Evaluation of Alternative Income Imputation Methods: the HILDA Experience

Rosslyn Starick[1] and Nicole Watson[2]
[1] Australian Bureau of Statistics (seconded to Melbourne Institute of Applied Economic and Social Research)
[2] Melbourne Institute of Applied Economic and Social Research, The University of Melbourne

June 2006

Abstract

This paper assesses various methods for imputing missing income data in a household-based longitudinal survey. Taking guidance from the experience of similar longitudinal studies, eight methods are evaluated in a simulation using data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The quality of the imputed data is evaluated against nine criteria. In a cross-sectional sense, many of the imputation methods perform well. However, the key concern is the quality of the cross-wave longitudinal imputations. When the methods are placed in a longitudinal context, the strengths and weaknesses of the methods become more apparent. The Little and Su method, implemented using imputation classes, performs the best for most income components. A simpler method is preferred where there fewer donors to draw on (such as for business income).

1. Introduction

All large-scale surveys have non-response and longitudinal surveys are no exception. There are three types of non-response in longitudinal surveys: firstly a respondent may not know or may not wish to answer a particular question during their interview (item non-response); secondly a person may not provide an interview in a certain wave but is interviewed in at least one other wave (wave non-response); and finally, a person may not provide an interview in any wave (unit non-response). As non-respondents may have different characteristics from respondents, ignoring their existence may result in biased population or regression estimates. Non-response is typically addressed by a combination of weighting and imputation. Weighting is usually used to address unit non-response and item non-response is usually dealt with by imputation (Lepkowski, 1989; Nordholt, 1998; Kalton and Brick, 2000; Dillman et al., 2002). Wave non-response may be viewed as a set of item non-response in the longitudinal record so imputation may be appropriate. Alternatively, it may be viewed as a missing unit for a wave and weighting adjustments may be appropriate (Kalton, 1986).

The imputation methods adopted in cross-sectional settings have been used for many years and are reasonably well understood, but less is known about the performance of imputation methods in a longitudinal context. Additional demands are placed on the imputation method in a longitudinal survey. What method preserves both the cross-sectional estimates and the estimates of change across waves? Should multiple wave non-response be imputed with the same donor or different donors at each wave? How do we best use subsequent and previous wave data in the imputation method?[1] How far into the future or into the past should we go?

The early experience of the HILDA Survey demonstrates one of the difficulties with imputation in the longitudinal context. The HILDA Survey is a large nationally-representative longitudinal survey of Australian households that commenced in 2001 (Watson and Wooden, 2004). Our focus is on income imputation as the income variables form a key part of the survey and are subject to substantial non-response.[2] A nearest neighbour regression method (described later in this paper)

[1] In a longitudinal survey, there is usually the opportunity to re-impute prior waves with each new release of the data.
[2] In wave 1, for example, 16 per cent of the respondents and 29 per cent of the households with a respondent were missing at least one income component.

was initially adopted to impute missing income data. This method led to an overstatement of the change in income between waves, even though the models used income information from other waves, other income variables from the wave being imputed and other respondent characteristics (Watson, 2004). Following the evaluation study which is the subject of this paper, a variant of the Little and Su method (Little and Su, 1989, also described later in this paper) was used in the latest release of the data, where donors are matched to recipients within imputation classes for some variables.

The evaluation study used data from the first three waves of the HILDA Survey to construct simulated data on which imputation methods were tested. Nine evaluation criteria were used to assess the predictive accuracy, distributional accuracy and estimation accuracy of each method. Eight imputation methods were tested, being one nearest neighbour method, four variants of the Little and Su method, and three carryover methods (all described below).

The various design features of the HILDA Survey that are relevant to the development of the imputation strategy include:

- The survey is household-based and, within some households, some but not all individuals provide an interview. For some analyses, the unit of interest will be the household and, for others, it will be the individual. Some individuals may never provide an interview, yet are part of a responding household for which total income is required.
- Weights are used to adjust for complete household-level unit and wave non-response. Imputation is used to complete the missing income data for person-level unit and wave non-response within a responding household and for person-level item non-response.
- The income module is included in the questionnaire at every wave. As all of the income components have a screener question to identify whether the respondent has income from the particular source or not, we almost always know whether any missing income is non-zero for a person who has been interviewed. Total individual income is calculated from the responses relating to each income component. Total household income is the sum of the incomes of the individuals in the household.
- The data are released annually and re-imputation occurs each time. This means that the imputation for any given wave can take advantage of the information in any future waves.

Our evaluation study is different from previous studies of income imputation methods in a longitudinal context in a number of ways. Previous studies have tended to focus on a single income variable, usually total income (for example, Tremblay, 1994; Quintano, et al, 2002; Frick and Grabka, 2003). As a result, these studies aim to identify one imputation strategy that provides the best results for this variable. Williams and Bailey (1996) did consider four income components but sought one imputation method that worked well for all. By contrast, we evaluate the performance of the imputation methods on all income components with a view to understanding which methods work ‘best’ for which components. We are open to using a variety of methods in the HILDA Survey if different methods work best for different variables. The evaluation criteria used in previous studies have varied, making a comparison of results difficult. Our approach is to use a relatively large number of evaluation criteria, although we of course do not claim to have used all possible criteria.

This paper describes a methodological evaluation framework for assessing a good imputation method (in section 2), details alternative imputation methods (in section 3) and presents a summary comparison of the performance of these methods (in section 4).

2. Evaluation Methodology

This section outlines how the evaluation study was designed, and, in particular, how the response mechanism was modeled in the simulated datasets. Also described are the evaluation criteria used in the paper.

2.1 Simulated Data

Ideally we want to compare the imputed value with the value the respondent would have reported if they had not refused or didn't know the value. However, as this is impossible, we simulated a series of datasets with missing values by taking what were actually complete cases and setting a proportion to missing. We use the term 'true' value in this paper to mean what a respondent actually reported before his/her data were set to missing.[3] We can then compare the imputed values from the various imputation methods with the true values.

More specifically, the simulated datasets comprise 8,720 people who responded and provided all income items in the waves they were eligible for. The sample of cases set to missing was based on a model of the response mechanism from the full HILDA dataset that assumed the missing values were missing at random (Rubin, 1976). That is, the probability that the income component was missing depended on a range of characteristics of the respondent but not on the value of the income component itself.

Modelling the Response Mechanism

Logistic models were constructed for each wave from the entire HILDA dataset to identify the relationship between the probability of reporting a particular income component and various explanatory variables.[4] Only cases that reported having that particular income were included in the models.

To take account of the dependence between income variables, the response mechanism was modeled sequentially through the list of components. As a first step, a model was constructed to predict the presence of a response to the question about wages and salaries. Then a model was constructed to predict the presence of a response to Government pensions, contingent on the response indicator for wages and salaries, and so on.

The probability that an individual provided an interview was also modeled at each wave.

Simulations

Ten datasets were simulated from the set of all complete cases by using the predicted probabilities of response to the various income components from the models described above.

In line with the proportion of non-responding persons observed in the entire HILDA data, a sample of non-responding persons from the set of complete cases was determined first. This was done by randomly assigning cases to responding or not, proportional to their predicted probability for being a non-respondent. All income values for the simulated non-respondents were set to missing.

The remaining cases became simulated respondents. The income components for a portion of these respondents were set to missing, in line with the missing rate in the entire HILDA dataset. This was done sequentially to mirror the dependant nature of the missing data. That is, a random sample of cases was set to missing for wages and salaries. Then, a random sample of cases was set to missing for Government pensions, which took into account previously simulated response indicators (in this case, previously simulated missing wages and salaries), and so on. The predicted value for all non-zero cases was calculated and only these cases were considered in simulating the missingness in that variable.

This procedure for creating a simulated dataset was carried out ten times to produce ten datasets for the evaluation, each with a different set of respondents with missing income variables and with a

[3] While what the respondent reports may not be completely correct, we cannot compare the imputed values to the correct values (as we do not know what they are), so we compare them to what was reported.

