Evaluation of Alternative Income Imputation Methods for the HILDA Survey

Rosslyn Starick

Australian Bureau of Statistics

Nicole Watson

The University of Melbourne

The Household, Income and Labour Dynamics in Australia (HILDA) Survey project was initiated and is funded by the Australian Government Department of Families, Community Services and Indigenous Affairs (FaCSIA) and is managed by the Melbourne Institute of Applied Economic and Social Research. The findings and views reported in this paper, however, are those of the authors and should not be attributed to FaCSIA, the Melbourne Institute or the Australian Bureau of Statistics.

Abstract

This paper assesses various methods for imputing income in a household-based longitudinal survey. Taking guidance from the experience of similar longitudinal studies, nine longitudinal methods are evaluated in a simulation study using five waves of data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The quality of the imputed data is evaluated by considering eleven criteria that measure the predictive, distributional and estimation accuracy of the cross-sectional estimates and the estimates of change between two waves. Many of the imputation methods perform well cross-sectionally, however, when the methods are placed in a longitudinal context, the strengths and weaknesses of the methods are more apparent. The Little and Su method that uses imputation classes performs the best overall.

Introduction

All large-scale surveys, including longitudinal surveys, have non-response and various weighting and imputation strategies are employed to address this problem. 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 either a set of item non-response in the longitudinal record or as a missing unit for a wave so either imputation or 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 survey 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, a large nationally-representative longitudinal survey of Australian households, demonstrates one of the difficulties with imputation in the longitudinal context.[2] The imputation in HILDA focuses on the income variables as they form a key part of the survey and are subject to substantial non-response.[3] A nearest neighbour regression method (described later in this paper) was initially adopted to impute missing income data. This method led to an overstatement of the change in income between waves, even though the regression models used income information in the other waves, other income variables from the wave being imputed and other respondent characteristics (Watson, 2004). In subsequent releases, a variant of the Little and Su method (Little and Su, 1989; also described later in this paper) was used, where donors are

[1] In a longitudinal survey, there is usually the opportunity to re-impute prior waves with each new release of the data.
[2] Further information about the HILDA Survey is given by Watson and Wooden (2004).
[3] 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.

matched to recipients within imputation classes. An evaluation study was undertaken to compare various imputation methods with a view to identifying the most suitable method for imputing income variables.

The evaluation study used data from the first five waves of the HILDA Survey to construct simulated data on which imputation methods were tested. Eleven evaluation criteria were used to assess the predictive accuracy, distributional accuracy and estimation accuracy of each method. Nine longitudinal imputation methods were tested, being one nearest neighbour method, four variants of the Little and Su method, three carryover methods and one hotdeck method (all described below). Two cross-sectional methods were assessed as the fallback option when the longitudinal methods cannot be used.

The various design features of the HILDA Survey 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 at each release. 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, and as a result seek a single imputation method that performs well for this variable alone (for example, Tremblay, 1994; Quintano, et al, 2002; Frick and Grabka, 2003). Williams and Bailey (1996) did consider four income components but also sought one imputation method that worked well for all. In 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 and give some consideration to why this might be the case. We are open to using a small number 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. In the HILDA evaluation study, a relatively large number of evaluation criteria are used, however we do not claim to have used all possible criteria.

This paper describes a methodological evaluation framework for assessing a good imputation method, details alternative imputation methods and presents a summary comparison of the performance of these methods in the HILDA Survey. Finally, we discuss the practicalities of each method, in terms of the programming complexity and the running times.

Evaluation Methodology

This section outlines how the evaluation study was designed. We first discuss how the response mechanism was modeled in the simulated datasets and then we describe the evaluation criteria used.

Simulated Data

Ideally we want to compare the imputed value with the value the respondent would have reported if they had not refused or did not 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 portion 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.[4] We can then compare the imputed values from the various imputation methods with the true values.

More specifically, the simulated datasets comprise 8193 people who responded and provided all income items in the first five waves of the HILDA Survey they were eligible for. The sample of cases set to missing were selected based on logistic models 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 depends 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 of the first five waves of the entire HILDA dataset to identify the relationship between the probability of reporting a particular income component and various explanatory variables.[5] Only cases that reported having that particular income were included in the models. To take account of the dependence of missingness between income components within a wave and across waves, the response mechanism was modeled sequentially through the list of income components across the waves. As a first step, a model was constructed to predict the presence of a response to the question about wages and salaries in wave 1. Then a model was constructed to predict the presence of a response to Government pensions in wave 1, contingent on the response indicator for wages and salaries in wave 1, and so on. The final model constructed was for private transfers in wave 5, which included a response indicator for all of the eight income components in waves 1 to 4 and the first seven income components in wave 5. The probability that an individual provided an interview was also modeled at each wave and included a response indicator for each prior wave.

Simulations

Thirty datasets were simulated from the set of all complete cases by using the above predicted probabilities to assign missingness to the various income components. The predicted probabilities were sequentially adjusted where missingness had been assigned to earlier income components.

In each wave, a sample of non-responding persons from the set of complete cases was determined first, in line with the proportion of non-responding persons observed in the entire HILDA data. This was done by randomly assigning cases to responding or not, proportional to their predicted

[4] While the respondent may not report income figures with complete accuracy, we can only compare the imputed values to what we know.

[5] The variables considered in the models include: age, marital status, highest level of education, labour force status, occupation status, multiple job holder, 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 education, time spent employed, and number of jobs held.

probability for being a non-respondent (which were adjusted for any missingness assigned in earlier waves). 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. Only non-zero cases were considered in simulating the missingness in that variable.

The missing data in the thirty simulated datasets were imputed via each imputation method and the resulting evaluation measures were standardized and averaged across the three broad accuracy classes – predictive, distribution and estimation.

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.[6] The donor characteristics provided are the mean value of the income component, the standard deviation as a multiple of the mean, the skew of the income component, and the correlation of the income component with age. In the top half of the table, the characteristics of respondents reporting non-zero incomes are provided as the donor pool can only includes these cases. For non-respondents in responding households, the donors can have either zero or non-zero amounts for the income component being imputed, so the characteristics reported in the second half of the table includes both types of cases. Some income components (such as business income, dividends and royalties, rental income and private transfers) are highly variable or highly skewed, and thus present a challenge for the imputing process. For business income, there are a large number of respondents that need to be imputed but only a small donor pool to draw from. When the cases with zero income for a component are included, the variability and the skew of the data are much greater. This table aids the interpretation of the results later in this paper.

Table 1: Characteristics of simulated datasets, averaged across waves 1 to 5

	Number imputed (recipients)	Potential donors					
		Number[a]	Recipnts / donors	Mean	Std dev / mean	Skew	Corr with age
Respondents (non-zero cases only)							
Wages and salaries	207	3468	0.06	37,329	0.83	3.9	0.24
Aust govt pensions	38	2200	0.02	8477	0.51	0.6	0.36
Business income	75	250	0.30	15,734	2.91	-1.3	0.00
Interest income	186	1159	0.16	2008	2.91	12.5	0.08
Dividends and royalties	154	1306	0.12	3056	4.19	11.4	0.11
Rent income	44	356	0.12	2144	4.14	1.6	0.19
Private transfers	24	152	0.16	4889	1.13	2.4	0.12
Non-respondents (zero and non-zero cases)							
Wages and salaries	450	5988	0.08	21,588	1.38	3.4	-0.18
Aust govt pensions	450	6158	0.07	3032	1.59	1.5	0.47
Business income	450	6120	0.07	636	15.38	-1.5	0.00
Interest income	450	6008	0.07	387	6.95	27.0	0.13
Dividends and royalties	450	6041	0.07	648	9.28	24.2	0.07
Rent income	450	6151	0.07	124	17.70	8.8	0.04
Private transfers	450	6171	0.07	120	9.54	15.5	-0.03
Total Financial Year income[b]	450	5601	0.08	27,956	1.07	4.3	-0.03

Notes: a. Number of potential donors depends on the imputation method being used. For respondents this is the number of donors with non-zero income amounts, but for non-respondents it is the number of donors with zero or non-zero income amounts.
b. For some methods, the donor used to impute total income also provides the various income components.

[6] 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.

While the simulated datasets are as realistic as possible to the HILDA environment, a difference in household size occurred. The average number of adults per household in the simulation datasets is 1.5 compared to 1.9 in the entire HILDA data. This is because non-respondents in 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'. It is not expected that this difference will substantially affect the results of this study.

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.

Seven of the eleven criteria used in this evaluation study are based on those proposed by Chambers (2000) for the Euredit Project (these are criteria 1, 2, 6, 8-11). We have included four additional criteria to help assess changes between waves (criteria 3 and 7) and how well relationships between variables have been maintained (criteria 4 and 5). The criteria measure predictive accuracy, distributional accuracy and estimation accuracy. When undertaking regression analysis, all eleven criteria are important. 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. Most of the criteria are applied to both the *level* of income at each wave and the *change* in income between waves. The exceptions are that criteria 3 and 7 apply only to the *change* in income between waves, and criteria 4 and 5 apply only to the *level* of income at each wave.

Apart from Criteria 7, the 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 to 5: Predictive Accuracy

The first five criteria assess the predictive accuracy of the imputation by considering how close the imputed value ($\hat{Y}$) is to the true value (Y^*). The imputation method should preserve the true values as far as possible.

The first criterion is the Pearson correlation between $\hat{Y}$ and Y^*:

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

where $\bar{Y}$ denotes the mean of *Y*-values. For data that are reasonably normal, this criteria provides a good measure of the imputation performance and a good imputation method will have *r* close to 1.

The second criterion uses a regression approach to evaluate the performance of the imputation method which is useful for highly skewed data. The imputed and true values are first transformed by taking the natural logarithm (*log (Y+1)*).[7] The transformed imputed values ($\hat{Y}_t$) are then

[7] 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.

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 t-test statistic closest to zero.

$$T = \frac{b-1}{\widehat{se}(b)} \tag{2}$$

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

The third criterion assesses the preservation of the change between waves by comparing the cross-wave correlations for the imputed and true values. The formulae for the absolute change in correlations of the imputed and true values between wave 1 and wave 2 is

$$d_{corr1,2} = \left| \frac{\sum_{i=1}^{n}\left(\hat{Y}_{i1} - \bar{\hat{Y}}_1\right)\left(\hat{Y}_{i2} - \bar{\hat{Y}}_2\right)}{\sqrt{\sum_{i=1}^{n}\left(\hat{Y}_{i1} - \bar{\hat{Y}}_1\right)^2 \sum_{i=1}^{n}\left(\hat{Y}_{i2} - \bar{\hat{Y}}_2\right)^2}} - \frac{\sum_{i=1}^{n}\left(Y_{i1}^* - \bar{Y}_1^*\right)\left(Y_{i2}^* - \bar{Y}_2^*\right)}{\sqrt{\sum_{i=1}^{n}\left(Y_{i1}^* - \bar{Y}_1^*\right)^2 \sum_{i=1}^{n}\left(Y_{i2}^* - \bar{Y}_2^*\right)^2}} \right| \tag{3}$$

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, d_{corr} will be close to zero).

The fourth and fifth criteria assess the preservation of the relationships between income variables. The two measures used are the Euclidean distance between the imputed and true data values in multi-dimensional space and the correlation between the variables for the imputed and true data.

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 mean of the Euclidean distances of the n imputed cases is then calculated.

$$mean\left(d_i\right) = \frac{1}{n}\sum_{i=1}^{n}\sqrt{\sum_{j=1}^{k}\left(y_{ij}^* - \hat{y}_{ij}\right)^2} \tag{4}$$

A good imputation method will have the lowest mean.

The last evaluation criterion assessing the predictive accuracy of the imputation compares the true correlations between the income variables being imputed with the correlations from the imputed data for each two-way combination of j = 1 to k (the formulae below is for variables 1 and 2).

$$\left| r_{Y_1^* Y_2^*} - r_{\hat{Y}_1 \hat{Y}_2} \right| = \left| \frac{\sum_{i=1}^{n}\left(Y_{i1}^* - \bar{Y}_1^*\right)\left(Y_{i2}^* - \bar{Y}_2^*\right)}{\sqrt{\sum_{i=1}^{n}\left(Y_{i1}^* - \bar{Y}_1^*\right)^2 \sum_{i=1}^{n}\left(Y_{i2}^* - \bar{Y}_2^*\right)^2}} - \frac{\sum_{i=1}^{n}\left(\hat{Y}_{i1} - \bar{\hat{Y}}_1\right)\left(\hat{Y}_{i2} - \bar{\hat{Y}}_2\right)}{\sqrt{\sum_{i=1}^{n}\left(\hat{Y}_{i1} - \bar{\hat{Y}}_1\right)^2 \sum_{i=1}^{n}\left(\hat{Y}_{i2} - \bar{\hat{Y}}_2\right)^2}} \right| \tag{5}$$

A good imputation method will have between-variable correlations close to the true between-variable correlations, so the difference will be close to zero.

Criteria 6 and 7: Distributional Accuracy

The next two criteria measure the distribution accuracy by considering whether the imputation method preserves the distribution of the true values.

The sixth criteria measures the distance between the empirical distribution functions for both the imputed and true values. The distance between these functions can be measured using the Kolmogorov-Smirnov distance:

$$d_{KS} = \max_j \left(\left| \frac{1}{n} \sum_{i=1}^{n} I\left(Y_i^* \le x_j\right) - \frac{1}{n} \sum_{i=1}^{n} I\left(\hat{Y}_i \le x_j\right) \right| \right) \tag{6}$$

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.

