



THE UNIVERSITY OF  
MELBOURNE

**HILDA PROJECT TECHNICAL PAPER SERIES**  
**No. 1/10, February 2010**

---

**HILDA Expenditure Imputation**

*Claire Sun*

---

**The HILDA Project was initiated, and is funded, by the Australian  
Government Department of Families, Housing, Community Services  
and Indigenous Affairs**

## **Acknowledgements**

I appreciate the assistance that Nicole Watson provided at various stage of the project. I am also grateful for the comments and suggestions provided by Roger Wilkins, Bruce Headey and the HILDA Technical Reference Group (John Henstridge, Stephen Horn, Frank Yu).

## Contents

<b>Introduction</b> .....	<b>1</b>
<b>Imputed Expenditure Variables Provided in Release 8 Datasets</b> .....	<b>2</b>
<b>Missing Data</b> .....	<b>3</b>
<b>Expenditure Imputation</b> .....	<b>5</b>
<i>Step 1: Identifying Longitudinal Households</i> .....	6
<i>Step 2: Identifying Lumpy Expenditure Items</i> .....	6
<i>Step 3: Population Carry-over Method</i> .....	7
<i>Step 4 and 5: Nearest Neighbour Regression Method</i> .....	10
<i>Step 6: Little and Su Method</i> .....	11
<b>Comparison with the Household Expenditure Survey (HES)</b> .....	<b>13</b>
<b>Quality of Imputation</b> .....	<b>17</b>
<i>Effects of Imputation on the Expenditure Distribution</i> .....	17
<i>Effects of Imputation on the Relationship Between Income and Expenditure</i> .....	20
<i>Weighted Average and Proportion of Mean Expenditure Imputed</i> .....	21
<i>Cross-wave Comparison</i> .....	22
<b>Childcare Expenditure</b> .....	<b>23</b>
<b>Concluding Remarks</b> .....	<b>24</b>
<b>References</b> .....	<b>25</b>
<b>Appendix 1: Overall Imputation Process for the Expenditure Items</b> .....	<b>27</b>
<b>Appendix 2: Variables Used in the Nearest Neighbour Regression Models</b> .....	<b>28</b>
<b>Appendix 3: Distribution of Expenditure Data Before and After Imputation, Waves 1 to 7</b> .....	<b>29</b>

## **Introduction**

The economic well-being of households is traditionally measured by the income they receive, however, a number of researchers have argued that consumption is a better measure of material well-being than income (for example, Ringen, 1988; Crossley and Pendakur, 2002). Headey (2008) has further argued that poverty measures should take into account of income, wealth and consumption.

The HILDA Survey has moved towards providing a full set of household financial accounts with the collection of income each wave, wealth on a four-year cycle starting in wave 2 and expenditure each wave from wave 5. Whilst we would ideally like to measure consumption (the ‘using up’ of goods and services) rather than expenditure (the amount paid for goods and services), this is a difficult concept to convey to respondents and it is not easy to measure actual consumption. Household expenditure on non-durable items (such as groceries, fuel and holiday expenditure) during the period is likely to have a close correspondence to consumption of these items, but estimation of consumption of durable items requires an estimate of the value of the services derived from the stock of durables held by the household (Wilkins et al. 2009). Expenditure on these durable items might only be a proxy measure of the consumption of these items.

We first collected detailed household expenditure information in wave 5. The list of items was expanded to include consumer durables in wave 6 and the definitions of some items were revised. It should be noted that the HILDA Survey does not attempt to measure all components of household expenditure and therefore, it does not provide a complete picture of household expenditure behaviour.

Until recently, the dominant view has been that a diary method is essential to achieve a valid measure of expenditure because without the assistance of the diary, the respondents would not be able to recall the expenditure correctly. The national expenditure surveys, like the ABS Household Expenditure Survey (HES), ask the respondents to fill in a detailed shopping diary over a certain period of time. However, the work done by Browning et al. (2003) suggests it is possible to derive an accurate recall-based measure of total expenditure by asking about an exhaustive list of highly disaggregated expenditure items. In fact, some items of expenditure are more validly reported in recall-based questions than a diary, because the respondents report their ‘usual spending’ on the item rather than spending during a period which may be atypical (Browning et al, 2003). The HILDA Survey asks retrospective questions on household expenditure, and these are predominantly included in the Self-Completion Questionnaire (SCQ). The people responsible for household bills are asked to fill in the household expenditure questions. As shown later in this paper, we have had reasonable success in measuring expenditure. Out of the 23 items we collected from wave 6 onwards, for 13 items, estimates of means from the HILDA data are shown to be close to means obtained from the HES.

As mentioned earlier, most expenditure items are collected in the SCQ from wave 5 onwards. We also collect some expenditure items (rent payments, mortgage repayments) in the Household Questionnaire (HQ) from wave 1. The expenditure data collected in the SCQ are more likely to be missing than most other data, mostly because some respondents do not return the SCQ. For the items collected in the SCQ, the proportion of households with missing expenditure information is around 17% across the four waves.

The missing expenditure items were first imputed in Release 8. Imputation on expenditure data was applied at the household level. The overall imputation process is similar to income and wealth imputation, where the Little and Su method, the nearest neighbour regression method and the population carry-over method are employed.

### Imputed Expenditure Variables Provided in Release 8 Datasets

This section lists all expenditure variables collected that have been imputed for Release 8. As with the income and wealth imputation, we have provided users with the pre-imputed variables, the post-imputed variables and a flag indicating whether the value is imputed or not. The post-imputed variables contain the reported value for cases where no imputation was required and the imputed value for those that do.

The expenditure imputation was undertaken at the household level. Table 1 provides an overview of the imputed variables for the household file. The first letter of the variable names in each table (represented as underscore ‘\_’) should be replaced by the letter corresponding to the wave. For example, ‘a’ for wave 1 and ‘b’ for wave 2 etc. As noted earlier, the scope of the expenditure variables vary over time:

- Usual payments on rent, usual repayments on first mortgage and second mortgage per month are collected in the HQ in every wave from wave 1.
- Weekly household expenditure on all groceries, groceries for food and drink, and meals eaten outside are collected in the HQ in waves 1, 3, 4 and 5.
- The annualised household expenditure items are derived from the variables collected in the SCQ from wave 5 onwards. The expenditure questions in the SCQ were revised in wave 6, which resulted in a slightly different set of expenditure components collected from wave 6 onwards.