[4] The variables considered in the models include: age, marital status, highest level of education, labour force status, occupation status, whether works in multiple jobs, usual hours worked, place of residence, value of the house, usual rent/mortgage repayments, whether speaks a language other than English, whether has a long term health condition, time since school spent working, time since school spent unemployed, and several variables that relate to the last financial year such as time spent in part time education, time spent employed, and number of jobs held.

different set of non-respondents. The missing data were imputed via each imputation method and the resulting evaluation measures were averaged to form a single set of results that are presented later in this paper.

Table 1 provides summary measures of the simulated datasets, including the number of cases that need to be imputed and various characteristics of the potential donors (who have non-zero income amounts).[5] The donor characteristics provided are the mean and median value of the income component, the standard deviation and inter-quartile range as a multiple of the mean and median respectively, and the correlation of the income component with age. Some income components (such as business income, dividends and royalties, and rental income) are highly variable and present a challenge for imputing well. Business income also has a large number of cases requiring imputation with a small donor pool to draw from. This table aids the interpretation of the results later in this paper.

While we aimed to make the simulated datasets as realistic as possible to the HILDA environment, some differences between the two datasets are observed. The average household size from the evaluation dataset is about 1.5 compared to the average household size of 1.9 (excluding children) in the real HILDA data. This is because part responding households (where some but not all adults provided an interview) could not be included in the simulation datasets as they were not 'complete'. The evaluation data consists of a slightly larger proportion of the older population and slightly less of the younger population than the full HILDA data (in the evaluation dataset, the average age of adult enumerated persons was 43.6 years compared to 43.3 years in the full HILDA dataset). This may be because the older population is more likely to report complete income data. Nevertheless, even with these limitations, the simulated datasets provide a very valuable basis for comparing imputation methods.

Table 1: Characteristics of simulated datasets, averaged across waves 1 to 3

	Number imputed (recipients)	*Potential donors*						
		Number [a]	*Recipnts / donors*	*Mean*	*Std dev / mean*	*Median*	*IQR / median*	*Corr with age*
Wages and salaries	293	4,047	0.07	36,185	0.83	32,153	1.01	0.22
Aust govt pensions	37	2,657	0.01	7,638	0.56	8,552	0.69	0.32
Business income	117	303	0.39	19,295	2.79	10,575	2.49	0.01
Interest income	272	1,293	0.21	2,233	3.03	562	2.59	0.07
Dividends and royalties	246	1,565	0.16	2,445	3.60	210	4.67	0.13
Rent income	80	405	0.20	2,861	5.59	1,801	3.34	0.12
Private pensions	25	498	0.05	19,438	2.13	11,852	1.64	0.11
Private transfers	34	167	0.20	4,913	1.18	3,367	1.65	0.15
Total FY income[b]	556	6,508	0.09	26,872	1.20	17,732	1.69	-0.04

Notes: a. Number of potential donors depends on the imputation method being used. The number reported in this table is the number of cases reporting the income item. This is the number of potential donors for the nearest neighbour regression method. Other imputation methods will have the same or fewer potential donors.

b. The number of potential donors for total financial year income is calculated as the number of cases providing complete income details. This is an underestimate. In reality, the total financial year income for non-responding people in responding households will be imputed after the responding persons have been imputed. All respondents (including those with some components imputed) will be considered as potential donors. As the characteristics of the potential donors will depend on the imputation method adopted, the characteristics reported in this table are of complete cases which provide a reasonable guide to those of all responding persons after imputation.

[5] The record with missing information is called the 'recipient' (i.e., it needs to be imputed). The 'donor' has complete information that is used to impute the recipient's missing value.

2.2 Evaluation Criteria

This section defines the evaluation criteria that provide the framework for comparing the imputation methods. A good imputation method must reproduce key statistical properties of the complete data. The evaluation criteria compare the imputed values with the true values in the simulated data.

The first five criteria are based on those proposed by Chambers (2000) for the Euredit Project. We have included four additional criteria that measure the predictive accuracy and distributional accuracy. When undertaking regression analysis, all nine criteria are important. They measure predictive accuracy (criteria 1, 2, 6, 8 and 9), distributional accuracy (criteria 3 and 7) and estimation accuracy (criteria 4 and 5). When producing aggregate estimates, distributional accuracy and estimation accuracy are important.

For a longitudinal survey it is important that the imputation method performs well both cross-sectionally and longitudinally. Criteria 1 to 5 are applied to both the *level* of income at each wave and the *change* in income between waves. Criteria 6 and 7 apply only to the *change* in income between waves. Criteria 8 and 9 apply only to the *level* of income at each wave.

Unless otherwise stated, all criteria are defined on the set of *n* imputed values within a dataset, rather than the set of all values. $\hat{Y}$ denotes the imputed version of variable *Y* and Y^* denote the true version of the same variable.

Criteria 1 and 2: Predictive Accuracy

The imputation method should preserve the true values as far as possible. The imputed value ($\hat{Y}$) should be close to the true value (Y^*). The Pearson correlation between $\hat{Y}$ and Y^* provides a good measure of the imputation performance for data that are reasonably normal.:

$$r_{\hat{Y}Y^*} = \frac{\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{\hat{Y}}\right)\left(Y_i^* - \bar{Y}^*\right)}{\sqrt{\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{\hat{Y}}\right)^2 \sum_{i=1}^{n}\left(Y_i^* - \bar{Y}^*\right)^2}} \qquad \text{(Criteria 1)}$$

where $\bar{Y}$ denotes the mean of *Y*-values. A good imputation method will have *r* close to 1.

A regression approach can be used to evaluate the performance of the imputation method for highly skewed data. The imputed and true values are first transformed by taking the natural logarithm (*log (Y+1)*).[6] The transformed imputed values ($\hat{Y}_t$) are then regressed against the transformed true values (Y_t^*) using a linear model

$$Y_t^* = \beta\hat{Y}_t + \varepsilon$$

For comparing imputation methods, the t-test statistic for $\beta = 1$ was calculated and the better imputation method will have the lower t-test statistic.

$$T = \frac{b-1}{\widehat{se}(b)} \qquad \text{(Criteria 2)}$$

where *b* denotes the estimated value of β and $\widehat{se}(b)$ is the estimated standard error of *b*.

Criteria 3: Distributional Accuracy

The imputation method should preserve the distribution of the true values. One measure of preservation of the distribution of the true values is the distance between the empirical distribution

[6] Only cases with non-negative incomes were included in the regression models for this criterion. Negative incomes occurred for business income, rental income and total income.

functions for both the imputed and true values. The empirical distribution functions for the cases imputed are

$$F_{Y^*n}(x) = \frac{1}{n}\sum_{i=1}^{n} I\left(Y_i^* \le x\right)$$

$$F_{\hat{Y}n}(x) = \frac{1}{n}\sum_{i=1}^{n} I\left(\hat{Y}_i \le x\right)$$

The distance between these functions can be measured using the Kolmogorov-Smirnov distance

$$d_{KS}\left(F_{Y^*n}, F_{\hat{Y}n}\right) = \max_j\left(\left|F_{Y^*n}\left(x_j\right) - F_{\hat{Y}n}\left(x_j\right)\right|\right) \quad \text{(Criteria 3)}$$

where the x_j values are the jointly ordered true and imputed values of Y. A good imputation method will have a small distance value.

Criteria 4 and 5: Estimation Accuracy

The imputation method should preserve the lower order moments of the distributions of the true values. The absolute difference between the moments of the empirical distribution for the true and imputed values is given by

$$m_k = \left|\frac{1}{n}\sum_{i=1}^{n}\left(Y_i^{*k} - \hat{Y}_i^k\right)\right| = \left|m\left(Y^{*k}\right) - m\left(\hat{Y}^k\right)\right| \qquad \text{for } k=1, 2, \ldots$$

In the evaluation study, criteria 4 and 5 measure the absolute difference in the means and variances:[7]

Absolute difference in mean (1st order moment):

$$m_1 = \left|m\left(Y^{*1}\right) - m\left(\hat{Y}^1\right)\right| \qquad \text{(Criteria 4)}$$

Absolute difference in variance (2nd order moment):

$$m_2 = \left|m\left(Y^{*2}\right) - m\left(\hat{Y}^2\right)\right| \qquad \text{(Criteria 5)}$$

A good imputation method will have a low absolute difference in moments.

