



THE UNIVERSITY OF
MELBOURNE

HILDA PROJECT TECHNICAL PAPER SERIES NO. 3/04, July 2004

Income and Wealth Imputation for Waves 1 and 2

Nicole Watson

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



Contents

INTRODUCTION	1
IMPUTATION METHOD.....	2
STEP 1 – IDENTIFY THE SCOPE OF THE MISSING DATA PROBLEM	2
STEP 2 – CONSTRUCT ESTIMATES FOR THE MISSING INFORMATION WHERE POSSIBLE..	2
STEP 3 – CONSTRUCT A REGRESSION MODEL FOR THE VARIABLE OF INTEREST	3
STEP 4 – IDENTIFY THE NEAREST NEIGHBOUR	4
STEP 5 – IMPUTE THE MISSING DATA	5
STEP 6 – CHECK THE QUALITY OF THE IMPUTATION	5
INCOME IMPUTATION	6
EXTENT OF MISSING INCOME DATA.....	6
QUALITY OF IMPUTATION	8
Effect of Imputation on Income Distribution.....	8
Within-sample 20 Per Cent Test	12
Comparison with External Data.....	14
Cross-Wave Comparison	16
POSSIBLE FUTURE ENHANCEMENTS TO THE IMPUTATION METHOD	20
WEALTH IMPUTATION.....	23
EXTENT OF MISSING WEALTH DATA	23
IMPUTATION PRINCIPLES	24
QUALITY OF IMPUTATION	25
IMPUTED VARIABLES PROVIDED IN WAVE 1 AND 2 DATASETS.....	26
CONCLUSION.....	29
REFERENCES	30
APPENDIX 1 – VARIABLES USED IN INCOME MODELS	31
APPENDIX 2 – VARIABLES USED IN WEALTH MODELS	40

Introduction

All surveys are confronted with the problem of non-response. In an earlier discussion paper, Watson and Wooden (2003) assessed the non-response problem in the HILDA Survey, reviewed various methods for dealing with the missing data, and proposed an imputation strategy. This strategy has been largely adopted for Release 2.0 of the HILDA data.¹ The scope of the imputation has been extended beyond income imputation to include the imputation of the wealth variables that were collected in wave 2. The purpose of this paper is to detail the imputation method used, and discuss the quality of the resulting imputation.

In brief, the imputation has been undertaken using a nearest neighbour regression method. The predicted values from a regression model for the variable of interest were used to identify the nearest case whose reported value could be inserted into the case with the missing value.

All imputation has been undertaken at the derived variable level, leaving the original data unchanged. In the main, both the pre-imputed and post-imputed variables are available in the datasets, along with an imputation flag, so that it is easy for the user to choose between using the pre-imputed data or the post-imputed data.

For respondents with item-nonresponse (i.e., where some questions during their interview were not answered), the income and wealth components have been imputed and the totals are the sum of the relevant components. These components and totals are available on the responding person file. However, for non-respondents within responding households just the income and wealth totals have been imputed. These totals for non-responding persons are available on the enumerated person file (along with the totals for responding persons.) Therefore, for income, only imputed totals are available at the household level on the household file. For wealth, the totals for the non-respondents are provided separately from the components summed across the respondents on the household file.

Given the limited resources available to undertake the imputation, we believe the imputation has improved the quality and usefulness of the cross-sectional results. However, our investigations have suggested that in considering change across the two waves the adopted imputation procedure is not performing as well as it could. There are a number of ways in which the imputation process could be improved and these are mentioned towards the end of the paper. We will be reviewing the imputation process and expect we will be introducing changes and modifications for the next data release.

We are grateful to Rob Bray and Stephen Horn from the Department of Family and Community Services who made comments on an earlier version of this paper and suggested improvements.

¹ Release 2.0 includes data for waves 1 and 2 and was released in January 2004.

Imputation Method

The income and wealth imputation for the HILDA Survey was implemented using a nearest neighbour regression method. A regression model for the variable of interest was used to identify a record with complete information for the variable of interest (called the donor) that was similar to the record with missing information (called the recipient). The donor's value is used to replace the recipient's missing value. Therefore, only real values reported by a respondent were used to impute missing cases. An important advantage of this method is that the variability of all imputed variables is generally maintained during the imputation process.²

The imputation process involved the following six steps.

Step 1 – Identify the scope of the missing data problem

Income and wealth variables are prime candidates for imputation as they have a relatively high rate of 'missingness' and the missingness is known to be non-random. The purpose of imputation is to correct the bias introduced into the estimates when working with incomplete data.

While an assessment of the missing data problem was undertaken earlier in Watson and Wooden (2003), the case has been restated in later sections of this paper. This was done for a number of reasons. The income model used in wave 1 has been revised and new income variables were created. Also the earlier paper focused only on wave 1 and we have now extended the imputation to wave 2 and to include wealth variables.

Due to the structure of the questionnaire, in all but a few cases we know when an individual received income or had wealth from a particular source or not.³ As a result, the missing variables are assumed to be non-zero. There are two exceptions here. The first is for some wealth variables where the screener questions did not preclude zero responses for items such as bank account balances, credit card debt and business assets. The second exception is in the imputation of total income and wealth for individuals who did not participate in an individual interview. For these cases their imputed income or wealth variables could be zero.

Step 2 – Construct estimates for the missing information where possible

There were a number of cases where a reasonable approximation of the missing information could be made based on the other information collected during the interview rather than imputing a value from elsewhere. These approximations or edits (as distinct from imputations) were generated in the following situations:

² When we impute, it appears in the dataset that we have more data points than we actually do have which will artificially reduce the standard errors. One solution to this problem is to calculate point estimates using post-imputed data, but calculate the standard errors using the pre-imputed data.

³ For a few cases, the respondent has refused or didn't know the answer to the screener question of whether they had income or wealth from a particular source.

- Current wages and salaries in wave 1. If financial year wages and salaries was reported along with how current wages and salaries compares to a year ago and the respondent was employed for all of the financial year, apply 75 per cent (to allow for different time periods involved) of the change from the financial year to get the current wages and salaries (affects 187 people in wave 1).⁴
- Current wages and salaries in wave 2. If current wages and salaries in wave 1 and financial year income in waves 1 and 2 are reported and the respondent was employed for both financial years, then apply the same ratio of current to financial year wages and salaries from wave 1 to wave 2 (affects 77 people in wave 2).
- Financial year wages and salaries in wave 1. If current income is reported along with how current wages and salaries compares to a year ago and the respondent was employed for all of the financial year, apply 75 per cent of the change to the current wages and salaries (affects 245 people in wave 1).
- Financial year wages and salaries in wave 2. If we have current wages and salaries in waves 1 and 2 and the respondent was employed for the full financial year, take 25 per cent of the increase or decrease from current income to get financial year income (affects 134 people in wave 2)
- Business income, interest, dividends and royalties, and rent. If both partners report having income from the same source and one knows the value but the other does not, then assume it is the same for both. (For business income, affects 34 people in wave 1 and 31 people in wave 2. For interest income, affects 76 people in wave 1 and 85 people in wave 2. For dividends and royalties, affects 120 people in wave 1 and 98 people in wave 2. For rent, affects 20 people in wave 1 and 25 people in wave 2.)

These cases were then removed from the subsequent steps in the imputation process.

Step 3 – Construct a regression model for the variable of interest

For each variable imputed, a regression model was developed using cases actually reporting a value for that variable. The primary aim of the regression was prediction rather than interpretation. While we have included variables thought to be important in predicting the various income and wealth components based on accepted economic theory, we did not limit the search for useful variables there. We sought to include any variables that might increase the predictive power of the model even if we could not readily explain why the variable was important.

The income and wealth variables have been transformed by taking the natural logarithm of the variables. Only cases with positive incomes were included in the

⁴ Only 75 per cent of the change for the year needs to be applied as the mid-point of the last financial year is the end of December and the mid-point of the interview dates is the end of September, resulting in a 9 month gap, not a 12 month gap.

income model (negative values occurred for business income, rental income and total income).

For the income models, a statistical package called MARS was used.⁵ This is an automatic regression package which finds the best model for the specified variable from the host of variables it is instructed to consider. Main effects and two-way interactions were considered. MARS provided a practical solution to the resource intensive problem of constructing good predictive models.

A full list of the variables considered in the income models is provided in Appendix 1, together with tables showing the variables kept in the final models. Two types of models were constructed for each wave: one set were constructed using only the information from same wave and another set were constructed using these variables plus the income information from the other wave where this was available.

For the wealth imputation, the choice of regression model was slightly more judgment-based than the automated process provided by the MARS program. However, similar to that methodology, a model with a large number of variables was considered initially for each regression. Insignificant variables were then excluded step-by-step to obtain a better regression model with care taken to retain variables that were expected to be of significance in explaining any specific left-hand side variable.

A full list of the variables considered and used in the wealth models is provided in Appendix 2. The wealth imputation used income data (including imputed income data) where necessary. This allowed us to use the same model for an individual variable across all persons or households. The one exception was that for enumerated persons, wave 1 information was used in a separate model if the person responded in wave 1.

Step 4 – Identify the nearest neighbour

The predicted value for all cases was calculated from the model and transformed back to the original scale.⁶

The cases were sorted by their predicted value. Where there were multiple cases with the same predicted value, they were sorted randomly within this predicted value. The cases with missing values were placed next to or near complete cases with similar predicted values, thus identifying the nearest neighbour.

This nearest neighbour is called the donor, and the record that is to be imputed is called the recipient.

⁵ See the Salford Systems website for an overview of the MARS package: www.salford-systems.com.

⁶ The standard correction to transform a variable with a normal distribution to one with a lognormal distribution was applied (using the formula provided in Greene 1993, p. 71). That is, the exponential of the predicted value from the model using the logged variable was multiplied by $e^{\sigma^2/2}$, where σ^2 is the variance of the logged residuals. Note that we could have identified the nearest neighbour equally well on the transformed scale, but it was easier to work in dollars rather than log dollars when developing the programs for the imputation system.

Step 5 – Impute the missing data

The actual value for the variable of interest of the donor with the predicted value which was the closest to the recipient's predicted value was inserted into the recipient's record.

A donor could only be used twice in the imputation of a particular variable. After this the case was set aside and the next nearest neighbour used.

Step 6 – Check the quality of the imputation

Once the imputation had been undertaken, a number of checks were made on the resulting data. These included:

- undertaking a within-sample 20 per cent test where the real values reported by a respondent were temporarily set to missing so that they could be compared to the results of the imputation procedure;
- comparison of the imputed data to benchmark information; and
- examining the effect of the imputation on the income distribution.

Sometimes these checks resulted in a revision to the imputation procedure (such as the inclusion of the estimation step where we were able to get better estimates another way).

The results of these final checks are reported later in this paper.

Income Imputation

Extent of Missing Income Data

A new income model was applied to waves 1 and 2 HILDA data for Release 2.0. This necessitated a change in the variables which needed to be imputed and the numbers involved. In brief, business income from incorporated businesses was added to wages and salaries, dividends from incorporated businesses were added to dividends variables, benefits were split between Australian and foreign sources, and other income was divided into a couple of different categories (one of which was irregular income which is now called 'windfall income').

Table 1 provides the revised counts of cases to be imputed for each income source in waves 1 and 2. For responding persons we have provided the proportion of missing cases from all non-zero cases. For these people we will only be imputing non-zero amounts. We either know they are non-zero due to the structure of the questionnaire or we assume they are non-zero where the 'don't know' or 'refused' occurred at the screener question (which happens rarely). For all enumerated persons (including respondents and non-respondents in responding households) and for households, we have provided the proportion of all cases (with zeros included) that is missing.

Several observations can be made about the figures presented in Table 1:

- Non-respondents comprise about 35 per cent of all enumerated people missing total financial year income (for wave 1, this is calculated as $(3212 - 2054)/3212$). We have far less information about these non-respondents on which we can make a meaningful imputation for the missing values. The information we do have for these people include: limited person details from the household form, household-level data, information about their partner if applicable, and income information from the other wave. Only total financial year income and windfall income is imputed for these persons so that these variables can be summed to the household level.
- Between waves 1 and 2, the proportion of missing income for both person and household level variables fell slightly. This is possibly because the respondents have become more comfortable with the survey and some less willing participants dropped out of the survey in wave 2.
- The variables with the highest proportion of missing cases include business income and investment income.
- The restructure of the income variables and the taxation model dictated which variables needed to be imputed. This has meant that an imputation system was devised for some variables with a small number of missing cases (such as benefits from foreign governments).
- Some components will be harder to impute than others. We should be able to make a reasonably good prediction for wages and salaries, but components such as business income, investment income and windfall income will be far more problematic.

Table 1: Number and proportion of cases with missing income data, waves 1 and 2^a

<i>Variable</i>	<i>Wave 1</i>		<i>Wave 2</i>	
	<i>Number of missing cases</i>	<i>Prop'n of cases, %</i>	<i>Number of missing cases</i>	<i>Prop'n of cases, %</i>
RESPONDING PERSONS (non-zero cases only)				
Current income				
Wages and salaries	462	6.0	310	4.2
Benefits	136	3.2	81	2.1
Financial year income				
Wages and salaries	666	7.9	550	6.9
Australian govt pensions	67	1.5	52	1.2
Foreign govt pensions	1	0.5	3	1.4
Business income	404	29.1	366	28.6
Investments				
Interest	661	19.5	596	18.6
Dividends and royalties	584	14.6	521	14.5
Rent	240	20.3	189	15.3
Private pensions	59	6.2	41	4.6
Private transfers	28	7.1	89	23.1
Total FY income ^b	2054	15.6	1817	14.7
Windfall income				
Windfall	32	4.1	31	2.9
ENUMERATED PERSONS (zero and non-zero cases)				
Total FY income	3212	21.2	2795	19.9
Windfall income	1190	7.9	1009	7.2
HOUSEHOLDS (zero and non-zero cases)				
Total FY income	2243	29.2	2009	27.7
Windfall income	838	10.9	723	10.0

Notes:

- a. The percentages reported in this table for responding persons are of all non-zero cases. This differs slightly from the wave 2 data quality paper for business and rental income where some people reporting zeros have been included as they could have received income from these sources (Watson and Wooden 2004).
- b. Total financial year income for respondents was calculated as the sum of components after imputation.