**Table 1: Imputed expenditure variables provided in the Release 8 household files**

	<i>Wave</i>	<i>Pre-imputed</i>	<i>Post-imputed</i>	<i>Flag</i>
<b>Household File</b>				
<i>Usual payments/repayments per month (Collected in the HQ)</i>				
Rent	1 - 8	_hsrnt	_hsrnti	_hsrntfg
First mortgage	1 - 8	_hsmg	_hsmgi	_hsmgfg
Second mortgage	1 - 8	_hssl	_hssli	_hsslfg
<i>Weekly household expenditure (Collected in the HQ)</i>				
All groceries	1, 3, 4, 5	_xpgroc	_xpgroci	_xpgrocf
Groceries for food and drink	1, 3, 4, 5	_xpfood	_xpfoodi	_xpfoodf
Meals eaten outside	1, 3, 4, 5	_xposml	_xposmli	_xposmlf
<i>Annualized household expenditure (Collected in the SCQ)</i>				
Groceries	5 - 8	_hxygroc	_hxygreci	_hxygrcf
Alcohol	5 - 8	_hxyalc	_hxyalci	_hxyalcf
Cigarettes and tobacco	5 - 8	_hxy cig	_hxy cigi	_hxy cigf
Public transport and taxis	5 - 8	_hxy pubt	_hxy pubti	_hxy pubtf
Meals eaten out	5 - 8	_hxy meal	_hxy mli	_hxy mlf
Leisure activities	5	_hxy hsg	_hxy hsgi	_hxy hsgf
Motor vehicle fuel	5 - 8	_hxy mvf	_hxy mvfi	_hxy mvff
Men's clothing and footwear	6 - 8	_hxy mcf	_hxy mcfi	_hxy mcf f
Women's clothing and footwear	6 - 8	_hxy wcf	_hxy wcfi	_hxy wcf f
Children's clothing and footwear	6 - 8	_hxy cc f	_hxy cc fi	_hxy cc f f
Clothing and footwear	5	_hxy clth	_hxy clti	_hxy cltf
Telephone rent and calls	5	_hxy tel	_hxy tli	_hxy tlf
Telephone rent and calls, internet charges	6 - 8	_hxy teli	_hxy tlii	_hxy tli f

**Table 1: (c'td)**

	<i>Wave</i>	<i>Pre-imputed</i>	<i>Post-imputed</i>	<i>Flag</i>
<b>Household File</b>				
<i>Annualized household expenditure (Collected in the SCQ)</i>				
Holidays and holiday travel costs	5 - 8	_hxyhol	_hxyholi	_hxyholf
Private health insurance	5 - 8	_hxyphi	_hxyphii	_hxyphif
Other insurances	6 - 8	_hxyoi	_hxyoii	_hxyoif
Fees paid to health practitioner	6 - 8	_hxyhlt	_hxyhlpi	_hxyhlpf
Medicines, prescriptions and pharmaceuticals	6 - 8	_hxyphrm	_hxyphmi	_hxyphmf
Health care	5	_hxyhlth	_hxyhthi	_hxyhthf
Electricity bills	5	_hxyelec	_hxyelei	_hxyelef
Gas bills	5	_hxygas	_hxygasi	_hxygasf
Other heating fuel	5	_hxyohf	_hxyohfi	_hxyohff
Electricity, gas bills and other heating fuel	6 - 8	_hxyutil	_hxyutli	_hxyutlf
Repairs, renovation and maintenance to home	5 - 8	_hxyhmrn	_hxyhmri	_hxyhmrnf
Motor vehicle repairs and maintenance	5 - 8	_hxyvmr	_hxyvmri	_hxyvmrf
Education fees	5 - 8	_hxyeduc	_hxyedci	_hxyedcf
Buying brand new vehicles	6 - 8	_hxyncar	_hxynceri	_hxyncrf
Buying used vehicles	6 - 8	_hxyucar	_hxyuceri	_hxyucrf
Computers and related services	6 - 8	_hxycomp	_hxycmpi	_hxycmpf
Audio visual equipment	6 - 8	_hxytvav	_hxytvi	_hxytvf
Household appliance	6 - 8	_hxywg	_hxywgi	_hxywgf
Furniture	6 - 8	_hxyfurn	_hxyfni	_hxyfnf

### Missing Data

As mentioned earlier, the household expenditure data in HILDA is collected in the HQ and the SCQ. The HQ is administered to one member of the household rather than individual household member per se. Household expenditure on a wide range of goods and services was first collected in the wave 5 SCQ. The list of items collected was revised and expanded to include consumer durables from wave 6. All persons completing a Person Questionnaire (PQ) are asked to complete an SCQ. While the person responsible for the household bills is asked to complete the household expenditure section in the SCQ, sometimes more than one person in a household provided answers. The household-level expenditure averages the responses across all individuals who provided a response to the questions, excluding those from dependent students who said they were not responsible for the household bills.

The percentage of cases with missing expenditure data for wave 1 to 8 is provided in Table 2 below. For the items collected in the HQ, the percentage of missingness is less than 2 per cent for every wave. The expenditure items collected in the SCQ have more missing cases than those collected in the HQ. The percentage of missing expenditure items for those collected in the SCQ is above 15 per cent for all four waves where this data was collected. The percentage of missing cases varies slightly among different components, but no items have particularly high item non-response compared to others. Unlike income and wealth data, there is no obvious declining trend of missing cases in the later waves observed for the expenditure data.