Criteria 6: Predictive accuracy of change

Another way to assess the preservation of the change between waves is the comparison of the cross-wave correlations for the imputed and true values. The formulae for the correlations between wave 1 and wave 2, for example, for the imputed and true values are

$$r_{\hat{Y}_1\hat{Y}_2} = \frac{\sum_{i=1}^{n}\left(\hat{Y}_{i1} - \hat{\bar{Y}}_1\right)\left(\hat{Y}_{i2} - \hat{\bar{Y}}_2\right)}{\sqrt{\sum_{i=1}^{n}\left(\hat{Y}_{i1} - \hat{\bar{Y}}_1\right)^2 \sum_{i=1}^{n}\left(\hat{Y}_{i2} - \hat{\bar{Y}}_2\right)^2}} \quad \text{(Criteria 6)}$$

$$r_{Y_1^*Y_2^*} = \frac{\sum_{i=1}^{n}\left(Y_{i1}^* - \bar{Y}_1^*\right)\left(Y_{i2}^* - \bar{Y}_2^*\right)}{\sqrt{\sum_{i=1}^{n}\left(Y_{i1}^* - \bar{Y}_1^*\right)^2 \sum_{i=1}^{n}\left(Y_{i2}^* - \bar{Y}_2^*\right)^2}}$$

[7] Skewness and kurtosis were also measured, but not reported in this paper (see Starick and Watson (2006) for details).

where Y_1 denotes the Y-values in wave 1 and Y_2 denotes the Y-values in wave 2. A good imputation method will have cross-wave correlations from the imputed data close to the true cross-wave correlations (that is, for waves 1 and 2, $r_{\hat{Y}_1\hat{Y}_2}$ should be close to $r_{Y_1^*Y_2^*}$).

Criteria 7: Distributional accuracy of change

In a longitudinal survey context, it is also important to assess the consistency of the income distribution between waves. This can be measured by comparing the income mobility in the dataset that includes the imputed values with that in the dataset that includes only true values (this measure includes all values in the data rather than those considered for imputation). The change in income decile group membership from one wave to another for each dataset is computed and the distribution for the dataset with imputation is tested for similarity to the distribution from the dataset of true values. A Chi-Square test is used where the observed cell frequencies are the cell frequencies from the dataset with imputed values and the expected cell frequencies are the true cell frequencies. The null hypothesis is

$H_0 : \hat{n}_{ij} = n_{ij}^*$ for all row i and column j.

The test statistic is

$$\chi^2 = \sum_{j=1}^{c}\sum_{i=1}^{r}\frac{\left(\hat{n}_{ij} - n_{ij}^*\right)^2}{n_{ij}^*} \qquad \text{(Criteria 7)}$$

The better imputation method will have the lower χ^2 statistic.

Criteria 8 and 9: Predictive accuracy between variables

It is important to assess the preservation of the relationships between income variables. One measure used is the Euclidean distance between the imputed and true data values in multi-dimensional space.

The Euclidean distance is calculated for each case for a set of income variables. Let k denote the number of income variables being imputed simultaneously. Let y_{ij}^* denote the true data value for observation i and the jth variable, where j = 1 to k and let $\hat{y}_{ij}$ denote the imputed data for the same observation i and variable j. The Euclidean distance between the true and imputed data values for observation i is

$$d_i\left(y_{ij}^*, \hat{y}_{ij}\right) = \sqrt{\sum_{j=1}^{k}\left(y_{ij}^* - \hat{y}_{ij}\right)^2}\ .$$

The mean of the Euclidean distances of the n imputed cases is then calculated.[8]

$$mean\left(d_i\right) = \frac{1}{n}\sum_{i=1}^{n} d_i \qquad \text{(Criteria 8)}$$

A good imputation method will have the lowest mean.

An alternative approach to assessing how well an imputation method preserves the relationships between income variables is to compare the true correlations between the income variables being imputed and with those from the imputed datasets. The formulae for the correlations between, for example, variable 1 and variable 2 for both the imputed and true values are

[8] The median and standard deviation of the Euclidian distances were also calculated but not reported in this paper (see Starick and Watson, 2006).

$$r_{\hat{Y}_1\hat{Y}_2} = \frac{\sum_{i=1}^{n}\left(\hat{Y}_{i1} - \hat{\bar{Y}}_1\right)\left(\hat{Y}_{i2} - \hat{\bar{Y}}_2\right)}{\sqrt{\sum_{i=1}^{n}\left(\hat{Y}_{i1} - \hat{\bar{Y}}_1\right)^2 \sum_{i=1}^{n}\left(\hat{Y}_{i2} - \hat{\bar{Y}}_2\right)^2}} \quad \text{(Criteria 9)}$$

$$r_{Y_1^*Y_2^*} = \frac{\sum_{i=1}^{n}\left(Y_{i1}^* - \bar{Y}_1^*\right)\left(Y_{i2}^* - \bar{Y}_2^*\right)}{\sqrt{\sum_{i=1}^{n}\left(Y_{i1}^* - \bar{Y}_1^*\right)^2 \sum_{i=1}^{n}\left(Y_{i2}^* - \bar{Y}_2^*\right)^2}}$$

where Y_1 denotes the Y-values of variable 1 and Y_2 denotes the Y-values of variable 2. A good imputation method will have between-variable correlations close to the true between-variable correlations.

3. Imputation Methods Tested

The following section describes the imputation methods considered in the evaluation study. The imputation methods adopted by large national household-based longitudinal surveys similar to HILDA Survey provided some guidance on which imputation methods were included in the evaluation study:

- In the British Household Panel Study, two main methods of imputation are used. For continuous variables, a nearest neighbour regression method is used, whilst for categorical variables, a hot deck method is used (Buck, 1997).
- The German Socio-Economic Panel predominantly uses an imputation method developed by Little and Su (1989). It is a simple stochastic longitudinal imputation method for repeated measures data. Three other methods are also used in certain circumstances (Frick and Grabka, 2003): mean substitution is used where the number of missing cases is small; median share is used where a link can be established between two variables; and regression based substitution is used for more complex income constructs.
- The Canadian Survey of Labour and Income Dynamics uses a last value carried forward method as the primary method. In the absence of data from the previous year, imputation using a nearest neighbour technique is employed.[9]
- The US Panel Study of Income Dynamics, in general uses hot deck procedures to impute missing data (Hofferth et al. 1998).
- The US Survey of Income and Program Participation currently uses two methods of imputation. Item non-response is imputed using a sequential hot deck imputation procedure (Pennell, 1993) and wave non-response is imputed using a longitudinal imputation procedure referred to as the random carryover method (Williams and Bailey, 1996).

The methods tested in this evaluation study (and are described in detail below) include:

- A nearest neighbour regression method;
- Four methods based on the Little and Su method, being with an without imputation classes and two variants to impute more than one income variable at a time; and
- Three carry-over methods, being the last value carried forward, the random carryover method, and the population carryover method.

[9] Information on the imputation method used in the Canadian Survey of Labour and Income Dynamics was obtained from the documentation about the SLID methodology from www.statcan.ca.

The two most obvious exclusions to this list of methods tested are the hotdeck method and multiple imputation. The hotdeck method was not included due to the time constraints of the project. It is expected that this method will perform similarly to the nearest neighbour method. Multiple imputation was not considered as a single imputation solution was required at this time for the HILDA public release file.[10]

Note that the nearest neighbour regression method, the carryover method and the first two Litte and Su methods are univariate imputation methods. That is, the imputation is applied one variable at a time. For the Little and Su methods, multiple wave missingness for a variable may be imputed simultaneously. The last two Little and Su methods considered are multivariate imputation methods, where two or more income variables are imputed simultaneously.