Quality of Imputation

Effect of Imputation on Income Distribution

The unweighted means, medians and standard deviations for the income variables before and after imputation for wave 1 are provided in Table 2. Similar statistics for wave 2 are provided in Table 3. The pre-imputation statistics exclude missing cases and the post-imputation statistics include them with the imputed value replacing their missing value.

We see that the distribution has changed little for the respondents when we consider just those cases that have income from a particular source. This is a positive result – we would not expect the imputation to greatly alter the distribution as we believe the item non-response occurs across the range of income rather than being concentrated in any one part of the distribution. However, had we considered the income distribution for all available cases (zeros and non-zeros) we would have generally seen an increase in the means, medians and standard deviations after imputation, simply because the proportion of non-zero cases has increased. Indeed, this effect can be seen in the enumerated person figures where all cases are included.

At the household level, the effect of imputation is more dramatic. In wave 1, the unweighted mean household income increased from \$47,980 before imputation to \$54,689 after imputation. A similarly large increase occurred in wave 2. There are two reasons for this result. The first is that larger households (who have the higher incomes) are more likely to be incomplete due to part household non-response. The second reason is that the less income a person has, the less likely they will receive income from multiple sources or have complex financial arrangements, thus increasing the likelihood of being able to report a complete set of income information (Watson and Wooden, 2002).

Table 2: Unweighted distribution of income data before and after imputation, wave 1

<i>Variable</i>	<i>Before imputation</i>			<i>After imputation</i>		
	<i>Mean</i>	<i>Median</i>	<i>Standard deviation</i>	<i>Mean</i>	<i>Median</i>	<i>Standard deviation</i>
RESPONDING PERSONS (non-zero cases only)						
Current income						
Wages and salaries	37,212	31,440	28,869	37,057	31,284	29,088
Benefits	8,662	8,812	4,181	8,622	8,812	4,193
Financial year income						
Wages and salaries	35,222	30,000	38,045	34,360	29,500	37,268
Australian govt pensions	6,750	7,692	4,316	6,735	7,670	4,311
Foreign govt pensions	4,427	3,406	3,665	4,404	3,353	3,669
Business income	16,776	10,400	35,756	18,429	11,697	39,963
Investments						
Interest	2,787	675	7,807	2,729	600	7,511
Dividends and royalties	2,224	200	8,433	2,240	200	8,244
Rent	3,702	1,421	25,302	3,484	1,200	23,253
Private pensions	16,043	11,246	20,504	16,130	11,027	20,794
Private transfers	4,773	3,250	5,576	4,895	3,380	6,046
Total FY income ^a	28,629	20,750	32,275	29,386	21,000	37,636
Windfall income						
Windfall	5,247	1,040	14,457	5,195	1,040	14,225
ENUMERATED PERSONS (zero and non-zero cases)						
Total FY income ^a	26,712	18,000	31,986	27,773	19,092	36,925
Windfall income	283	0	3,557	287	0	3,477
HOUSEHOLD (zero and non-zero cases)						
Total FY income ^a	47,980	37,000	45,052	54,689	42,659	58,061
Windfall income	524	0	5,063	566	0	4,971

Notes:

- a. Total income in this table is the sum of the income components – it does not include Family Tax Benefit Part A or Part B, or Child Care Benefit.

Table 3: Unweighted distribution of income data before and after imputation, wave 2

<i>Variable</i>	<i>Before imputation</i>			<i>After imputation</i>		
	<i>Mean</i>	<i>Median</i>	<i>Standard deviation</i>	<i>Mean</i>	<i>Median</i>	<i>Standard deviation</i>
RESPONDING PERSONS (non-zero cases only)						
Current income						
Wages and salaries	37,986	33,370	28,797	37,706	32,952	28,723
Benefits	9,058	9,255	4,180	9,031	9,255	4,233
Financial year income						
Wages and salaries	35,880	31,000	33,200	35,093	30,000	33,242
Australian govt pensions	7,481	8,320	4,371	7,463	8,268	4,374
Foreign govt pensions	4,697	3,500	4,807	4,689	3,500	4,775
Business income	20,849	12,867	50,923	20,664	12,400	46,109
Investments						
Interest	2,265	500	6,438	2,294	500	6,303
Dividends and royalties	3,053	220	12,661	3,111	250	12,264
Rent	3,357	2,500	14,159	3,391	2,244	14,153
Private pensions	20,378	12,000	49,751	21,019	12,000	50,065
Private transfers	4,899	3,600	5,552	5,176	3,640	5,975
Total FY income ^a	30,062	21,407	36,190	31,094	22,022	37,889
Windfall income						
Windfall	17,303	2,000	59,871	17,167	2,000	59,409
ENUMERATED PERSONS (zeros and non-zero cases)						
Total FY income ^a	28,188	19,132	35,789	29,049	20,000	37,018
Windfall income	1,383	0	17,559	1,457	0	17,718
HOUSEHOLDS (zeros and non-zero cases)						
Total FY income ^a	50,659	38,601	54,128	56,209	43,000	57,810
Windfall income	2,578	0	24,756	2,820	0	24,888

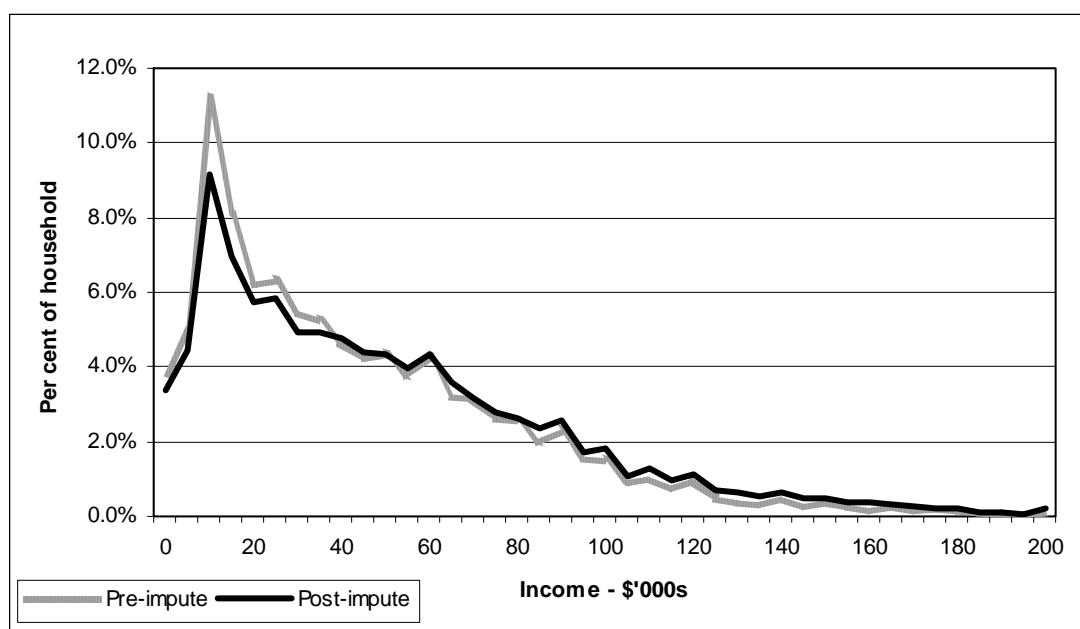
Notes:

- a. Total income in this table is the sum of the income components – it does not include Family Tax Benefit Part A or Part B, or Child Care Benefit.

Weighted pre-imputation and post-imputation statistics were also calculated, but the general thrust of the observations is unchanged, so the tables are not reproduced here.

Another way to view the impact of the imputation on the income distribution is via a graphical presentation. Figure 1 illustrates the change in the main section of the household income distribution for wave 1 as a result of the imputation. This time, weights have been applied (though the unweighted results are very similar). The grey shaded line shows the income distribution for total financial year income prior to imputation (using 70.8 per cent of the responding households). The black line shows the revised income distribution after the imputation has been undertaken. In comparing the two distributions, we see that the proportion of low income households has been pulled down by the imputation and the proportion of high income households has been pushed up. The corresponding graph for wave 2 is almost identical, so is not provided here.

Figure 1: Weighted distribution of total financial year household income, wave 1



Within-sample 20 Per Cent Test

One way to test that the imputation method is producing feasible results is to set aside a proportion of cases, replace the actual values with missing values, run the cases through the imputation procedure, and then compare the imputed values with the actual values. Ideally this should be done at the beginning of the imputation process, but in this case the test began after the regression models had been fitted (i.e., after step 3) to greatly reduce the time taken to run the test and the programming complexity involved. We do not expect this to affect the results of the test very much.

For each income component, a random sample of 20 per cent of the cases was selected and those cases with a non-zero amount were set to missing.⁷ The actual values were stored in a separate variable. The random samples were drawn independently of each other. Imputation of the missing values was then undertaken and the imputed and actual values compared.

Table 4 shows the results of this test for waves 1 and 2. For each wave, the number of cases included in the test is reported, together with the mean of the actual and imputed values. The fourth column for each wave provides the p-value from a test of whether the differences between the mean and actual values are significantly different from zero.

For all but one variable, the imputed values are not significantly different from the actual values. The one variable that is significant at the 5 per cent level is income from rental properties for wave 1. Given we are testing 26 variables, we expect on average for at least one variable to be significant at the 5 per cent level by chance alone even if there were no real differences. Therefore, the fact that we have found one is not cause for concern.

⁷ Note that this 20 per cent test does not test the assumption that the missing values are non-zero, nor does it test the ability of the imputation process to correct for non-random missingness. It is simply testing the nearest neighbour donor method in correcting for random missingness.

Table 4: Outcome of 20 percent sample test on non-zero cases, waves 1 and 2

Variable	Wave 1				Wave 2			
	<i>n</i>	<i>Actual</i>	<i>Imputed</i>	<i>p-value</i> ^a	<i>n</i>	<i>Actual</i>	<i>Imputed</i>	<i>p-value</i> ^a
RESPONDING PERSONS								
Current income								
Wages and salaries	1433	36,164	36,538	0.5003	1404	38725	39194	0.4054
Benefits	796	8,561	8,791	0.1053	807	9156	9067	0.5360
Financial year income								
Wages and salaries	1534	36,660	36,945	0.8363	1516	35211	35299	0.8951
Australian govt pensions	930	6,838	6,795	0.7780	918	7468	7400	0.5768
Foreign govt pensions	39	4,963	4,739	0.7300	43	4265	3738	0.6032
Business income	184	17,192	16,368	0.7403	175	19348	17477	0.5745
Investments								
Interest	550	2,732	2,949	0.3863	518	2161	2639	0.0919
Dividends and royalties	691	2,568	2,229	0.3336	641	3093	3286	0.7806
Rent	207	407	4,740	0.0254	197	3194	17	0.1530
Private pensions	203	16,344	16,023	0.8445	180	18231	23334	0.4063
Private transfers	71	4,681	5,869	0.0904	56	7014	4876	0.1000
Windfall income								
Windfall	148	5,337	5007	0.8200	206	15297	21797	0.1375

Notes:

- a. The p-value is the probability of the difference being at least as large as that observed under the assumption that the difference has mean zero and standard deviation of observed sample. A finite population correction factor has been applied as the test sample is 20 per cent of available cases. That is, we are applying a two-tailed test where

$$t = \frac{\bar{x}_{diff}}{\sqrt{\frac{1}{n} \left(1 - \frac{n}{N}\right) s_{diff}^2}}$$

has t-distribution with n-1 degrees of freedom. N is the total number of non-zero non-missing cases and n is the number of non-zero cases in the 20-per cent test with actual values that have been set to missing.

Comparison with External Data

A further way to test the plausibility of our imputed data is to compare the HILDA estimates with accepted external data. The Survey of Income and Housing Costs (SIHC), conducted by the Australian Bureau of Statistics (ABS), provides us with a generally suitable comparison. We are hampered a little in our comparison in that the most recent information from the ABS relates to the 99/00 financial year. Wave 1 of the HILDA Survey relates to the 00/01 financial year. We have applied inflation factors to the ABS figures to get approximate 00/01 estimates.⁸

Before discussing the comparisons of the SIHC and HILDA estimates, it is worth noting a few differences between the surveys.

- The interviews for the 2000/01 SIHC were conducted by the ABS in approximately equal number each month during the financial year 2000/01 (ABS cat. no. 6523.0). The financial year income information collected, however, relates to the preceding (1999/00) financial year. In contrast, the vast majority of the HILDA interviews are conducted between August and December each year. As a result, the average recall period for SIHC respondents is longer than for the HILDA Survey respondents.
- The definition of regular and irregular income is not as clear in the HILDA Survey as it is in the SIHC. We expect that the HILDA Survey will have slightly more irregular components added to wages and salaries. We have also attempted to disentangle regular and irregular sources of income after the interview, whereas the SIHC does this during the interview.

The first column of Table 5 shows the SIHC person-level means for each income source and the second column has these figures inflated up to approximate 2000/01 figures. The next two columns are the weighted person-level means before and after imputation from the HILDA Survey. The final two columns show the differences between the HILDA means and the SIHC means for 2000/01 financial year, both before and after the HILDA imputation. Similarly, Table 6 provides the comparison for wave 2.

There are several observations to be made about these tables:

- The HILDA means after imputation are higher than those before imputation. This is expected given we are including both zero and non-zero cases in the construction of these means and our imputed values are all non-zero.
- The estimate of business income is vastly improved by the imputation process.
- Wages and salaries income and investment income are raised by around \$100 and \$300 respectively following imputation. The large difference between

⁸ Note that our inflation factors from 99/00 are unusually complicated by the introduction of the GST in July 2000. The assumptions made about the inflation factors are documented as notes in Tables 5 and 6.