In a longitudinal survey context, it is also important to assess the consistency of the income distribution between waves. The seventh criterion compares the income mobility in the dataset that includes the imputed values with 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 of the dataset with imputation is tested for similarity to the distribution of the dataset of true values. A Chi-Square test is used where the observed cell frequencies are those from the imputed dataset 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

$$\chi^2 = \sum_{j=1}^{10} \sum_{i=1}^{10} \frac{\left(\hat{n}_{ij} - n_{ij}^*\right)^2}{n_{ij}^*} \tag{7}$$

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

Criteria 8 to 11: Estimation Accuracy

The final four criteria measure the estimation accuracy of the imputation methods by assessing whether the lower order moments of the distributions of the true values are preserved. Criteria 8 to 11 measure the absolute difference between the true and imputed cases in the mean, variance, skewness and kurtosis. A good imputation method will have a low absolute difference in moments.

$$m_1 = \left| \frac{1}{n} \sum_{i=1}^{n} \left(Y_i^* - \hat{Y}_i\right) \right| \tag{8}$$

$$m_2 = \left| \frac{1}{n} \sum_{i=1}^{n} \left(Y_i^{*2} - \hat{Y}_i^2\right) \right| \tag{9}$$

$$m_3 = \left| \frac{1}{n} \sum_{i=1}^{n} \left(Y_i^{*3} - \hat{Y}_i^3\right) \right| \tag{10}$$

$$m_4 = \left| \frac{1}{n} \sum_{i=1}^{n} \left(Y_i^{*4} - \hat{Y}_i^4\right) \right| \tag{11}$$

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 the 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) in their Cross-National Equivalence File. 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 the 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.[8]
- 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 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;
- A hotdeck method;
- Four methods based on the Little and Su method, being with and 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.

The most obvious exclusion from this list of methods tested is multiple imputation. This method was not considered as a single imputation solution is required for the HILDA data release file at this time.[9]

In certain situations where the Little and Su methods and the carry-over methods do not work on their own (for example, the respondent is only interviewed in one wave and does not provide information about an income component), a cross-sectional method needs to be used. Two fall-back options were assessed: a nearest neighbour method and a hotdeck method. Only information about

[8] 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.

[9] 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.

the respondent from within the wave was used to form hotdeck classes or as covariates in the regression equation for the nearest neighbour method.

Note that the nearest neighbour regression method, the hotdeck method, the carryover method and the first two Little 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 is imputed simultaneously. The last two Little and Su methods considered are multivariate imputation methods, where two or more income variables are imputed simultaneously.

Longitudinal 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 reported value for the variable being imputed replaces the missing value of the recipient.

For each wave and for each variable imputed, log-linear regression models using information from the same wave as well as information from other waves (if available) were constructed. Over 30 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). A stepwise elimination process in SAS was used to identify the key variables for each variable, wave and simulation.

The predicted values from these regression models for the variable being imputed were used to identify the nearest case (donor *d*) whose reported value (Y_d) could be inserted into the case with the missing value ($\hat{Y}_i = Y_d$). Donor *d* has the closest predicted value to the respondent *i*, that is $\left|\hat{\mu}_i - \hat{\mu}_d\right| \leq \left|\hat{\mu}_i - \hat{\mu}_p\right|$ for all respondents *p* (potential donors) where $\hat{\mu}_i$ is the predicted mean of *Y* for individual *i* that needs to be imputed, and Y_d is the observed value of *Y* for respondent *d*.

For respondents, the missing income is imputed for each variable. For non-respondents, only donors for total income are identified and the income components are taken from the same donor as total income.

For wages and salaries, government pensions, rental income, an additional restriction that the donor and recipient fall within the same age class (15-19, 20-24, 25-34, 35-44, 45-54, 55-64, 65+) was also applied.[10] For interest income, dividends and royalties, and private transfers, the age classes the donors and recipients were matched within were (15-24, 25-54, 55+). No age class restrictions were applied for business income. Total income for non-respondents had the more detailed age class restrictions applied.

This method provides one of two fallback solutions when the Little and Su and carryover methods considered in this study cannot be used.[11] When this fallback solution is adopted, the nearest neighbour regression method excludes income reported by respondents in other waves from the regression model, so we have termed this fallback method a 'cross-sectional nearest neighbour regression method'.

[10] Age groups were used to create the imputation classes because it is a simple characteristic known for almost all donors and recipients and helped avoid imputing unrealistic income amounts, especially for the younger age groups.

[11] 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 initial imputed values.

Longitudinal Hotdeck Method

The hotdeck method randomly matches suitable donors to recipients within imputation classes. The donor's reported value for the variable being imputed replaces the missing value of the recipient.

Up to 15 categorical variables were used to define the imputation classes for each income component. Suitable classes where derived from subject matter knowledge and investigative regression analysis using the data in simulation 1.[12] These classes were then used in all waves and all simulations. The variables considered in the formation of the imputation classes were the categorical equivalent of those considered in the nearest neighbour modeling process.

Where there were not sufficient donors within a class, the imputation classes were sequentially folded back, removing the least important class variable first until a suitable donor was found. When more than one donor could be matched to a recipient *i* within an imputation class *c*, a donor *d* was selected randomly (the class of the donor and the recipient are the same, i.e. $c_i = c_d$). The donor's reported value was inserted into the recipient's missing value $\hat{Y}_i = Y_d$. A hotdeck macro (HESIMPUT) written by the Methodology Division of the Australian Bureau of Statistics was used.

This method provides an alternative fallback solution when the Little and Su and carryover methods considered in this study cannot be used. When this fallback solution is adopted, it only has cross-sectional information from which to form the imputation classes (that is the income bands from other waves is not available), so we have termed this fallback method a 'cross-sectional hotdeck method'.

Basic Little and Su Method

The imputation method proposed by Little and Su (1989) will be referred to as the 'basic Little and Su method' to distinguish it from the modified version using imputation classes which will be referred to as the 'Little and Su method with imputation classes'.

The basic Little and Su method incorporates (via a multiplicative model) the trend across waves (column effect), the recipient's departure from the trend in the waves where the income component has been reported (row effect), and a residual effect donated from another respondent with complete income information for that component (residual effect). The model is of the form

$$imputation = (roweffect)(columneffect)(residualeffect).$$

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.

[12] The sample in simulation 1 was randomly divided into 10 parts and log-linear regression models were fitted (via a stepwise process) 10 times each wave, dropping 1/10th of the sample each time. Those variables most frequently included in the regression models each wave and in each replica of the sample were considered as imputation classes.

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

The missing value Y_{ij} was imputed by

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

where the three terms in brackets 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 or hotdeck method. In addition, the donors must have non-zero row effects to avoid divisions by zero.

Little and Su Method with Imputation Classes

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 basic characteristics of the donors and recipients. Donors and recipients were matched within longitudinal imputation classes defined by the following age ranges in the latest wave: 15-19, 20-24, 25-34, 35-44, 45-54, 55-64, 65+. The column and row effects are calculated within each imputation class and donors are matched to recipients which share the same imputation class.

Little and Su Method Using Key Variables

Two multivariate methods are considered in this evaluation study and these aim to impute two or more income components over time 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. Any missing items for related variables 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. Deciding that wages and salaries is more important than Government pensions, 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.

Little and Su Method Across Multiple Variables

The second multivariate method 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 donor is identified.

Let Y_k denote the income variables being imputed, k = 1, …, K. For each Y_k, compute the column effects c_{kj} and the row effects $\bar{Y}_k^{(i)}$. 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 d is found.

The missing value Y_{kij} is imputed by $\hat{Y}_{kij} = Y_{kdj} \frac{\bar{Y}_k^{(i)}}{\bar{Y}_k^{(d)}}$. That is, the row effects that are specific to variable k for individual i and d are used (the combined row effect is just used to identify the nearest suitable donor). As for the previous multivariate method, the income components imputed simultaneously were firstly wages and salaries together with Government pensions, and then dividends, royalties, interest income together with rent income.

Last Value Carried Forward

A last value carried forward method 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 or hotdeck method is used.

Random Carryover Method

The random carryover method imputes single missing wave data that is bounded on both sides by an interviewed wave (Williams and Bailey, 1996). 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 for one wave where only one wave of data was available. For example, the last value was carried forward where only wave 1 information was available or the next value was carried backwards where only wave 5 information was available. Where no information was available in surrounding waves, the nearest neighbour method or the hotdeck method was used to impute these waves. A value was only carried forward or backwards one wave only.

Population Carryover Method

A variation of the random carryover method was also implemented in the evaluation study and 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 is created which equals 1 when the reported change between waves *t* and *t+1* is smaller than the reported change between waves *t* and *t-1* for the complete cases; and 0 otherwise. The proportion *p*, of the interviewed sample where the change between waves *t* and *t+1* is smaller than the change between waves *t* and *t-1* is then determined. Either the preceding wave or the subsequent wave donates the imputed amount reflecting the probabilities associated with the occurrence of change between waves found in the complete cases.

Comparison of Imputation Methods

Of the nine imputation methods described above, the Little and Su methods and the carryover methods are examined both when the NNRM fallback method is used and when the hotdeck fallback method is used, thus resulting in sixteen imputation methods being implemented in this study. These imputation methods are compared via the eleven evaluation criteria. The performance of the imputation methods for each variable is first considered in a cross-sectional context and then in a longitudinal context. The longitudinal performance of the method is more important in a longitudinal survey than its cross-sectional performance. Finally, an overall summary of the methods is provided.

To help draw conclusions from the many criteria and methods considered, the evaluation measures[13] were standardized and these standardized scores were averaged within the three classes of evaluation measures – predictive, distributional and estimation.[14] These averaged scores are reported cross-sectionally (for the level of income) and longitudinally (for the change in level between waves) for both respondents and non-respondents in the first table provided for each variable in the following sections. The standardization was undertaken across the 30 replicates, eight variables, five waves and 16 methods, but was done separately for respondents and non-respondents as the scores tend to be substantially different for these two groups. Methods with low averaged standardized scores are better than methods with high scores (the lowest scores is indicated with a bold entry in the tables). By averaging the scores within the three classes of evaluation measures, we have treated the measures within each class equally. The advantage of averaging standardized scores over a simple ranking system is that it allows us to see how close each method is to alternative methods and we can compare the performance of the methods across variables.

The methods were compared for each two-way combination via a t-test to identify significant differences in the predictive, distributional and estimation summary scores. The results of these tests are reported in the second table provided for each variable in the following sections. , the results of t-tests to compare the predictive, distributional and estimation summary measures for each combination of two methods are summarized. The lower section of the table to the left of the diagonal (unshaded section) report the results for cross-sectional estimates and the upper section of the table to the right of the diagonal (shaded section) report the results for estimates of change (the longitudinal estimates). P, D and E denote that the methods were significantly different at the 1% level. p, d, e denotes significant differences at the 5% level.[15] The results presented in the two tables

[13] Box plots of the evaluation scores for the 30 simulations and 5 waves for each variable, respondent group and method tested are found in Appendix 1 of Starick and Watson (2007).

[14] Distributional accuracy is only measured by Criteria 6 (Kolmogorov-Smirnov distance) with the exception of the longitudinal distributional accuracy of total income (measured by Criteria 7 chi-square statistic). The distributional accuracy reported in Tables 2.1-8 include only Criteria 6. The results for Criteria 7 are reported separately.

[15] No correction has been made because we are undertaking multiple comparisons. If we adjust the significance level according to the Bonferroni correction for 120 tests within each variable, respondent group and type of estimate, a 5% level would be reduced to a 0.04% level, which seems to be too conservative.

are considered together in the following discussion. The labels used for the methods (to save space in the second table for each variable) are:

- nnrml – Longitudinal NNRM (using cross-sectional NNRM fallback);
- ls – Basic Little and Su method (using cross-sectional NNRM fallback);
- lsc – Little and Su method with imputation classes (using cross-sectional NNRM fallback);
- lsk – Little and Su method via a key variable (using cross-sectional NNRM fallback);
- lsd – Little and Su method via a distance function (using cross-sectional NNRM fallback);
- lvcf – Last value carried forward method (using cross-sectional NNRM fallback);
- rco – Random carryover method (using cross-sectional NNRM fallback);
- pco – Population carryover method (using cross-sectional NNRM fallback);
- hdl – Longitudinal hotdeck method (using cross-sectional hotdeck method fallback);
- hdls – Basic Little and Su method (using cross-sectional hotdeck method fallback);
- hdlsc – Little and Su method with imputation classes (using cross-sectional hotdeck method fallback);
- hdlsk – Little and Su method via a key variable (using cross-sectional hotdeck method fallback);
- hdlsd – Little and Su via a distance function (using cross-sectional hotdeck method fallback);
- hdlvcf – Last value carried forward method (using cross-sectional hotdeck method fallback);
- hdrco – Random carryover method (using cross-sectional hotdeck method fallback); and
- hdpco – Population carryover method (using cross-sectional hotdeck method fallback).

Only significant differences are highlighted in this discussion.

Wages and Salaries

Table 2.1 provides the average standardized scores for wages and salaries and Table 2.2 shows which methods are significantly different from each other.

The methods using the cross-sectional nearest neighbour regression method (NNRM) as the fallback method almost always performed marginally better for wages and salaries (though not always significantly better) than their counterparts using the cross-sectional hotdeck fallback method. Therefore, the following discussion focuses on the methods using the NNRM fallback method.

Cross-sectionally, the carryover methods perform the best for respondents on predictive and distributional accuracy (few methods differ on estimation accuracy). The longitudinal NNRM is a close second, ahead of the other methods on distributional accuracy. This is followed closely by the Little and Su method that uses imputation classes, which is ahead of the other Little and Su methods on predictive accuracy. The longitudinal hotdeck method was the poorest performer on predictive accuracy and was amongst the worst performers on distributional and estimation accuracy.

Longitudinally, the differences between the methods are more apparent. The carryover methods perform well on the predictive measures, but are very poor distributionally, as would be expected given they impute either zero change or a large change between waves. The Little and Su methods perform well, with the one using imputation classes performing best with respect to distributional accuracy.

For non-respondents, the performance of the methods is more disperse but similar overall conclusions are drawn. The population carryover method performs well cross-sectionally, though performs very poorly on the distribution of change. The Little and Su methods perform reasonably well cross-sectionally and better than the other methods longitudinally, with the Little and Su method using imputation classes being slightly ahead of the other methods in this class.

Overall, for wages and salaries, the Little and Su method using imputation classes is the best performer longitudinally, particularly on maintaining the distribution of change, and performs reasonably well on a cross-sectional basis.

