooth kernel function to weight the contributions of the different waves across the window – the log standardised scores of the method using the smooth kernel function are almost identical to those giving equal weight to the waves for the equivalent length of the window. We also find that including the cross-sectional nearest neighbour imputes in the calculation of the row effects tends to significantly reduce the accuracy of the imputation of these variables, both cross-sectionally and longitudinally, regardless of window length. It can sometimes improve one dimension of accuracy but this is usually at the expense of significantly reducing other dimensions. For example, use of the cross-sectional nearest neighbour imputes in calculating the row effects improves the cross-sectional distribution accuracy for wages and salaries of respondents but significantly reduces the longitudinal predictive and distribution accuracy.

For business income of respondents, we find some mixed results regarding the best length of the window. For cross-sectional estimates, the full window (of 14 waves) does not perform as well as those with shorter windows on two of the three accuracy dimensions, however for all three of the longitudinal accuracy dimensions the full window performs the best. As the longitudinal dimension is more important than the cross-sectional dimension in a longitudinal study, we would choose the full window over the shorter variants investigated here. Further, the use of the cross-sectional nearest neighbour imputes in the row effects is not beneficial for the imputation of respondents.

The results for the sum of the three income components show the impact of using the same method on all components. Here the 3-wave window performs the best for respondents (irrespective of how the contribution of the waves are weighted). This shows the importance of considering multiple variables when assessing imputation methods as some methods will be better for some variables than for others – only looking at total income (or the sum of the largest components) would have missed these nuances.

For non-respondents, we see that the full 14-wave window performs better than the other methods involving shorter windows, particularly on longitudinal predictive and distribution accuracy. For wages and salaries and benefit income, the traditional Little and Su method outperforms the method that uses the cross-sectional nearest neighbour imputes in the calculation of the row effects. However, for business income, the use of these cross-sectional imputes is beneficial on two of the three cross-sectional accuracy dimensions but does not significantly improve longitudinal accuracy.

These results suggest that a combination method involving the 3-wave window and the 14-wave window would be beneficial to the accuracy of the imputes. A 3-wave window would apply to respondents who are missing wages and salaries or benefit income and a 14-wave window would apply to business income of respondents and all income components for non-respondents. This combination method results in very similar accuracy measures as those reported for the corresponding method and variable in Table 2. The largest impact is on the longitudinal distributional accuracy of wages and salaries where the score worsens (i.e., increases) by 1.8 percent. Nevertheless, none of the changes in the accuracy scores are significantly different from the 3-wave window results for wages and salaries and benefit income for respondent, nor are they significantly different from the 14-wave window results for these variables for non-respondents. This suggests that by combining these two methods, we can realise the benefits of each without making any aspect of the imputation accuracy significantly worse.

Table 2: Average log standardised scores cross-sectional (X) and longitudinal (L) scores of predictive (P), distribution (D) and estimation (E) accuracy