**Table 2: Percentage of cases with missing expenditure data, wave 1 - 8**

Variable	Wave							
	1	2	3	4	5	6	7	8
<b>Household (zero and non-zero cases)</b>								
<i>Usual payments/repayments per month (collected in the HQ)</i>								
Rent	1.9	1.5	1.4	1.3	1.2	0.9	1.2	1.4
First mortgage	0.3	0.4	0.4	0.3	0.3	0.4	0.5	0.5
Second mortgage	0.7	0.5	0.5	0.5	0.5	0.5	0.6	0.6
<i>Weekly household expenditure (collected in the HQ)</i>								
All groceries	1.2	-	1.0	0.9	0.9	-	-	-
Groceries for food and drink	2.0	-	1.7	1.2	1.2	-	-	-
Meals eaten outside	0.9	-	1.0	0.9	0.8	-	-	-
<i>Annualised household expenditure (collected in the SCQ)</i>								
Groceries	-	-	-	-	15.1	14.5	16.5	18.0
Alcohol	-	-	-	-	15.9	15.4	17.1	18.8
Cigarettes and tobacco	-	-	-	-	16.4	16.2	17.8	19.0
Public transport and taxis	-	-	-	-	16.6	16.9	18.4	19.5
Meals eaten out	-	-	-	-	15.1	15.1	16.8	18.5
Leisure activities	-	-	-	-	15.9	-	-	-
Motor vehicle fuel	-	-	-	-	15.6	14.6	16.7	18.4
Men's clothing and footwear	-	-	-	-	-	15.7	17.5	19.1
Women's clothing and footwear	-	-	-	-	-	16.4	18.1	19.3
Children's clothing and footwear	-	-	-	-	-	17.2	18.4	20.2
Clothing and footwear	-	-	-	-	16.6	-	-	-
Telephone rent and calls	-	-	-	-	16.0	-	-	-
Telephone rent and calls, internet charges	-	-	-	-	-	14.7	16.7	18.2
Holidays and holiday travel costs	-	-	-	-	15.8	15.1	17.3	18.9
Private health insurance	-	-	-	-	16.3	15.6	17.5	19.2
Other insurances	-	-	-	-	-	15.7	17.7	19.5
Fees paid to health practitioner	-	-	-	-	-	16.1	17.7	19.5
Medicines, prescriptions and pharmaceuticals	-	-	-	-	-	16.0	17.8	19.5
Health care	-	-	-	-	17.3	-	-	-
Electricity bills	-	-	-	-	16.8	-	-	-
Gas bills	-	-	-	-	16.9	-	-	-
Other heating fuel	-	-	-	-	17.1	-	-	-
Electricity, gas bills and other heating fuel	-	-	-	-	19.3	16.1	17.7	19.5
Repairs, renovation and maintenance to home	-	-	-	-	17.0	16.2	17.7	19.5
Motor vehicle repairs and maintenance	-	-	-	-	16.6	15.8	17.3	19.3
Education fees	-	-	-	-	16.4	16.4	17.8	19.7
Buying brand new vehicles	-	-	-	-	-	16.8	18.3	20.0
Buying used vehicles	-	-	-	-	-	16.6	18.1	19.9
Computers and related services	-	-	-	-	-	15.8	17.7	19.5
Audio visual equipment	-	-	-	-	-	16.2	17.9	19.6
Household appliance	-	-	-	-	-	16.4	17.9	19.7
Furniture	-	-	-	-	-	16.7	18.4	20.3

The expenditure items collected in the SCQ are more prone to be missing primarily because an SCQ was not obtained from the relevant person in the household. Indeed, over 90 per cent of the missingness in the SCQ expenditure items is due to lack of return of the SCQ rather than missing responses on a returned SCQ. As shown in Table 3 below, in wave 5, 645 out of 7,125 responding household (just over 9 per cent) had no SCQ returned from anyone in the household, and this number increased to 821 (just below 12 per cent of responding household) in wave 8. For the

households where all the household members returned their SCQ, around 5 per cent had all the expenditure items missing (i.e. every one in the household skipped the expenditure section).

**Table 3: Self Completion Questionnaire (SCQ) response rate (wave 5 onwards)**

	Wave			
	5	6	7	8
<i>Person level</i>				
Responding person	12,759	12,905	12,789	12,785
No Matching SCQ	1,294	1,189	1,409	1,591
SCQ response rate	89.9	90.8	89.0	87.6
<i>Household level</i>				
Responding households	7,125	7,139	7,063	7,066
Households where everyone returned SCQ	6,161	6,249	6,014	5,884
<i>No expenditure data provided</i>	312	324	317	334
<i>Some but not all expenditure data provided</i>	866	1,066	965	896
<i>All expenditure data provided</i>	4,983	4,859	4,732	4,654
Households where at least one person returned SCQ but not all SCQ returned	319	305	334	361
<i>No expenditure data provided</i>	69	66	63	60
<i>Some but not all expenditure data provided</i>	58	70	68	63
<i>All expenditure data provided</i>	192	169	203	238
Households with no one returned SCQ	645	585	715	821

### Expenditure Imputation

The expenditure imputation is done at the household level. In general, the expenditure imputation is quite similar to imputation for wealth and income. The imputation methods used to impute the expenditure data are:

- Little and Su Method
- Nearest Neighbour Regression Method
- Population Carryover Method

More details on the imputation methods used in the HILDA survey can be found in Hayes and Watson (2009).

The overall imputation steps are:

1. Create a longitudinal household identifier.
2. Identify the lumpy expenditure items.
3. Carry-over zeros: The population carryover method is used to determine zero and non-zero expenditure flags for non-lumpy expenditure items prior to any other imputation.
4. Run the nearest neighbour regression imputation process to identify households where zero is a sensible impute (essentially a filter process deciding if the record has the expense or not).
5. Rerun the nearest neighbour regression imputation method to impute all households that require non-zero expenditure amounts.
6. Use the Little and Su imputation method to identify a suitable longitudinal donor for records that can be longitudinally linked and have at least one wave



Lumpy items are treated differently in the imputation system in order to preserve the irregular nature of these expenditure items. Figure 5 in Appendix 1 illustrates the overall imputation process, and how we decide whether a household has the expenditure or not.

### ***Step 1: Identifying Longitudinal Households***

The longitudinal imputation methods require the unit to have a longitudinal identifier. In HILDA, we do not define households over time through a common identifier, hence households need to be linked before longitudinal imputation can be performed at the household level. A longitudinal household identifier was created to link households from one wave to another for the expenditure imputation. Any changes in the household membership will result in expenditure changes, so households are only linked if there is no change of the household member(s) between two waves<sup>1</sup>. So, birth of a new child, death of a household member, split or merger of household members across waves will all result in non-linking as these events were considered to have an effect on household expenditure.

The percentage of households can be longitudinally linked for expenditure imputation is presented in Table 4. The diagonal top half of the table presents the percentage of linked households across all waves from the start to the end wave relative to all households in the start wave. The diagonal bottom half of the table presents percentages relative to the end wave. The percentages tend to be larger for the bottom diagonal as the number of households at later waves is generally smaller (hence, the denominator is smaller). A higher percentage of households are linked when only a one-wave step is involved.