Table 2.1: Average value of standardised predictive (P), distributional (D) and estimation (E) evaluation measures for wages and salaries

	Cross-sectional				*Longitudinal*			
Base method: Longitudinal method	*P*	*D*	*E*	*Ave*	*P*	*D*	*E*	*Ave*
Respondents								
NNRM: NNRM Longitudinal	-0.30	**-0.97**	-0.17	-0.48	0.46	-0.13	**-0.04**	0.10
NNRM: Basic Little & Su	-0.25	-0.84	-0.23	-0.44	0.16	-0.44	-0.01	-0.10
NNRM: Little & Su w Imp Classes	-0.36	-0.88	-0.22	-0.49	0.16	**-0.63**	-0.02	**-0.17**
NNRM: Little & Su Key Var	-0.26	-0.85	-0.23	-0.44	0.16	-0.44	-0.01	-0.10
NNRM: Little & Su Distance	-0.24	-0.84	-0.23	-0.44	0.18	-0.43	-0.01	-0.09
NNRM: LVCF	-0.43	-0.92	-0.19	-0.51	-0.23	2.22	0.05	0.68
NNRM: Random Carryover	-0.47	-0.96	-0.23	-0.55	-0.24	1.16	-0.01	0.30
NNRM: Population Carryover	**-0.47**	-0.96	**-0.24**	**-0.55**	**-0.24**	1.20	-0.01	0.32
HD: Hotdeck Longitudinal	-0.11	-0.83	-0.15	-0.37	0.51	-0.24	-0.01	0.08
HD: Basic Little & Su	-0.23	-0.84	-0.23	-0.43	0.15	-0.46	0.00	-0.10
HD: Little & Su w Imp Classes	-0.33	-0.88	-0.21	-0.47	0.17	-0.63	-0.01	-0.16
HD: Little & Su Key Var	-0.23	-0.84	-0.23	-0.43	0.15	-0.46	0.00	-0.10
HD: Little & Su Distance	-0.22	-0.84	-0.23	-0.43	0.19	-0.44	-0.01	-0.09
HD: LVCF	-0.28	-0.91	-0.21	-0.46	-0.14	2.28	0.03	0.72
HD: Random Carryover	-0.36	-0.91	-0.23	-0.50	-0.20	1.21	-0.03	0.33
HD: Population Carryover	-0.37	-0.91	-0.23	-0.50	-0.20	1.25	-0.03	0.34
Non-respondents								
NNRM: NNRM Longitudinal	-0.24	-0.17	-0.12	-0.18	-0.11	2.48	0.07	0.81
NNRM: Basic Little & Su	-0.52	0.05	-0.23	-0.24	-0.32	0.38	-0.05	0.00
NNRM: Little & Su w Imp Classes	**-0.55**	-0.01	**-0.27**	-0.27	-0.35	**0.34**	-0.06	**-0.02**
NNRM: Little & Su Key Var	-0.47	0.12	-0.22	-0.19	-0.32	0.39	**-0.06**	0.00
NNRM: Little & Su Distance	-0.47	0.32	-0.13	-0.10	-0.32	0.40	-0.03	0.01
NNRM: LVCF	-0.30	-0.18	-0.13	-0.20	-0.21	2.37	0.05	0.73
NNRM: Random Carryover	-0.53	-0.25	-0.22	-0.33	**-0.38**	1.72	0.04	0.46
NNRM: Population Carryover	-0.53	-0.25	-0.22	**-0.33**	-0.37	1.73	0.04	0.46
HD: Hotdeck Longitudinal	-0.06	0.10	0.21	0.08	-0.16	2.33	0.13	0.77
HD: Basic Little & Su	-0.22	0.24	-0.02	0.00	-0.14	1.10	0.08	0.35
HD: Little & Su w Imp Classes	-0.21	0.12	-0.10	-0.06	-0.14	1.11	0.05	0.34
HD: Little & Su Key Var	-0.13	0.28	-0.01	0.05	-0.13	1.11	0.08	0.35
HD: Little & Su Distance	-0.14	0.37	0.04	0.09	-0.13	1.15	0.09	0.37
HD: LVCF	-0.15	-0.19	-0.14	-0.16	-0.19	2.33	0.06	0.73
HD: Random Carryover	-0.44	-0.27	-0.21	-0.30	-0.34	1.65	0.07	0.46
HD: Population Carryover	-0.43	**-0.27**	-0.21	-0.31	-0.34	1.66	0.07	0.46

Table 2.2: Significant differences between predictive (P), distributional (D) and estimation (E) evaluation measures for wages and salaries

Respondents	nnrml	ls	lsc	lsk	lsd	lvcf	rco	pco	hdl	hdls	hdlsc	hdlsk	hdlsd	hdlvcf	hdrco	hdpco
nnrml		PD	PD	PD	PD	PD	PD	PD	d	PD	PD	PD	PD	PD	PD	PD
ls	D		D			PD	PD	PD	PD		D			PD	PD	PD
lsc	D	P		D	D	PD	PD	PD	PD	D		D	D	PD	PD	PD
lsk	D		P			PD	PD	PD	PD		D			PD	PD	PD
lsd	D		P			PD	PD	PD	PD		D			PD	PD	PD
lvcf	Pd	PD		PD	PD		D	D	PD	PD	PD	PD	PD	p	D	D
rco	P	PD	PD	PD	PD				PD	PD	PD	PD	PD	pD		
pco	P	PD	PD	PD	PD				PD	PD	PD	PD	PD	pD		
hdl	PD	P	P	P	P	PD	PDe	PDe		PD	PD	PD	PD	PD	PD	PD
hdls	D		P			PD	PD	PD	P		D			PD	PD	PD
hdlsc	D	p		p	p	p	PD	PD	P	p		D	D	PD	PD	PD
hdlsk	D		P			PD	PD	PD	P		p			PD	PD	PD
hdlsd	D		P			PD	PD	PD	p		P			PD	PD	PD
hdlvcf	D	D		d	d	P	Pd	Pd	PD	D		D	d		D	D
hdrco	D	PD		Pd	Pd		Pd	Pd	PD	PD		PD	Pd			
hdpco	D	PD		Pd	Pd		pd	Pd	PDe	PD		PD	Pd			

Non-respondents	nnrml	ls	lsc	lsk	lsd	lvcf	rco	pco	hdl	hdls	hdlsc	hdlsk	hdlsd	hdlvcf	hdrco	hdpco
nnrml		PDe	PDE	PDE	PDe	p	PD	PD		D	D	D	D		PD	PD
ls	PDE					pDe	De	D	PDE	PDE	PDe	PDE	PDE	PDe	De	De
lsc	PDE					PDe	De	De	PDE	PDE	PDe	PDE	PDE	PDe	DE	De
lsk	PDE	P	PD			pDe	De	De	PDE	PDE	PDe	PDE	PDE	PDe	De	De
lsd	PD	PDE	PDE	DE		pD	D	D	PDE	PDe	PD	PDe	PDe	PD	De	D
lvcf	p	PDE	PDE	PDE	PD		PD	PD		D	D	D	D		PD	PD
rco	PDE	D	D	PD	PDE	PdE			PD	PD	PD	PD	PD	PD		
pco	PDE	D	D	PD	PDE	PdE			PD	PD	PD	PD	PD	PD		
hdl	PDE	PE	PdE	PE	PDE	PDE	PDE	PDE		D	D	D	D		PD	PD
hdls	DE	PDE	PDE	PDE	PE	PDE	PDE	PDE	PDE					D	PD	PD
hdlsc	D	PE	PDE	PE	PD	PD	PDE	PDE	PE	De				D	PD	PD
hdlsk	PDE	PDE	PDE	PDE	PE	PDE	PDE	PDE	PDE	P	PDE			D	PD	PD
hdlsd	PDE	PDE	PDE	PDE	PE	PDE	PDE	PDE	PDE	PD	PDE			D	PD	PD
hdlvcf	P	PDE	PDE	PDE	PD	P	Pe	Pe	PDE	PDE	PD	DE	DE		PD	PD
hdrco	PDe	PD	PD	D	De	PDe	P	P	PDE	PDE	PDE	PDE	PDE	Pde		
hdpco	PDE	PD	PD	D	De	PDe	P	P	PDE	PDE	PDE	PDE	PDE	Pde		

Australian Government Pensions

Table 3.1 provides the average standardized scores for Australian Government pensions and Table 3.2 shows which methods are significantly different from each other.

The imputation methods for cross-sectional estimates of Government pensions for respondents can only be distinguished on their predictive or estimation accuracy. The longitudinal NNRM is the poorest performer and the population and random carryover methods that use the cross-sectional hotdeck method as their fallback method are equally the best. From the middle of the pack, the Little and Su method with imputation classes (with the nearest neighbour regression method fallback option) is slightly ahead of the other Little and Su methods.

When considering the estimates of change between waves, we find that the carryover methods perform surprisingly well on predictive accuracy but are very poor in terms of distributional

accuracy. The Little and Su methods offer a reasonable compromise in the trade-off between predictive and distributional accuracy. While the longitudinal hotdeck method is better than the longitudinal NNRM, the other methods are indistinguishable between which fallback method they adopt.

For non-respondents, the methods generally perform very different from each other on all three broad areas measured. The methods using the cross-section NNRM as their fallback option are substantially better than those using the fallback hotdeck method. The carryover methods perform the best cross-sectionally, but the Little and Su methods are better longitudinally. Assuming predictive accuracy is a little less important than distributional and estimation accuracy, the basic Little and Su and the one using imputation classes perform equally well and offer the best compromise on estimates of change.

Table 3.1: Average value of standardised predictive (P), distributional (D) and estimation (E) evaluation measures for Australian Government pensions

	Cross-sectional				*Longitudinal*			
Base method: Longitudinal method	P	D	E	Ave	P	D	E	Ave
Respondents								
NNRM: NNRM Longitudinal	-0.01	0.20	-0.31	-0.04	0.03	0.98	-0.28	0.25
NNRM: Basic Little & Su	-0.09	0.16	-0.34	-0.09	-0.21	0.38	-0.28	-0.04
NNRM: Little & Su w Imp Classes	-0.15	**0.13**	-0.34	-0.12	-0.21	0.31	-0.28	-0.06
NNRM: Little & Su Key Var	-0.09	0.16	-0.34	-0.09	-0.21	0.38	-0.28	-0.04
NNRM: Little & Su Distance	-0.07	0.16	-0.34	-0.08	-0.21	0.38	-0.28	-0.04
NNRM: LVCF	-0.28	0.29	-0.34	-0.11	-0.47	3.09	-0.24	0.79
NNRM: Random Carryover	**-0.41**	0.14	-0.37	**-0.21**	-0.48	1.88	-0.29	0.37
NNRM: Population Carryover	-0.41	0.14	**-0.37**	-0.21	**-0.48**	1.87	-0.29	0.37
HD: Hotdeck Longitudinal	-0.07	0.24	-0.33	-0.06	-0.24	0.55	**-0.29**	0.00
HD: Basic Little & Su	-0.05	0.18	-0.34	-0.07	-0.18	0.41	-0.28	-0.01
HD: Little & Su w Imp Classes	-0.12	0.14	-0.34	-0.11	-0.20	**0.28**	-0.28	**-0.07**
HD: Little & Su Key Var	-0.05	0.18	-0.34	-0.07	-0.18	0.41	-0.28	-0.01
HD: Little & Su Distance	-0.03	0.18	-0.34	-0.06	-0.18	0.41	-0.28	-0.01
HD: LVCF	-0.16	0.27	-0.35	-0.08	-0.39	3.10	-0.24	0.82
HD: Random Carryover	-0.33	0.16	-0.37	-0.18	-0.44	1.87	-0.29	0.38
HD: Population Carryover	-0.33	0.16	-0.37	-0.18	-0.44	1.86	-0.29	0.38
Non-respondents								
NNRM: NNRM Longitudinal	0.08	-0.16	-0.48	-0.18	0.11	1.59	-0.39	0.43
NNRM: Basic Little & Su	-0.23	0.54	-0.46	-0.05	-0.15	0.81	**-0.43**	0.08
NNRM: Little & Su w Imp Classes	-0.25	0.47	-0.47	-0.08	-0.17	0.79	-0.43	0.06
NNRM: Little & Su Key Var	-0.39	1.32	-0.38	0.18	**-0.53**	1.17	-0.37	0.09
NNRM: Little & Su Distance	**-0.40**	1.32	-0.38	0.18	-0.52	1.03	-0.37	0.05
NNRM: LVCF	-0.07	-0.30	-0.50	-0.29	-0.11	0.62	-0.42	0.03
NNRM: Random Carryover	-0.33	**-0.34**	**-0.50**	**-0.39**	-0.30	0.40	-0.42	-0.11
NNRM: Population Carryover	-0.33	-0.34	**-0.50**	-0.39	-0.30	0.40	-0.42	-0.11
HD: Hotdeck Longitudinal	0.19	-0.25	-0.49	-0.18	0.06	1.32	-0.42	0.32
HD: Basic Little & Su	0.13	0.90	-0.36	0.23	0.10	1.31	-0.37	0.35
HD: Little & Su w Imp Classes	0.13	0.86	-0.37	0.21	0.09	1.26	-0.37	0.32
HD: Little & Su Key Var	0.20	0.05	-0.44	-0.06	0.06	0.80	-0.41	0.15
HD: Little & Su Distance	0.19	0.05	-0.44	-0.07	0.06	0.87	-0.41	0.17
HD: LVCF	0.10	-0.16	-0.47	-0.18	-0.09	0.66	-0.43	0.05
HD: Random Carryover	-0.21	-0.25	-0.48	-0.32	-0.28	**0.35**	-0.43	**-0.12**
HD: Population Carryover	-0.21	-0.26	-0.48	-0.32	-0.28	0.35	-0.43	-0.12

Compared to wages and salaries, it is harder to impute Government pensions well. On both predictive accuracy and distributional accuracy, the standardized scores for Government pensions fall short of the wages and salaries experience. In terms of estimation accuracy, however, the methods perform well for benefit income, presumably because benefit income is not as skewed as wages and salary income.