3.1 Nearest Neighbour Regression Method

The nearest neighbour regression method (also known as predictive mean matching (Little, 1988)) seeks to identify the 'closest' donor to each record that needs to be imputed via the predicted values from a regression model for the variable to be imputed. The donor's actual value for the variable being imputed replaces the missing value of the recipient. This method is used in the British Household Panel Study and was the first method implemented in the HILDA Survey (Watson, 2004).

For each wave and for each variable imputed, regression models using information from the same wave as well as information from other waves (if available) were constructed. Over 100 variables were considered for inclusion in the income models covering demographic characteristics, employment characteristics, the respondent's partner's characteristics (if the respondent had a partner), the respondent's partner's income, and income reported by the respondent in other waves (where this is known). Following the procedure adopted in the HILDA Survey, a statistical package called MARS (Multivariate Adaptive Regression Splines) was used to construct the income models.[11] MARS is an automatic regression package which finds the best model for the specified variable from the host of variables provided. Main effects and two-way interactions were considered. MARS provided a practical solution to the resource intensive problem of constructing good predictive models.

The predicted values from a regression model for the variable being imputed were used to identify the nearest case whose reported value could be inserted into the case with the missing value:

$$\hat{Y}_i = Y_l$$

where $\left|\hat{\mu}_i - \hat{\mu}_l\right| \leq \left|\hat{\mu}_i - \hat{\mu}_p\right|$ for all respondents *p* (potential donors), $\hat{\mu}_i$ is the predicted mean of *Y* for individual *r* that needs to be imputed, and Y_l is the observed value of *Y* for respondent *l* (the donor).

This method provides a fall back solution when the other imputation methods considered in this paper cannot be used.[12]

3.2 Basic Little and Su Method

The imputation method proposed by Little and Su (1989), and used by the German Socio-Economic Panel, will be referred to as the 'basic Little and Su method' to distinguish it from the modified version implemented in Releases 3.0 and 4.0 of the HILDA data. The modified version will be referred to as the 'extended Little and Su method'.

---

[10] When standard statistical software can routinely include multiple imputation in many procedures and commands and our users are versed in these methods, then a multiple imputation solution will need to be considered.

[11] See the Salford Systems website for an overview of the MARS package: www.salford-systems.com.

[12] For example, the Little and Su method (described later in this paper) cannot be used if a respondent has not reported any income data for the variable being imputed and the nearest neighbour method is used to provide the first imputed value.

The basic Little and Su method incorporates trend and individual level information into the imputed amounts by using a multiplicative model based on row (person) and column (wave) effects. The model is of the form

imputation = (row effect) x (column effect) x (residual).

The column (wave) effects are calculated by

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

where $\bar{Y} = \frac{1}{m}\sum_j \bar{Y}_j$

for each wave j = 1, …, m. $\bar{Y}_j$ is the sample mean of variable Y for wave j, based on complete cases and $\bar{Y}$ is the global mean of variable Y based on complete cases.

The row (person) effects are calculated by

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

for both complete and incomplete cases. Here, the summation is over recorded waves for case i; m_i is the number of recorded waves; Y_{ij} is the variable of interest for case i, wave j; and c_j is the simple wave correction from the column effect.

The cases were ordered by $\bar{Y}^{(i)}$, and incomplete case i is matched to the closest complete case, say l.

The missing value Y_{ij} was imputed by

$$\hat{Y}_{ij} = \left[\bar{Y}^{(i)}\right]\left[c_j\right]\left[\frac{Y_{lj}}{\bar{Y}^{(l)}c_j}\right]$$

$$= Y_{lj}\frac{\bar{Y}^{(i)}}{\bar{Y}^{(l)}}$$

where the three terms in square parentheses represent the row, column, and residual effects. The first two terms estimate the predicted mean, and the last term is the stochastic component of the imputation from the matched case.

It is important to note that due to the multiplicative nature of the Little and Su method, a zero individual effect will result in a zero imputed value. However, it is quite valid to have an individual reporting zero income in previous waves and then report that they have income but either don't know its value or refuse to provide it. The individual's effect would be zero and any imputed amount via the Little and Su method would also be zero, which we know is not true. Therefore, recipients with zero individual effects are imputed using the nearest neighbour regression method.

In addition, the donors must have non-zero individual effects to avoid divisions by zero.

3.3 Extended Little and Su Method

Ideally, the donor and the recipient should have similar characteristics that are associated with the variable being imputed. The basic Little and Su method, therefore, was extended to take into account the characteristics of the donors and recipients. Donors and recipients were matched within imputation classes defined by the following age ranges: 15-19, 20-24, 25-34, 35-44, 45-54, 55-64,

65+.[13] The extended Little and Su method was implemented in Release 3.0 and 4.0 of the HILDA data for many of the income components and the remainder were implemented using the basic Little and Su method (i.e., without the imputation classes).

The column (wave) effects are calculated as

$$c_{hj} = \frac{\bar{Y}_{hj}}{\bar{Y}_h}$$

where $\bar{Y}_h = \frac{1}{m}\sum_j \bar{Y}_{hj}$

for each wave $j = 1, \ldots, m$, and for each age group $h = 1, \ldots, c$. $\bar{Y}_{hj}$ is the sample mean of variable Y for wave j, and age group h based on complete cases and $\bar{Y}_h$ is the global mean of variable Y for age group h based on complete cases.

The row (person) effects are calculated as

$$\bar{Y}_h^{(i)} = \frac{1}{m_i}\sum_j \frac{Y_{hij}}{c_{hj}}$$

for both complete and incomplete cases. Here, the summation is over recorded waves for case i; m_i is the number of recorded waves; Y_{hij} is the variable of interest for case i, wave j, age group h; and c_{hj} is the simple wave correction (i.e., the column effect).

The cases were ordered by $\bar{Y}_h^{(i)}$, and incomplete case i is matched to the closest complete case, say l within age group h.

The missing value Y_{hij} was imputed by

$$\hat{Y}_{hij} = \left[\bar{Y}_h^{(i)}\right]\left[c_{hj}\right]\left[\frac{Y_{hlj}}{\bar{Y}_h^{(l)} c_{hj}}\right]$$

$$= Y_{hlj}\frac{\bar{Y}_h^{(i)}}{\bar{Y}_h^{(l)}}$$

where the three terms in square parentheses represent the row, column, and residual effects. The first two terms estimate the predicted mean, and the last term is the stochastic component of the imputation from the matched case.

3.4 Last Value Carried Forward

A last value carried forward method, similar to that used in the Canadian Survey of Labour and Income Dynamics, was assessed. Where reported information from the previous wave is available, this is used to fill in the missing variable. That is, the missing value Y_{ij} for case i, wave j is imputed by $\hat{Y}_{i,j} = Y_{i,j-1}$. Where reported information from the previous wave is absent, the nearest neighbour regression method is used.

[13] Age groups were used to create the imputation classes because it is a simple characteristic known for almost all donors and recipients. For a few cases, age was missing and was therefore imputed from a family with a similar relationship structure to the missing case.

3.5 Random Carryover Method

The random carryover method (Williams and Bailey, 1996) imputes single missing wave data that is bounded on both sides by an interviewed wave. This means that this method does not impute data where there are two or more consecutive missing waves, nor does it impute the first or last wave.

Under this method, the value from either the preceding or subsequent wave is donated to the recipient. The choice between these two possibilities is made randomly: a value r is randomly assigned to each case for each missing item, where $r = 0$ or 1. If $r = 0$ then the imputed value comes from the preceding wave. If $r = 1$ then the imputed value comes from the subsequent wave.

In the evaluation study, we also carried forward or backwards information where only one wave of data was available. That is, the last value was carried forward where only wave 1 information was available or the next value was carried backwards where only wave 3 information was available. Where no information was available, the nearest neighbour method was used to impute the first wave and this was then carried forward to the other waves.

3.6 Population Carryover Method

A variation of the random carryover method was also implemented in the evaluation study and this is referred to as the 'population carryover method' (Williams and Bailey, 1996). Rather than choosing a donor by assigning a random value r, a donor is determined by reflecting the population changes in the reported income amounts between waves.