**Table 4: Percentage of linked household for expenditure imputation**

		<i>End Wave</i>							
		1	2	3	4	5	6	7	8
<i>Start Wave</i>	1	.	74	58	48	42	35	31	27
	2	79	.	75	58	50	42	36	32
	3	62	76	.	74	60	48	41	36
	4	53	61	75	.	74	57	48	41
	5	45	51	60	72	.	72	57	48
	6	38	42	48	56	71	.	76	61
	7	33	37	41	47	58	77	.	76
	8	30	32	36	40	48	62	76	.

### ***Step 2: Identifying Lumpy Expenditure Items***

For some expenditure items, we do not expect a typical household to have such expense each year. We need to identify these lumpy expenditure items prior to imputation and treated them differently in the imputation process. In order to identify the lumpy items, we examine the number of times a household has reported a zero or

<sup>1</sup> For the purpose of wealth imputation, households were linked if there is no split or merger of the household, and any additional household members were children (defined for wealth imputation to be less than 18), and any missing household members were either children or deceased.

non-zero amount for a certain expenditure component during the waves the data was collected.

For the households that reported at least one non-zero value, the percentage of households that reported one, two or three or more zero values is presented in Table 5. Expenditure on new vehicles, used vehicles, public transport, computers, white goods and furniture was categorized as lumpy items, because more than 30 per cent of the households (out of these have reported at least one non-zero amount) reported a non-zero figure only once. These items were treated differently in the imputation system so that the irregular nature of these expenditure items can be preserved.

**Table 5: Percentage of households by number of times a household reported a non-zero amount for the expenditure component**

<i>Variable</i>	Once	Twice	3 times or more
Groceries	0.1	0.4	99.5
Alcohol	9.6	10.0	80.4
Cigarettes and tobacco	18.4	14.5	67.0
Public transport and taxis	31.8	20.7	47.4
Meals eaten out	7.4	8.0	84.5
Motor vehicle fuel	2.4	3.4	94.2
Men's clothing and footwear	12.9	16.9	70.3
Women's clothing and footwear	9.8	14.9	75.3
Children's clothing and footwear	17.1	12.9	70.0
Telephone rent and calls, internet charges	1.2	3.3	95.6
Holidays and holiday travel costs	13.9	11.7	74.3
Private health insurance	7.3	5.9	86.8
Other insurances	4.6	6.1	89.2
Fees paid to health practitioner	11.3	13.9	74.8
Medicines, prescriptions and pharmaceuticals	4.2	8.6	87.1
Electricity, gas bills and other heating fuel	0.5	1.1	98.3
Repairs, renovation and maintenance to home	13.5	14.2	72.3
Motor vehicle repairs and maintenance	3.8	7.1	89.1
Education fees	19.8	14.8	65.4
Buying brand new vehicles	79.7	18.5	1.8
Buying used vehicles	74.0	22.0	4.0
Computers and related services	30.1	33.2	36.7
Audio visual equipment	50.2	35.4	14.4
Household appliance and white goods	56.4	31.8	11.8
Furniture	47.9	35.0	17.1

### ***Step 3: Population Carry-over Method***

#### *Screener questions*

For the items collected in the HQ, there is some information available to determine whether the household has the certain expenditure component or not. For the mortgage repayments (both first mortgage and second mortgage), there was a question asking whether the household has the mortgage or not. The “yes” answer to the question was used to restrict the imputation to non-zero cases only. When collecting expenditure on groceries in the HQ, we firstly asked about total groceries expenditure, and asked about the amount spent on food in the next question. If the person answered the HQ did not provide an answer on the total weekly grocery amount, but stated a non-zero amount for the food expenditure, this would restrict the weekly grocery imputation to be non-zero amounts only. Table 6 shows the number of cases where we

had information to restrict the imputation to be non-zero amounts for the expenditure items collected in the HQ.

**Table 6: Non-zero restrictions on expenditure variables to be imputed (variables collected in the HQ)**

<i>Variables</i>		<i>Wave</i>							
		1	2	3	4	5	6	7	8
First mortgage repayments	Require imput'n	144	106	100	92	85	62	88	102
	Require non-zero imput'n	132	99	96	86	76	54	78	92
Second mortgage repayments	Require imput'n	57	38	39	34	39	35	39	42
	Require non-zero imput'n	45	34	30	28	29	30	29	33
Weekly household expenditure on grocery	Require imput'n	90	-	71	66	66	-	-	-
	Require non-zero imput'n	6	-	6	3	5	-	-	-

For the expenditure components collected in the SCQ, there is a screener question asked whether the household has such expense or not. If the person who answered the SCQ stated the household did not have such expense, zero was derived for the item. If the person stated the household had the expense but did not provide a value for the component, the imputation was restricted to be non-zero amounts only. However, the information obtained from the screener questions is very limited because most missingness in the SCQ is due to lack of the SCQ being returned. The amount of information available to restrict the imputation to non-zero amounts is presented in Table 7 below. There are very few cases (less than 10) where we have information available to restrict the imputation to non-zero amounts only.

**Table 7: Non-zero restrictions on expenditure variables to be imputed (for variables collected in the SCQ)**

	<i>Wave 5</i>		<i>Wave 6</i>		<i>Wave 7</i>		<i>Wave 8</i>	
	Require Imput'n	Require non-zero imput'n	Require Imput'n	Require non-zero imput'n	Require Imput'n	Require non-zero imput'n	Require Imput'n	Require non-zero imput'n
<b>Household-level expenditure</b>								
Groceries	42	2	51	1	55	0	45	0
Alcohol	99	0	116	0	99	0	106	1
Cigarettes and tobacco	135	0	171	1	146	0	119	0
Public transport and taxis	147	0	220	0	186	1	155	2
Meals eaten out	45	1	92	1	75	0	79	0
Leisure activities	99	1						
Motor vehicle fuel	77	2	57	0	66	0	77	1
Men's clothing and footwear	-	-	132	0	128	0	125	1
Women's clothing and footwear	-	-	185	0	166	1	142	0
Children's clothing and footwear			241	1	188	0	199	1
Clothing and footwear	153	1	-	-	-	-	-	-
Telephone rent and calls	107	3	-	-	-	-	-	-
Telephone rent and calls, internet charges	-	-	61	1	68	0	62	0
Holidays and holiday travel costs	92	2	95	0	109	1	112	1