For Australian Government pensions overall, the Little and Su method with imputation classes provides a reasonably good imputation solution across both respondents and non-respondents and for cross-sectional estimates and estimates of change.

Table 3.2: Significant differences between predictive (P), distributional (D) and estimation (E) evaluation measures for Australian Government pensions

Respondents	nnrml	ls	lsc	lsk	lsd	lvcf	rco	pco	hdl	hdls	hdlsc	hdlsk	hdlsd	hdlvcf	hdrco	hdpco
nnrml		PD	PD	PD	PD	PDe	PD	PD	PD	PD	PD	PD	PD	PDe	PD	PD
ls	pe					PDe	PD	PD						PDe	PD	PD
lsc	Pe					PDe	PD	PD	D					PD	PD	PD
lsk	pe					PDe	PD	PD						PDe	PD	PD
lsd	e		p			PDe	PD	PD						PDe	PD	PD
lvcf	PE	P	Pd	P	P		DE	DE	PDE	PD	PDe	PD	PD		DE	DE
rco	PE	PE	PE	PE	PE	Pde			PD	PD	PD	PD	PD	DE		
pco	PE	PE	PE	PE	PE	Pde			PD	PD	PD	PD	PD	De		
hdl			p			P	PE	PE			D			PDE	PD	PD
hdls	e		P			P	PE	PE						PD	PD	PD
hdlsc	PE					Pd	PE	PE						PDe	PD	PD
hdlsk	e		P			P	PE	PE						PD	PD	PD
hdlsd	e		P			P	PE	PE			p			PD	PD	PD
hdlvcf	PE		d		p	P	Pde	Pde		p		p	P		De	De
hdrco	PE	PE	PE	PE	PE	e	p	p	PE	PE	Pe	PE	PE	Pe		
hdpco	PE	PE	PE	PE	PE	de	p	p	PE	PE	Pe	PE	PE	Pe		

Non-respondents	nnrml	ls	lsc	lsk	lsd	lvcf	rco	pco	hdl	hdls	hdlsc	hdlsk	hdlsd	hdlvcf	hdrco	hdpco
nnrml		PDE	PDE	PDe	PDe	PDE	PDE	PDE	De	De	De	D	D	PDE	PDE	PDE
ls	PDE			PDE	PDE	D	PD	PD	PD	PDE	PDE	Pe	Pe	d	PD	PD
lsc	PD			PDE	PDE	D	PD	PD	PD	PDE	PDE	P	P	pd	PD	PD
lsk	PDE	PDE	PDE			PDE	PDE	PDE	PdE	P	P	PDE	PDE	PDE	PDE	PDE
lsd	PDE	PDE	PDE			PDE	PDE	PDE	PDE	PD	PD	PDE	PdE	PDE	PDE	PDE
lvcf	PDE	PDE	PDE	PDE	PDE		PD	PD	PD	PDE	PDE	PD	PD		PD	PD
rco	PDE	PDE	PDE	PDE	PDE	P			PD	PDE	PDE	PD	PD	PD		
pco	PDE	PDE	PDE	PDE	PDE	P			PD	PDE	PDE	PD	PD	PD		
hdl	Pd	PDE	PDE	PDE	PDE	Pe	PdE	PdE		E	E	D	D	PD	PD	PD
hdls	pDE	PDE	PDE	PDE	PDE	PDE	PDE	PDE	PDE			DE	DE	PDE	PDE	PDE
hdlsc	PDE	PDE	PDE	PD	PD	PDE	PDE	PDE	PDE			DE	DE	PDE	PDE	PDE
hdlsk	PDE	PDE	PDE	PDE	PDE	PDE	PDE	PDE	DE	PDE	PDE			Pd	PD	PD
hdlsd	PDE	PDE	PDE	PDE	PDE	PDE	PDE	PDE	DE	PDE	PDE			PD	PD	PD
hdlvcf		PDe	PD	PDE	PDE	PDE	PDE	PDE	Pde	DE	DE	PDE	PDE		PD	PD
hdrco	Pd	DE	DE	PDE	PDE	PE	PdE	PdE	P	PDE	PDE	PDE	PDE	Pd		
hdpco	Pd	DE	DE	PDE	PDE	PE	PdE	PdE	P	PDE	PDE	PDE	PDE	Pd		

Business Income

Table 4.1 provides the average standardized scores for business income and Table 4.2 shows which methods are significantly different from each other.

In the first table we notice that the standardized scores for business income of respondents are generally higher than those for wages and salaries (Table 2.1) and benefits (Table 3.1) for predictive accuracy and estimation accuracy, but tends to fall between the two for distributional accuracy. As business income is highly skewed it is understandable that the methods do not perform well on estimation accuracy. For non-respondents, however, the methods tend to perform well distributionally for business income compared to the previous two variables considered, but are not as good on predictive and estimation accuracy. The improvement with respect to distributional accuracy is most likely due to the much higher proportion of zero values in the business income distribution (that is it is easier to impute zeros as most non-respondents would have zero business income than it is to impute the value of the business income when it is known to be non-zero as in the case of respondents).

For respondents, the basic Little and Su method performs well both cross-sectionally and longitudinally.[16] The Little and Su methods using the NNRM fallback option do not perform any better or worse than those using the hotdeck fallback option. The carryover methods, however, are better using the hotdeck fallback option. The random carryover method and the population carryover method perform better than the basic Little and Su method on distributional accuracy for cross-sectional estimates but are worse for the estimates of change. The longitudinal NNRM is the worst performer in imputing business income out of all of the methods tested.

For non-respondents, the population carryover or random carryover methods stand out as the better performers, primarily because of better predictive accuracy. These methods are also better for distributional accuracy for estimates of change compared to all other methods and for level estimates compared to the performance of some methods (the longitudinal NNRM, the last value carried forward method – irrespective of the fallback method used – and the longitudinal hotdeck method).

Overall, for business income, the basic Little and Su method was the best for respondents and the population or random carryover method was better for non-respondents.

[16] Note that the Little and Su method using a key variable or a distance function to find a suitable donor mirrors the performance of the Basic Little and Su method for business income and private transfers because these methods do not operate any differently from the basic Little and Su method for these variables.

Table 4.1: Average value of standardised predictive (P), distributional (D) and estimation (E) evaluation measures for business income

	Cross-sectional				*Longitudinal*			
Base method: Longitudinal method	P	D	E	Ave	P	D	E	Ave
Respondents								
NNRM: NNRM Longitudinal	0.21	0.25	0.51	0.33	0.80	0.40	0.64	0.61
NNRM: Basic Little & Su	-0.12	0.03	**0.21**	0.04	0.43	-0.03	0.29	0.23
NNRM: Little & Su w Imp Classes	0.03	-0.05	1.17	0.38	0.55	**-0.12**	0.86	0.43
NNRM: Little & Su Key Var	-0.12	0.03	**0.21**	0.04	0.43	-0.03	0.29	0.23
NNRM: Little & Su Distance	-0.12	0.03	**0.21**	0.04	0.43	-0.03	0.29	0.23
NNRM: LVCF	0.14	-0.08	0.44	0.17	0.46	0.42	0.46	0.45
NNRM: Random Carryover	0.03	-0.19	0.39	0.08	0.42	0.24	0.44	0.37
NNRM: Population Carryover	0.03	-0.19	0.39	0.07	0.42	0.24	0.44	0.37
HD: Hotdeck Longitudinal	0.17	-0.02	0.27	0.14	0.68	0.31	0.26	0.42
HD: Basic Little & Su	**-0.17**	0.08	0.21	0.04	0.35	-0.03	**0.25**	**0.19**
HD: Little & Su w Imp Classes	0.02	-0.03	1.19	0.39	0.49	-0.11	0.89	0.42
HD: Little & Su Key Var	**-0.17**	0.08	0.21	0.04	0.35	-0.03	**0.25**	**0.19**
HD: Little & Su Distance	**-0.17**	0.08	0.21	0.04	0.35	-0.03	**0.25**	**0.19**
HD: LVCF	0.02	-0.19	0.29	0.04	0.34	0.31	0.29	0.31
HD: Random Carryover	-0.08	-0.24	0.26	-0.02	**0.31**	0.17	0.28	0.25
HD: Population Carryover	-0.08	**-0.24**	0.26	**-0.02**	0.31	0.17	0.28	0.25
Non-respondents								
NNRM: NNRM Longitudinal	0.73	-0.71	0.22	0.08	0.37	-0.63	**0.26**	0.00
NNRM: Basic Little & Su	0.35	-0.78	**0.12**	-0.10	0.40	-0.60	0.31	0.04
NNRM: Little & Su w Imp Classes	0.41	-0.76	0.20	-0.05	0.41	-0.61	0.57	0.13
NNRM: Little & Su Key Var	0.35	-0.78	**0.12**	-0.10	0.40	-0.60	0.31	0.04
NNRM: Little & Su Distance	0.35	-0.78	**0.12**	-0.10	0.40	-0.60	0.31	0.04
NNRM: LVCF	0.49	-0.74	0.24	0.00	0.32	-0.68	0.31	-0.01
NNRM: Random Carryover	**0.02**	-0.78	0.14	-0.20	0.18	-0.71	0.30	-0.08
NNRM: Population Carryover	0.02	-0.78	0.14	**-0.20**	**0.18**	-0.71	0.30	-0.08
HD: Hotdeck Longitudinal	0.73	-0.72	0.32	0.11	0.36	-0.63	0.31	0.01
HD: Basic Little & Su	0.45	-0.78	0.23	-0.03	0.53	-0.52	0.38	0.13
HD: Little & Su w Imp Classes	0.53	-0.77	0.31	0.02	0.58	-0.53	0.68	0.24
HD: Little & Su Key Var	0.45	-0.78	0.23	-0.03	0.53	-0.52	0.38	0.13
HD: Little & Su Distance	0.45	-0.78	0.23	-0.03	0.53	-0.52	0.38	0.13
HD: LVCF	0.51	-0.75	0.26	0.01	0.31	-0.68	0.27	-0.03
HD: Random Carryover	0.05	-0.79	0.17	-0.19	0.22	**-0.73**	0.26	-0.08
HD: Population Carryover	0.05	**-0.79**	0.17	-0.19	0.22	**-0.73**	0.26	**-0.08**

Table 4.2 Significant differences between predictive (P), distributional (D) and estimation (E) evaluation measures for business income

Respondents	nnrml	ls	lsc	lsk	lsd	lvcf	rco	pco	hdl	hdls	hdlsc	hdlsk	hdlsd	hdlvcf	hdrco	hdpco
nnrml		PDE	PD	PDE	PDE	Pe	PDe	PDe	E	PDE	PD	PDE	PDE	PE	PDE	PDE
ls	PDE		e			De	De	De	PD		e			D	D	D
lsc	PDe	pE		e	e	D	D	D	De	pe		pe	pe	pDe	PDe	PDe
lsk	PDE		pE			De	De	De	PD		e			D	D	D
lsd	PDE		pE			De	De	De	PD		e			D	D	D
lvcf	D	PdE	e	PdE	PdE		D	D	PE	DE	D	DE	DE	e	pDe	pDe
rco	PD	pDe	De	pDe	pDe	d			Pe	De	D	De	De	e	e	e
pco	PD	pDe	De	pDe	pDe	d			Pe	De	D	De	De	e	e	e
hdl	DE	P	pE	P	P	e	PD	PD		PD	pDe	PD	PD	P	PD	PD
hdls	PDE		PdE			PDE	PDe	PDe	P		e			D	D	D
hdlsc	PDe	pE		pE	pE	pe	De	De	PE	PdE		e	e	De	pDe	pDe
hdlsk	PDE		PdE			PDE	PDe	PDe	P		PdE			D	D	D
hdlsd	PDE		PdE			PDE	PDe	PDe	P		PdE			D	D	D
hdlvcf	PDE	PD	DE	PD	PD	pd			PD	PD	DE	PD	PD		D	D
hdrco	PDE	D	DE	D	D	PDe	p	p	PD	D	DE	D	D	p		
hdpco	PDE	D	DE	D	D	PDe	p	p	PD	D	DE	D	D	p		

Non-respondents	nnrml	ls	lsc	lsk	lsd	lvcf	rco	pco	hdl	hdls	hdlsc	hdlsk	hdlsd	hdlvcf	hdrco	hdpco
nnrml							pD	pD		pD	pD	pD	pD	d	pD	pD
ls	PD					D	PD	PD		d	pd	d	d	D	pD	pD
lsc	Pd					D	PD	PD		D	D	D	D	D	pD	pD
lsk	PD					D	PD	PD		d	pd	d	d	D	pD	pD
lsd	PD					D	PD	PD		d	pd	d	d	D	pD	pD
lvcf	P	P		P	P					PD	PD	PD	PD		D	D
rco	PD	P	P	P	P	Pd			pD	PD	PD	PD	PD			
pco	PD	P	P	P	P	Pd			pD	PD	PD	PD	PD			
hdl		PDe	P	PDe	PDe	P	PD	PD		pD	PD	pD	pD	d	D	D
hdls	PD					d	P	P	PD					PD	PD	PD
hdlsc	PD	P	p	P	P		P	P	Pd					PD	PD	PD
hdlsk	PD					d	P	P	PD					PD	PD	PD
hdlsd	PD					d	P	P	PD					PD	PD	PD
hdlvcf	Pd	P		P	P		P	P	P						D	D
hdrco	PD	P	P	P	P	PD			PD	P	P	P	P	Pd		
hdpco	PD	P	P	P	P	PD			PD	P	P	P	P	Pd		

Interest Income

Table 5.1 provides the average standardized scores for interest income and Table 5.2 shows which methods are significantly different from each other.

For interest income of respondents, the Little and Su methods perform better than the carryover methods on predictive accuracy of both the level estimates and estimates of change and on the distributional accuracy of the estimates of change, but not on the distributional accuracy of level estimates. There is no difference in performance between the four different Little and Su methods tested. The methods using the NNRM fallback method perform better than those using the hotdeck fallback method. Neither the longitudinal methods (NNRM or hotdeck) perform well.