	Wages and Salaries								Government benefits							
Method	XP	XD	XE	LP	LD	LE	X	L	XP	XD	XE	LP	LD	LE	X	L
Respondents																
LS 14	1.51*	1.41*	1.45*	1.99*	1.61*	1.91	1.46*	1.84*	1.47	1.67*	1.49*	1.64	1.75	2.14	1.54*	1.85
LS NN 14	1.54*	1.35*	1.45*	2.04*	1.65*	1.93	1.44*	1.87*	1.50*	1.69*	1.49*	1.68	1.79*	2.18	1.56*	1.88
LS 3	1.49*	**1.27**	1.42	1.96	1.51	1.86	1.39	1.77	1.46	**1.59**	1.46	**1.64**	**1.72**	2.16	**1.50**	**1.84**
LS 5	1.48*	1.28	1.42	2.04*	1.61*	1.86	1.39	1.84*	1.46	1.60	1.45	1.69*	1.76	2.12	1.51	1.86
LS 7	1.47	1.30*	1.42	2.08*	1.67*	1.88	1.40	1.87*	1.46	1.61	1.46	1.71*	1.81*	2.16	1.51	1.90*
LS 9	**1.45**	1.31*	1.43*	2.09*	1.70*	1.91	1.40	1.90*	1.46	1.61	1.46	1.72*	1.81*	2.14	1.51	1.89
LS NN 3	1.57*	1.27	**1.42**	2.04*	1.59*	**1.85**	1.42*	1.83*	1.52*	1.60	1.45	1.72*	1.77*	2.14	1.52*	1.88
LS NN 5	1.55*	1.27	1.42	2.07*	1.65*	1.87	1.42*	1.86*	1.51*	1.61	1.46	1.75*	1.80*	**2.11**	1.52*	1.89
LS NN 7	1.54*	1.29*	1.42	2.11*	1.69*	1.87	1.42*	1.89*	1.51*	1.60	1.45	1.76*	1.82*	2.13	1.52	1.91*
LS NN 9	1.55*	1.30*	1.43	2.12*	1.70*	1.89	1.42*	1.90*	1.50*	1.60	**1.45**	1.76*	1.82*	2.14	1.52	1.90*
LS SK 3	1.49*	1.27	1.42	**1.95**	**1.49**	1.86	**1.39**	**1.76**	1.47	1.60	1.46	1.64	1.73	2.16	1.51	1.84
LS SK 5	1.48	1.28	1.42	2.01*	1.58*	1.88	1.39	1.82*	1.46	1.61	1.46	1.65	1.73	2.17	1.51	1.85
LS SK 7	1.49*	1.30*	1.43	2.06*	1.64*	1.89	1.41	1.86*	1.46	1.60	1.46	1.72*	1.79*	2.16	1.51	1.89
LS SK 9	1.48	1.31*	1.43*	2.08*	1.68*	1.90	1.41	1.89*	**1.45**	1.62	1.46	1.71*	1.82*	2.16	1.51	1.89
Non-respondents																
LS 14	1.42	1.58*	1.42	1.31	**1.79**	**2.21**	1.48*	**1.77**	1.74*	1.52*	1.45	1.60	**1.76**	**2.19**	1.57*	**1.85**
LS NN 14	**1.42**	1.62*	1.47*	**1.31**	1.82	2.28	1.50*	1.81	**1.70**	1.58*	1.46*	**1.59**	1.78	2.25	1.58*	1.88
LS 3	1.45*	1.49	1.42	1.53*	2.13*	2.27	1.45	1.98*	1.75*	1.49	1.44	1.74*	1.99*	2.34	1.56	2.02*
LS 5	1.44*	1.50	1.42	1.56*	2.20*	2.31	1.45	2.02*	1.74*	1.48	1.44	1.75*	2.01*	2.30	1.55	2.02*
LS 7	1.43	1.51*	1.42	1.56*	2.21*	2.32	1.45	2.03*	1.73*	**1.48**	1.44	1.75*	2.03*	2.33	**1.55**	2.04*
LS 9	1.43	1.52*	1.42	1.55*	2.21*	2.29	1.46	2.02*	1.73*	1.49	1.44	1.75*	2.02*	2.31	1.55	2.03*
LS NN 3	1.45*	1.51*	1.43*	1.56*	2.25*	2.27	1.46	2.02*	1.73*	1.52*	1.45*	1.75*	2.06*	2.35*	1.57*	2.05*
LS NN 5	1.44*	1.51	1.43*	1.56*	2.25*	2.30	1.46	2.04*	1.73*	1.51*	1.45*	1.75*	2.06*	2.35*	1.56	2.06*
LS NN 7	1.43	1.51*	1.43*	1.57*	2.24*	2.31	1.46	2.04*	1.72	1.50	1.45*	1.75*	2.06*	2.34	1.55	2.05*
LS NN 9	1.43*	1.52*	1.43*	1.56*	2.22*	2.30	1.46*	2.02*	1.72	1.52*	1.45*	1.74*	2.04*	2.36*	1.56	2.05*
LS SK 3	1.45*	**1.48**	**1.42**	1.53*	2.11*	2.24	**1.45**	1.96*	1.75*	1.49	1.44	1.74*	2.00*	2.33	1.56	2.02*
LS SK 5	1.44*	1.50	1.42	1.56*	2.18*	2.28	1.45	2.01*	1.74*	1.48	**1.44**	1.75*	2.00*	2.30	1.55	2.02*
LS SK 7	1.43	1.52*	1.42	1.56*	2.20*	2.30	1.46	2.02*	1.73*	1.48	1.44	1.75*	2.03*	2.32	1.55	2.03*
LS SK 9	1.43	1.53*	1.42	1.55*	2.20*	2.31	1.46	2.02*	1.73*	1.49	1.44	1.75*	2.02*	2.37*	1.55	2.05*