**Table 7: (C'td)**

	Wave 5		Wave 6		Wave 7		Wave 8	
	Require Imput'n	Require non- zero imput'n	Require Imput'n	Require non- zero imput'n	Require Imput'n	Require non- zero imput'n	Require Imput'n	Require non- zero imput'n
<b>Household-level expenditure</b>								
Private health insurance	127	3	127	0	125	4	134	1
Other insurances	-	-	134	0	138	0	154	2
Fees paid to health practitioner	-	-	163	2	141	2	151	3
Medicines, prescriptions and pharmaceuticals	-	-	156	2	145	2	153	3
Health care	293	7	-	-	-	-	-	-
Electricity bills	162	5	-	-	-	-	-	-
Gas bills	171	6	-	-	-	-	-	-
Other heating fuel	184	0	-	-	-	-	-	-
Electricity, gas bills and other heating fuel	-	-	160	1	139	1	155	1
Repairs, renovation and maintenance to home	181	4	171	0	142	1	150	1
Motor vehicle repairs and maintenance	150	6	141	2	113	0	136	1
Education fees	133	1	188	0	148	0	168	3
Buying brand new vehicles	-	-	215	1	179	0	191	1
Buying used vehicles	-	-	196	1	167	0	184	1
Computers and related services	-	-	143	0	136	0	150	2
Audio visual equipment	-	-	172	1	151	2	157	1
Household appliance	-	-	185	2	166	1	142	1
Furniture	-	-	205	1	192	1	212	1

For the non-lumpy items, the population carry-over method was used first to determine whether the household has the expenditure or not. The percentage of zeros imputed by the population carryover method for each expenditure variable is shown in Table 8.

The population carryover method did not impute all the zeros possible. In the subsequent steps, the households who did not have a zero/non-zero determination from this method could receive a zero imputed via the nearest neighbour regression method or the Little and Su method.

**Table 8: Percentage of household with zeros imputed via population carryover method, wave 1-8**

Variable	Wave							
	1	2	3	4	5	6	7	8
<i>Household (zero and non-zero cases)</i>								
<b>Usual payments/repayments per month</b>								
First mortgage	2.8	3.8	0	1.1	2.4	3.2	3.4	1.0
Second mortgage	12.3	7.9	10.3	8.8	17.9	5.7	12.8	2.4
<b>Weekly household expenditure</b>								
All groceries	0	-	2.8	1.5	3.0	-	-	-

**Table 8: (c'td)**

Variable	Wave							
	1	2	3	4	5	6	7	8
<i>Household (zero and non-zero cases)</i>								
<b>Weekly household expenditure</b>								
Groceries for food and drink	0	-	1.7	3.6	2.4	-	-	-
Meals eaten outside	4.5	-	22.1	15.2	21.4	-	-	-
<b>Annualised household expenditure</b>								
Groceries	-	-	-	-	0.2	0.8	0.6	0.6
Alcohol	-	-	-	-	9.9	19.6	16.6	12.5
Cigarettes and tobacco	-	-	-	-	20.6	33.4	33.9	22.2
Meals eaten out	-	-	-	-	6.2	12.5	9.0	6.2
Motor vehicle fuel	-	-	-	-	5.6	8.6	7.6	6.7
Men's clothing and footwear	-	-	-	-	-	13.3	20.5	14.6
Women's clothing and footwear	-	-	-	-	-	10.7	17.2	10.4
Children's clothing and footwear	-	-	-	-	-	21.9	33.9	22.3
Telephone rent and calls, internet charges	-	-	-	-	-	1.8	3.2	1.6
Holidays and holiday travel costs	-	-	-	-	12.9	22.4	20.4	14.3
Private health insurance	-	-	-	-	17.0	28.0	27.1	18.5
Other insurances	-	-	-	-	-	6.8	10.2	7.7
Fees paid to health practitioner	-	-	-	-	-	11.6	14.8	10.2
Medicines, prescriptions and pharmaceuticals	-	-	-	-	-	4.6	7.5	4.1
Electricity, gas bills and other heating fuel	-	-	-	-	-	2.8	3.4	2.5
Repairs, renovation and maintenance to home	-	-	-	-	17.5	25.4	26.3	17.2
Motor vehicle repairs and maintenance	-	-	-	-	8.3	13.3	12.7	8.7
Education fees	-	-	-	-	21.3	34.9	32.5	23.5

#### **Step 4 and 5: Nearest Neighbour Regression Method**

The nearest neighbour regression method was applied so that every record requiring imputation for each variable got imputed. Both the population carryover method used in the previous step and the Little and Su method in step 6 have limitations that restrict them to only impute those households can be longitudinal linked. In situations where the other methods cannot be applied the results from the nearest neighbour regression method are used.

For the expenditure imputation, the nearest neighbour regression method was run twice. It was run first to determine which cases should be imputed with zero or non-zero amounts (i.e., whether the household had the expenditure or not). Logistic regression models were constructed each wave for the expenditure variables. Over 30 household-level variables were considered for inclusion in the expenditure models covering household demographic characteristics, household income. The variables included in the regression models are listed in Appendix 2. Only the zero amounts from this step were retained.

Then, the nearest neighbour regression method was run again to determine the non-zero amounts to be imputed for those cases deemed to have non-zero values from the previous run. Log-linear regression models were constructed. The variables used in the models were the same as these used in the previous run, and a backwards elimination process in SAS was used to select the variables. The unusual values (extremely large or extremely small values) were identified and excluded from the donor pools before the donor selection process. In addition, each complete record was limited to being used as a donor twice in the procedure. This restriction avoided the possibility of unusual values being imputed too often.

As a result of running the nearest neighbour method twice, the donors were selected in two stages and the regression models were created from different pools of data. The zero selection stage allowed all records to be included while the next stage restricted the cases to a subset with non-zero expenditure values.

For the household with all SCQ expenditure data missing, donors were identified utilizing the sum of the expenditure items collected in the SCQ and the imputed expenditure components were all taken from a single donor in order to retain the correlations between the components. For the households where only a few expenditure items were missing, any missing expenditure component was imputed separately.

#### *Imputation classes*

Household expenditure is likely to be associated with household income. For the expenditure imputation, the equivalised household disposable income<sup>2</sup> bands together with some household characters were used as imputation classes.

For most variables, the age group<sup>3</sup> of the highest income earner in the household together with equivalised disposable income band were used as imputation class for both stages.

Deviations from these imputation classes were made for certain variables. For expenditure on men's, women's and children's clothing and footwear, whether the household has male, female or child residents together with the equivalised disposable income band were used as imputation classes in both steps. Only the equivalised disposable income band was used as imputation classes in step 5 for mortgage repayments (first mortgage and second mortgage), expenditure on private health insurance, other heating fuels, new and used vehicle.