SIHC and HILDA estimates for wages and salaries is thus *not* a result of the imputation.

- There is a large jump in the HILDA estimate of windfall income between wave 1 and 2. This is presumably due to a change in the questionnaire where inheritances and bequests are explicitly asked for in the final question on financial year income from wave 2 onwards. These amounts can be very large and were presumably grossly under-reported in wave 1.

Further discussion of the representativeness of the HILDA income data is provided in Watson and Wooden (2004).

Table 5: Survey of Income and Housing Costs and HILDA Survey, financial year person-level means compared for wave 1

	<i>Survey of Income and Housing Costs</i>		<i>HILDA wave 1 (2000/01)</i>		<i>Difference from SIHC (HILDA-SIHC)</i>	
	<i>99/00</i>	<i>Approx 00/01^a</i>	<i>Without imputation</i>	<i>With imputation</i>	<i>Without imputation</i>	<i>With imputation</i>
Wages and salaries	18,510	19,528	20,955	21,098	1,427	1,570
Benefits ^b	2,312	2,451	2,202	2,219	-249	-232
Business income	1,737	1,780	1,159	1,726	-621	-54
Investment income	1,049	1,075	1,322	1,564	247	489
Sum of above components	23,608	24,834	25,638	26,607	804	1,773
Other regular income ^c	652	668	1,164	1,237	N/A	N/A
Windfall income	N/A	N/A	302	311	N/A	N/A

Notes:

- SIHC estimates for 00/01 financial year are calculated from 99/00 by applying:
 - 5.5% increase to wages and salaries (being the Average Weekly Earnings increase for all employees from 99/00 to 00/01);
 - 6.0% increase to benefits (being the Consumer Price Index increase from 99/00 to 00/01 which includes the GST effect); and
 - 2.5% to other income components (being the Consumer Price Index increase from September 2000 to September 2001 to avoid the effect of the introduction of the GST).
- \$403 in Family Tax Benefit has been removed from the SIHC estimates (as this is calculated separately in HILDA). Neither SIHC nor HILDA estimates include Child Care Benefit.
- Income from other sources cannot be directly compared with the ABS as the HILDA Survey has not clearly differentiated regular from irregular components. We have only assumed which sources are more likely to be regular and placed them in the 'other regular' category. Those more likely to be irregular are placed in 'windfall' income.

Source: The ABS data were provided by Roger Wilkins and come from the Survey of Income and Housing costs, 2000/2001, confidentialised unit record file (cat. no. 6541.0.30.001). Both the ABS and HILDA estimates are weighted.

Table 6: Survey of Income and Housing Costs and HILDA Survey, financial year person-level means compared for wave 2

	<i>Survey of Income and Housing Costs</i>		<i>HILDA wave 2 (2001/02)</i>		<i>Difference from SIHC (HILDA-SIHC)</i>	
	<i>99/00</i>	<i>Approx 01/02^a</i>	<i>Without imputation</i>	<i>With imputation</i>	<i>Without imputation</i>	<i>With imputation</i>
Wages & salaries	18,510	20,342	21,700	21,819	1,358	1,477
Benefits ^b	2,312	2,555	2,540	2,557	-15	2
Business income	1,737	1,838	1,381	1,885	-457	47
Investment income	1,049	1,110	1,305	1,659	195	549
Sum of above components	23,608	25,845	26,926	27,920	1,081	2,075
Other regular income ^c	652	690	1,551	1,677	N/A	N/A
Windfall income	N/A	N/A	1,405	1,428	N/A	N/A

Notes:

- a. SIHC estimates for 01/02 financial year are calculated from 99/00 by applying:
 - i. 9.9% increase to wages and salaries (being the Average Weekly Earnings increase for all employees from 99/00 to 01/02);
 - ii. 10.5% increase to benefits (being the Consumer Price Index increase of 6.0% from 99/00 to 00/01 which includes the GST effect, and the Average Weekly Earnings increase for all employees from 00/01 to 01/02); and
 - iii. 5.8% to other income components (being the Consumer Price Index increase from September 2000 to September 2002 to avoid the effect of the introduction of the GST).
- b. See note 2 for Table 5.
- c. See note 3 for Table 5.

Source: See source for Table 5.

Cross-Wave Comparison

The previous sections on the quality of the imputation have not raised serious concerns about the income imputation. However, all of this analysis was of a cross-sectional nature where we were only considering the within-wave effects. Now we turn our attention to the longitudinal component of the income imputation.

The principal aim of the HILDA Survey is to collect data to enable the measurement of changes over time. We thus need to be aware of the impact of the imputation on estimates of change.

While reported income information from the other wave was included in the regression models, along with wave 1 post-imputation values for the wave 2 models, this did not guarantee a strong concordance between the two values. Presented below in Table 7 are the correlations between waves for each of the income components.

Only respondents with non-zero income from the particular source for both waves are considered here. The first two columns present the correlations and number of units contributing to the correlations for cases where both waves did not require imputation. The middle two columns are for cases where imputation was undertaken in one of the two waves. The last two columns are for cases where both waves were imputed.

For the most part, we see a marked decline in correlation as the number of waves imputed increases. Total income, for example, has a correlation of 0.7 between cases where all components of income are reported in both waves. Where one wave had to be imputed, the correlation fell to 0.5. When two waves were imputed, the correlation was just 0.3.

Before commenting on what these correlations might mean for longitudinal analysis, it is worth considering why data might be missing. We suspect that people with less stable employment situations would be less likely to tell us their financial year income (they may have had multiple jobs over the year, or worked part of the year, or received benefits for part of the year). People who have very regular income sources would be more likely to know the amount (such as income from one job or from a stable benefit situation). Therefore, we would expect some decline in the correlation between years for a particular source. However, we did not expect the decline to be as large as that shown in Table 7.

Table 7: Correlation between income for wave 1 and 2 respondents, by source (non-zero cases only)

	<i>No imputation</i>		<i>One wave imputed</i>		<i>Both waves imputed</i>	
	<i>Corr</i>	<i>N</i>	<i>Corr</i>	<i>N</i>	<i>Corr</i>	<i>N</i>
<i>Current income</i>						
Wages & salaries	0.78	5435	0.74	394	0.38	44
Benefits	0.56	3028	0.51	100	0.35	9
<i>Financial year income</i>						
Wages & salaries	0.77	5869	0.46	589	0.42	91
Benefits	0.55	3361	0.31	67	0.04	6
Business	0.67	481	0.25	261	-0.08	81
Investment	0.50	3110	0.44	903	0.13	310
Other income	0.20	655	0.29	68	-0.01	5
Total	0.70	8354	0.50	2161	0.29	576
Windfall income	0.57	236	0.14	9	-	-

Alternatively, and most obviously, the imputation process itself might be the source of the low year-on-year correlations. We know from work undertaken by Solon (1989) on measurement error, that it most likely biases downward the estimated correlation between two variables. The greater the measurement error, the greater the bias will be in the correlations. This provides an analogy for the imputation ‘error’ we have introduced when we impute a value that is different from the true unknown value. The correlations between waves will be lower and we will exaggerate the extent of income mobility.

Another way to look at this problem is to consider the degree of imputation required, rather than simply using the number of waves imputed. In Table 8 below we restrict our attention to total financial year income and include all enumerated adults in both waves (responding and non-responding). The enumerated sample is split into the following seven groups, depending on whether they provided an interview and income details for their main and secondary incomes:

- No imputation required – the respondent provided full income details in both waves;
- Imputation to secondary income in one wave – the respondent provided full details of their main income for both waves, but secondary income needed imputing in one wave;
- Imputation to secondary income in both waves – the respondent provided full details of their main income for both waves, but secondary income needed imputing in both waves;
- Imputation to main income in one wave – the respondent provided full details of their main income in one wave, but did not provide full details in the other wave;
- Imputation to main income source in both waves – the respondent did not provide full details of their main income in either wave;
- Imputation for unit non-response in one wave – the respondent provided an interview in one wave, but did not provide an interview in the other; and
- Imputation for unit non-response in both waves – the individual did not provide an interview in either wave.⁹

With each step down this list of seven groups, we are imputing more of the individual’s total income and have less on which to base the imputation. Table 8 shows that the correlations between total financial year income is around 0.7 where we either don’t need to impute any income or are just imputing secondary income. The correlations fall to around 0.3 when we impute main income or where there is unit non-response in one wave only. The correlations drop further to around 0.05 when we have to impute main income in both waves, or where there is unit non-response in both waves.

⁹ The differentiation between ‘main’ and ‘secondary’ income was made on the post-imputed data. The ‘main’ income source is the income component that has the greatest value of the five components: wages and salaries; benefits; business income; investments; and other sources. The ‘secondary’ income includes all the remaining sources combined together.

Table 8: Correlation between total financial year income waves 1 and 2 enumerated adults by degree of imputation

<i>Degree of imputation required</i>	<i>Corr</i>	<i>N</i>
None required	0.71	9160
Only secondary income in one wave	0.69	1178
Only secondary income in both waves	0.65	177
Main income in one wave	0.27	1256
Main income in both waves	0.05	222
Unit non-response in one wave	0.28	508
Unit non-response in both waves	0.07	455

Next we consider the effect the imputation has on the estimates of income mobility between the two waves.¹⁰ Table 9 shows the proportion of the population that have shifted income deciles between waves 1 and 2 by the degree of imputation required. This time, the degree of imputation is calculated at the household level – the first five categories are for fully responding households in both waves and the last two are for partially responding households in at least one wave. Households fall into the category that matches the worst situation of any of the individuals in the household. For example, in a household of two adults where one is a non-respondent in one wave and the other is missing their main income in both waves, the household would fall into the sixth category of ‘unit non-response in one wave’.

Where no imputation was required for any of the household members, 37.6 per cent of people had no change to their income decile. There is a general decline of this figure with each increase in the degree of imputation required – indeed, for people in households with non-respondents in both waves the proportion with no change is just 15.9 per cent. The main exception here is the group for which secondary income was imputed in both waves and the number in this group is very small.

If we focus just on very strong shifts between deciles, we see that only 3.3 per cent of people in households with no imputation moved five or more deciles. In contrast to this, 13.7 per cent of people in partially non-responding households in both waves moved five or more deciles.

It is clear from this table, that the greater the degree of imputation, the greater the income mobility estimated.

The lack of correlation and greater income mobility across waves for imputed cases is clearly a problem for analysis of change. However, should the researcher exclude the imputed cases from their analyses, they may overstate the case of no change. This problem certainly needs to be revisited, and minimised wherever possible, in future releases of the HILDA data.

¹⁰ Rob Bray from the Department of Family and Community Services suggested this useful extension to the analysis of the impact of the income imputation across waves.

Table 9: Income decile shifts between waves 1 and 2 by degree of imputation^a

<i>Degree of imputation required</i>	<i>n^b</i>	<i>Number of deciles shifted</i>									
		<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>
None required	9,299	37.6	36.6	13.6	5.9	2.8	1.5	1.0	0.5	0.2	0.1
Only secondary income in one wave	1,797	30.4	37.0	14.7	6.6	6.7	2.1	1.0	1.2	0.3	0.0
Only secondary income in both waves	266	45.1	25.2	15.0	4.5	3.5	4.1	1.0	0.6	0.0	1.0
Main income in one wave	2,609	26.5	30.1	17.9	10.9	5.7	4.1	1.6	1.5	1.1	0.6
Main income in both waves	499	21.3	20.6	18.6	13.8	9.8	6.6	4.6	2.6	0.4	1.9
Unit non-response in one wave	1,480	22.0	29.2	18.5	12.9	4.8	4.4	4.6	1.4	1.5	0.8
Unit non-response in both waves	1,214	15.9	28.0	20.5	9.4	12.4	6.1	3.9	1.8	1.1	0.8

Notes:

- a. The disposable household income is equalised using the OECD method where each person in the household is given a score (1 for the first adult, 0.5 for subsequent adults and 0.3 for each child < 18 years old) and the income is divided by the sum of these scores. Income deciles are assigned to each household by sorting the households by equalised disposable income and allocating 10 per cent of the weighted number of households to each decile. The household income deciles are assigned to each enumerated person in the household. The longitudinal enumerated person weights are used to calculate the number of deciles the population has moved between waves 1 and 2.
- b. n is the number of enumerated persons falling into each imputation category.

Possible Future Enhancements to the Imputation Method

The preceding analysis of the quality of the imputation has raised a number of issues regarding the imputation method that should be addressed in future releases (assuming resources are available):

- *Changes in the variables included in the regression models over time.* A comparison of the variables included in the regression models for waves 1 and 2 in Appendix 1 shows that there are a number of variables which appear in wave 1 and not in wave 2 and visa versa. With an additional wave of data, this should be reviewed with the view to constructing a more stable set of models.
- *Some zero values reported where only non-zero values imputed.* For business income and rental income, the preceding skips in the questionnaire do not necessarily restrict the income from that source to be non-zero. However, only non-zero income was imputed. For the most part, it is very likely the imputed amount should be non-zero, but this needs to be reviewed.
- *A small number of don't knows or refused at the screener questions.* There are a small number of cases that gave a 'don't know' or 'refused' response at the screener question for whether they had any income from a particular source. It is possible that some of these may actually have no income from that source,

but the imputed value would have been non-zero. This also needs to be reviewed in future data releases.