Table 2 *continued*

Method	Business Income XP	XD	XE	LP	LD	LE	X	L	Sum of income components XP	XD	XE	LP	LD	LE	X	L
Respondents																
LS 14	1.51*	1.80*	**1.82**	**1.51**	**1.81**	**2.84**	1.71	**2.05**	1.59*	1.43*	1.44*	2.05	1.55*	**2.03**	1.49*	1.88*
LS NN 14	1.47	1.87*	1.90	1.55*	1.86*	2.91	1.75	2.11	1.52*	1.32*	1.44	2.09*	1.57*	2.05	1.43*	1.90*
LS 3	1.46	1.77	1.95*	1.57*	1.87*	3.00	1.73	2.15	1.48	1.24	1.43	2.03	1.44	2.04	1.39	1.84
LS 5	1.47	1.75	1.91	1.59*	1.87*	3.01	1.71	2.15*	1.48	1.26*	1.43	2.11*	1.54*	2.03	1.39	1.89*
LS 7	1.46	1.75	1.93*	1.60*	1.87*	3.06*	1.71	2.18*	1.46	1.29*	1.43	2.14*	1.60*	2.07	1.39	1.94*
LS 9	1.46	1.75	1.91	1.58*	1.86*	3.03	1.71	2.16*	**1.45**	1.30*	1.43	2.16*	1.62*	2.07	1.40	1.95*
LS NN 3	1.48	1.81*	2.02*	1.63*	1.92*	3.08*	1.77*	2.21*	1.57*	1.24	1.44*	2.11*	1.52*	2.07	1.42*	1.90*
LS NN 5	1.48	1.80*	1.98*	1.64*	1.93*	3.02	1.76*	2.20*	1.55*	1.26	1.44	2.14*	1.57*	2.06	1.42*	1.92*
LS NN 7	1.49*	1.80*	1.98*	1.64*	1.94*	3.04	1.76*	2.21*	1.53*	1.27*	1.44*	2.18*	1.60*	2.07	1.41*	1.95*
LS NN 9	1.49*	1.82*	1.98*	1.64*	1.94*	3.02	1.76*	2.20*	1.53*	1.28*	1.44*	2.19*	1.62*	2.09	1.42*	1.97*
LS SK 3	1.46	1.77	1.96*	1.57*	1.86*	3.03	1.73	2.16*	1.48	**1.24**	1.43	**2.02**	**1.42**	2.03	**1.39**	**1.82**
LS SK 5	1.47	1.75	1.93*	1.58*	1.86*	3.03	1.72	2.16*	1.47	1.26*	**1.43**	2.07*	1.50*	2.05	1.39	1.87
LS SK 7	1.46	**1.75**	1.93*	1.60*	1.86*	3.05*	1.71	2.17*	1.47	1.29*	1.43	2.13*	1.57*	2.04	1.40	1.91*
LS SK 9	**1.45**	1.75	1.91	1.58*	1.86*	3.02	**1.71**	2.15*	1.47	1.31*	1.44	2.15*	1.62*	2.09	1.40*	1.95*
Non-respondents																
LS 14	2.00*	1.38	1.74*	1.82	1.53	2.66	1.71	2.00	1.27*	1.65*	1.42*	1.33	**1.87**	2.32	1.45*	1.84
LS NN 14	**1.95**	1.39*	**1.68**	**1.78**	**1.50**	**2.56**	**1.67**	**1.95**	**1.23**	1.71*	1.43*	**1.31**	1.90	**2.22**	1.45*	**1.81**
LS 3	2.02*	1.37	1.72	1.93*	1.61*	2.74	1.70	2.09*	1.31*	**1.49**	1.42*	1.50*	2.16*	2.38	1.41	2.01*
LS 5	2.01*	1.37	1.72	1.93*	1.62*	2.74	1.70	2.10*	1.30*	1.50	1.41	1.52*	2.21*	2.35	1.40	2.03*
LS 7	2.01*	1.37	1.70	1.93*	1.62*	2.73	1.69	2.09*	1.29*	1.50	1.41	1.53*	2.22*	2.32	**1.40**	2.03*
LS 9	2.01*	1.38	1.70	1.92*	1.62*	2.74	1.70	2.09*	1.28*	1.52*	1.41	1.52*	2.22*	2.31	1.40	2.02*
LS NN 3	2.01*	1.39*	1.74*	1.93*	1.61*	2.72	1.71*	2.09*	1.30*	1.50	1.42	1.51*	2.23*	2.29	1.41	2.01*
LS NN 5	2.00*	1.39	1.73	1.93*	1.62*	2.74	1.71	2.10*	1.30*	1.50	1.42	1.52*	2.24*	2.30	1.40	2.02*
LS NN 7	2.00*	1.38	1.72	1.93*	1.61*	2.69	1.70	2.08*	1.28*	1.50	1.41	1.53*	2.24*	2.29	1.40	2.02*
LS NN 9	2.00*	1.40*	1.72	1.92*	1.61*	2.72	1.71	2.08*	1.28*	1.52*	1.41	1.52*	2.23*	2.29	1.40	2.01*
LS SK 3	2.02*	**1.36**	1.71	1.93*	1.62*	2.73	1.70	2.09*	1.32*	1.49	1.42*	1.50*	2.15*	2.40*	1.41	2.01*
LS SK 5	2.01*	1.37	1.71	1.93*	1.62*	2.74	1.70	2.09*	1.30*	1.50	1.41	1.52*	2.20*	2.36	1.40	2.03*
LS SK 7	2.01*	1.36	1.70	1.93*	1.61*	2.69	1.69	2.08*	1.29*	1.50	1.41	1.53*	2.21*	2.32	1.40	2.02*
LS SK 9	2.01*	1.38	1.70	1.92*	1.62*	2.68	1.70	2.07*	1.28*	1.53*	**1.41**	1.53*	2.21*	2.33	1.41	2.02*