#### ***Step 6: Little and Su Method***

The Little and Su method was applied where possible. This method incorporates (via a multiplicative model) the trend across waves (column effect), the recipient's departure from the trend (row effect), and a residual effect donated from another case with complete expenditure information for that component (residual effect) – see Hayes and Watson (2009) for details of this method. Only cases that have been enumerated in more than one wave, longitudinally linked, and have at least one wave of non-zero data available can be imputed via this method. The nearest neighbour regression imputed values from the previous step were used when calculating the column and row effects.

For the lumpy items, when selecting the donor for the Little and Su method, the donor must have the same zero pattern as recipients. See example 1 below (which contains hypothetical data), household 1 reported they spent nothing on purchasing new

---

<sup>2</sup> The equivalised disposable household income is derived by calculating an equivalence factor and then dividing the household disposable income by the factor. The equivalence factor is calculated using the modified OECD equivalence scale which is widely accepted among Australian analysts of income distribution. According to ABS release Category 6503.0, the equivalence scale is built up by allocating points to each person in a household. Taking the first adult in the household as having a weight of 1 point, each additional person who is 15 years or older is allocated 0.5 points, and each child under the age of 15 is allocated 0.3 points. Equivalised household disposable income is derived by dividing total household disposable income by a factor equal to the sum of the equivalence points allocated to the household members.

<sup>3</sup> The age group is classified as: <19, 20-24, 25-34, 35-44, 45-54, 55-64, 65+.

vehicles in wave 6, in wave 7 the household spent \$25,000 and the household did not give an answer in wave 8. When selecting a donor to impute the missing amount in wave 8, the donors are restricted to those that had a zero amount in wave 6 and a non-zero amount in wave 7. The wave 8 amount can be zero or non-zero. The final donor (household 4) was selected based on the row effects calculated.

**Example 1: Imputation for lumpy items (hypothetical data)**

Record require imputation				Potential donors			
Household	wave 6	wave 7	wave 8	Household	Wave 6	wave 7	wave 8
1	0	25,000	missing	2	0	32,000	0
				3	0	28,000	0
				4	0	20,000	10,000

The percentage of missing cases imputed by each imputation method is illustrated in Table 9. The households which cannot be linked between waves were imputed by the nearest neighbour regression method. For the housing expenditure variables (rent payment, mortgage repayment and second mortgage repayment), which have been collected in 8 waves so far, the majority of cases were imputed by the Little and Su method. For the expenditure items where we only have three waves of data available, like those items collected in the SCQ from wave 6 onwards, more than half of the cases were imputed by the nearest neighbour regression method.

**Table 9: Percentage of missing cases imputed by imputation method, (wave 1-8)**

<i>Imputation method</i>	<i>Wave</i>							
	1	2	3	4	5	6	7	8
<b>Housing expenditure variables (collected in wave 1-8 Household Questionnaire)</b>								
Nearest Neighbour	37.8	10.9	21.1	21.1	15.5	25.4	15.0	38.1
Little & Su	57.3	85.1	76.5	76.2	78.4	71.4	80.0	60.8
Carryover	4.9	4.0	2.4	2.7	6.1	3.2	5.0	1.1
<b>Weekly household expenditure variables (collected in wave 1, 3, 4, and 5 Household Questionnaire)</b>								
Nearest Neighbour	56.2	-	27.6	23.3	35.6	-	-	-
Little & Su	42.8	-	65.0	70.2	56.6	-	-	-
Carryover	1.0	-	7.4	6.5	7.8	-	-	-
<b>Annualised household expenditure variables (collected in the Self Completion Questionnaire from wave 5)</b>								
Nearest Neighbour	-	-	-	-	62.8	42.6	39.2	52.8
Little & Su	-	-	-	-	27.3	40.5	44.9	36.1
Carryover	-	-	-	-	9.9	17.0	16.0	11.1
<b>Annualised household expenditure variables (collected in the Self Completion Questionnaire from wave 6)</b>								
Nearest Neighbour	-	-	-	-	-	69.7	57.9	66.9
Little & Su	-	-	-	-	-	24.8	33.8	27.6
Carryover	-	-	-	-	-	5.5	8.3	5.5

*Imputation class*

The Little and Su imputation method used the age group of the highest income earner as an imputation class. Donors and recipients were matched within longitudinal imputation classes assigned based on date of birth. Age group 15-19 corresponded to people born between 1988-1992, age group 20-24 born between 1983-1987, etc. The column and row effects were calculated within each imputation class and donors were matched to recipients which share the same imputation class.

## **Comparison with the Household Expenditure Survey (HES)**

Due to the concerns about whether expenditure can be accurately collected on a recall basis, it is worth examining how well we measure the household expenditure in the HILDA Survey. One way to assess the measurement validity is to compare the HILDA estimates with external benchmarks. The latest Household Expenditure Survey (HES), conducted by the Australian Bureau of Statistics (ABS) in 2003 to 2004, provides us with a generally suitable comparison.

Before detailing the comparison of the HES and the HILDA estimates, it is worth noting several differences between the surveys.<sup>4</sup>

1. The collection method – HES mainly employs a diary method where respondents record their actual expenditure over a two-week period, beginning the day after interview. Estimates for infrequently or more expensive items are derived from the Household Questionnaire in HES, which collects expenditure information on a recall basis for varying periods. The HILDA Survey collects household expenditures on all items on a recall basis.
2. The recall period – For items that are collected on a recall basis in both surveys, the recall period can differ. For example, HES respondents were asked to recall the spending on household appliances or furniture over the last 3 months and for items such as insurance and utilities bills, respondents were asked for the value of their last payment and the length of the time to which it related. On the other hand, HILDA respondents reported expenditure on household appliances, furniture, insurance, and electricity and gas bills over the last 12 months and telephone and mobile bills on a monthly basis.
3. The reference period – For the HES, interviewing was conducted throughout the 12 months of the 03-04 financial year. The total period covered by expenditure estimates from the HES is a function of the recall or reporting period and the timing of interview. For HILDA, most the data was collected between August to December each year, and the respondents are asked for ‘the best estimate of the average amount spent on that item’.
4. The method to derive household expenditure – In HES, a personal diary is administered to all usual residents aged 15 years and over in household to record their expenditure over a two week period. Household expenditure for the items collected in the diary is the sum of amounts reported by each household member. In HILDA, we asked the person who is responsible for the household bills to fill in the expenditure questions and make his/her best estimates on the average household spending. If more than one person in the household reported household expenditure, the household expenditure is derived by taking the average of the amounts reported (with the exception of dependant students not responsible for the bills).
5. The item classification – For items recorded in the HES diary, the classification of these items into categories was decided at the data entry stage. However, in HILDA, the respondents decide what to include for each item according to the explanatory notes provided. For example, when collecting groceries, the HILDA question asks “weekly expenses on groceries (include

---

<sup>4</sup> All the information about HES is obtained from the ‘Household Expenditure Survey and Survey of Income and Housing User Guide 2003-04’ (Cat. No. 6503.0) published by ABS.



food, cleaning products, pet food and personal care products. Do not include alcohol or tobacco)". In HES, the respondents were asked to record everything they spent money on during the two-week period, and the data entry operator uses a computer system to classify whether the item is "food and non-alcoholic beverages" or "Alcoholic beverages" etc.