For non-respondents, the random or population carry over methods using either fallback method perform the best, always on distributional grounds but sometimes on predictive accuracy as well.

Table 5.1: Average value of standardised predictive (P), distributional (D) and estimation (E) evaluation measures for interest income

	Cross-sectional				*Longitudinal*			
Base method: Longitudinal method	P	D	E	Ave	P	D	E	Ave
Respondents								
NNRM: NNRM Longitudinal	0.18	-0.50	-0.02	-0.11	0.87	-0.33	0.08	0.21
NNRM: Basic Little & Su	-0.18	-0.54	**-0.15**	-0.29	**0.26**	**-0.64**	**0.02**	**-0.12**
NNRM: Little & Su w Imp Classes	-0.17	-0.53	-0.11	-0.27	0.27	-0.61	0.07	-0.09
NNRM: Little & Su Key Var	-0.17	-0.54	-0.15	-0.29	0.26	-0.64	0.02	-0.12
NNRM: Little & Su Distance	-0.18	-0.55	-0.15	**-0.29**	0.26	-0.64	0.02	-0.12
NNRM: LVCF	0.13	-0.61	-0.05	-0.18	0.57	0.12	0.11	0.27
NNRM: Random Carryover	-0.03	**-0.69**	-0.10	-0.27	0.49	-0.16	0.07	0.13
NNRM: Population Carryover	-0.03	-0.69	-0.10	-0.27	0.49	-0.16	0.07	0.14
HD: Hotdeck Longitudinal	0.25	-0.23	0.00	0.01	0.97	-0.15	0.10	0.31
HD: Basic Little & Su	-0.20	-0.49	-0.13	-0.27	0.27	-0.62	0.02	-0.11
HD: Little & Su w Imp Classes	-0.16	-0.47	-0.09	-0.24	0.30	-0.59	0.11	-0.06
HD: Little & Su Key Var	-0.19	-0.48	-0.13	-0.27	0.27	-0.63	0.02	-0.11
HD: Little & Su Distance	**-0.20**	-0.48	-0.13	-0.27	0.27	-0.62	0.02	-0.11
HD: LVCF	0.16	-0.50	0.00	-0.11	0.63	0.10	0.22	0.32
HD: Random Carryover	-0.01	-0.57	-0.05	-0.21	0.54	-0.18	0.16	0.17
HD: Population Carryover	-0.01	-0.57	-0.05	-0.21	0.54	-0.18	0.16	0.17
Non-respondents								
NNRM: NNRM Longitudinal	0.21	-0.43	-0.01	-0.08	0.25	0.28	0.02	0.18
NNRM: Basic Little & Su	-0.05	-0.05	-0.16	-0.09	0.15	0.32	-0.07	0.13
NNRM: Little & Su w Imp Classes	-0.04	-0.01	-0.12	-0.06	0.16	0.30	-0.06	0.13
NNRM: Little & Su Key Var	-0.07	-0.05	**-0.17**	-0.09	0.14	0.33	**-0.09**	0.13
NNRM: Little & Su Distance	-0.07	-0.20	-0.15	-0.14	0.10	0.18	-0.07	0.07
NNRM: LVCF	0.05	-0.47	-0.03	-0.15	0.17	-0.14	0.04	0.02
NNRM: Random Carryover	-0.12	-0.55	-0.11	-0.26	**0.04**	-0.28	0.00	**-0.08**
NNRM: Population Carryover	**-0.12**	-0.55	-0.11	**-0.26**	0.04	**-0.28**	0.00	-0.08
HD: Hotdeck Longitudinal	0.19	-0.44	-0.03	-0.09	0.29	0.28	0.04	0.20
HD: Basic Little & Su	0.06	-0.17	-0.10	-0.07	0.28	0.45	0.03	0.25
HD: Little & Su w Imp Classes	0.09	-0.03	-0.09	-0.01	0.32	0.55	0.02	0.30
HD: Little & Su Key Var	0.06	-0.16	-0.11	-0.07	0.28	0.44	0.03	0.25
HD: Little & Su Distance	0.06	-0.19	-0.10	-0.08	0.25	0.43	0.03	0.24
HD: LVCF	0.09	-0.49	-0.04	-0.15	0.25	-0.10	0.10	0.08
HD: Random Carryover	-0.08	**-0.55**	-0.10	-0.25	0.14	-0.25	0.04	-0.02
HD: Population Carryover	-0.08	-0.55	-0.10	-0.25	0.15	-0.25	0.04	-0.02

Table 5.2: Significant differences between predictive (P), distributional (D) and estimation (E) evaluation measures for interest income

Respondents	nnrml	ls	lsc	lsk	lsd	lvcf	rco	pco	hdl	hdls	hdlsc	hdlsk	hdlsd	hdlvcf	hdrco	hdpco
nnrml		PD	PD	PD	PD	PD	PD	PD	D	PD	PD	PD	PD	PD	Pd	Pd
ls	Pe					PD	PD	PD	PD					PDE	PDe	PDe
lsc	P					PD	PD	PD	PD					PDe	PD	PD
lsk	Pe					PD	PD	PD	PD					PDE	PDe	PDe
lsd	Pe					PD	PD	PD	PD					PDE	PDe	PDe
lvcf	D	Pe	Pd	P	P		D	D	PD	PD	PD	PD	PD		D	D
rco	PD	PD	PD	pD	PD	P			P	PD	PD	PD	PD	PDe		
pco	PD	PD	PD	pD	PD	P			P	PD	PD	PD	PD	PDe		
hdl	D	PDE	PD	PDE	PDE	pD	PD	PD		PD	PD	PD	PD	PD	P	P
hdls	Pe					PD	PD	PD	PDe					PDE	PDe	PDe
hdlsc	P				d	PD	pD	pD	PD					PD	PD	PD
hdlsk	Pe					PD	PD	PD	PDe					PDE	PDe	PDe
hdlsd	Pe					PD	PD	PD	PDe					PDE	PDe	PDe
hdlvcf		PE	P	PE	PE	d	PD	PD	D	Pe	P	Pe	Pe		D	D
hdrco	P	Pe	P	Pe	Pe	P	D	D	PD	Pd	Pd	Pd	Pd	P		
hdpco	P	Pe	P	Pe	Pe	P	D	D	PD	P	Pd	Pd	Pd	P		

Non-respondents	nnrml	ls	lsc	lsk	lsd	lvcf	rco	pco	hdl	hdls	hdlsc	hdlsk	hdlsd	hdlvcf	hdrco	hdpco
nnrml				e	p	D	PD	PD		D	D	D	D	D	D	D
ls	PDE				d	De	D	D	pe	pde	PD	pe	e	DE	De	De
lsc	PDe				d	D	D	D		d	pD	d	d	DE	De	De
lsk	PDE				d	De	D	D	pe	pe	PDe	pe	e	DE	De	De
lsd	PDE	D	D	D		De	D	D	Pe	PDe	PD	PDe	pDe	PDE	De	De
lvcf	P	pDe	pD	PDE	PDe		pD	pD	D	D	pD	D	D		D	D
rco	PD	D	pD	D	D	PD			PD	PD	PD	PD	PD	PD		
pco	PD	D	pD	D	D	PD			PD	PD	PD	PD	PD	PD		
hdl		PDE	PD	PDE	PDe	P	PD	PD		D	D	D	d	D	pD	pD
hdls	PD	pd	pD	Pd	P	D	PD	PD	PD					D	pD	pD
hdlsc	PD	P	P	P	PD	D	PD	PD	PD	D			d	D	PD	PD
hdlsk	PD	pd	pD	Pd	P	D	PD	PD	PD		D			D	pD	pD
hdlsd	PD	PD	pD	PD	P	D	PD	PD	PD		D			D	D	D
hdlvcf	P	PDe	PD	PDe	PDe		Pd	Pd	P	D	D	D	D		D	D
hdrco	PD	D	D	D	D	PD			PD	PD	PD	PD	PD	Pd		
hdpco	PD	D	D	D	D	PD			PD	PD	PD	PD	PD	Pd		

Dividends and Royalties

Table 6.1 provides the average standardized scores for dividends and royalties and Table 6.2 shows which methods are significantly different from each other. The differences between the methods for dividends and royalties come from the measures for predictive and distributional accuracy as estimation accuracy is rarely different between the methods.

While the Little and Su methods using the hotdeck fallback option for level estimates for the respondents are better on predictive accuracy, the Little and Su methods using the NNRM fallback option are better on distributional accuracy. The performance of the carryover methods is reasonably close to the Little and Su methods but the longitudinal NNRM and the longitudinal hotdeck method are both poor performers.

When considering estimates of change for respondents, the Little and Su using imputation classes and the hotdeck fallback option performs the best, being a small margin ahead of the other Little and Su methods on predictive accuracy. The carryover methods are poor at achieving distributional accuracy and the longitudinal NNRM and the longitudinal hotdeck methods are both poor at predictive accuracy.

For non-respondents, the population and random carryover methods perform better than the rest and there is no difference between these methods based on which fallback option is chosen.

Overall for dividends and royalties, a Little and Su method would be suitable for respondents. We select the basic Little and Su method using the NNRM fallback option because the extra complexity of the other methods is not justified and the NNRM fallback method is the most likely choice for other variables. For non-respondents, a carryover method would be suitable.

Table 6.1: Average value of standardised predictive (P), distributional (D) and estimation (E) evaluation measures for dividends and royalties

	Cross-sectional				*Longitudinal*			
Base method: Longitudinal method	P	D	E	Ave	P	D	E	Ave
Respondents								
NNRM: NNRM Longitudinal	0.39	-0.48	0.05	-0.01	1.52	0.05	0.40	0.66
NNRM: Basic Little & Su	-0.04	**-0.67**	0.00	-0.23	0.73	-0.61	**0.26**	0.12
NNRM: Little & Su w Imp Classes	-0.04	-0.65	0.01	-0.23	0.67	-0.63	0.32	0.12
NNRM: Little & Su Key Var	-0.03	**-0.67**	0.00	-0.23	0.73	-0.61	**0.26**	0.12
NNRM: Little & Su Distance	-0.04	**-0.67**	0.00	-0.23	0.73	-0.61	**0.26**	0.12
NNRM: LVCF	0.14	-0.51	0.02	-0.12	0.84	0.27	0.37	0.50
NNRM: Random Carryover	0.05	-0.63	-0.01	-0.20	0.77	-0.05	0.36	0.36
NNRM: Population Carryover	0.05	-0.63	-0.01	-0.20	0.77	-0.04	0.36	0.36
HD: Hotdeck Longitudinal	0.12	-0.07	0.11	0.05	1.13	-0.02	0.37	0.49
HD: Basic Little & Su	-0.18	-0.59	**-0.02**	-0.26	0.71	-0.61	0.27	0.12
HD: Little & Su w Imp Classes	-0.15	-0.57	0.02	-0.24	**0.60**	**-0.63**	0.32	**0.10**
HD: Little & Su Key Var	-0.17	-0.59	**-0.02**	-0.26	0.71	-0.61	0.27	0.12
HD: Little & Su Distance	**-0.18**	-0.59	**-0.02**	**-0.26**	0.71	-0.61	0.27	0.12
HD: LVCF	0.01	-0.20	0.06	-0.05	0.68	0.24	0.36	0.42
HD: Random Carryover	-0.09	-0.43	0.01	-0.17	0.61	-0.07	0.35	0.30
HD: Population Carryover	-0.09	-0.43	0.01	-0.17	0.61	-0.07	0.35	0.30
Non-respondents								
NNRM: NNRM Longitudinal	0.27	-0.39	0.02	-0.03	0.18	0.76	0.12	0.35
NNRM: Basic Little & Su	-0.04	0.39	**-0.01**	0.11	0.16	0.84	0.14	0.38
NNRM: Little & Su w Imp Classes	-0.04	0.51	0.01	0.16	0.24	0.81	0.14	0.40
NNRM: Little & Su Key Var	-0.08	-0.19	0.00	-0.09	0.08	0.28	0.12	0.16
NNRM: Little & Su Distance	-0.08	-0.19	0.00	-0.09	0.08	0.29	0.12	0.17
NNRM: LVCF	0.10	-0.38	0.04	-0.08	0.05	0.05	0.19	0.10
NNRM: Random Carryover	-0.11	-0.46	0.00	**-0.19**	**-0.08**	-0.15	0.16	-0.02
NNRM: Population Carryover	-0.11	-0.46	0.00	-0.19	-0.08	-0.15	0.16	-0.02
HD: Hotdeck Longitudinal	0.25	-0.33	0.08	0.00	0.22	0.93	0.16	0.44
HD: Basic Little & Su	0.03	0.07	0.01	0.04	0.26	0.89	0.15	0.44
HD: Little & Su w Imp Classes	0.04	0.27	0.06	0.12	0.35	1.06	0.16	0.52
HD: Little & Su Key Var	0.03	-0.06	0.01	-0.01	0.27	0.81	0.15	0.41
HD: Little & Su Distance	0.03	-0.06	0.01	-0.01	0.27	0.81	0.15	0.41
HD: LVCF	0.12	-0.43	0.04	-0.09	0.06	-0.01	0.14	0.06
HD: Random Carryover	**-0.11**	**-0.48**	0.06	-0.18	-0.04	**-0.19**	**0.12**	**-0.04**
HD: Population Carryover	-0.11	-0.48	0.06	-0.18	-0.04	-0.19	0.12	-0.04

Table 6.2: Significant differences between predictive (P), distributional (D) and estimation (E) evaluation measures for dividends and royalties