- *Inconsistencies between the detailed level data and the imputed variable.* We have adopted the practice of providing users with the data as reported at the detailed level and the post-imputed variable at a less detailed level. This restricts us from overwriting the underlying data of the recipient with the donor's information as is often done in other surveys. Some inconsistencies may be identified between the detailed level data and the imputed variable. To illustrate this, consider the following example: a person with two jobs may know their current wages from their main job, but may not know their wages from their second job. We would have imputed current wages and salaries, but would not have placed a restriction on the imputed amount to be more than their wages from their main job. Another example is not using the profit or loss information for business and rental income when this is known. The extent of such inconsistencies should be investigated and where possible resolved. In the first example, it may be more appropriate to apply the hourly rate of pay from the main job to the second job if the hours are known rather than to impute it.
- *Review whether to impute at the derived variable level only.* Currently only the derived variables for income and wealth are imputed. This means that any use of the finer level data collected in the questionnaire is subject to missingness. Imputation at the finer will depend on user demand, resources available and whether a suitable approach can be identified.
- *Donor identification.* The current method identifies the donor based on the predicted value of the regression model. The donor and the recipient may not match on a number of key variables (but we generally expect them to be close). We could improve this method by taking a more common sense donor that matches on a number of key variables from the neighbourhood of close donors.
- *Treatment of negative incomes.* Negative incomes have been mistreated in the modelling process (by being discarded). This does not mean that negative values have not been imputed. A neighbour with a positive predicted value but a negative actual value could be used to impute a missing case. This could be greatly improved by modelling whether an individual has a positive or negative income, and then modelling the amount to get a more realistic predicted value.
- *It may be better to impute the change rather than the level of income.* To address the issue of correlations between waves deteriorating with each wave imputed, it may be more appropriate to impute the change from one wave to the next rather than trying to impute the level of income at each wave and then infer the change. One potential problem with this approach is that it may lead to some unrealistic imputed values as we would not be imputing a response actually reported by an individual.

- *Use the same donor for multiple variables within one wave.* It may be appropriate for some variables to use the same donor to impute for multiple ‘missingness’. This will reduce the modelling work required and will improve the correlations between the variables imputed.
- *Use the same donor for multiple waves.* It may be appropriate to use the same donor across multiple waves where the recipient needs to be imputed in multiple waves.
- *Extend the imputation system to include income information from future waves.* It will be necessary to revise the income imputation to include information from wave 3 and possibly beyond into the imputation models for waves 1 and 2. Therefore, regardless of whether the preceding modifications are implemented, the income imputation in the next release will be different to Release 2.0.
- *Investigate alternative imputation methods.* Ideally, we would want to implement several imputation methods and compare their strengths and weaknesses before adopting the one best suited to the HILDA environment. Now that we have a more detailed understanding of the nearest neighbour method, it would be beneficial to further investigate several other methods and compare the results.

Wealth Imputation¹¹

Extent of Missing Wealth Data

Table 10 summarises the extent to which data from the wealth module were missing due to non-response. Only 61 per cent of households provided all wealth data. The total wealth of 34 per cent of the households could not be calculated as a responding individual had not provided complete information. In the remaining 5 per cent of households, total wealth could not be calculated solely due to the presence of non-respondents. When we consider the wealth components at the individual level, missing observations represented less than 11 per cent of all valid cases. Therefore, after imputation, a large part of a household's total wealth is based on actual data rather than imputed data.

Table 10: Number and proportion of cases with missing wealth data, Wave 2

<i>Wealth component</i>	<i>Missing cases (no.)^a</i>	<i>Valid cases (no.)^b</i>	<i>% of valid cases missing^c</i>	<i>% of all cases missing</i>
HOUSEHOLDS – Wealth components from the Household Questionnaire (HQ)				
Housing equity	531	5176	10.2	7.4
Equities	455	2978	15.0	6.3
Other cash-type investments	29	241	8.3	0.4
Trusts	123	390	29.0	1.7
Childrens' bank accounts	85	1399	5.8	1.2
Life insurance policies	200	794	24.1	2.8
Vehicles	145	6355	2.2	2.0
Collectibles	150	1050	12.0	2.1
Net business worth	231	1090	20.6	3.2
Total of HQ wealth components	1433	7245	19.7	19.7
RESPONDING PERSONS – Wealth components from the Person Questionnaire (PQ)				
Bank accounts	905	12825	7.0	6.9
Superannuation ^d	939	8843	10.5	7.2
Credit card debt	160	7448	2.1	1.2
Personal loans and other debts	174	3679	4.6	1.3
Total of PQ wealth components	1887	13041	14.5	14.5
HOUSEHOLDS				
Total household wealth	2846	7245	39.3	39.3

Notes:

- A 'missing case' is any observation where the respondent was unable to either indicate whether they had an asset or liability of the type in question, or were unable to provide a value for that asset or liability.
- A 'valid case' is any observation where the respondent reported owning the asset in question, having a credit card or having personal loans or debts.
- The figures reported in this column do not exactly equal 'missing cases' divided by 'valid cases'. This is because for all components there are a small number of cases where respondents did not answer the key screening question.
- In the case of superannuation assets, respondents were asked first to indicate which of seven broad monetary bands represented the current value of their superannuation. They were then asked to estimate the exact value of these assets within that band. For the purposes of this table we have only treated as missing those cases where individuals could not or would not choose a category. There are a total of 582 cases where a range was provided but not an exact value within that range and these cases have been imputed along with the other 'missing cases'.

¹¹ This section was prepared by Ellis Connolly, Kylie Smith and Marion Kohler of Economic Group, Reserve Bank of Australia.

Also note that missing cases are a higher proportion of valid cases than of all cases. This reflects the fact that households and persons were far more likely to know whether or not they have an asset than know the value of that asset. Therefore, the imputation is likely to lead to higher average asset and debt values for the entire data set.

Imputation principles

Due to the much lower level of missing data at the disaggregated level, missing wealth data were imputed at the detailed component level. Aggregate household wealth numbers were then obtained by adding the imputed and the actual data.

The one exception to this rule – similarly to that used for the income imputation – was for wealth data for non-responding persons. Since we have very little information on these persons directly, only total assets and total debts were imputed. These were then added to the total household financial assets and total household debts, respectively, of the enumerated person's household.

For many of the missing values we knew whether or not a household has an asset (or debt).¹² The regression was therefore estimated using only data from those households that have the asset (or debt) in question – this allowed us to avoid functional form problems arising from including a large number of zeros in the regression. Imputed values were therefore all non-zero for most types of wealth. In cases where a household or person can have an asset or debt with a zero value, such as bank accounts, business assets or credit card debt, it was possible that a zero value was imputed. For business debt, we did not know whether or not a household owning a business had business debt. Therefore, a large number of zeros were included in the regression for imputing business debt.

In order to obtain models with a high predictive value, imputed income data and wave 1 information, where available, were also used in the wealth imputation.

The imputed numbers were checked for implausible values on the basis of net wealth. In the few cases where the donor chosen lead to implausible or internally inconsistent values, another neighbouring donor was chosen. This usually affected only a handful of households. One notable case where a number of different donors were chosen was in the superannuation regression where non-retirees provided a range, but no specific estimate of the value. In 2.3 per cent of these non-retiree cases, the chosen donor was outside the range given. These were re-imputed with the extra restriction implied by the ranges.

¹² In a small number of cases where this information is also missing we assumed that a household owned the asset or debt in question.

Quality of Imputation

The quality of the imputation is likely to be affected by two factors: the quality of the regression model that was used; and any shortcomings of the imputation method (that is, a regression-based nearest-neighbour technique).

As can be expected, the fit of regression models for micro-economic data varies widely. The regressions for housing assets and debts were reasonably good, as were the regressions for retired superannuation, credit card debt and cash investments. The regressions for business assets and debt gave some cause for concern. This is because one potential indicator of the value of business assets, business income, could not be used due to serious concerns about reporting errors (a number of households showed a mismatch between whether they reported owning a business and receiving business income, which are surveyed in the Household and Person Questionnaires, respectively). Also, the lack of a screening question concerning business debt (i.e., a question whether or not households had business debt) resulted in the imputation of business debt being less accurate. Other regressions which had low predictive power related to smaller items, such as HECS loans or childrens' bank accounts.

As the choice of imputation methodology for the wealth data was guided by the methodology chosen for the income imputation, fewer quality tests were conducted for the wealth imputation than for the income imputation. However, a within-sample 10 per cent test showed that the imputed values lined up reasonably well with the actual values. The difference was especially small for the larger value items, such as property assets. Not surprisingly, the differences were larger for those variables where the regression model had a low predictive power, but fortunately these tended to be smaller items on the households' balance sheets. Given that it is a micro-data survey, the HILDA data (based on both imputed and unimputed values) compares surprisingly well with aggregate benchmarks, such as ratios of non-financial to financial assets, or gearing ratios. For more detail, see also Reserve Bank of Australia *Bulletin*, April 2004 and the quality paper for Wave 2 (Watson and Wooden 2004).

Imputed Variables Provided in Wave 1 and 2 Datasets

Where possible, we have sought to provide users with the pre-imputed variables (i.e. as reported variables), the post-imputed variables and a flag indicating which values are reported and which are imputed. While users only need the pre- and post-imputed variables or the post-imputed and the flag variables, we thought the extra flexibility of all three variables would be of assistance to users. The post-imputed variables contain the reported value for cases where no imputation was required.

An overview of the imputed income variables is provided in Table 11 and the imputed wealth variables are listed in Table 12. The first letter of the income variable names in Table 11 (represented as an underscore ‘_’) should be replaced by the letter corresponding to the wave (‘a’ for wave 1 and ‘b’ for wave 2).

Table 11: Imputed income variables provided in Release 2.0

	<i>Pre-imputed</i>	<i>Post-imputed</i>	<i>Flag</i>
Responding person file			
Current income			
Wages and salaries	_wsce	_wscei	_wscef
Benefits	_bnc	_bneci	_bncef
Financial year income^a			
Wages and salaries	_wsfe	_wsfe_i	_wsfef
Australian govt pensions	_bnfaup	_bnfaupi	_bnfaupf
Foreign govt pensions	_bnffp	_bnffpi	_bnffpf
Business income	_bifn, _bifp	_bifin, _bifip	_biff
Investments ^b	_oifinvn, _oifinvp	_oifinin, _oifinip	_oifinf
Private pensions	_oifpp	_oifppi	_oifppf
Private transfers	_oifpt	_oifpti	_oifptf
Total FY income ^c	Not provided	_tifefn, _tifefp	_tifeff
Windfall income	_oifwfl	_oifwfli	_oifwflf
Enumerated person file			
Total FY income ^c	Not provided	_tifefn, _tifefp	_tifeff
Windfall income	Not provided	_oifwfli	_oifwflf
Household file			
Total FY income ^d	Not provided	_hifefn, _hifefp	_hifeff
Windfall income	Not provided	_hifwfl	_hifwflf

Notes:

- Several sub-totals also provided on dataset (by summing imputed components): Australian pensions (_bnfatot including child care benefit and family tax benefit – relevant imputation flag will need to be created by user), market income (_tifmktn, _tifmktp, with flag _tifmktf), private income (_tifprin, _tifprip, with flag _tifprif).
- In the datasets, investment income is the combination of interest, dividends/royalties and rent. These were not meant to be provided separately, but two of the three imputed components have been included on the file by mistake (interest _oiinti, and dividends/royalties _oidvryi).
- The following variables use total person financial year income (_tifefn, _tifefp) in their calculations: income tax (_txinc), medicare (_txmed), total taxes (_txtot), disposable income (_tifdin, _tifdip). Use _tifeff as imputation flag for these variables.
- The following variables sum imputed person level information to household level: household total taxes (_hiftax), disposable income (_hifdin, _hifdip). Use _hifeff as imputation flag for these variables.

Table 12: Imputed wealth variables provided in Release 2.0

	<i>Pre-imputed</i>	<i>Post-imputed</i>	<i>Flag</i>
Responding person file			
Assets			
Joint bank accounts	bpwjbank	bpwjbani	bpwjbanf
Own bank accounts	bpwobank	bpwobani	bpwobanf
Superannuation – retirees	bpwsupr	bpwsupri	bpwsuprf
Superannuation – non-retirees	bpwsupwk	bpwsupwi	bpwsupwf
Debts			
HECS debt	bpwhecdt	bpwhecdi	bpwhecdf
Joint credit cards	bpwjccdt	bpwjccdi	bpwjccdf
Own credit cards	bpwoccdt	bpwoccdi	bpwoccdf
Other personal debt	bpwothdt	bpwothdi	bpwothdf
Enumerated person file			
Total person assets	Not provided	bpwassei	bpwassef
Total person debts	Not provided	bpwdebt	bpwdebt
Household file			
Assets			
Joint bank accounts*	bhwjbank	bhwjbani	bhwjbanf
Own bank accounts*	bhwobank	bhwobani	bhwobanf
Children’s bank accounts	bhwcbank	bhwcbani	bhwcbanf
Superannuation – retirees*	bhwsupr	bhwsupri	bhwsuprf
Superannuation – non-retirees*	bhwsuhwk	bhwsuhwi	bhwsuhwf
Business assets	bhwbusva	bhwbusvi	bhwbusvf
Cash investment	bhwcain	bhwcaini	bhwcainf
Equity investment	bhweqinv	bhweqini	bhweqinf
Collectables	bhwcoll	bhwcolli	bhwcollf
Home asset	bhwhmval	bhwhmvai	bhwhmvaf
Other property assets	bhwopval	bhwopvai	bhwopvaf
Life insurance	bhwinsur	bhwinsui	bhwinsuf
Trust funds	bhwtrust	bhwtrusi	bhwtrusf
Vehicles value	bhwvech	bhwvechi	bhwvechf
Total household assets	bhwasset	bhwassei	bhwassef
Debts			
HECS debt*	bhwhecdt	bhwhecdi	bhwhecdf
Joint credit cards*	bhwjccdt	bhwjccdi	bhwjccdf
Own credit cards*	bhwoccdt	bhwoccdi	bhwoccdf
Other personal debt*	bhwothdt	bhwothdi	bhwothdf
Business debt	bhwbusdt	bhwbusdi	bhwbusdf
Home debt	bhwhmdt	bhwhmdti	bhwhmdtf
Other property debt	bhwopdt	bhwopdti	bhwopdtf
Total household debts	bhwdebt	bhwdebt	bhwdebt

Notes: * Care should be taken when using these variables at the household level. These household variables are calculated as the sum of the equivalent wealth component for responding persons *only*. If non-responding adults exist in these household, no attempt to apportion their imputed total assets and debts to the person level components has been made, resulting in an *underestimate* of these components at the household level.