6. What is measured – HES primarily adopts an acquisition approach, where the full cost payable by the household of acquiring a good or service within a given period is collected. The full cost is collected regardless of whether the household actually paid for the good or services within the period. In the HILDA Survey, we asked for 'the average amount spent' on each item (which is close to the payment approach which asks for the payment made by the household within a given period). For the items which are normally acquired, paid for and consumed within a relatively short period of time, the two approaches will be similar. For durable items and items purchased on credit that are not fully consumed or paid during the recall period, the two approaches will result in differences. For example, when asking the expenditure on white goods, HES collected total costs of them (regardless whether the total amount is paid by the household) and estimated the expenditure for the reference period. By contrast, in HILDA, we asked the respondents to estimate the average household spending on white goods. If the item is paid by instalments, the household will probably report the amount paid during the reference period.

The differences in approach will likely lead to differences in the distributions of the expenditure items, but for many items, the mean value should in principle be the same. By comparing HILDA and HES estimates of mean expenditure, we can obtain some indication of how well the expenditure data is measured in HILDA.

To compare with the HES estimates, the wave 6 HILDA estimates are used for most items. For mortgage and rental payments which are collected in every wave, the wave 4 figures are used as they are closer to the HES collection dates. The HILDA figures have been adjusted by the growth of real net disposable income per capita. The items collected on weekly and monthly basis in wave 6 of HILDA have been deflated by 6.5% to count for the increase in the real income from December quarter 2004 (mid-point of the HES data collection) to September quarter 2006 (mid-point of the HILDA data collection). For the annual expenses collected in wave 6, the figures were deflated by 5.7% to count for the increase in the real income between December quarter 2004 and March quarter 2006. Mortgage repayments and rent payments have been deflated by 1.6% to count the income growth from December quarter 2004 to September quarter 2004.

Table 10 below shows the comparison between HILDA estimates and the HES estimates on various weekly expenditure items. The pre- and post-imputed HILDA estimates are presented. As we can see the imputation does not make a big difference to the estimates. A band of plus or minus 10 per cent is used to roughly judge whether the estimates are well-measured. Most items collected in the HILDA are reasonably close to HES. Expenditure on the following items differs by more than 10 per cent between the two surveys:

- Motor vehicle fuel
- Clothing and footwear

- Holidays and holiday costs
- Private health insurance
- Medicines, prescriptions and pharmaceuticals
- Repairs, renovation and maintenance to home
- Education fees
- Buying vehicles (new and used)
- Household appliance
- Furniture

The expenditure on groceries from the HILDA Survey is slightly over 10 per cent higher than that obtained from HES. This probably can be explained by the significant increase of food prices from 2003 to 2006. The motor vehicle fuel costs reported in the HILDA Survey is more than 25% higher than HES. However, the deflated wave 5 figure is the same as what HES reported (which is 33 dollars). This suggests that the rise in the petrol price can lead to the jump of the household expenditure on fuel. For the question on the private health insurance, we did not specify whether to include or exclude the government rebate in the HILDA Survey. Hence, some respondents may report the out-of-pocket amounts by deducting 30 per cent government rebate. Where in HES, it asked the value of the last payment, which is the amount the insurance company charged. The HILDA figure on expenditure on repairs, renovation and maintenance to home is more than HES reported, this is probably because we did not provide guidelines on what to include and the items HILDA respondents included might vary from HES classification. For the big durable items like white goods, cars, and furniture, the expenditure reported in the HILDA Survey and HES differs by more than 20 per cent, which indicates these items can be difficult for respondents to recall the actual amounts. When comparing with HES, we have aggregate ‘computers and related devices’ and ‘audio visual equipment’ as the differences between two might be a bit ambiguous for respondents (for example, an ipod is suggested on the questionnaire as computer devices whereas many people would put it under audio visual equipment).

The aggregated expenditure in HILDA differs by only 1.6 per cent from the HES total for the same items. The sum of the expenditure items collected in HILDA is \$743 per week after adjusting for the income growth<sup>5</sup>. This is 83% of the total expenditure collected in HES on all goods and services<sup>6</sup>. Although there are some discrepancies between HILDA estimates and HES estimates (10 out of 23 aggregated items collected in HILDA from wave 6 onwards appeared to differ more than 10 per cent with HES estimates), overall we appear to have had reasonable success in terms of capturing expenditure information on a standard recall-based questionnaire<sup>7</sup>. More detailed analysis on comparing HILDA estimates with HES can be found in Wilkins and Sun, 2010.

---

<sup>5</sup> The sum of the same items collected in the HES is \$706 per week (ABS CAT. No. 6535.055.001)

<sup>6</sup> Total household goods and services expenditure recorded in HES was \$893 per week. This figure includes payments of mortgage interest but not principal. For comparison with HILDA, average weekly amount of principal repayments in 2003-2004 \$36 was added.

<sup>7</sup> The well-measured items comprise 55% of total household expenditure on goods and services.