An indicator variable c is created which equals 1 when the reported change between waves 2 and 3 is smaller than the reported change between waves 1 and 2 for the complete cases; and 0 otherwise. The proportion p, of the interviewed sample where the change between waves 2 and 3 is smaller than the change between waves 1 and 2 is then determined. Whether the preceding wave or the subsequent wave donates the imputed amount is determined by reflecting the probabilities associated with the occurrences of change between waves found in the complete cases.

3.7 Multivariate Imputation Method Option 1

The multivariate imputation methods aim to impute more than one variable simultaneously in order to maintain the covariance structure between the variables. The first option considered assumes one income component is more important than another and uses the basic Little and Su method to find a suitable donor based on the most important component. The remaining missing items are imputed using the information from the same donor.

For example, suppose we have a case which has missing values for wages and salaries and for Government pensions. We decided that wages and salaries is more important than Government pensions. So, we imputed wages and salaries using the Little and Su method and, using the same donor we also impute Government pensions.

Because of the complexities of finding a suitable donor with the right amount of non-zero non-missingness for each recipient, this method was tested by imputing wages and salaries together with Government pensions and by imputing dividends, royalties, interest and rent together.

3.8 Multivariate Imputation Method Option 2

The second multivariate option considered is also based on the basic Little and Su method and involves calculating a combined row effect that is a function of the row effects for each income variable. The cases are then ordered by this combined row effect and the nearest neighbour is identified.

Let Y_k denote the income variables being imputed, $k = 1, \ldots, K$. For each Y_k, compute the column effects c_{kj} and the row effects $\overline{Y}_k^{(i)}$. A combined row effect is calculated as the Euclidean distance between the row effects of the incomplete case i and each potential donor p of the variables imputed together. For two variables, the combined row effect is

$$\bar{Y}^{(i)} = d\left(\bar{Y}_1^{(i)}, \bar{Y}_2^{(i)}\right) = \sqrt{\left(\bar{Y}_1^{(i)} - \bar{Y}_1^{(p)}\right)^2 + \left(\bar{Y}_2^{(i)} - \bar{Y}_2^{(p)}\right)^2}$$

And for three variables, the combined row effect is

$$\bar{Y}^{(i)} = d\left(\bar{Y}_1^{(i)}, \bar{Y}_2^{(i)}, \bar{Y}_3^{(i)}\right) = \sqrt{\left(\bar{Y}_1^{(i)} - \bar{Y}_1^{(p)}\right)^2 + \left(\bar{Y}_2^{(i)} - \bar{Y}_2^{(p)}\right)^2 + \left(\bar{Y}_3^{(i)} - \bar{Y}_3^{(p)}\right)^2}$$

The cases are ordered by this combined row effect, and the nearest suitable donor *l* is found.

The missing value Y_{kij} is imputed by

$$\hat{Y}_{kij} = Y_{klj} \frac{\bar{Y}_k^{(i)}}{\bar{Y}_k^{(l)}}$$

That is, the row effects that are specific to variable *k* for individual *i* and *l* are used (the combined row effect is just used to identify the nearest suitable donor). As for the multivariate option 1, the income components imputed simultaneously were wages and salaries together with Government pensions, and dividends, royalties, interest income together with rent income.

4. Comparison of Imputation Methods

The eight imputation methods implemented in this evaluation study are compared via the nine evaluation criteria. The performance of the imputation methods is first considered in a cross-sectional context and then in a longitudinal context. For a longitudinal survey, the longitudinal performance of the method is more important than the cross-sectional performance.

To help draw conclusions from the many criteria and methods considered, the imputation methods were ranked from 1 to 8 within each income item within the first five evaluation criteria – where 1 denotes the best performance and 8 denotes the worst performance. The ranks for each evaluation criteria were then summed for each imputation method.[14] This approach means that the first five evaluation criteria are treated as if they are of equal importance. The best a method can score is 5 (that is, they ranked first in each of the five criteria) and the worst ranked score a method can achieve is 40 (by being ranked eighth for all five criteria).

Table 2 presents the summed ranks across the first five evaluation criteria for each income item for each wave and the average summed ranks across the three waves. Bold table entries indicate which imputation method performed better (the lower the sum, the better the performance). Cross-sectionally, the nearest neighbour regression method often performs the worst of all the methods and, when it is not the worst, it is usually among the poorer performers. The Little and Su methods perform well for variables that had a large number of donors and were not highly variable (such as waves and salaries). The carryover methods performed well for variables that were either highly variable (such as business income, interest, dividends and royalties, and rent) or had a small number of cases to be imputed (such as private pensions and transfers). The carryover methods also worked well for Government pension income, but the extended Little and Su method provides a good alternative due to a strong correlation between this income component and age (the basic Little and Su method performed amongst the worst for this variable). For non-responding persons, the carryover methods or the Extended Little and Su method performed well.

[14] The mean and standard deviation of the evaluation measures across the 10 simulation datasets can be obtained from Starick and Watson (2006, Appendix 1).

Table 2: Sum of the Ranked Cross-Sectional Evaluation Measures for FY Income Variables, Criteria (1) to (5)

Variable	*W1*	*W2*	*W3*	*Ave*
RESPONDING PERSONS				
Wages and salaries				
NNRM	29	37	33	33.0
Basic Little & Su	22	19	19	20.0
Extended Little & Su	**9**	17	30	18.7
LVCF	29	21	25	25.0
Random Carryover	25	26	19	23.3
Population Carryover	25	**15**	19	19.7
Multivariate Option 1	16	25	16	19.0
Multivariate Option 2	19	20	**14**	**17.7**
Aust govt pensions				
NNRM	14	22	33	23.0
Basic Little & Su	32	33	18	27.7
Extended Little & Su	19	**10**	22	17.0
LVCF	**11**	25	19	18.3
Random Carryover	23	18	**13**	18.0
Population Carryover	23	14	**13**	**16.7**
Multivariate Option 1	28	34	30	30.7
Multivariate Option 2	23	22	27	24.0
Business income				
NNRM	15	22	27	21.3
Basic Little & Su	24	23	24	23.7
Extended Little & Su	29	35	33	32.3
LVCF	19	15	16	16.7
Random Carryover	**12**	15	**6**	11.0
Population Carryover	**12**	**9**	**6**	**9.0**
Multivariate Option 1	24	23	24	23.7
Multivariate Option 2	24	23	24	23.7
Interest income				
NNRM	33	35	34	34.0
Basic Little & Su	27	18	24	23.0
Extended Little & Su	21	31	15	22.3
LVCF	31	19	23	24.3
Random Carryover	**14**	**16**	**14**	**14.7**
Population Carryover	**14**	19	**14**	15.7
Multivariate Option 1	19	20	25	21.3
Multivariate Option 2	15	21	26	20.7
Dividends and royalties				
NNRM	28	32	33	31.0
Basic Little & Su	**12**	26	**14**	17.3
Extended Little & Su	25	32	30	29.0
LVCF	28	18	25	23.7
Random Carryover	22	**6**	15	**14.3**
Population Carryover	22	9	15	15.3
Multivariate Option 1	18	33	23	24.7
Multivariate Option 2	19	24	20	21.0