Note that in addition to total household assets and debts, several sub-totals and totals are also provided on dataset (by summing imputed components):

- business equity,
- investment equity,
- home equity,
- other property equity,
- total property equity,
- total credit card debt,
- total superannuation,
- total bank accounts,
- total property debt,
- total property value,
- household financial assets,
- household non-financial assets,
- net worth, and
- total assets and debts of non-respondents in responding households.

All relevant imputation flags have been provided – see HILDA wave 2 coding framework for details. These subtotals (except for household financial assets) exclude wealth imputations on non-responding persons, for whom only totals for debt and asset were imputed. (The non-responding person totals are included in the totals for household financial assets, net worth, total assets and debt.)

Conclusion

This paper has detailed the imputation methodology applied for waves 1 and 2 of the HILDA Survey.

An analysis of the quality of the income imputation reveals that the imputation is probably too variable when considering changes in income over time, but when using just cross-section data, it is acceptable. We expect that the imputation method will be revised prior to the next data release.

For the wealth imputation, only a small number of the wealth imputation models had low predictive power (such as, business assets and debt, children's bank accounts and HECS debt). However, these items tend to play a less prominent role in households' balance sheets compared with large items such as property, for which the imputation models produce acceptable results. The ratios based on aggregate wealth data from the HILDA Survey including imputed data compare reasonably well with aggregate benchmarks.

Any feedback from users on the imputed variables and any suggestions for improvements are most welcome. Please direct this feedback to Nicole Watson via email: n.watson@unimelb.edu.au.

References

Greene, W. H. (1993), *Econometric Analysis* (3rd edition), Prentice Hall, New Jersey.

Solon, G. (1989), 'The Value of Panel Data in Economic Research', in *Panel Surveys*, edited by Kasprzyk, D., Duncan, G., Kalton, G., and Singh, M.P., Wiley, New York.

Watson, N, and Wooden, M, (2002), 'Assessing the Quality of the HILDA Survey Wave 1 Data', HILDA Project Technical Paper Series No. 4/02, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Watson, N, and Wooden, M, (2003), 'Towards an Imputation Strategy for Wave 1 of the HILDA Survey', HILDA Project Discussion Paper Series No. 1/03, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Watson, N, and Wooden, M, (2004), 'Assessing the Quality of the HILDA Survey Wave 2 Data', HILDA Project Discussion Paper Series No. 4/04, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Appendix 1 – Variables Used in Income Models

The variables initially considered in all income models include:

Demographic characteristics	Demographic characteristics (c'td)	Employment characteristics	Partner characteristics (if applicable)	Wave 1 income (if available – imputed for Wave 2 models)
Age	Value of house	Usual hours worked in all jobs	Whether have partner	Current wages and salaries
Sex	Amount paid in mortgage	Occupational status	Age	Current benefits
Whether of pension age	Amount paid in rent	Occupation - 2 digit (present or most recent)	Sex	FY wages and salaries
Highest level of education	Number of motorbikes in HH	Industry – 2 digit (present or most recent)	Highest level of education	FY Aust govt pensions
Approximate number of years spent in education	Number of cars in HH	Labour force status	Approximate number of years spent in education	FY foreign govt pensions
Relationship in household	Whether eldest when growing up	Whether supervised other employees	Usual hours worked in all jobs	FY business income
Marital status	Number of siblings	Estimate of hours worked in last year	Labour force status	FY interest
Whether Aboriginal or Torres Strait Islander	Presence of long term health condition	Workplace size of main job	Occupational status	FY dividends/royalties
Number children aged 0 in HH	Hours spent caring	Tenure with current employer	Estimate of hours worked in last year	FY rent
Number children aged 1 to 4 in HH	Number of non-resident children aged 0 to 14	Tenure in current occupation	Proportion of last FY spent in employment	FY private pensions
Number children aged 5 to 14 in HH	Number of non-resident children aged 15+	Whether multiple job holder	Proportion of last FY spent in FT study	FY private transfers
Number children aged 0 to 14 in HH		Contract type	Proportion of last FY spent in PT study	FY total income
Number of other adults in HH		Type of employer's business	Proportion of last FY spent not in labour force	FY windfall
Whether dependent student		Proportion of last FY spent in employment	Proportion of last FY spent in unemployment	
Whether non-dependent child		Proportion of last FY spent in full-time study	Number of jobs held in calendar period	Wave 2 income (if available)
Remoteness area		Proportion of last FY spent in part-time study	Presence of long term health condition	Current wages and salaries
SEIFA index of educational disadvantage		Proportion of last FY spent not in labour force	First language spoken was language other than English	Current benefits
SEIFA index of economic resources		Proportion of last FY spent in unemployment		FY wages and salaries
SEIFA index of disadvantage		Number of jobs held in calendar period		FY Aust govt pensions
Time spent in Australia		Not employed		FY foreign govt pensions
Broad country of birth		Time since school spent not in labour force		FY business income
First language spoken was language other than English		Time since school spent in job		FY interest
Fathers broad occupation			Partners income (if available)	FY dividends/ royalties
Whether father employed when r aged 14			Current wages and salaries	FY rent
Whether father unemployed when r growing up			Current benefits	FY private pensions
Mothers broad occupation			FY wages and salaries	FY private transfers
HH expenditure on food			FY Aust govt pensions	FY total income
HH expenditure on groceries			FY foreign govt pensions	FY windfall
HH expenditure on meals outside home			FY business income	
Number of bedrooms in house			FY interest	
Whether renting, purchasing, owning or other			FY dividends/royalties	
			FY rent	
			FY private pensions	
			FY private transfers	
			FY total income	
			FY windfall	

Table A1: Variables used in regression model for wave 1 income imputation where wave 2 income was known

	<i>Current</i>		<i>Financial year income</i>									<i>Total</i>	<i>Windfall</i>
	<i>Wages and salaries</i>	<i>Benefits</i>	<i>Wages and salaries</i>	<i>Aust govt pensions</i>	<i>Foreign govt pensions</i>	<i>Business income</i>	<i>Interest</i>	<i>Dividends and royalties</i>	<i>Rent</i>	<i>Private pensions</i>	<i>Private transfers</i>		
Wave 2 income information													
Current wages and salaries	X		X										
FY wages and salaries	X		X										
FY Aust govt pensions		X		X									
FY business income						X							
FY interest							X						
FY dividends/ royalties								X					
FY rent									X				
FY private pensions										X			
FY private transfers											X		
FY total income												X	
FY windfall													X
Wave 1 income information													
Current wages and salaries			X										
Current benefits	X												
FY wages and salaries	X												
FY Aust govt pensions		X	X										
FY business income	X												
FY dividends/ royalties							X						
Demographic characteristics													
Age	X		X	X								X	X
Highest level of education			X										
Relationship in household												X	
Number children aged 0 in HH												X	
Remoteness area				X								X	
Time spent in Australia					X								
Employment characteristics													
Usual hours worked in all jobs	X		X										
Occupation - 2 digit (present or most recent)	X		X								X		
Industry – 2 digit (present or most recent)	X					X	X						X
Labour force status												X	
Estimate of hours worked in last year			X	X									

Table A1 (c'td)

	<i>Current</i>		<i>Financial year income</i>									<i>Total</i>	<i>Windfall</i>
	<i>Wages and salaries</i>	<i>Benefits</i>	<i>Wages and salaries</i>	<i>Aust govt pensions</i>	<i>Foreign govt pensions</i>	<i>Business income</i>	<i>Interest</i>	<i>Dividends and royalties</i>	<i>Rent</i>	<i>Private pensions</i>	<i>Private transfers</i>		
Employment characteristics (c'td)													
Contract type						X							
Type of employer's business	X												
Proportion of last FY spent in employment			X										
Proportion of last FY spent not in labour force			X									X	
Not employed						X						X	
Time since school spent not in labour force			X										
Time since school spent in job												X	
Partner characteristics													
Current benefits		X								X			
FY Aust govt pensions				X									
FY business income						X							
FY interest							X						
FY dividends/ royalties							X	X					
FY rent									X				
FY total income												X	
Labour force status												X	
Adjusted R-squared	0.82	0.26	0.76	0.29	0.00	0.21	0.40	0.39	0.18	0.19	0.40	0.55	0.26

Table A2: Variables used in regression model for wave 1 income imputation where wave 2 income was unknown

	<i>Current</i>		<i>Financial year income</i>								<i>Total</i>	<i>Windfall</i>
	<i>Wages and salaries</i>	<i>Benefits</i>	<i>Wages and salaries</i>	<i>Aust govt pensions</i>	<i>Foreign govt pensions</i>	<i>Business income</i>	<i>Interest</i>	<i>Dividends and royalties</i>	<i>Rent</i>	<i>Private pensions</i>		
Wave 1 income information												
Current wages and salaries			X							X		
Current benefits	X			X								
FY wages and salaries	X			X		X				X		
FY Aust govt pensions		X	X									
FY business income	X											
FY interest									X			
FY dividends/ royalties							X					
Demographic characteristics												
Age	X		X	X			X			X		X
Sex										X		
Relationship in household		X										X
Remoteness area				X								X
SEIFA index of educational disadvantage												X
Time spent in Australia					X							X
Broad country of birth												X
HH expenditure on food outside home											X	
Value of house											X	
Amount paid in rent											X	
Number of other adults in HH											X	
Employment characteristics												
Usual hours worked in all jobs	X	X	X									
Occupation - 2 digit (present or most recent)	X		X	X		X				X		X
Industry – 2 digit (present or most recent)	X		X	X			X	X		X		X
Labour force status												X
Whether supervised other employees									X			
Estimate of hours worked in last year	X		X			X						
Workplace size of main job		X										
Type of employer's business						X		X				
Proportion of last FY spent in employment			X	X						X		

Table A2 (c'td)

	<i>Current</i>		<i>Financial year income</i>								<i>Total</i>	<i>Windfall</i>	
	<i>Wages and salaries</i>	<i>Benefits</i>	<i>Wages and salaries</i>	<i>Aust govt pensions</i>	<i>Foreign govt pensions</i>	<i>Business income</i>	<i>Interest</i>	<i>Dividends and royalties</i>	<i>Rent</i>	<i>Private pensions</i>			<i>Private transfers*</i>
Employment characteristics (c'td)													
Proportion of last FY spend not in labour force			X										
Not employed	X											X	
Partner characteristics													
Current benefits		X											
FY Aust govt pensions		X		X									
FY business income						X							
FY interest							X						
FY dividends/ royalties								X					
FY rent									X				
FY total income												X	
Windfall income													X
Sex												X	
Highest level of education		X											
Labour force status												X	
Occupational status										X			
Estimate of hours worked in last year												X	
Adjusted R-squared	0.82	0.27	0.70	0.29	0.06	0.05	0.19	0.26	0.07	0.25	0.00	0.44	0.32

Notes: * No variables used in this model.

Table A3: Variables used in regression model for wave 2 income imputation where wave 1 income was known or imputed

	<i>Current</i>		<i>Financial year income</i>									<i>Total</i>	<i>Windfall</i>
	<i>Wages and salaries</i>	<i>Benefits</i>	<i>Wages and salaries</i>	<i>Aust govt pensions</i>	<i>Foreign govt pensions</i>	<i>Business income</i>	<i>Interest</i>	<i>Dividends and royalties</i>	<i>Rent</i>	<i>Private pensions</i>	<i>Private transfers</i>		
Wave 1 income information (imputed where necessary)													
Current benefits				X								X	
Current wages and salaries	X	X	X									X	X
FY wages and salaries	X	X	X								X	X	
FY Aust govt pensions				X								X	X
FY foreign pensions					X								
FY business income						X							
FY interest							X					X	
FY dividends/ royalties								X				X	
FY rent									X				
FY private pensions										X			X
FY private transfers											X		
FY total income			X					X				X	
FY windfall													X
Wave 1 income information													
Current wages and salaries		X	X	X									
Current benefits	X			X									
FY wages and salaries	X												
FY Aust govt pensions		X			X								
FY business income	X												
FY dividends/ royalties											X		
Demographic characteristics													
Age	X				X							X	X
Whether of pension age	X												
Highest level of education			X										X
Relationship in household											X		
Remoteness area		X											
SEIFA index of educational disadvantage												X	
SEIFA index of Economic Resources					X								
First language spoken was language other than English		X											
Fathers broad occupation					X						X		
Mothers broad occupation											X		X

Table A3 (c'td)

	<i>Current</i>		<i>Financial year income</i>									<i>Total</i>	<i>Windfall</i>
	<i>Wages and salaries</i>	<i>Benefits</i>	<i>Wages and salaries</i>	<i>Aust govt pensions</i>	<i>Foreign govt pensions</i>	<i>Business income</i>	<i>Interest</i>	<i>Dividends and royalties</i>	<i>Rent</i>	<i>Private pensions</i>	<i>Private transfers</i>		
Demographic characteristics (c'td)													
Value of house											X		X
Amount paid in rent					X								
Whether eldest when growing up											X		
Hours spent caring											X		
Employment characteristics													
Usual hours worked in all jobs	X												
Occupation - 2 digit (present or most recent)	X	X		X		X					X		
Industry – 2 digit (present or most recent)	X			X		X	X				X		X
Estimate of hours worked in last year	X		X	X		X							
Tenure with current employer	X												
Multiple job holder	X												
Whether multiple job holder	X												
Proportion of last FY spent in employment			X										
Proportion of last FY spent not in labour force											X		
Proportion of last FY spent in unemployment													X
Partner characteristics													
FY Aust govt pensions				X									X
FY foreign pensions					X								
FY business income						X							X
FY interest							X						
FY dividends/ royalties								X					
FY rent									X				
FY windfall income													X
Sex												X	
Highest level of education			X										X
Labour force status		X											
Adjusted R-squared	0.80	0.33	0.75	0.54	0.42	0.31	0.42	0.46	0.38	0.07	0.38	0.58	0.28

Table A4: Variables used in regression model for wave 2 income imputation where wave 1 income was unknown