Respondents	nnrml	ls	lsc	lsk	lsd	lvcf	rco	pco	hdl	hdls	hdlsc	hdlsk	hdlsd	hdlvcf	hdrco	hdpco
nnrml		PDe	PD	PDe	PDe	PD	P	P	P	PDe	PD	PDe	PDe	PD	Pd	Pd
ls	PD					D	D	D	PD					D	D	D
lsc	PD					PD	D	D	PD					D	D	D
lsk	PD					D	D	D	PD					D	D	D
lsd	PD					D	D	D	PD					D	D	D
lvcf	P	PD	PD	PD	PD		D	D	PD	pD	PD	pD	pD	P	PD	PD
rco	PD					d			P	D	PD	D	D	D	P	P
pco	PD					d			P	D	PD	D	D	D	P	P
hdl	PD	PDe	PDe	PDe	PDe	De	DE	DE		PD	PD	PD	PD	PD	P	P
hdls	Pde	Pd	P	Pd	Pd	P	P	P	PDE					D	D	D
hdlsc	Pd	pD	pd	pD	pD	P	P	P	PDe					D	D	D
hdlsk	Pde	pd	p	Pd	pd	P	P	P	PDE					D	D	D
hdlsd	Pde	Pd	P	Pd	Pd	P	P	P	PDE					D	D	D
hdlvcf	PD	D	D	D	D	PD	D	D	pd	PDe	PD	PDe	PDe		D	D
hdrco	P	D	D	D	D	P	pD	pD	PDE	D	D	D	D	D		
hdpco	P	D	D	D	D	P	pD	pD	PDE	D	D	D	D	D		

Non-respondents	nnrml	ls	lsc	lsk	lsd	lvcf	rco	pco	hdl	hdls	hdlsc	hdlsk	hdlsd	hdlvcf	hdrco	hdpco
nnrml				D	D	D	PD	PD	D	d	pD			D	PD	PD
ls	PD			D	D	D	PD	PD			pD			D	PD	PD
lsc	PD			D	D	pD	PD	PD	d		D			pD	PD	PD
lsk	PD	D	D			D	pD	pD	D	pD	PD	pD	pD	D	D	D
lsd	PD	D	D			D	pD	pD	D	pD	PD	pD	pD	D	D	D
lvcf	P	PD	PD	PD	PD		D	D	pD	PD	PD	PD	PD		D	D
rco	Pd	D	D	D	D	Pd			PD	PD	PD	PD	PD	D		
pco	P	D	D	D	D	Pd			PD	PD	PD	PD	PD	D		
hdl		PD	PD	PD	PD	P	PD	PD			d	d	d	pD	PD	PD
hdls	PD	D	D	PD	PD	D	PD	PD	PD		d			pD	PD	PD
hdlsc	PD	p	pD	PD	PD	D	PD	PD	PD	D		D	D	PD	PD	PD
hdlsk	PD	D	D	Pd	Pd	D	PD	PD	PD	d	D			PD	PD	PD
hdlsd	PD	D	D	pd	Pd	D	PD	PD	PD	d	D			PD	PD	PD
hdlvcf	P	PD	PD	PD	PD		P	P	PD	pD	D	pD	pD		D	D
hdrco	PD	D	D	D	D	PD			PD	PD	PD	PD	PD	P		
hdpco	PD	D	D	D	D	PD			PD	PD	PD	PD	PD	P		

Rental Income

Table 7.1 provides the average standardized scores for rental income and Table 7.2 shows which methods are significantly different from each other.

None of the methods tested for rental income of respondents stand out. The basic Little and Su method performs well in terms of predictive and estimation accuracy for cross-sectional estimates for respondents, but is poor on distributional accuracy. For estimates of change, this method performs well (even distributionally). The population or carryover methods perform reasonably well cross-sectionally, but as occurs with the other variables, the distributional accuracy for estimates of change are quite poor. The methods using the hotdeck fallback option perform better than those using the NNRM fallback option in terms of distributional accuracy. The longitudinal NNRM and

the longitudinal hotdeck both perform reasonably well for cross-sectional estimates but are very poor when it comes to distributional accuracy of estimates of change.

For non-respondents, the population or random carryover methods perform well overall, both cross-sectionally and longitudinally.

Table 7.1: Average value of standardised predictive (P), distributional (D) and estimation (E) evaluation measures for rental income

	Cross-sectional				*Longitudinal*			
Base method: Longitudinal method	P	D	E	Ave	P	D	E	Ave
Respondents								
NNRM: NNRM Longitudinal	0.02	0.22	-0.16	0.03	0.37	0.84	-0.13	0.36
NNRM: Basic Little & Su	**-0.28**	0.68	-0.18	0.07	**0.04**	0.42	-0.16	0.10
NNRM: Little & Su w Imp Classes	-0.13	0.65	0.01	0.18	0.24	0.45	-0.04	0.22
NNRM: Little & Su Key Var	-0.26	0.62	-0.11	0.08	0.12	0.45	-0.06	0.17
NNRM: Little & Su Distance	-0.27	0.66	-0.18	0.07	0.07	0.44	**-0.16**	0.12
NNRM: LVCF	-0.10	0.18	-0.16	-0.03	0.28	0.64	-0.10	0.27
NNRM: Random Carryover	-0.14	0.23	-0.14	-0.02	0.26	0.66	-0.10	0.28
NNRM: Population Carryover	-0.14	0.23	-0.14	-0.02	0.26	0.66	-0.10	0.28
HD: Hotdeck Longitudinal	-0.06	0.34	-0.18	0.04	0.37	0.75	-0.14	0.33
HD: Basic Little & Su	-0.27	0.41	**-0.18**	-0.01	0.08	**0.32**	-0.12	**0.09**
HD: Little & Su w Imp Classes	-0.14	0.46	-0.05	0.09	0.15	0.38	-0.07	0.15
HD: Little & Su Key Var	-0.25	0.40	-0.10	0.02	0.16	0.32	-0.03	0.15
HD: Little & Su Distance	-0.26	0.41	-0.18	-0.01	0.09	0.33	-0.12	0.10
HD: LVCF	-0.15	0.13	-0.13	-0.05	0.29	0.56	-0.05	0.27
HD: Random Carryover	-0.21	**0.10**	-0.12	-0.08	0.27	0.54	-0.05	0.26
HD: Population Carryover	-0.21	**0.10**	-0.12	**-0.08**	0.27	0.54	-0.05	0.26
Non-respondents								
NNRM: NNRM Longitudinal	0.22	-0.77	-0.01	-0.19	0.23	-0.50	-0.09	-0.12
NNRM: Basic Little & Su	0.06	-0.76	**-0.11**	-0.27	0.25	-0.45	-0.09	-0.10
NNRM: Little & Su w Imp Classes	0.10	-0.72	0.00	-0.21	0.24	-0.45	0.00	-0.07
NNRM: Little & Su Key Var	-0.05	-0.77	-0.01	-0.28	0.07	**-0.65**	-0.09	**-0.22**
NNRM: Little & Su Distance	-0.01	-0.79	-0.07	-0.29	0.19	-0.54	**-0.11**	-0.16
NNRM: LVCF	0.05	-0.78	-0.05	-0.26	0.14	-0.60	-0.07	-0.18
NNRM: Random Carryover	-0.13	-0.78	-0.05	-0.32	0.04	-0.62	-0.07	-0.22
NNRM: Population Carryover	-0.13	-0.78	-0.05	-0.32	**0.04**	-0.62	-0.07	-0.22
HD: Hotdeck Longitudinal	0.28	-0.78	-0.05	-0.18	0.26	-0.43	-0.10	-0.09
HD: Basic Little & Su	0.13	-0.72	-0.06	-0.22	0.38	-0.32	-0.03	0.01
HD: Little & Su w Imp Classes	0.12	-0.73	-0.03	-0.21	0.38	-0.31	-0.03	0.01
HD: Little & Su Key Var	0.13	-0.76	-0.04	-0.23	0.34	-0.36	-0.05	-0.02
HD: Little & Su Distance	0.10	-0.73	-0.06	-0.23	0.34	-0.33	-0.04	-0.01
HD: LVCF	0.03	**-0.80**	-0.03	-0.27	0.18	-0.62	-0.07	-0.17
HD: Random Carryover	-0.13	-0.79	-0.06	-0.33	0.08	-0.64	-0.08	-0.21
HD: Population Carryover	**-0.13**	-0.79	-0.06	**-0.33**	0.08	-0.64	-0.08	-0.21

Table 7.2: Significant differences between predictive (P), distributional (D) and estimation (E) evaluation measures for rental income

Respondents	nnrml	ls	lsc	lsk	lsd	lvcf	rco	pco	hdl	hdls	hdlsc	hdlsk	hdlsd	hdlvcf	hdrco	hdpco
nnrml		PD	pDe	PD	PD	d	p	p		PD	PD	PD	PD	De	De	De
ls	PD		PE			Pd	PDe	PDe	PD		E	pe		PE	PE	PE
lsc	PDE	PE			PE	d	d	d	pDe	pe			pe			
lsk	PD		Pe			Pd	pd	pd	PD					P	P	P
lsd	PD		PE			Pd	Pde	Pde	PD		E	e		PE	PE	PE
lvcf	P	PD	DE	PD	PD					PD	pD	pD	PD			
rco	P	PD	DE	PD	PD					PD	D	D	PD			
pco	P	PD	DE	PD	PD					PD	D	D	PD			
hdl		PD	DE	PD	PD	d	p	p		PD	PD	PD	PD	dE	dE	dE
hdls	PD	D	PDE	D	D	PD	PD	PD	P					PDe	PDe	PDe
hdlsc	PDE	PDE	d	pd	PDE	DE	DE	DE	E	PE				pd	pd	pd
hdlsk	PD	D	PD	D	D	PD	PD	PD	P		p			pD	pD	pD
hdlsd	PD	D	PDE	D	D	PD	PD	PD	P		PE			PDe	PDe	PDe
hdlvcf	P	PDe	DE	pD	PDe				pD	PDe	De	PD	PDe			
hdrco	Pd	De	pDE	D	De	P	d	d	PDe	De	D	D	De			
hdpco	Pd	De	pDE	D	De	P	d	d	PDe	De	D	D	De			

Non-respondents	nnrml	ls	lsc	lsk	lsd	lvcf	rco	pco	hdl	hdls	hdlsc	hdlsk	hdlsd	hdlvcf	hdrco	hdpco
nnrml			e	pD		D	PD	PD		pD	pD	D	D	D	pD	pD
ls	Pe		e	PD	D	D	PD	PD		pD	pD	d	D	D	pD	PD
lsc	PD	de		pDe	De	D	PD	PD	e	pD	pD	d	D	D	pD	pD
lsk	P	Pe	Pd		D				PD	PD	PD	PD	PD			
lsd	P		PD			d	pD	pD	D	PD	PD	pD	pD	D	D	D
lvcf	P		D	P					D	PD	PD	PD	PD			
rco	P	P	PD		P	P			PD	PD	PD	PD	PD	p		
pco	P	P	PD		P	P			PD	PD	PD	PD	PD	p		
hdl		P	PD	P	P	P	P	P		D	D	d	D	D	PD	PD
hdls	PD	d		Pd	PD	D	PD	PD	PD					PD	PD	PD
hdlsc	Pd			Pd	PD	D	PD	PD	PD					PD	PD	PD
hdlsk	P		d	P	P	p	P	P	P	d	d			pD	PD	PD
hdlsd	Pd			P	PD	D	PD	PD	PD					pD	PD	PD
hdlvcf	P	d	D	p			P	P	P	PD	PD	P	pD			
hdrco	P	P	PD	p	P	P			P	PD	PD	P	PD	P		
hdpco	P	P	PD	p	P	P			P	PD	PD	P	PD	P		

Private Transfers

Table 8.1 provides the average standardized scores for private transfers and Table 8.2 shows which methods are significantly different from each other.

For respondent private transfers, the decision of which imputation method to use comes down to a trade-off between the quality of the cross-sectional estimates and the longitudinal estimates. The random and population carryover methods perform well on distributional accuracy for level estimates, but are middle of the road on longitudinal distributional accuracy. In contrast, the Little and Su methods are better on longitudinal distributional accuracy, but are appreciably worse on cross-sectional distributional accuracy. While the carryover methods perform statistically better than the Little and Su methods on cross-sectional estimation accuracy, the actual improvement is not great. Therefore, given longitudinal estimates are more important to this study than cross-

sectional estimates, the Little and Su method should be selected for this variable. There is no significant difference in the performance of the Little and Su methods tested.

For non-respondents, the random or population carryover methods perform the best.

Table 8.1: Average value of standardised predictive (P), distributional (D) and estimation (E) evaluation measures for private transfers