Variable	*W1*	*W2*	*W3*	*Ave*
RESPONDING PERSONS				
Rent income				
NNRM	23	18	21	20.7
Basic Little & Su	21	26	34	27.0
Extended Little & Su	26	**11**	24	20.3
LVCF	23	21	12	18.7
Random Carryover	**19**	21	**11**	**17.0**
Population Carryover	**19**	23	**11**	17.7
Multivariate Option 1	24	28	37	29.7
Multivariate Option 2	**19**	32	25	25.3
Private pensions				
NNRM	30	30	25	28.3
Basic Little & Su	**14**	25	25	21.3
Extended Little & Su	21	26	29	25.3
LVCF	29	14	13	18.7
Random Carryover	18	11	**9**	12.7
Population Carryover	18	**9**	**9**	**12.0**
Multivariate Option 1	**14**	25	25	21.3
Multivariate Option 2	**14**	25	25	21.3
Private transfers				
NNRM	24	32	33	29.7
Basic Little & Su	26	16	18	20.0
Extended Little & Su	22	26	23	23.7
LVCF	25	18	26	23.0
Random Carryover	**5**	26	**12**	14.3
Population Carryover	**5**	**15**	**12**	**10.7**
Multivariate Option 1	26	16	18	20.0
Multivariate Option 2	26	16	18	20.0
Total FY income				
NNRM	**19**	34	30	27.7
Basic Little & Su	28	19	31	26.0
Extended Little & Su	**19**	25	32	25.3
LVCF	21	12	13	15.3
Random Carryover	22	20	**6**	16.0
Population Carryover	22	**12**	**6**	**13.3**
Multivariate Option 1	24	29	34	29.0
Multivariate Option 2	**19**	29	23	23.7
NON-RESPONDING PERSONS				
Total FY income				
NNRM	33	29	36	32.7
Basic Little & Su	16	21	17	18.0
Extended Little & Su	15	29	22	22.0
LVCF	31	23	29	27.7
Random Carryover	25	14	18	19.0
Population Carryover	25	**12**	18	18.3
Multivariate Option 1	15	22	**16**	**17.7**
Multivariate Option 2	**14**	30	19	21.0

Table 3 shows the results for the first five criteria in the longitudinal context, again using the ranking approach described above.[15] Longitudinally, the nearest neighbour regression method performed poorly for all income variables with the exception of Government pensions. However, as this method performed so poorly in the cross-sectional sense, it is not seriously considered when an alternative method (such as the extended Little and Su method) can provide a good cross-sectional

[15] For responding persons, the measures were computed on cases where they were respondents in at least one of the two waves; and for non-responding persons, the measures were computed on cases where they were non-respondents in at least one of the two waves. This means that a person could contribute to the results for both responding persons and non-responding persons.

and a good longitudinal solution. The Little and Su methods were much preferable to the carryover methods in predicting the change in levels between waves for wages and salaries, interest income, dividends and royalties, private transfers and total income for non-responding persons. The carryover methods performed the best for business income, and private pensions. The carryover methods, by their nature, impute zero change between waves so will underestimate the true amount of change. However, for certain types of income which can be quite large and variable (such as business income and private pensions), it is better to impute missing values from oneself (and underestimate the change) rather than from a donor (and overestimate the change).

Table 3: Sum of the Ranked Longitudinal Evaluation Measures for FY Income Variables, Criteria (1) to (5)

Variable	*W2-W1*	*W2-W2*	*W3-W1*	*Ave*
RESPONDING PERSONS				
Wages and salaries				
NNRM	31	30	22	27.7
Basic Little & Su	19	**18**	22	19.7
Extended Little & Su	**14**	25	**15**	**18.0**
LVCF	24	19	28	23.7
Random Carryover	22	31	24	25.7
Population Carryover	21	**18**	25	21.3
Multivariate Option 1	25	**18**	17	20.0
Multivariate Option 2	24	21	23	22.7
Aust govt pensions				
NNRM	**12**	24	22	19.3
Basic Little & Su	27	21	24	24.0
Extended Little & Su	20	22	28	23.3
LVCF	19	**18**	**16**	**17.7**
Random Carryover	28	28	17	24.3
Population Carryover	29	24	17	23.3
Multivariate Option 1	23	24	28	25.0
Multivariate Option 2	22	19	23	21.3
Business income				
NNRM	24	26	30	26.7
Basic Little & Su	22	23	**16**	20.3
Extended Little & Su	26	25	31	27.3
LVCF	22	18	19	19.7
Random Carryover	**13**	**13**	17	**14.3**
Population Carryover	14	14	**16**	14.7
Multivariate Option 1	22	23	**16**	20.3
Multivariate Option 2	22	23	**16**	20.3
Interest income				
NNRM	25	32	33	30.0
Basic Little & Su	21	23	22	22.0
Extended Little & Su	21	20	**16**	19.0
LVCF	31	24	23	26.0
Random Carryover	23	**18**	19	20.0
Population Carryover	29	19	20	22.7
Multivariate Option 1	16	25	23	21.3
Multivariate Option 2	**14**	19	20	**17.7**
Dividends and royalties				
NNRM	31	30	32	31.0
Basic Little & Su	20	21	**16**	**19.0**
Extended Little & Su	**15**	25	19	19.7
LVCF	26	20	26	24.0
Random Carryover	23	21	18	20.7
Population Carryover	30	**18**	18	22.0
Multivariate Option 1	20	25	23	22.7
Multivariate Option 2	**15**	20	23	19.3

Variable	*W2-W1*	*W2-W2*	*W3-W1*	*Ave*
RESPONDING PERSONS				
Rent income				
NNRM	20	26	24	23.3
Basic Little & Su	**15**	**14**	26	**18.3**
Extended Little & Su	16	19	25	20.0
LVCF	29	18	17	21.3
Random Carryover	22	23	15	20.0
Population Carryover	24	20	**14**	19.3
Multivariate Option 1	28	32	32	30.7
Multivariate Option 2	26	28	23	25.7
Private pensions				
NNRM	23	24	24	23.7
Basic Little & Su	27	26	18	23.7
Extended Little & Su	15	28	37	26.7
LVCF	17	**9**	17	14.3
Random Carryover	**14**	14	**14**	14.0
Population Carryover	15	12	**14**	**13.7**
Multivariate Option 1	27	26	18	23.7
Multivariate Option 2	27	26	18	23.7
Private transfers				
NNRM	20	24	38	27.3
Basic Little & Su	21	**16**	15	17.3
Extended Little & Su	28	30	**10**	22.7
LVCF	**13**	18	29	20.0
Random Carryover	27	29	19	25.0
Population Carryover	14	**16**	20	**16.7**
Multivariate Option 1	21	**16**	15	17.3
Multivariate Option 2	21	**16**	15	17.3
Total FY income				
NNRM	32	18	35	28.3
Basic Little & Su	**16**	30	**14**	20.0
Extended Little & Su	18	31	23	24.0
LVCF	20	22	25	22.3
Random Carryover	19	**12**	19	**16.7**
Population Carryover	20	15	19	18.0
Multivariate Option 1	29	25	22	25.3
Multivariate Option 2	26	27	18	23.7
NON-RESPONDING PERSONS				
Total FY income				
NNRM	26	30	36	30.7
Basic Little & Su	17	**12**	**16**	**15.0**
Extended Little & Su	**15**	18	19	17.3
LVCF	21	20	21	20.7
Random Carryover	30	21	21	24.0
Population Carryover	28	28	21	25.7
Multivariate Option 1	**15**	24	20	19.7
Multivariate Option 2	28	27	21	25.3

The performance of some methods was inconsistent. The Little and Su methods tend to have a more variable performance than the carryover methods, both cross-sectionally and longitudinally. For example, while the extended Little and Su method performed the best for wages and salaries on average, it performed poorly for the wave 3 cross-section and the change between waves 2 and 3.

Within the Little and Su methods, the extended Little and Su method tended to work best for wages and salaries, interest and rent. For wages and salaries, there is a large pool of donors to choose from, the income component correlate well with age, and it is not highly variable. It is not obvious why interest and rent perform better with age imputation classes. Nevertheless, the age groupings used in the imputation classes improved the matching of donors and recipients. The basic Little and Su method tends to work best for dividends and royalties and total income for non-responding persons.[16] These items do not correlate particularly well with age. Based on criteria 1 to 5, the multivariate option 1 consistently performed worse than the second multivariate option, with the exception of wages and salaries in the longitudinal context (in Table 3). Compared to the basic Little and Su method (on which these multivariate methods are based), the multivariate option 2 (combined row effect) performed better cross-sectionally, except for dividends and royalties. Longitudinally, the multivariate option 2 performs similarly to the basic Little and Su option. However, as the extended Little and Su method performs better than the basic Little and Su method for a number of these variables, a version of the multivariate option 2 should be tested based on the extended Little and Su method.