	<i>Current</i>		<i>Financial year income</i>									<i>Total</i>	<i>Windfall</i>
	<i>Wages and salaries</i>	<i>Benefits</i>	<i>Wages and salaries</i>	<i>Aust govt pensions</i>	<i>Foreign govt pensions*</i>	<i>Business income</i>	<i>Interest</i>	<i>Dividends and royalties</i>	<i>Rent</i>	<i>Private pensions</i>	<i>Private transfers</i>		
Wave 2 income information													
Current wages and salaries		X	X										
Current benefits	X			X									
FY wages and salaries	X			X				X					
FY Aust govt pensions		X	X			X				X			
FY interest							X						
FY dividends/ royalties								X					
Demographic characteristics													
Age	X		X			X	X	X		X	X	X	X
Sex												X	
Relationship in household				X				X				X	
SEIFA index of economic resources											X		
SEIFA index of disadvantage												X	
Time spent in Australia								X					
Value of house												X	
Amount paid in mortgage												X	
Amount paid in rent												X	
Number of children aged 1 to 4												X	
Whether eldest when growing up											X		
Presence of long term health condition												X	
Employment characteristics													
Usual hours worked in all jobs	X												
Occupational status	X												
Occupation - 2 digit (present or most recent)								X	X		X		
Industry – 2 digit (present or most recent)	X	X				X		X		X			
Estimate of hours worked in last year	X		X					X					
Tenure with current employer	X												
Contract type								X					
Proportion of last FY spent in employment			X	X									

Table A4 (c'td)

	<i>Current</i>		<i>Financial year income</i>								<i>Total</i>	<i>Windfall</i>	
	<i>Wages and salaries</i>	<i>Benefits</i>	<i>Wages and salaries</i>	<i>Aust govt pensions</i>	<i>Foreign govt pensions*</i>	<i>Business income</i>	<i>Interest</i>	<i>Dividends and royalties</i>	<i>Rent</i>	<i>Private pensions</i>			<i>Private transfers</i>
Partner characteristics													
Current benefits												X	
FY Aust govt pensions				X								X	
FY business income						X						X	
FY interest							X						
FY dividends/ royalties							X						
FY rent								X					
FY private pensions								X					
Age												X	
Adjusted R-squared	0.81	0.30	0.67	0.39	0.00	0.12	0.24	0.38	0.38	0.15	0.13	0.38	0.24

Notes: * No variables used in this model.

Appendix 2 – Variables Used in Wealth Models

The variables considered in all wealth models include:

Household Income (logged)

Positive total income, and income squared
 Negative total income, and income squared
 Wages, and wages squared
 Benefits, and benefits squared
 Negative business income, and business squared
 Positive business income, and business squared
 Dividends, and dividends squared
 Interest, and interest squared
 Negative rent, and rent squared
 Positive rent, and rent squared
 Other, and other squared
Household Characteristics
 Year when moved
 Reason moved recently (closer to amenities, better neighbourhood, bigger place, to be closer to family, for lifestyle, related to marriage, smaller place, work reasons, business reasons)
 Own share in inv property
 Share owned in inv property
 Own share in home
 Share owned in home
 Have home loan
 Have second home loan
 Home loan from friend
 Previously owned a home
 Have life tenure
 Own home
 Rent
 Involved in rent-buy scheme
 Mortgage payments
 Payment schedule (ahead of schedule, behind schedule, on schedule)
 Have border in house

House Characteristics

Dwelling type (separate house, semi-detached 1 storey, semi-detached 2+ storey, flat shop, house shop, semi-detached attached to shop, flat in 1 storey block, flat in 2 storey block, flat in 3 storey block, flat in 4-9 storey block, flat in 10+ storey block, flat attached to house, caravan/houseboat/tent/cabin, other/public/no value)
 Number of bedrooms
 Home condition (excellent, good, average, poor, very poor, no value)
 Rental characteristics (rent a caravan, community rent, rent from employer, rent from govt, rent from real estate agent)
Family Characteristics
 Family type (single, couple with children, couple without children, mixed family, single parent, multifamily)
 Number of adults
 Average adult age
 Average child age
 Number employed
 Number of females
 Number of males
 Number of persons
 Number who speak poor English
 Number unhealthy persons
 Number of children
 Number born overseas

Other Assets/debt

Has ordinary vehicle
 Has recreational vehicle
 Has other vehicle
 Has credit card
 Has bonds
 Has business debt
 Has business
 Has life insurance policy
 Has trusts
 Sole beneficiaries of trusts
 Has Housing Investment Property
 Has Investment Property Loan
 Has HECS debt
 Has personal debt
 Number with personal debt
 Has ever had shares
 Has collectibles
 Retired with super
 Working with personal super
 Don't know working with personal super
 Has personal bank account
 Has joint bank account
 Has ATM access
 Number of businesses
 Have unincorporated Business
 Own share in business
 Reported value of 1 business
 Reported value of 2 businesses
 Reported value of 3 businesses
 Reported value of all businesses
 Pay off credit card (hardly ever pay off, not very often, about half the time, most months, always or almost always)

Household Reference Person (HRP)

Health
 Age, and age squared
 Gender
 Living at home
 Disabled
 Indigenous
 Children
 Like more children
 Marital status (defacto, divorced, married, never married, separated, widowed)
 Years since school
 Years in work
 Years not in labour force
 Years retired
 Years unemployed
 Years in Australia
 Years since moved out of home
 Under employed
 Over employed
 Union member
 Don't know if union member
 Top qualification (uni, diploma, year11, year12, other)
 Finished year 12
 Speak English
HRP Well-being
 Financial
 Amount of free time
 Home in which live
 Employment
 Neighbourhood
 Feeling part of community
 Safety
 Life

HRP Employment

Occupation (manager, tradesperson, labourer, clerical, professional, intermediate, no value)
 Full time employed
 Part time employed
 Unemployed
 Not in labour force (carer, home duties, illness, other, retired, student, travel, voluntary)
 Likelihood of losing job
 Likelihood of finding job
 Lose job voluntarily
 Weeks in occupation
 Number in workplace <20
 Number in workplace >20
 Level of job satisfaction
 Employment type (fixed, casual, permanent)
 Private/govt employment (private for profit, private not for profit, government)
 Employment benefits (paid holidays, paid sick days)
HRP History
 Parents divorced at some stage
 Number of siblings
 Have siblings
 Family status when 14 (living with own mother and father, living with father and stepmother, living with mother and stepfather, living with father only, living with mother only, living other)
 Region of birth(Australia, America, North Africa and Middle East, North East Asia, Non-Australia, Western Europe, New Zealand, Oceania, South Central Asia, South East Asia, South East Europe, Sub-Saharan Africa, UK and Ireland)

HRP Parent

Father's characteristics (father had job, don't know if father had job)

Father's occupation (manager, professional, tradesperson, clerical, intermediate, labourer)

Father's unemployment (father experienced unemploy, don't know if exper unemploy)

Mother's characteristics (mother had job, don't know if mother had job)

Mother's occupation (manager, professional, tradesperson, clerical, intermediate, labourer)

Parent's birthplace (Australia, America, North Africa and Middle East, North East Asia, New Zealand, Oceania, South Central Asia, South East Asia, North west Europe, South East Europe, Sub-Saharan Africa, UK and Ireland, no value)

Household Location (Broad)

Remoteness area (metropolitan, inner regional, outer regional, remote, very remote)

State (ACT, NSW, NT, Qld, SA, Tas, Vic, WA)

City (capital city, Adelaide, Brisbane, Hobart, Melbourne, Perth, Sydney)

Household Location (Detailed)

City ring (Inner Adelaide, Middle Adelaide, Outer Adelaide, Inner Hobart, Middle Hobart, Outer Hobart, Inner Perth, Middle Perth, Outer Perth, Inner Sydney, Middle Sydney, Outer Sydney, Inner Brisbane, Middle Brisbane, Outer Brisbane, Inner Melbourne, Middle Melbourne, Outer Melbourne)

Detailed Regions (These were derived from the postcode data and hence are not provided in greater detail here.)

Person Specific Characteristics

Age

Age squared

Speaks Language Other than English

Person Living Status (couple with dependents, couple without dependents, single parent, lone person)

Family Size

Has an ATM card

Years Worked Squared

Person Location (VIC non-capital city, NSW non-capital city, QLD non-capital city, SA non-capital city, TAS non-capital city, WA non-capital city, NT non-capital city)

Person Income

Positive Total Income

Negative Total Income

Wages and Salary

Government Income

Interest Income

Dividends Income

Positive Business Income

Negative Business Income

Positive Rental Income

Negative Rental Income

Person Assets/debt**Superannuation**

If employer makes super contributions

Percentage Employer Contribution

Only Receive Employer Contributions

Percentage Person Contrib of Wages

Don't know if personal contrib to own super

Don't know percentage contrib to own super

Don't know type of largest Super fund

Partner Makes Contribution to Super fund

Accumulation Fund

Defined Benefit Fund * Age

Defined Benefit Fund * Income

Accumulation Fund * Income

Super Range (\$5000 or less, \$5001 -

\$20000, \$20001 - \$50000, \$50001 -

\$100000, \$100001 - \$200000, \$200001

- \$500000)

Commonwealth Public Servant

HECS

Amount Annual Heecs Have to pay (logged)

Don't have to pay off any Heecs Annually

Started Uni 1997 or Earlier

Started Uni 1997 or Earlier * age

Bachelors Degree

Honours Degree

Household Total Income More \$100000

Other Personal Debt

Repayments Being Made on Personal debt

Table A5: Variables Used in Household and Non- Response Regressions

Home value				Home loan		
Share owned home	semi-detach shop	Age squared	Safety	positive total income	Other vehicle	Amt of free time
Positive total income	flat in 1 storey block	gender	living father & stepmother	Neg total income	Personal debt	Home in which live
Positive total income squared	flat in 2 storey block	Living at home	living mother only	Benefits	Hardly ever pay off	Number of siblings
Neg total income	flat in 3 storey block	Indigenous	America (hrp)	Neg business income	Not very often	Living other
Wages	flat in 4-9 storey block	Like more children	North East Asia (hrp)	Dividends	Most months	North East Asia (hrp)
Benefits	caravan etc	Divorced	South East Asia (hrp)	Interest	single parent	South East Europe (hrp)
Neg business income	Number of bedrooms	Separated	South East Europe (hrp)	other	Number of adults	Sub-Saharan Africa (hrp)
Neg business income squared	Home condition good	Occ - tradesperson (hrp)	don't know father job	Year when moved	Average adult age	Father had job
Positive business income	Home condition average	Occ - labourer (hrp)	Occ - tradesperson (f)	Better neighbourhood	Number of persons	don't know father job
Interest	Home condition poor	Occ - clerical (hrp)	Occ - clerical (f)	Bigger place	Number unhealthy persons	Occ - clerical (f)
Interest squared	Has inv Property Loan	Occ - professional (hrp)	Occ - intermediate (f)	Related to marriage	gender	Mother had job
Neg rent	Bonds	Occ - intermediate (hrp)	Occ - labourer (f)	Smaller place	Living at home	Occ - professional (m)
Neg rent squared	trusts	Part time employed	Occ - professional (m)	Have home loan	children	Occ - labourer (m)
Positive rent	Ever had shares	Unemployed	Occ - clerical (m)	Have second home loan	Never Married	North Africa & Middle East
Positive rent squared	Collectibles	NILF - carer	Occ - labourer (m)	home loan from friend	Occ - no value (hrp)	North East Asia
Other	Ordinary vehicle	NILF - illness	South East Europe	Previously owned a home	NILF - carer	South East Europe
Year when moved	Other vehicle	NILF - Retired	Inner regional	On schedule	NILF - home duties	Sub-Saharan Africa
Closer to amenities	credit card	NILF - travel	outer regional	Behind schedule	NILF - illness	Inner regional
Better neighbourhood	Personal debt	NILF - voluntary	remote	House shop	NILF - other	outer regional
Bigger place	Retired with super	Weeks in Occ	ACT	flat in 2 storey block	NILF - Retired	Middle Hobart
To be closer to family	ATM access	Number in workplace >20	NT	caravan etc	NILF - voluntary	Middle Perth
Related to marriage	Hardly ever pay off	Years in work	Qld	Other/public/no value	Weeks in Occ	Inner Sydney
Smaller place	Not very often	Years unemployed	SA	Home condition excellent	Number in workplace >20	Inner Brisbane
Work reasons	About half the time	Under employed	Vic	Home condition good	Over employed	Middle Melbourne
Business reasons	couple with children	Over employed	WA	Home condition average	don't know union member	Adj R-sq 0.49
Have second home loan	couple without children	Casual	Detailed regions	Home condition poor	Casual	
home loan from friend	mixed family	Paid holidays	Inner Adelaide	Home condition no value	Private not for profit	
Previously owned a home	single parent	Top qualification - diploma	Middle Adelaide	Has inv Property Loan	top qualification - year11	
On schedule	Average adult age	top qualification - year11	Outer Adelaide	Business	Top qualification - year12	
Behind schedule	Number employed	Top qualification - year12	Inner Hobart	trusts	top qualification - other	
Have border in house	Number of females	Home in which live	Middle Perth	Ever had shares	Finished year 12	
Semi-detached 1 storey	age	Neighbourhood	Adj R-sq 0.66	Ordinary vehicle	Financial	