	Cross-sectional				*Longitudinal*			
Base method: Longitudinal method	P	D	E	Ave	P	D	E	Ave
Respondents								
NNRM: NNRM Longitudinal	0.24	1.27	-0.28	0.41	0.22	1.38	-0.22	0.46
NNRM: Basic Little & Su	-0.27	1.22	-0.28	0.22	-0.16	0.71	-0.24	0.10
NNRM: Little & Su w Imp Classes	-0.30	1.20	-0.27	0.21	-0.19	**0.60**	-0.23	**0.06**
NNRM: Little & Su Key Var	-0.27	1.22	-0.28	0.22	-0.16	0.71	-0.24	0.10
NNRM: Little & Su Distance	-0.27	1.22	-0.28	0.22	-0.16	0.71	-0.24	0.10
NNRM: LVCF	-0.01	0.97	-0.30	0.22	-0.11	1.08	-0.24	0.25
NNRM: Random Carryover	-0.20	**0.78**	**-0.31**	**0.09**	-0.16	0.98	-0.24	0.19
NNRM: Population Carryover	-0.20	**0.78**	**-0.31**	**0.09**	-0.16	0.98	-0.24	0.19
HD: Hotdeck Longitudinal	0.07	1.23	-0.25	0.35	0.19	1.37	-0.22	0.45
HD: Basic Little & Su	-0.32	1.24	-0.27	0.22	-0.14	0.62	-0.25	0.08
HD: Little & Su w Imp Classes	**-0.39**	1.16	-0.27	0.16	**-0.21**	0.63	-0.24	0.06
HD: Little & Su Key Var	-0.32	1.24	-0.27	0.22	-0.14	0.62	-0.25	0.08
HD: Little & Su Distance	-0.32	1.24	-0.27	0.22	-0.14	0.62	-0.25	0.08
HD: LVCF	0.04	1.00	-0.29	0.25	-0.15	1.09	-0.26	0.23
HD: Random Carryover	-0.16	0.84	-0.30	0.13	-0.16	0.99	**-0.27**	0.19
HD: Population Carryover	-0.16	0.84	-0.30	0.13	-0.16	0.99	**-0.27**	0.19
Non-respondents								
NNRM: NNRM Longitudinal	1.24	-0.93	-0.04	0.09	0.91	-0.81	0.23	0.11
NNRM: Basic Little & Su	0.62	-0.90	-0.06	-0.12	0.87	-0.84	0.27	0.10
NNRM: Little & Su w Imp Classes	0.65	-0.89	-0.06	-0.10	0.79	-0.84	0.26	0.07
NNRM: Little & Su Key Var	0.62	-0.90	-0.06	-0.12	0.87	-0.84	0.27	0.10
NNRM: Little & Su Distance	0.62	-0.90	-0.06	-0.12	0.87	-0.84	0.27	0.10
NNRM: LVCF	0.72	-0.95	-0.01	-0.08	0.56	-0.93	0.22	-0.05
NNRM: Random Carryover	0.10	-0.98	**-0.14**	-0.34	0.39	-0.95	0.16	-0.13
NNRM: Population Carryover	0.10	**-0.98**	-0.14	-0.34	0.39	-0.95	0.16	-0.13
HD: Hotdeck Longitudinal	0.56	-0.95	-0.04	-0.15	0.54	-0.95	**0.11**	-0.10
HD: Basic Little & Su	0.11	-0.96	-0.06	-0.30	0.71	-0.94	0.20	-0.01
HD: Little & Su w Imp Classes	0.13	-0.95	-0.03	-0.28	0.62	-0.94	0.19	-0.04
HD: Little & Su Key Var	0.11	-0.96	-0.06	-0.30	0.71	-0.94	0.20	-0.01
HD: Little & Su Distance	0.11	-0.96	-0.06	-0.30	0.71	-0.94	0.20	-0.01
HD: LVCF	0.01	-0.94	0.08	-0.28	0.32	-0.98	0.17	-0.16
HD: Random Carryover	**-0.49**	-0.97	-0.01	**-0.49**	0.21	-0.99	0.20	-0.19
HD: Population Carryover	-0.49	-0.97	-0.01	-0.49	**0.21**	**-0.99**	0.20	**-0.19**

Table 8.2: Significant differences between predictive (P), distributional (D) and estimation (E) evaluation measures for private transfers

Respondents	nnrml	ls	lsc	lsk	lsd	lvcf	rco	pco	hdl	hdls	hdlsc	hdlsk	hdlsd	hdlvcf	hdrco	hdpco
nnrml		PD	PD	PD	PD	Pd	PD	PD		PD	PD	PD	PD	Pde	PDE	PDE
ls	P					D	D	D	PD					D	D	D
lsc	P					D	D	D	PD					D	De	De
lsk	P					D	D	D	PD					D	D	D
lsd	P					D	D	D	PD					D	D	D
lvcf	Pd	Pd	P	Pd	Pd				Pd	D	D	D	D		e	e
rco	PD	De	De	De	De	P			PD	D	D	D	D			
pco	PD	De	De	De	De	P			PD	D	D	D	D			
hdl	Pe	P	P	P	P	dE	PDE	PDE		PD	PD	PD	PD	PdE	PDE	PDE
hdls	P					Pd	De	De	P					D	D	D
hdlsc	P	p		p	p	P	PDe	PDe	P					D	De	De
hdlsk	P					Pd	De	De	P					D	D	D
hdlsd	P					Pd	De	De	P					D	D	D
hdlvcf	Pd	Pd	P	Pd	Pd		P	P	dE	Pd	P	Pd	Pd			
hdrco	PD	De	pDe	De	De	p			PDE	PDe	PDe	PDe	PDe	P		
hdpco	PD	De	pDe	De	De	p			PDE	PDe	PDe	PDe	PDe	P		

Non-respondents	nnrml	ls	lsc	lsk	lsd	lvcf	rco	pco	hdl	hdls	hdlsc	hdlsk	hdlsd	hdlvcf	hdrco	hdpco
nnrml						PD	PD	PD	pDe	pD	PD	pD	pD	PD	PD	PD
ls	Pd					PD	PD	PD	DE	D	pD	D	D	PD	PD	PD
lsc	PD					pD	PD	PD	DE	D	D	D	D	PD	PD	PD
lsk	Pd					PD	PD	PD	DE	D	pD	D	D	PD	PD	PD
lsd	Pd					PD	PD	PD	DE	D	pD	D	D	PD	PD	PD
lvcf	P	D	D	D	D				e					pD	PD	PD
rco	PDe	PD	PD	PD	PD	PDE				P	p	P	P	D	D	D
pco	PDe	PD	PD	PD	PD	PDE				P	p	P	P	D	D	D
hdl	Pd	D	D	D	D		PDe	PDe						d	D	D
hdls	Pd	PD	PD	PD	PD	P	D	D	P					PD	PD	PD
hdlsc	P	PD	PD	PD	PD	P	De	De	P					PD	PD	PD
hdlsk	Pd	PD	PD	PD	PD	P	D	D	P					PD	PD	PD
hdlsd	Pd	PD	PD	PD	PD	P	D	D	P					PD	PD	PD
hdlvcf	Pe	PDE	PDE	PDE	PDE	P	DE	DE	Pe	E	e	E	E			
hdrco	PD	PD	PD	PD	PD	P	PE	PE	P	P	P	P	P	PD		
hdpco	PD	PD	PD	PD	PD	P	PE	PE	P	P	P	P	P	PD		

Total Financial Year Income

Table 9.1 provides the average standardized scores for total financial year income (which is the sum of the imputed components) and Table 6.2 shows which methods are significantly different from each other.

As we are likely to select a variety of methods to impute the income components, looking at the quality of the imputation estimates at the total financial year income level will indicate what methods we would have chosen if we had only looked at total income. Obviously, the results for total financial year income are heavily driven by wages and salaries as the key contributor.

It is also interesting to note that the standardized summary statistics for total income are more negative (indicating the accuracy is better) for predictive and distributional accuracy than many of the components. For respondents, this is most likely an artifact of imputing only a portion of total

income as many components may not needed to be imputed. For non-respondents, a very good summary score may suggest that it is easier to estimate the total income a person receives than it is to estimate the components in which they receive it.

For respondents, the basic Little and Su method performs the best for cross-sectional estimates. Aside from a small improvement in predictive accuracy for the carryover methods, the methods using the NNRM fallback method perform equally well to those using the hotdeck fallback method.

Table 9.1: Average value of standardised predictive (P), distributional (D) and estimation (E) evaluation measures for total financial year income

	Cross-sectional				*Longitudinal*			
Base method: Longitudinal method	P	D	E	Ave	P	D	E	Ave
Respondents								
NNRM: NNRM Longitudinal	-0.85	-1.14	0.00	-0.66	1.09	-0.65	0.41	0.28
NNRM: Basic Little & Su	-0.92	-1.18	-0.11	-0.74	0.07	-1.03	0.54	-0.14
NNRM: Little & Su w Imp Classes	-0.86	-1.17	0.36	-0.56	0.26	-1.08	0.84	0.01
NNRM: Little & Su Key Var	-0.92	-1.18	-0.09	-0.73	0.08	-1.03	0.60	-0.12
NNRM: Little & Su Distance	-0.93	-1.18	-0.11	-0.74	0.07	-1.03	0.53	-0.14
NNRM: LVCF	-0.76	-1.13	-0.04	-0.65	-0.20	-0.32	0.50	-0.01
NNRM: Random Carryover	-0.73	**-1.18**	-0.06	-0.66	-0.15	-0.64	0.47	-0.11
NNRM: Population Carryover	-0.74	-1.18	-0.06	-0.66	-0.16	-0.63	0.47	-0.11
HD: Hotdeck Longitudinal	**-0.94**	-1.13	-0.08	-0.71	0.81	-0.72	**0.34**	0.14
HD: Basic Little & Su	-0.91	-1.17	-0.16	**-0.74**	0.04	-1.02	0.43	**-0.18**
HD: Little & Su w Imp Classes	-0.80	-1.15	0.39	-0.52	0.22	**-1.08**	0.85	0.00
HD: Little & Su Key Var	-0.90	-1.16	-0.14	-0.73	0.06	-1.02	0.49	-0.16
HD: Little & Su Distance	-0.92	-1.16	**-0.16**	-0.74	0.05	-1.02	0.43	-0.18
HD: LVCF	-0.80	-1.11	-0.11	-0.68	**-0.25**	-0.28	0.36	-0.06
HD: Random Carryover	-0.83	-1.17	-0.11	-0.70	-0.20	-0.63	0.41	-0.14
HD: Population Carryover	-0.83	-1.16	-0.11	-0.70	-0.21	-0.62	0.41	-0.14
Non-respondents								
NNRM: NNRM Longitudinal	-1.10	-0.07	0.00	-0.39	-0.24	2.51	**0.06**	0.78
NNRM: Basic Little & Su	-1.21	0.09	-0.02	-0.38	-0.28	0.71	0.23	0.22
NNRM: Little & Su w Imp Classes	-1.22	-0.08	0.03	-0.42	-0.22	0.54	0.47	0.26
NNRM: Little & Su Key Var	**-1.29**	0.08	-0.11	-0.44	-0.31	0.53	0.20	**0.14**
NNRM: Little & Su Distance	-1.27	0.12	-0.08	-0.41	-0.31	**0.53**	0.21	0.14
NNRM: LVCF	-1.05	-0.03	0.02	-0.36	-0.28	3.87	0.08	1.22
NNRM: Random Carryover	-1.22	-0.15	-0.11	-0.50	**-0.40**	2.30	0.08	0.66
NNRM: Population Carryover	-1.22	-0.16	-0.11	-0.50	-0.40	2.32	0.08	0.67
HD: Hotdeck Longitudinal	-0.93	0.13	0.30	-0.17	-0.25	2.33	0.16	0.74
HD: Basic Little & Su	-1.04	0.46	0.36	-0.07	-0.18	1.65	0.32	0.60
HD: Little & Su w Imp Classes	-1.05	0.24	0.34	-0.16	-0.13	1.63	0.59	0.69
HD: Little & Su Key Var	-1.07	0.36	0.28	-0.14	-0.19	1.60	0.27	0.56
HD: Little & Su Distance	-1.06	0.46	0.33	-0.09	-0.19	1.66	0.29	0.58
HD: LVCF	-0.97	-0.06	-0.06	-0.37	-0.29	3.78	0.10	1.20
HD: Random Carryover	-1.17	-0.18	-0.14	-0.50	-0.38	2.21	0.13	0.65
HD: Population Carryover	-1.17	**-0.18**	**-0.14**	**-0.50**	-0.38	2.22	0.12	0.66

Table 9.2: Significant differences between predictive (P), distributional (D) and estimation (E) evaluation measures for total financial year income

Respondents	nnrml	ls	lsc	lsk	lsd	lvcf	rco	pco	hdl	hdls	hdlsc	hdlsk	hdlsd	hdlvcf	hdrco	hdpco
nnrml		PD	PDE	PD	PD	PD	P	P	PD	PD	PDE	PD	PD	PD	P	P
ls	De		Pde			PD	PD	PD	PDe		pDe			PD	PD	PD
lsc	dE	E		pd	Pde	PDe	PDE	PDE	PDE	PDE		Pde	PDE	PDE	PDE	PDE
lsk	D		E			PD	PD	PD	PDe		pD			PDe	PD	PD
lsd	De		E			PD	PD	PD	PD		pDe			PD	PD	PD
lvcf		PD	pDE	PD	PD		D	D	PD	PD	PDe	PD	PD		D	D
rco	PD	P	PE	P	P	D			Pd	PD	PDE	PD	PD	D		
pco	pD	P	PE	P	P	D			Pd	PD	PDE	PD	PD	D		
hdl	p	D	DE	D	D	P	PD	PD		PD	PDE	PD	PD	PD	Pd	PD
hdls	dE		E			Pde	Pe	Pe	d		pDE			PD	PD	PD
hdlsc	E	pE		pE	pE	E	dE	dE	PE	pE		pDe	pDE	PDE	PDE	PDE
hdlsk	E		E			Pd	P	P	d		E			PD	PD	PD
hdlsd	E		E			Pe	Pe	Pe	d		pE			PD	PD	PD
hdlvcf	e	PD	DE	PD	PD		D	D	P	PD	dE	pD	PD		D	D
hdrco	de	p	E	p	p	d	p	p	Pd		E		p	D		
hdpco	de	p	E	p	p	d	p	p	Pd		E		p	D		

Non-respondents	nnrml	ls	lsc	lsk	lsd	lvcf	rco	pco	hdl	hdls	hdlsc	hdlsk	hdlsd	hdlvcf	hdrco	hdpco
nnrml		DE	D	De	De	D	P	P		DE	pDe	DE	DE	D	Pd	Pd
ls	PD		D	D	D	De	pDe	pDe	D	D	PD	D	D	De	pD	pD
lsc	P	D				D	PD	PD	D	D	D	D	D	D	PD	PD
lsk	PDe	P	pD			D	D	D	D	pD	PD	pD	pD	D	D	D
lsd	PD	p	D			D	D	D	D	pD	PD	pD	pD	D	D	D
lvcf	p	PD	P	PDE	PDe		pD	pD	D	pDE	PDe	DE	DE		pD	pD
rco	PdE	De	d	PD	D	PDE			P	PDE	PDe	PDE	PDE	pD		
pco	PdE	De	d	PD	D	PDE			P	PDE	PDe	PDE	PDE	pD		
hdl	PDE	PE	PDE	PE	PE	PDE	PDE	PDE		DE	pD	De	De	D	P	P
hdls	PDE	PDE	PDE	PDE	PDE	DE	PDE	PDE	PD					pDE	PDE	PDE
hdlsc	DE	PDE	PDE	PDE	PDE	DE	PDE	PDE	Pd	D				PDe	PDe	PDe
hdlsk	DE	PDE	PDE	PDE	PDE	DE	PDE	PDE	PD		D			pDE	PDe	PDe
hdlsd	DE	PDE	PDE	PDE	PDE	DE	PDE	PDE	PD		D			pDE	PDE	PDE
hdlvcf	P	PD	P	PD	PD	Pe	PD	PD	pDE	pDE	PDE	PDE	PDE		D	D
hdrco	PDE	DE	De	PD	PD	PDE			PDE	PDE	PDE	PDE	PDE	PDe		
hdpco	PDE	DE	De	PD	PD	PDE			PDE	PDE	PDE	PDE	PDE	PDe		

For estimates of change, the performance of the methods is a little more mixed. The basic Little and Su method (using either fallback method) provides a reasonable balance between the three accuracy measures. Compared to the basic Little and Su method, the Little and Su method with imputation classes performs poorer for both predictive and estimation accuracy but better for distributional accuracy. The three carryover methods perform better than the basic Little and Su method on predictive accuracy but are poorer in distributional accuracy.