Minimum score on accuracy dimension for respondents and for non-respondents are in bold. * indicates methods significantly different from method with minimum score.

LS w = Little and Su method with w-wave window, where w=3, 5, 7, 9 and 14. LS NN w=Little and Su method using cross-sectional nearest neighbour imputes with w-wave window, LS SK w=Little and Su method with w-wave smooth kernel window.

Conclusions

This investigation of alternative variants of the Little and Su method has not identified a single solution for all income components. We find that it is better to use a 3-wave window with equal weights given to each wave in the imputation of wages and salaries and benefit income of respondents. For business income of respondents and for all income components of non-respondents it is better to use the full 14-wave window (as is the current practice). This is likely because business income of respondents and the income of non-respondents (which can include zero and non-zero amounts) are quite variable and it is better to use as much of the available information as possible. There is little support for the use of the cross-sectional nearest neighbour imputes in the calculation of the row effect or to have a smooth kernel function to weight the contribution of different waves.

This approach does add more complexity to the imputation programs, but it will stabilise the imputation of wages and salaries and benefit income for respondents in all but the most recent waves. Total income may still be subject to change in early waves for some respondents as business income may change. Also, given we are not recommending one imputation method to be used for all income components, we will need to determine what method will be used for the other income components not examined here.

It is expected that the results produced using the HILDA Survey data would translate to other longitudinal surveys. Ideally, this should be confirmed through replication of this evaluation study.

References

Chambers, R. (2001). *Evaluation criteria for statistical editing and imputation.* National Statistics Methodological Series No. 28, Office for National Statistics.

Chambers, R. & Zhao, X. (2003). Evaluation of Edit and Imputation Methods Applied to the UK Annual Business Inquiry (Chapter 4.2). In J. Charlton, *Towards Effective Statistical Editing and Imputation Strategies – Findings of the EUREDIT Project* (Volumes 1 and 2). Retrieved from www.cs.york.ac.uk/euredit

Haisken-DeNew, J. P. & Frick, J. R. (eds.) (2005). *Desktop Companion to the German Socio-Economic Panel*. Berlin: German Institute for Economic Research (DIW Berlin).

Hayes, C. & Watson, N. (2009). *HILDA Imputation Methods*. HILDA Technical Paper 2/09, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Knies, G. (ed.) (2015). *Understanding Society: The UK Household Longitudinal Study Waves 1-5 User Manual*. Institute for Social and Economic Research, University of Essex.

Lipps, O. (2010). *Income Imputation in the Swiss Household Panel 1999-2007*. FORS Working Papers 2010-1. Swiss Foundation for Research in the Social Sciences, University of Lausanne.

Little, R. J. A. & Su, H. L. (1989). Item Non-Response in Panel Surveys. In D. Kasprzyk, G.J. Duncan, G. Kalton, M. P. Singh, *Panel Surveys*. New York: John Wiley and Sons.

Mullan, K., Daraganova, G., and Baker, K. (2015). *Imputing Income in the Longitudinal Study of Australian Children*. LSAC Technical Paper 14, Australian Institute of Family Studies.

Rubin, D. B. (1976). Inference and Missing Data. *Biometrika, 63*(3), 581-592.

Little, R. J. A. & Rubin, D. B. (2002). *Statistical Analysis with Missing Data,* Second Edition. New York: John Wiley and Sons.

Wand, M. P. & Jones, M. C. (1995). *Kernel Smoothing*. London: Chapman and Hall.

Watson, N (2012). *Longitudinal and Cross-sectional Weighting Methodology for the HILDA Survey*. HILDA Technical Paper 2/12, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Watson, N. & Starick, R. (2011). Evaluation of Alternative Income Imputation Methods for a Longitudinal Survey. *Journal of Official Statistics, 27*(4), 693-715.

Williams, T. R. & Bailey, L. (1996). *Compensating for Missing Wave Data in the Survey of Income and Program Participation (SIPP)*. Proceedings of the American Statistical Association, Survey Research Methods Section, Chicago.