Inv property			Inv property loan		Collectibles	
Share owned in inv property	Collectibles	Inner regional	Share owned in inv property	Over employed	Benefits	Tas
positive total income	Ordinary vehicle	outer regional	positive total income	Union member	Positive business income	Brisbane
Neg total income	Recreational vehicle	very remote	positive total income squared	Finished year 12	Dividends	Adj R-sq 0.25
Benefits	Other vehicle	ACT	Benefits	Parents divorced at some stage	Interest	
Neg business income	Retired with super	Qld	Neg business income	living mother & stepfather	positive rent	
Neg business income squared	Joint bank account	SA	Neg rent	living father only	other	
Dividends	Not very often	Tas	positive rent	Living other	Closer to amenities	
Interest	About half the time	Brisbane	Year when moved	America (hrp)	Business reasons	
Interest squared	multifamily	Adj R-sq 0.42	To be closer to family	Sub-Saharan Africa (hrp)	Have home loan	
Neg rent	Number of adults		Own share in inv property	Mother had job	Previously owned a home	
Neg rent squared	Number of persons		Own share in home	Occ - professional (m)	flat in 1 storey block	
positive rent	age		Have home loan	Occ - labourer (m)	Has Housing inv Property	
positive rent squared	Like more children		Previously owned a home	South East Asia	credit card	
Year when moved	Divorced		Have border in house	Sub-Saharan Africa	HECS debt	
Closer to amenities	Never Married		Other/public/no value	Inner regional	couple without children	
To be closer to family	Separated		rent	very remote	Number who speak poor english	
Related to marriage	Widowed		Rent from govt	NT	Number unhealthy persons	
Work reasons	Occ - tradesperson (hrp)		Home condition no value	Qld	health	
Own share in inv property	Occ - labourer (hrp)		Recreational vehicle	SA	Occ - clerical (hrp)	
Own share in home	NILF - illness		Personal debt	Tas	Occ - no value (hrp)	
Has inv Property Loan	NILF - other		Don't know working with personal super	Adelaide	Likelihood of finding job	
Have second home loan	Number in workplace >20		Personal bank account	Melbourne	Weeks in Occ	
Previously owned a home	Years unemployed		ATM access	Adj R-sq 0.38	Under employed	
Behind schedule	Union member		couple with children		Paid holidays	
Semi-detached 1 storey	Government		Number who speak poor english		Paid sick days	
Semi-detached 2+ storey	Top qualification - year12		Number unhealthy persons		Employment	
flat in 4-9 storey block	Finished year 12		Number born overseas		Feeling part of community	
caravan etc	Neighbourhood		Living at home		living father & stepmother	
Have life tenure	Feeling part of community		Separated		South East Europe (hrp)	
Rent from real estate agent	living father & stepmother		Unemployed		UK & Ire& (hrp)	
Business	Living other		NILF - voluntary		Occ - tradesperson (f)	
trusts	Occ - labourer (f)		Number in workplace >20		Don't know if father experienced unemployment	

Shares value			Trusts value		
positive total income	HECS debt	South Central Asia (hrp)	Sole beneficiaries of trusts	couple with children	Occ - intermediate (f)
positive total income squared	Retired with super	Occ - professional (f)	wages	couple without children	Occ - tradesperson (m)
Neg total income	Personal bank account	Mother had job	wages squared	mixed family	Occ - intermediate (m)
Benefits	ATM access	Occ - tradesperson (m)	Benefits	single parent	South East Asia
Dividends	Hardly ever pay off	America	Neg business income	Number of adults	North west Europe
Dividends squared	Not very often	Oceania	Positive business income	Average adult age	UK & Irel&
Interest	About half the time	South Central Asia	Dividends	Number employed	very remote
To be closer to family	multifamily	No value	Neg rent	age	Qld
Work reasons	Number of adults	outer regional	other	Age squared	WA
Own share in inv property	age	Qld	Closer to amenities	Disabled	Brisbane
Own share in home	Occ - tradesperson (hrp)	SA	Smaller place	Indigenous	Perth
Have home loan	Occ - clerical (hrp)	Tas	rent	Never Married	Adj R-sq 0.4
On schedule	Occ - professional (hrp)	Vic	Bonds	Widowed	
Behind schedule	Occ - intermediate (hrp)	WA	Business debt	Occ - tradesperson (hrp)	
semi-detach shop	NILF - carer	Hobart	Business	Occ - professional (hrp)	
flat attached to house	NILF - Retired	Adj R-sq 0.41	Ever had shares	NILF - Retired	
Home condition excellent	NILF - student		Own share in inv property	Likelihood of losing job	
Home condition good	NILF - travel		Have home loan	Likelihood of finding job	
Home condition average	Weeks in Occ		Have second home loan	Number in workplace >20	
Home condition poor	Years since school		Previously owned a home	Level of job satisfaction	
rent	Years retired		On schedule	Over employed	
Involved in rent-buy scheme	Years unemployed		House shop	Government	
Rent from govt	Years since moved out of home		flat in 1 storey block	Paid sick days	
Rent from real estate agent	Top qualification - year12		flat attached to house	Top qualification - year12	
Rent a caravan	top qualification - other		Other/public/no value	Parents divorced at some stage	
Home condition no value	Speak english		Home condition no value	living mother & stepfather	
Business debt	Financial		Recreational vehicle	Living other	
Business	Safety		Personal bank account	Western Europe (hrp)	
Life insurance policy	Have siblings		Joint bank account	South Central Asia (hrp)	
trusts	Living other		Own share in business	South East Asia (hrp)	
Collectibles	America (hrp)		Hardly ever pay off	Father had job	
credit card	Oceania (hrp)		About half the time	Occ - clerical (f)	

Children a/c value		Bonds value		Business Assets		
positive total income	Occ - tradesperson (hrp)	positive total income	Disabled	Have unincorporated Business	Occ - tradesperson (hrp)	don't know mother job
Neg total income	Part time employed	Neg total income	children	Neg business income	Occ - labourer (hrp)	Occ - professional (m)
Benefits	NILF - carer	wages	Defacto	Positive business income	Occ - clerical (hrp)	Occ - tradesperson (m)
Positive business income	Years in work	Benefits	Divorced	Dividends	Occ - professional (hrp)	North Africa & Middle East
Positive business income sq'd	Over employed	Dividends	Separated	positive rent	Unemployed	Oceania
positive rent	Financial	Interest	Occ - tradesperson (hrp)	other	NILF - home duties	Inner regional
Year when moved	Employment	Interest squared	Occ - labourer (hrp)	Closer to amenities	NILF - Retired	outer regional
Closer to amenities	Parents divorced at some stage	Own share in inv property	Occ - clerical (hrp)	Business reasons	Likelihood of losing job	remote
Has Housing inv Property	living father & stepmother	Have border in house	Occ - professional (hrp)	House shop	Lose job voluntarily	NT
Has inv Property Loan	North Africa & Middle East (hrp)	Home condition excellent	Occ - intermediate (hrp)	Has inv Property Loan	Weeks in Occ	WA
Business	Father had job	Home condition good	Part time employed	Business debt	Number in workplace >20	Brisbane
Life insurance policy	don't know father job	Home condition average	NILF - home duties	Life insurance policy	Level of job satisfaction	Hobart
trusts	Don't know if father experienced	Has Housing inv Property	NILF - Retired	Don't know working with	Don't know if father	Melbourne
Ever had shares	unemployment	Has inv Property Loan	Over employed	personal super	experienced unemployment	Perth
Collectibles	Mother had job	Business debt	Union member	Ever had shares	Union member	Adj R-sq 0.31
credit card	North Africa & Middle East	trusts	Paid holidays	Ordinary vehicle	Casual	
HECS debt	South East Europe	Collectibles	Paid sick days	Other vehicle	Top qualification - diploma	
Number with personal debt	Sub-Saharan Africa	Personal debt	Top qualification - year12	credit card	top qualification - year11	
Working with personal super	UK & Irel&	Number with personal debt	Financial	HECS debt	Speak english	
ATM access	No value	Hardly ever pay off	Have siblings	trusts	Financial	
Hardly ever pay off	Qld	Most months	New Zeal& (hrp)	ATM access	Feeling part of community	
Not very often	SA	couple without children	Oceania (hrp)	Not very often	Safety	
About half the time	Vic	mixed family	Mother had job	About half the time	Parents divorced at some	
Most months	WA	single parent	Occ - professional (m)	mixed family	stage	
multifamily	Melbourne	multifamily	Occ - intermediate (m)	Number of adults	living father only	
Average child age	Perth	Average adult age	New Zeal&	Average adult age	North Africa & Middle	
Number of children	Adj R-sq 0.24	Number employed	outer regional	Number employed	East (hrp)	
Average adult age		Number of females	WA	Number who speak poor english	Non-Australia (hrp)	
health		Number of persons	Adelaide	Number born overseas	New Zeal& (hrp)	
Age		Number unhealthy persons	Brisbane	gender	South East Europe (hrp)	
Age squared		Age	Perth	Indigenous	Occ - professional (f)	
Widowed		Age squared	Adj R-sq 0.61	Never Married	Years in work	
				Separated	Mother had job	

Business Debt		Vehicles			
Neg business income	Occ - tradesperson (hrp)	positive total income	trusts	Defacto	Non-Australia (hrp)
Neg business income squared	Part time employed	positive total income squared	Ever had shares	Widowed	New Zeal& (hrp)
Interest	Unemployed	Benefits squared	Collectibles	Occ - tradesperson (hrp)	South East Asia (hrp)
positive rent	Lose job voluntarily	Neg business income	Ordinary vehicle	Occ - labourer (hrp)	South East Europe (hrp)
positive rent squared	Weeks in Occ	Neg business income squared	Recreational vehicle	Occ - clerical (hrp)	Father had job
other	Number in workplace >20	Interest	Other vehicle	Occ - professional (hrp)	Occ - professional (f)
Smaller place	Level of job satisfaction	Neg rent	credit card	Part time employed	Occ - tradesperson (f)
Work reasons	Years since school	positive rent	HECS debt	Unemployed	Occ - clerical (f)
home loan from friend	Years retired	Year when moved	Personal debt	NILF - travel	Occ - intermediate (f)
Previously owned a home	Union member	To be closer to family	Number with personal debt	NILF - voluntary	Occ - labourer (f)
On schedule	don't know union member	Related to marriage	Retired with super	Weeks in Occ	Father experienced unemployment
Behind schedule	Government	Own share in inv property	Working with personal super	Level of job satisfaction	America
Have border in house	Financial	Have home loan	Joint bank account	Years since school	North East Asia
caravan etc	Living other	On schedule	Own share in business	Years in work	South East Asia
Home condition no value	South Central Asia (hrp)	Have border in house	Hardly ever pay off	Years unemployed	outer regional
Ordinary vehicle	UK & Irel& (hrp)	Semi-detached 1 storey	couple with children	Years in Australia	ACT
Other vehicle	Occ - clerical (f)	Flat shop	couple without children	Over employed	SA
Bonds	Mother had job	semi-detach shop	mixed family	Government	Tas
trusts	Occ - intermediate (m)	flat in 2 storey block	single parent	Paid holidays	Vic
Working with personal super	South Central Asia	Number of bedrooms	multifamily	Paid sick days	WA
Don't know working with personal super	outer regional	Home condition excellent	Number of adults	top qualification - year11	Melbourne
Joint bank account	SA	Home condition good	Average adult age	Finished year 12	Perth
Hardly ever pay off	Vic	Home condition average	Number of persons	Speak english	Adj R-sq 0.43
Not very often	Melbourne	Home condition poor	Number who speak poor english	Financial	
Most months	Adj R-sq 0.18	Have life tenure	Number unhealthy persons	Amt of free time	
single parent		rent	health	Home in which live	
Number who speak poor english		Involved in rent-buy scheme	age	Neighbourhood	
health		Rent a caravan	Age squared	Feeling part of community	
age		Home condition no value	gender	Number of siblings	
Age squared		Business debt	Living at home	living father & stepmother	
Indigenous		Has Housing inv Property	children	Living other	
Separated		Business	Like more children	America (hrp)	

Life Insurance Policies			Assets: No W1		Debt: No W1	
Neg total income wages	Own share in business	Paid sick days	Age	Benefits (hh)	Age	About half the time (hrp)
Neg business income	Hardly ever pay off	Top qualification - diploma	Gender	dividends (hh)	Gender	Most months (hrp)
Dividends	Not very often	Top qualification - year12	Disabled	Interest income (hh)	Disabled	Benefits (hh)
Interest	About half the time	Finished year 12	Personal Bank account	Neg rent income (hh)	Joint bank account	Interest income (hh)
Interest squared	single parent	Amt of free time	Joint bank account	Wages (hh)	Own credit card	Other income (hh)
positive rent	multifamily	Parents divorced at some stage	working super	Indigenous (hrp)	HECS debt	Wages (hh)
other	Average adult age	America (hrp)	Personal debt	Years since school (hrp)	Personal debt	Finished year 12 (hrp)
Closer to amenities	Number employed	North Africa & Middle East (hrp)	Positive total income	Years in work (hrp)	Positive total income	Years in work (hrp)
To be closer to family	Number of females	North East Asia (hrp)	Neg total income	Years unemployed (hrp)	Neg total income	Underemployed (hrp)
Related to marriage	Number of persons	New Zeal& (hrp)	Number of adults	Paid sick leave (hrp)	Number of persons	Permanent (hrp)
Work reasons	Number who speak poor english	don't know father job	Average adult age	Weeks in occ (hrp)	Average adult age	Paid sick leave (hrp)
Business reasons	Number unhealthy persons	Occ - professional (f)	vehicle	Separated (hrp)	vehicle	Widowed (hrp)
Have second home loan	age	Occ - tradesperson (f)	Recreational vehicle	Widowed (hrp)	Bonds	defacto (hrp)
Previously owned a home	Age squared	Occ - labourer (f)	Other vehicle	Never married (hrp)	Ever owned shares	Don't know Union member (hrp)
Have border in house	gender	America	Business	Diploma (hrp)	Owned home previously	Satisfied home (hrp)
Flat shop	children	Inner regional	Ever owned shares	Just year 11 (hrp)	Business debt	Satisfied employment (hrp)
flat in 3 storey block	Like more children	outer regional	Share in inv property	Don't know union member (hrp)	Inner regional	Satisfied financial (hrp)
flat in 4-9 storey block	Defacto	remote	second home loan	Satisfied with home (hrp)	outer regional	Satisfied safety (hrp)
caravan etc	Never Married	NT	Friend loan	Satisfied financial (hrp)	Remote	Like more children (hrp)
Home condition excellent	Separated	Qld	Rent	Satisfied neighbourhood (hrp)	ACT	Adj R-sq 0.22
Home condition good	NILF - home duties	Tas	Remote	Like more children (hrp)	QLD	
Home condition average	NILF - illness	Vic	NT	Adj R-sq 0.49	SA	
Have life tenure	NILF - other	WA	SA		VIC	
Has Housing inv Property	NILF - Retired	Hobart	WA		Inner Adelaide	
Business debt	Level of job satisfaction	Melbourne	Inner Adelaide		Middle Adelaide	
trusts	Years in work	Perth	Middle Adelaide		Inner Hobart	
Ordinary vehicle	Years since moved out of home	Adj R-sq 0.23	Outer Adelaide		Middle Hobart	
credit card	Under employed		Middle Hobart		Outer Perth	
HECS debt	Union member		Middle Perth		Inner Melbourne	
Number with personal debt	don't know union member		Outer Perth		Middle Melbourne	
Joint bank account	Casual		Inner Sydney		Hardly ever pay off (hrp)	
	Government		Outer Brisbane		Not very often (hrp)	