For non-respondents, the population and random carryover method is the best for cross-sectional estimates but are very poor in maintaining the distributional accuracy for estimates of change. The basic Little and Su method provides a better compromise between estimates of change and level estimates – it is much better than the carryover methods in maintaining longitudinal distributional accuracy but trades off some predictive accuracy. It is also slightly poorer than the carryover estimates in cross-sectional predictive accuracy.

For total financial year income, we have the additional input in Table 10 from Criteria 7 (which is the chi-square test statistic comparing the movement between income deciles across waves for real and imputed income). The experience of respondents is similar to non-respondents. The Little and Su methods are generally better than the random or population carry over methods in maintaining the distribution of change based on Criteria 7. The longitudinal NNRM was least likely to maintain the distribution of change.

Overall, had we wanted to select one method to apply across all variables, we would have choosen the basic Little and Su method for both respondents and non-respondents, based on the results for total income.

Table 10: Chi-square test statistics on total financial year income deciles (criteria 7)

Imputation Method	W1 to W2	W2 to W3	W3 to W4	W4 to W5	W1 to W5
Respondents					
NNRM: NNRM Longitudinal	83.6	73.0	67.3	53.3	54.9
NNRM: Basic Little & Su	49.2	**35.4**	**33.9**	33.6	**38.5**
NNRM: Little & Su w Imp Classes	45.8	37.0	36.7	35.8	38.7
NNRM: Little & Su Key Var	48.5	35.9	34.6	**33.4**	38.7
NNRM: Little & Su Distance	49.3	37.0	34.9	34.8	38.6
NNRM: LVCF	58.3	49.3	52.9	42.8	51.0
NNRM: Random Carryover	47.4	43.2	42.5	36.6	43.8
NNRM: Population Carryover	47.4	43.3	42.7	37.2	43.8
HD: Hotdeck Longitudinal	74.9	57.9	49.4	53.2	55.7
HD: Basic Little & Su	47.7	37.2	35.0	34.7	39.4
HD: Little & Su w Imp Classes	**43.8**	37.4	35.7	35.7	38.6
HD: Little & Su Key Var	46.9	37.4	35.6	34.9	39.2
HD: Little & Su Distance	47.8	38.8	35.2	35.8	38.9
HD: LVCF	60.3	49.9	51.4	42.4	55.0
HD: Random Carryover	44.0	40.8	39.5	36.5	43.5
HD: Population Carryover	44.1	40.8	39.3	36.8	43.5
Non-respondents					
NNRM: NNRM Longitudinal	410.5	332.2	317.9	275.2	319.8
NNRM: Basic Little & Su	180.0	158.1	140.6	**130.1**	147.6
NNRM: Little & Su w Imp Classes	153.7	**144.5**	140.0	130.7	**130.3**
NNRM: Little & Su Key Var	182.8	156.7	148.9	138.2	136.2
NNRM: Little & Su Distance	178.7	153.5	145.2	134.7	148.0
NNRM: LVCF	345.3	271.7	210.7	201.6	344.5
NNRM: Random Carryover	137.9	166.5	**138.3**	140.6	178.1
NNRM: Population Carryover	**137.8**	166.3	139.2	140.2	178.1
HD: Hotdeck Longitudinal	337.5	372.2	335.8	285.2	283.0
HD: Basic Little & Su	246.9	209.4	182.5	171.3	160.6
HD: Little & Su w Imp Classes	209.8	191.6	180.0	165.7	150.5
HD: Little & Su Key Var	246.5	205.9	188.0	179.0	161.9
HD: Little & Su Distance	256.5	212.0	186.7	177.3	163.7
HD: LVCF	336.5	247.1	218.5	172.1	322.6
HD: Random Carryover	152.8	147.2	170.2	137.8	171.3
HD: Population Carryover	152.6	146.1	171.0	136.5	171.3

Summary of Performance of Methods

For cross-sectional estimates, the random or population carryover methods often perform the best, but perform very poorly on the distributional accuracy of change between waves. The Little and Su method usually provided a reasonable compromise between the accuracy of level estimates versus estimates of change, particularly for respondents.

Where there is a reasonably good correlation between the imputation class variable used in the Little and Su method (being age ranges) and the variable being imputed, the Little and Su variant that uses imputation classes performed better than the other Little and Su methods (such as for wages and salaries, and Australian Government pensions). When the imputation class variable was weakly associated with the variable to be imputed, the basic Little and Su method performed the best. That is, adding unhelpful imputation classes can make the method perform worse than having no imputation classes at all (especially when the donor pool is small).

The added complexity of the multivariate Little and Su methods were not justified as they often did not perform better than the basic Little and Su method or the Little and Su method with imputation classes.

For non-respondents, it was clear that the Little and Su method did not perform as well as the random or population carryover method for some variables. While the carryover methods are more likely to understate change and overstate correlation between waves, it may be preferable to overstating change occurs with some of the other methods (Herringa and Lepkowski, 1986). We suspect the carryover methods are better for non-respondents because accurately imputing zero amounts via the Little and Su method is difficult, particularly for variables that have a high proportion of zeros. It may be that a mixture of the two methods will work well for these variables – the carryover method could be used to determine whether the case should be zero or non-zero and a Little and Su method could be used to determine a suitable imputation value.

Of the carryover methods, there is little to distinguish the performance of the random carryover method and the population carryover method but the last value carried forward method was always poorer than these two. As the population carryover method attempts to take into account the shift in the income values between waves of unimputed cases, it is preferred over the random carryover method.

The longitudinal nearest neighbour regression method and the longitudinal hotdeck method performed reasonably well for cross-sectional estimates of wages and salaries and Government pensions, but were reasonably poor performers when considering estimates of change. They were also poor contenders for both cross-sectional and longitudinal accuracy for the other five income components.

In terms of which cross-sectional fallback method was used, the Little and Su methods tested were often indistinguishable in performance between the nearest neighbour regression fallback method and the hotdeck fallback method. For the carryover methods, the methods using the nearest neighbour regression method as their fallback method often performed better on distributional accuracy than those using the hotdeck fallback method, though sometimes the hotdeck fallback method is better on predictive accuracy.

Table 11 provides a summary of the imputation method recommended for each variable.

Table 11: Recommended imputation method (all using NNRM fallback option), by variable

Variable	Respondents	Non-Respondents
Wages and salaries	Little and Su method using imputation classes	Little and Su method using imputation classes
Australian Government pensions	Little and Su method using imputation classes	Little and Su method using imputation classes
Business income	Basic Little and Su method	Population carryover method
Interest income	Basic Little and Su method	Population carryover method
Dividends and royalties	Basic Little and Su method	Population carryover method
Rental income	Basic Little and Su method	Population carryover method
Private transfers	Basic Little and Su method	Population carryover method
Total financial year income	Total of components	Total of components

Practical Considerations

It is important to consider the complexity of an imputation system to program or explain to users when determining which imputation method to adopt. There would be little sense in adopting an imputation method that is extremely complex to program or explain if it performed only marginally ahead of a simpler method. With a complex method there are greater overheads in developing and maintaining the programs used in the production system than with a simpler system. It is also harder for users to determine what impact the imputation may have on their analysis when a more complex imputation method is chosen.

Of the four basic imputation methods that were considered in this evaluation, the carry over methods are by far the simplest to program, extremely quick to run and very easy to understand.

The nearest neighbour regression method and the hotdeck method have similar large setup requirements in terms of constructing and checking numerous variables that describe the respondent's circumstances used in the regression models or imputation classes. To help ensure consistency when constructing these variables, both of these methods have been developed for a long longitudinal file (where the waves are stacked on top of each other). We use a SAS macro developed by the ABS for the hotdeck imputation to match donors to recipients and this saved us some programming time. Nevertheless, this macro was developed in the 1990s so may not be as efficient or user friendly as it could be (for example, we needed to extend it to cater for variable names longer than eight characters). Both the nearest neighbour regression method and the hotdeck method are reasonably straightforward to explain to users.

By comparison, the data preparation work for the Little and Su method is very easy as it only involves the income variables and small number of other variables, but the programming of the algorithm to select the donor is more complex than the nearest neighbour regression method. For example, matching donors and recipients by appropriate income response categories is difficult if we want to maximize the pool of potential donors to be used. This is probably more difficult in the HILDA Survey because we usually know a missing income amount for a respondent is non-zero and need to find a suitable donor with an appropriate non-zero amount in the appropriate waves. The Little and Su method is the hardest method to explain to users.

Aside from the programming complexity of the methods, the time taken to run the program will also impact on the usability of a particular method. Table 12 shows the number of minutes it took to

run the imputation programs for all income variables in simulation 1.[17] While none of these methods take an extremely long time to run, a reasonably long run time can limit the amount of checking and rerunning of the programs that can be undertaken in a production cycle. The hotdeck methods were time intensive, with the longitudinal hotdeck method taking 114 minutes, the longest of all methods. The basic hotdeck method (which does not include any longitudinal information) took a similar amount of time to run as the longitudinal nearest neighbour regression method. The Little and Su methods tended to take 30 to 45 minutes, though this could be reduced to 5 to 8 minutes with the use of imputation classes (as the search for a suitable donor was restricted to a smaller group of respondents). The carryover methods were exceptionally quick, taking just a third of a minute to run all three methods.

Table 12: Time taken to run imputation programs on simulation 1, minutes

	Base = Nearest Neighbour Regression Method	*Base = Hotdeck*
Fallback base method	3	30
Longitudinal version of base method	52	126
Basic Little & Su	45	14
Little & Su w Imp Class	8	3
Little & Su Key Var	42	13
Little & Su Distance	46	12
LVCF	0.1	0.1
Random Carryover	0.1	0.1
Population Carryover	0.1	0.1

Conclusions

An assessment of the performance of alternative imputation methods was conducted using data from the first five 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 criteria for each income item.

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 have a large pool of donors and are well correlated with age (such as wages and salaries, and Government pensions), the Little and Su method with imputation classes is recommended for both respondents and non-respondents. For all other income components for respondents that are not well correlated with age or have a smaller donor pool, the basic Little and Su method works well. For non-respondents, some further investigation should be undertaken to test a combination of the carryover method (to determine zeros and non-zeros) together with the basic Little and Su method (to determine the non-zero amount). We have avoided using a mixture of methods based on the nearest neighbour regression method and the hotdeck method, due to the more complex development work required for these methods.

[17] The time taken for the later simulations was quicker because of the way SAS processes repeated segments of code.

The evaluation framework was useful in comparing the different methods. It was important to consider at least three or four of the largest components to total financial year income as they had different characteristics that were more suited to different methods. It was also important to study the effect of the imputation method on both respondents and non-respondents (though this aspect is only relevant to household surveys where total household income is calculated). Summarising the results into the three accuracy components – predictive, distribution, and estimation – was also useful as some methods performed extremely well on some aspects yet poorly on others. The number of measures in each of these dimensions could possibly be reduced. Determining which imputation methods were significantly better or worse was helpful in focusing our attention on only the important differences between the methods. It was difficult to compare so many methods in one evaluation, but now this has been done any future comparisons for the HILDA Survey can be made against the 'best' method identified in this study.

This project has highlighted a number of possible areas for future work to improve our understanding of income imputation in a longitudinal survey such as the HILDA Survey. Firstly, we should investigate alternative imputation classes for use in the Little and Su method for variables not associated with age. Secondly, a combination of the population carryover method and the basic Little and Su method should be investigated for non-respondents. Thirdly, other imputation methods should be investigated using this evaluation framework, such as other multivariate imputation methods (such as the hierarchical imputation method used in the Euredit Project (Pannekoek, 2002)) or methods that use more information about the level, trend and variability around the trend than the Little and Su method uses. Finally, the response mechanism could be modified to one that is not missing at random to determine how much this matters in the evaluation of the imputation methods (an example is given by Champney and Bell (1982)).

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.

Champney, T.F. and Bell, R. (1982). Imputation of Income: A Procedural Comparison. Proceedings of the Survey Research Methods Section, American Statistical Association, 431-436.

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, ed. R.M. Groves, D.A. Dillman, J.L. Eltinge and R.J.A. Little. New York: Wiley.

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. Berlin: DIW.

Heeringa, S.G., and Lepkowski, J.M. (1986). Longitudinal Imputation for the SIPP. Proceedings of the Survey Research Methods Section, American Statistical Association, 206-210.

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, 2, 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, ed. D. Rose. London: Routledge.

Lepkowski, J.M., (1989). Treatment of Wave Nonresponse in Panel Surveys. In Panel Surveys, ed. D. Kasprzyk, G.J. Duncan, G. Kalton, and M.P. Singh. New York: Wiley.

Little, R.J.A. (1988). Missing Data Adjustments in Large Surveys. Journal of Business and Economic Statistics, 6, 287-296.

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

Nordholt, E.S., (1998). Imputation: Methods, Simulation Experiments and Practical Examples. International Statistical Review,66, 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, 63, 581-590.

Starick, R. and Watson, N. (2007). An Evaluation of Alternative Income Imputation Methods in the HILDA Survey. HILDA Project Technical Paper Series, No. 1/07, 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, 809-814.

Watson, N. (2004). Income and Wealth Imputation for Waves 1 and 2. HILDA Project Technical Paper Series No. 3/04. Melbourne: 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, 7, 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, 305-310.