Within the carryover methods, the population carryover method performed the best for business income, private pensions and private transfers. These items have either a small number of cases to be imputed (as is the case for private pensions, private transfers) or have a small pool of donors to choose from and are highly variable (as for business income).

Government pensions could be equally well imputed using either the population carryover method or the extended Little and Su method. There are only a small number of cases to be imputed, the income is not highly variable and it is correlated well with age.

Next, we look at the cross-wave correlations produced by the imputed data and compare these to the true data (Table 4). Bold table entries indicate which correlation coefficients derived from the imputed data are closest to the true correlation coefficients. Based on cross-wave correlations, both the basic and extended Little and Su methods perform better most of the time for change estimates. The remainder of the time the carryover methods perform better, especially for movement estimates from wave 1 to wave 3. However, for most of the time, the carryover methods produce cross-wave correlations that are unrealistically high. By contrast, the cross-wave correlations for the nearest neighbour method are unrealistically low.

The next evaluation criterion addresses the distributional consistency between waves by considering the change in income decile group membership from one wave to another. Based on this evaluation measure, the results (in Table 5) clearly show that the nearest neighbour regression method does not preserve the consistency of the income distribution between waves (the calculated χ^2 exceeds the critical value).[17] This is also true for the carryover methods, except for the change between wave 1 and wave 3. However, the Little and Su methods (including the multivariate methods) do preserve the consistency of the income distribution between waves.

The final two evaluation criteria measure how well an imputation method preserves the relationships between income variables. Table 6 presents the mean Euclidean distance between the

[16] Although the results in Tables 2 and 3 indicate that the multivariate option 1 performed the best, the total financial year income for non-responding persons was not actually imputed simultaneously with any other income item. The basic Little and Su method was applied to the non-responding persons where the donors have come from the responding persons, some of which have been imputed using the multivariate option 1. If this method is not going to be adopted for the components, then the non-responding persons cannot be imputed via this method.

[17] The critical value of χ^2 for $\alpha = 0.05$ and 81 degrees of freedom is 101.88.

true and imputed data values. Lower measures indicate the better the imputation method (in other words, the closer the distance between the true and imputed values in 2-dimensional or 3-dimensional space). The results show that the carryover methods performed better. As the imputed value is being carried over from the same case, the relationships between income variables are inherently being preserved. The multivariate option 2 generally preserves the relationships better than option 1 and always better than the basic Little and Su method. Sometimes the extended Little and Su method outperforms option 2.

Lastly, we look at the correlations between income variables and assess how well the relationships are preserved using this evaluation measure (Table 7). Due to weak relationships between some variables, it is difficult to comments on these results. The multivariate imputation options do not preserve the relationships any better than the univariate methods as measured by this criteria.

Table 4: Cross-Wave Correlations, Total FY Income (Evaluation Measure 6)

Sample	*true*	*NNRM*	*Basic L&S*	*Extended L&S*	*LVCF*	*RCOM*	*PCOM*	*Multi Option 1*	*Multi Option 2*
Wave 1 to Wave 2									
Responding persons	*0.72*	0.54	**0.71**	0.67	0.66	0.79	0.79	0.69	0.69
Non-responding persons	*0.73*	0.51	0.82	0.82	**0.73**	0.95	0.94	0.85	0.81
Wave 2 to Wave 3									
Responding persons	*0.72*	0.51	0.69	**0.71**	0.70	0.85	0.85	0.68	**0.71**
Non-responding persons	*0.66*	0.26	**0.66**	0.72	0.59	0.96	0.97	0.71	0.74
Wave 1 to Wave 3									
Responding persons	*0.74*	0.51	0.70	0.65	0.57	**0.72**	**0.72**	0.71	**0.72**
Non-responding persons	*0.77*	0.44	0.80	**0.78**	0.58	0.80	0.80	0.81	0.81

Table 5: Chi-Square Test Statistics on Total FY Income Deciles (Evaluation Measure 7)

Imputation Method	*W1 to W2*	*W2 to W3*	*W1 to W3*
NNRM	354.86	431.80	156.88
Basic Little & Su	57.67	**61.52**	65.86
Extended Little & Su	52.36	62.74	**54.31**
LVCF	107.90	103.70	98.81
Random Carryover	162.86	136.91	89.30
Population Carryover	181.61	125.25	89.30
Multivariate Option 1	58.19	62.72	62.74
Multivariate Option 2	**52.19**	63.42	61.05

Table 6: Preservation of Relationships Between FY Income Variables, Mean Euclidean Distances for Responding Persons (Evaluation Measure 8)

Imputation Method	*W1*	*W2*	*W3*	*W2-W1*	*W3-W2*	*W3-W1*
Wages and salaries, Aust govt pensions						
NNRM	11,186	12,521	9,801	12,134	11,390	10,831
Basic Little & Su	10,192	8,447	9,354	9,198	8,760	10,006
Extended Little & Su	9,519	8,099	9,558	8,670	8,695	9,707
LVCF	11,186	8,786	8,424	9,978	8,532	10,119
Random Carryover	**8,067**	**7,536**	**8,189**	7,341	7,587	**8,051**
Population Carryover	**8,067**	**7,356**	**8,189**	**7,240**	**7,474**	**8,051**
Multivariate Option 1	10,084	8,494	9,198	9,152	8,735	9,809
Multivariate Option 2	10,123	8,374	9,148	9,098	8,670	9,837
Interest income, Dividends and royalties, Rent income						
NNRM	4,745	3,491	4,060	4,177	3,846	4,636
Basic Little & Su	4,011	2,819	3,943	3,403	3,190	4,017
Extended Little & Su	4,198	3,048	3,842	3,613	3,271	4,078
LVCF	4,745	2,755	3,137	3,733	2,922	4,216
Random Carryover	**3,447**	**2,491**	**3,034**	**2,830**	**2,585**	**3,111**
Population Carryover	**3,447**	2,528	**3,034**	2,855	2,610	**3,111**
Multivariate Option 1	4,306	3,424	4,018	3,916	3,604	4,178
Multivariate Option 2	3,837	3,186	3,778	3,491	3,345	3,819

Table 7: Preservation of Correlations Between FY Income Variables, Responding Persons (Evaluation Measure 9)

Imputation Method	*W1*	*W2*	*W3*
Wages and salaries, Aust govt pensions			
true	-0.35	-0.39	-0.41
NNRM	**-0.36**	-0.35	-0.34
Basic Little & Su	-0.31	-0.35	**-0.35**
Extended Little & Su	-0.31	-0.35	-0.33
LVCF	**-0.36**	-0.34	-0.30
Random Carryover	-0.33	-0.35	-0.30
Population Carryover	-0.33	**-0.36**	-0.30
Multivariate Option 1	-0.32	-0.34	**-0.35**
Multivariate Option 2	-0.30	-0.35	**-0.35**
Interest income, Dividends and royalties			
true	0.20	0.14	0.17
NNRM	**0.20**	0.09	0.09
Basic Little & Su	**0.20**	**0.13**	0.14
Extended Little & Su	0.15	0.16	0.12
LVCF	**0.20**	0.11	**0.16**
Random Carryover	0.21	0.17	0.19
Population Carryover	0.21	0.18	0.19
Multivariate Option 1	0.22	**0.13**	0.14
Multivariate Option 2	0.23	0.15	0.10

Imputation Method	*W1*	*W2*	*W3*
Interest income, Rent income			
true	0.15	0.04	0.03
NNRM	0.04	0.00	-0.03
Basic Little & Su	0.04	0.05	-0.01
Extended Little & Su	0.05	0.06	0.00
LVCF	0.04	**0.03**	-0.02
Random Carryover	0.07	**0.03**	-0.01
Population Carryover	0.07	0.05	-0.01
Multivariate Option 1	0.05	-0.06	0.00
Multivariate Option 2	**0.08**	0.07	0.01
Dividends and royalties, Rent income			
true	0.04	0.04	0.01
NNRM	0.07	-0.02	-0.02
Basic Little & Su	0.06	0.01	-0.02
Extended Little & Su	0.02	0.02	-0.01
LVCF	0.07	0.01	0.00
Random Carryover	0.07	**0.03**	**0.01**
Population Carryover	0.07	**0.03**	**0.01**
Multivariate Option 1	**0.04**	0.01	-0.01
Multivariate Option 2	0.10	0.05	**0.01**

5. Conclusions

An assessment of the performance of alternative imputation methods was conducted using data from the first three waves of the HILDA Survey. A set of evaluation criteria, based on the statistical properties of a good imputation method, were used to compare these imputation methods.