Assets: W1			Debt: W1		
Age	Business	Satisfied home (hrp)	Age	Positive total income (hrp)	Satisfied Safety (hrp)
Gender	Ever had shares	Satisfied financial (hrp)	Gender	Neg total income (hrp)	<u>Adj R-sq 0.32</u>
Disabled	Share in inv property	Satisfied safety (hrp)	Wages W1	Numer of persons	
Wages W1	Business debt	Like more children (hrp)	Positive Business income W1	Vehicles	
Benefits W1	Second home loan	Adj R-sq 0.55	Interest income W1	Previously owned home	
Interest income W1	Friend loan		Positive total income W1	Business debt	
Positive total income W1	Rent		Negative total income W1	Inner regional	
Neg total income W1	NT		Speak English W1	remote	
Separated W1	SA		Separated W1	Very remote	
Divorced W1	Middle Adelaide		Part-time Employed W1	ACT	
Widowed W1	Outer Adelaide		Unemployed W1	NT	
Never married W1	Middle Hobart		Retired W1	QLD	
Part-time Employed W1	Outer Hobart		Home duties W1	SA	
Unemployed W1	Inner Perth		Student W1	VIC	
Student W1	Middle Sydney		Indigenous W1	Middle Adelaide	
Years retired W1	Outer Brisbane		Finished year 12 W1	Middle Hobart	
Just year 12 W1	Middle Melbourne		Diploma W1	Inner Sydney	
Just year 11 W1	Most months (hrp)		Just year 11 W1	Inner Melbourne	
Years since school W1	Benefits (hh)		Other qualification W1	Middle Melbourne	
Years in work W1	Positive business income (hh)		Years in work W1	Hardly ever pay off (hrp)	
Paid holidays W1	dividends (hh)		Permanent W1	Not very often (hrp)	
government W1	Interest income (hh)		Satisfied financial W1	About half the time (hrp)	
Satisfied financial W1	rent income (hh)		Credit card W1	Most months (hrp)	
Credit card W1	Wages (hh)		hardly ever pay off W1	Indigenous (hrp)	
hardly ever pay off W1	Indigenous (hrp)		Not very often W1	Years in work (hrp)	
Joint bank account	Years in work (hrp)		Half the time W1	Years unemployed (hrp)	
Retired super	Lose job (hrp)		Most months W1	Underemployed (hrp)	
Working super	Paid Holidays (hrp)		Joint bank account	Casual (hrp)	
Personal debt	Paid sick leave (hrp)		Own credit card	Permanent (hrp)	
Positive total income (hrp)	Weeks in occ (hrp)		Joint credit card	Just year 11 (hrp)	
Neg total income (hrp)	Satisfied job (hrp)		HECS debt	Don't know union member (hrp)	
Number of adults	Widowed (hrp)		Personal debt	Satisfied financial (hrp)	

Table A6: Variables Used in Person Regressions

Own Bank Accounts			Joint Bank Accounts		Retired Superannuation	
Wages and Salary	NT	Recreational vehicle	Wages and Salary	Personal bank account	Government Income	America
Government Income	ACT	trusts	Government Income	credit card	Other Income	Sub-Saharan Africa
Interest Income	Vic	Ever had shares	Interest Income	Retired with super	Positive Rental Income	Related to marriage
Positive Business Income	SA	Own share in inv property	Positive Rental Income	Personal debt	Positive Total Income	Not very often
Positive Total Income	Living other	Own share in home	Single parent	Home in which live	Couple without dependents	About half the time
Couple without dependents	Parents divorced at some stage	Have home loan	gender	Financial Satisfaction	Single parent	Ordinary vehicle
Single parent	Years since moved out of home	Have second home loan	Age Squared	Safety Satisfaction	gender	Recreational vehicle
gender	Number of siblings	home loan from friend	Separated	Feeling part of community	age	Other vehicle
age	Father experienced unemployment	rent	Divorced	Amount of free time	Age Squared	Business debt
Widowed	Melbourne	Involved in rent-buy scheme	Unemployed	New Zealand	Speak english	trusts
Defacto	Adelaide	Adj R-sq 0.41	NILF - illness	Oceania (parents)	Top qualification - diploma	Have second home loan
Never Married	VIC - Non-capital City		NILF - carer	Sub-Saharan Africa (parents)	top qualification - year11	Previously owned a home
Like more children	SA - Non-capital City		Years retired	Oceania	Years Worked Squared	Have life tenure
Unemployed	Has an atm card		Indigenous	UK and Ireland	Family Size	rent
NILF - Retired	Joint bank account		Years unemployed	South East Europe	children	Adj R-sq 0.46
NILF - home duties	credit card		Over employed	South Central Asia	multifamily	
NILF - student	Retired with super		Government	To be closer to family	Qld	
NILF - illness	HECS debt		Weeks in occupation	Hardly ever pay off	WA	
NILF - voluntary	Personal debt		Family Size	Not very often	Living at home	
Indigenous	Home in which live		children	Ordinary vehicle	Number of siblings	
top qualification - year11	Financial		Life Satisfaction	Other vehicle	Years in Australia	
top qualification - other	Safety		multifamily	Bonds	SA - Non-capital City	
Years unemployed	Amount of free time		ACT	Business	credit card	
Permanent	North Africa and Middle East		Vic	Ever had shares	HECS debt	
Paid sick days	North East Asia		SA	Have home loan	Financial	
Government	South Central Asia		Tas	Have second home loan	South Central Asia (parents)	
Weeks in occupation	To be closer to family		Living with father and	Previously owned a home	America (parents)	
Occupation - intermediate	Hardly ever pay off		stepmother at 14	rent	Sub-Saharan Africa (parents)	
Occupation - labourer	Not very often		Living with mother only at 14	Adj R-sq 0.27	New Zealand	
Level of job satisfaction	About half the time		Father experienced		UK and Ireland	
Life	Most months		unemployment		North Africa and Middle East	
multifamily	Ordinary vehicle		Melbourne		North East Asia	
			VIC - Non-capital City			
			Inner regional			

Working Superannuation			Own Credit Card Debt		Joint Credit Card Debt	
Wages and Salary	Occupation - labourer	Super \$5000 or Less	Parents divorced at some stage	remote	Government Income	Interest Income
Government Income	Occupation - no value	Super \$5001 - \$20000	Speaks Language Other than English	very remote	Feeling part of community	UK and Ireland (parents)
Other Income	Family Size	Super \$20001 - \$50000	Interest Income	Personal bank account	Satisfaction	South East Asia (parents)
Interest Income	Life	Super \$50001 - \$100000	Dividends Income	Personal debt	Couple with dependents	America
Negative Business Income	ACT	Super \$100001 - \$200000	Positive Total Income	Home in which live	Couple without dependents	Bigger place
Positive Total Income	SA	Super \$200001 - \$500000	Negative Total Income	Financial	gender	Smaller place
Negative Total Income	Tas	Adj R-sq 0.84	Single parent	North East Asia (parents)	Separated	Business debt
Couple with dependents	Adelaide		gender	South Central Asia (parents)	Never Married	trusts
Couple without dependents	SA - Non-capital City		age	Hardly ever pay off	NILF - student	Ever had shares
Gender	remote		Age Squared	Not very often	NILF - voluntary	Have second home loan
Age	very remote		Government Income	About half the time	Finished year 12	Have life tenure
Age Squared	Has an atm card		Widowed	Most months	Top qualification - year12	Involved rent-buy scheme
Disabled	credit card		Unemployed	Other vehicle	Years unemployed	Hardly ever pay off
Widowed	Personal debt		NILF - Retired	Business	Under employed	Not very often
Like more children	Financial		NILF - home duties	Own share in home	Paid holidays	About half the time
Unemployed	Safety		NILF - carer	Have second home loan	Paid sick days	Most months
NILF - home duties	South Central Asia (parents)		NILF - travel	Have life tenure	children	Adj R-sq 0.65
NILF - student	Work reasons		Years retired	rent	Life Satisfaction	
NILF - illness	Ordinary vehicle		Likelihood of losing job	Involved rent-buy scheme	SA	
NILF - carer	Business debt		Weeks in occupation	Adj R-sq 0.64	Living with mother and stepfather	
NILF - travel	Business		Level of job satisfaction		Living with father only	
NILF - voluntary	Previously owned a home		multifamily		Living with mother only	
top qualification - year11	If employer makes super contributions		Vic		Parents divorced at some stage	
Years since school	Don't know if personal contributions		Qld		Father had job	
Years in work	to own super fund		SA		Father experienced unemployment	
Years Worked Squared	Don't know percentage contributions		Wages and Salary		outer regional	
Years unemployed	to own super fund		Mother had job		remote	
Under employed	Don't know type of largest Super fund		Melbourne		Personal debt	
Lose job voluntarily	Partner Makes Contribution to Super fund		Brisbane		Financial Satisfaction	
Paid holidays	Defined Benefit Fund * Income		VIC - Non-capital City			
Occupation - tradesperson	Accumulation Fund * Income		Inner regional			
Occupation - intermediate						

Working Superannuation (No Range)			HECS Debt		Personal Debt	
Wages and Salary	Years unemployed	Feeling part of community	Wages and Salary	Tas	Wages and Salary	Life
Government Income	Casual	South Central Asia (parents)	Government Income	Years since moved out of home	Government Income	SA
Interest Income	Permanent	New Zealand	Other Income	Number of siblings	Dividends Income	WA
Dividends Income	Lose job voluntarily	North East Asia	Interest Income	Father had job	Positive Business Income	Tas
Positive Business Income	Likelihood of finding job	America	Dividends Income	Adelaide	Negative Business Income	America
Negative Business Income	Paid holidays	Work reasons	Positive Business Income	Hobart	Positive Rental Income	Living with father only
Positive Total Income	Government	Business reasons	Positive Total Income	remote	Negative Rental Income	Living with mother only
Negative Total Income	Number in workplace >20	Hardly ever pay off	gender	Has an atm card	Couple with dependents	South East Europe
Couple with dependents	Weeks in occupation	Recreational vehicle	age	Working with personal super	Couple without dependents	Number of siblings
Couple without dependents	Occupation - tradesperson	Other vehicle	Age Squared	Home in which live	gender	VIC - Non-capital City
Speaks Language Other than English	Occupation - no value	Business debt	Speaks Language Other than English	Amount of free time	Don't know if union member	Living with mother and stepfather
gender	Union member	Business trusts	Disabled	UK and Ireland (hrp)	Age Squared	Inner regional
age	Family Size	Ever had shares	Unemployed	For lifestyle	Never Married	remote
Age Squared	children	Own share in inv property	NILF - home duties	Recreational vehicle	Part time employed	very remote
NILF - illness	multifamily	Have home loan	NILF - student	Other vehicle	Unemployed	credit card
Disabled	NT	Have second home loan	NILF - voluntary	trusts	NILF - student	Retired with super
Divorced	Parents divorced at some stage	Commonwealth Public Servant	Indigenous	Own share in home	NILF - voluntary	HECS debt
Widowed	Years since moved out of home	If employer makes super contributions	Finished year 12	Previously owned a home	Top qualification - diploma	Employment
Defacto	Living at home	Percentage Employer Contribution	Top qualification - diploma	Involved in rent-buy scheme	Top qualification - year12	Financial
Like more children	Years in Australia	Only Receive Employer Contrib	top qualification - other	How much Annual Hecs Have to pay (logged)	top qualification - year11	Feeling part of community
Part time employed	SA - Non-capital City	Percentage Person Contrib of Wages	Years since school	Household Total Income is	top qualification - other	Whether Repayments Being
NILF - student	WA - Non-capital City	Don't know if pers contrib to own super	Under employed	More than \$100000	Years in work	Made on Personal debt
Single parent	Inner regional	Don't know percent contrib to own super	Likelihood of losing job	Started Uni 1997 or Earlier	Permanent	Parents divorced at some stage
NILF - carer	outer regional	Don't know percent contrib to own super	Lose job voluntarily	Started Uni 1997 or Earlier * age	Likelihood of losing job	WA - Non-capital city
NILF - voluntary	remote	Full time employed	Likelihood of finding job	Bachelors Degree	Paid sick days	Sub-Saharan Africa
Finished year 12	very remote	Don't know type of largest Super fund	Paid holidays	Honours Degree	Occupation - clerical	Business reasons
Top qualification - diploma	Has an atm card	Accumulation Fund	Weeks in occupation	Don't have to pay off any Hecs Annually	Occupation - intermediate	Ordinary vehicle
Top qualification - year12	Joint bank account	Partner Makes Contribution to Super	Occupation - clerical		Occupation - labourer	Other vehicle
top qualification - year11	credit card	Accumulation Fund * Income	Occupation - intermediate children	Adj R-sq 0.27	Union member	Bonds
Years since school	HECS debt	Defined Benefit Fund * Age	Qld		Age	trusts
Years in work	Personal debt		SA		Level of job satisfaction	Own share in home
Years Worked Squared	Financial				health	Have second home loan
						Oceana (parents)
						Adj R-sq 0.22
		Adj R-sq 0.58				