The results of this evaluation study did not demonstrate that one imputation method performed consistently better against each of the evaluation measures for each income item and in each wave.

The evidence shows that different imputation methods performed better for different income items. Using a variety of imputation methods best suited to each variable should produce superior results to the use of one imputation method for all variables. For items that are not highly variable, have a large pool of donors and are well correlated with age (such as wages and salaries), the extended Little and Su method is recommended. For items that are highly variable and have a limited pool of donors (such as business income), a population carryover method works well. For items that have only a few missing cases (such as private transfers and private benefits), a population carryover method provides a suitable imputation method that is very simple to implement. The results for Government pensions, interest income, dividends and royalties, and rent are more mixed. Due to the problem with the carryover method overstating correlations between waves (discussed in more detail below), our preference is to use a Little and Su method. Given the high correlation between age and the value of Government pensions, the extended Little and Su method should be used. The results from the evaluation study show that the extended Little and Su method is slightly better than the basic Little and Su method for interest and rent, but that the basic Little and Su method is better for dividends and royalties.

It should be noted that a significant disadvantage of the carryover methods is that they are more likely to understate change and therefore overstate the correlation between waves (because the value from one of the surrounding waves is used for the missing wave). However, this method may be preferable to gross overstatement of change for some variables resulting from other methods (Heeringa and Lepkowski, 1986). For example, the carryover methods performed the best for

business income even when considering criteria 1 to 5 on the change in business income between waves.

The multivariate imputation methods did not yield superior results, though there is some evidence to suggest that if they were based on the extended Little and Su method they would perform better. However, it is likely that the improvements to the quality of the imputation would be small and would not justify the additional complexity required.

This project has highlighted a number of areas for future work to improve our understanding of income imputation in a longitudinal survey such as the HILDA Survey:

- Modify the construction of the simulated datasets to ensure they mirror key characteristics in the main datasets (such as household size).
- Incorporate other evaluation measures, such as:
    - t-test for whether mean of the imputed cases is the same as the true values;
    - Bias and relative bias, standard deviation of the bias to assess its size and direction (we have calculated the absolute value of the bias in criteria 1).
- Evaluate other imputation methods using this evaluation framework, such as:
    - the hotdeck imputation of level;
    - the hotdeck imputation of change (as used by the New Zealand Survey of Family, Income and Employment);
    - improvements to the extended Little and Su method in the formation of imputation classes;
    - a variation to the nearest neighbour method to impute change rather than level; and
    - other multivariate imputation methods, for example the hierarchical imputation method that was used in the Euredit Project (Pannekoek, 2002).
- Summarise the results differently. For example, the Euredit project undertook a second stage of analysis of the evaluation measures using ANOVA modelling (Chambers and Zhao, 2003).
- Vary the response mechanism (an example is given by Champney and Bell (1982)). The imputation methods could be re-evaluated under a response mechanism that is not missing at random to determine how much the response mechanism matters.

References

Buck, N. (1997), 'Imputation for Missing Income Data in a Panel Study', Paper presented at the IASS/IAOS Satellite Meeting on Longitudinal Studies, Jerusalem, 27-31 August, 1997 (Draft Paper, ESRC Research Centre on Micro-Social Change, University of Essex).

Chambers, R. (2000), 'Evaluation Criteria for Statistical Editing and Imputation', Working Paper for the Euredit Project on the Development and Evaluation of New Methods for Editing and Imputation, University of Southampton, Southampton, UK.

Chambers, R. and Zhao, X. (2003), *Evaluation of Edit and Imputation Methods Applied to the UK Annual Business Inquiry*, Volume 1, Euredit Deliverable D6.2, Chapter 4.2.

Champney, T.F. and Bell, R. (1982), 'Imputation of Income: A Procedural Comparison', *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp431-436.

Dillman, D.A., Eltinge, J.L., Groves, R.M. and Little, R.J.A (2002) 'Survey Nonresponse in Design, Data Collection and Analysis', in *Survey Nonresponse*, edited by Groves, R.M., Dillman, D.A., Eltinge, J.L. and Little, R.J.A., Wiley, New York.

Frick, J. and Grabka, M. (2003), 'Missing Income Data in the German SOEP: Incidence, Imputation and its Impact on the Income Distribution', DIW Discussion Paper Series No. 376, DIW, Berlin.

Heeringa, S.G., and Lepkowski, J.M. (1986), 'Longitudinal Imputation for the SIPP', *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp206-210.

Hofferth, S., Stafford, F.P., Yeung, W.J., Duncan, G.J., Hill, M.S., Lepkowski, J., Morgan, J.N. (1998), 'A Panel Study of Income Dynamics: Procedures and Codebooks – Guide to the 1993 Interviewing Year', Institute for Social Research, The University of Michigan.

Kalton, G. (1986), 'Handling Wave Nonresponse in Panel Surveys', *Journal of Official Statistics*, Vol. 2, No. 3, pp. 303-314.

Kalton, G., and Brick, M., (2000), 'Weighting in Household Panel Surveys' in *Researching Social and Economic Change: the Uses of Household Panel Studies*, edited by Rose, D., Routledge, London.

Lepkowski, J.M., (1989), 'Treatment of Wave Nonresponse in Panel Surveys' in *Panel Surveys*, edited by Kasprzyk, D., Duncan, G.J., Kalton, G, Singh, M.P., Wiley, New York.

Little, R.J.A. (1988), *Missing Data Adjustments in Large Surveys*, Journal of Business & Economic Statistics, Vol. 6, No. 3, pp. 287-296.

Little, R.J.A., and Su, H.L. (1989) 'Item Non-response in Panel Surveys' in *Panel Surveys*, edited by Kasprzyk, D., Duncan, G., and Singh, M.P., Wiley, New York.

Nordholt, E.S., (1998), 'Imputation: Methods, Simulation Experiments and Practical Examples', *International Statistical Review*, Vol. 66, pp. 157-180.

Pannekoek, J. (2002), *(Multivariate) Regression and Hot Deck Imputation Methods*, Euredit Deliverable 5.1.1, Statistics Netherlands.

Pennell, S.G. (1993) 'Cross-Sectional Imputation and Longitudinal Editing Procedures in the Survey of Income and Program Participation', Institute for Social Research, The University of Michigan.

Quintano, C., Castellano, R., and Regoli, A. (2002), 'A Mixed Imputation Procedure in a Split Panel Survey', Paper presented at International Conference on Improving Surveys, Copenhagen, 25-28 August.

Rubin, D.B. (1976) 'Inference and Missing Data', *Biometrika*, vol. 63, pp. 581-590.

Starick, R. and Watson, N. (2006, forthcoming), 'An Evaluation of Alternative Income Imputation Methods in the HILDA Survey', HILDA Project Technical Paper Series, Melbourne Institute of Applied Economic and Social Research, The University of Melbourne.

Tremblay, A. (1994), 'Longitudinal Imputation of SIPP Food Stamp Benefits', *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp. 809-814.

Watson, N. (2004), 'Income and Wealth Imputation for Waves 1 and 2', HILDA Project Technical Paper Series No. 3/04, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Watson, N. and Wooden, M. (2004), 'The HILDA Survey: A Summary', *Australian Journal of Labour Economics*, Vol. 7, No. 2, pp. 117-124.

Williams, T.R., and Bailey, L. (1996), *Compensating for Missing Wave Data in the Survey of Income and Program Participation (SIPP)*, Proceedings of the Survey Research Methods Section, American Statistical Association, pp. 305-310.