**Table 10: Comparison between HES and HILDA mean estimates**

Expenditure Items	HES	HILDA (post- imputed)	HILDA (pre- imputed)	Diff (%)	Comments
Groceries	131	146	148	11.2%	HES=food and non-alcoholic beverages(03)-meals eaten outside(0311)+cleaning products(0801010201/301/401/501/901)+pet food(1104010000/101/200/201/202/203/299)+personal care products(120101)
Alcohol	23	23	23	0.8%	HES=alcohol beverages(04)
Cigarettes and tobacco	12	12	13	-1.4%	HES=tobacco products(05)
Public transport and taxis	6	6	7	0.2%	HES=public transport fares(100107)+taxi fares(1001080101)
Meals eaten out	42	39	39	-6.8%	HES=meals and fast food (0311). The HILDA estimates is obtained from the wave 5 HQ.
Motor vehicle fuel	33	42	42	26.7%	HES=motor vehicle fuel, lubricants, and additives(100103)
Clothing and footwear	35	29	30	-16.6%	HES=clothing and footwear(06). HILDA estimates is obtained from wave 6 SCQ. In wave 5, we asked expenditure on clothing and footwear, and from wave 6 onwards, we asked expenditure on men's, women's and children's clothing and footwear separately. This help
Telephone rent and calls, Internet charges	30	31	31	3.1%	HES=telephone and facsimile charges(080103)+internet charges(1102999902/03)
Holidays and holiday travel costs	35	42	45	21.4%	HES=holidays(1103)
Private health insurance	18	14	15	-23.4%	HES=accident and health insurance(0901)
Other insurances	21	20	21	-5.8%	HES=compulsory insurance of motor vehicle, motor cycle, caravan and trailer(1001040102/302)+other insurance of motor vehicle, motor cycle, caravan and trailer(1001040201/401)+house and contents insurance(010104)
Fees paid to health practitioner	14	14	14	-3.0%	HES=health practitioner's fees (0902)
Medicines, prescriptions and pharmaceuticals	9	7	7	-18.4%	HES=Medicines and pharmaceutical products(090301)
Electricity, gas bills and other heating fuel	23	21	22	-7.8%	HES=Domestic fuel and power(02)
Repairs, renovation and maintenance to home	21	39	44	87.3%	HES=Repairs and maintenance payments to contractors(010105)+Repairs and maintenance (materials only)(010106)
Motor vehicle repairs and maintenance	14	15	16	8.2%	HES=crash repairs(1001060101)+vehicle servicing(1001060201)
Education fees	18	15	15	-17.0%	HES=education fees for primary and secondary school(130202)+education fees excluding primary and secondary school fees(130203)
Buying vehicles (new +used)	49	62	67	26.9%	HES=motor vehicle purchase(100101)+other vehicle purchase(100102)

**Table 10: (C'td)**

Expenditure Items	HES	HILDA (post- imputed)	HILDA (pre- imputed)	Diff (%)	Comments
Electronic devices (computer and audio visual equipments)	15	16	17	6.3%	HES=Home computer equipment (including pre-packaged software)(110102)+Audio-visual equipment and parts(110101); HILDA=Computers and related devices+ Audio visual equipment
Household appliance	12	8	9	-30.5%	HES=household appliance(0703) HES=bedroom, lounge/dining room, outdoor, garden and other furniture(0701010201-0701010501)
Furniture	16	10	11	-36.6%	HES=Mortgage repayments - principal component (selected dwelling)(15010101)+Mortgage repayments - interest component (selected dwelling)(010102)
Mortgage payment	82	84	80	2.1%	
Rent	47	47	56	-0.6%	HES=Rent payments(010101)

## Quality of Imputation

### *Effects of Imputation on the Expenditure Distribution*

The households that do not provide answers to the expenditure items are likely to have systematic differences from the group that answers every question. Excluding these cases from analysis can have negative impacts on the representativeness of the results and the number of cases researchers can use in their analysis. Table 11 compares the unweighted distribution of the expenditure variables before and after imputation for wave 8. Similar tables for other waves are provided in Appendix 3. The imputation has a relatively small impact on most of the expenditure components. The differences between mean and median for pre-imputed and post-imputed data are really small. For most expenditure components, the means are slightly lower after imputation. The lower mean values are partially due to more zeros in the post-imputed data. As shown in Table 12, the percentage of zeros in the post-imputed data is somewhat higher than it in the pre-imputed data for most variables.

**Table 11: Wave 8 unweighted distribution of expenditure data (non-zero cases) before and after imputation**

Variable	Before Imputation			After Imputation		
	Mean	Median	Standard Deviation	Mean	Median	Standard Deviation
<b>Households (non-zero only)</b>						
Rental payments	928	847	544	928	841	552
Mortgage repayments (first and second)	1,992	1,629	2,111	1,982	1,608	2,082
Groceries	9,016	7,821	5,288	8,865	7,821	5,248
Alcohol	2,088	1,564	2,152	2,104	1,564	2,147
Cigarettes and tobacco	2,842	2,607	2,005	2,840	2,607	2,006
Public transport and taxis	1,487	1,043	2,615	1,512	1,043	2,772
Meals eaten out	2,864	2,346	3,075	2,847	2,089	3,050
Motor vehicle fuel	2,988	2,400	4,640	2,943	2,400	4,376
Clothing and footwear	2,023	1,200	2,952	1,986	1,200	2,892

Telephone rent and calls, internet charges	2,006	1,380	2,890	2,046	1,320	3,014
-----------------------------------------------	-------	-------	-------	-------	-------	-------

**Table 11: (c'td)**

Variable	Before Imputation			After Imputation		
	Mean	Median	Standard Deviation	Mean	Median	Standard Deviation
<b>Households (non-zero only)</b>						
Holidays and holiday travel costs	4,289	2,000	8,079	4,283	2,000	7,924
Private health insurance	1,713	1,500	1,194	1,698	1,500	1,191
Other insurances	1,414	1,150	1,631	1,394	1,100	1,578
Fees paid to health practitioner	1,116	500	1,731	1,093	500	1,719
Medicines, prescriptions and pharmaceuticals	483	300	1,509	470	280	1,395
Electricity, gas bills and other heating fuel	1,395	1,100	5,202	1,365	1,100	4,727
Repairs, renovation and maintenance to home	4,382	1,000	18,863	4,278	1,000	18,332
Motor vehicle repairs and maintenance	1,118	800	2,497	1,114	751	2,382
Education fees	2,832	1,000	4,987	2,859	1,000	4,996
buying vehicles(new and used)	16,076	10,000	22,626	16,059	10,000	22,160
computers and related services	998	600	1,157	994	600	1,160
Audio visual equipment	1,195	600	1,553	1,192	600	1,559
household appliance	1,295	800	1,851	1,308	800	2,025
furniture	1,751	1,000	3,318	1,731	1,000	3,217

**Table 12: Percentage of zeros before and after imputation (wave 5- 8)**

Variable	Wave 5		Wave 6		Wave 7		Wave 8	
	Pre-imputed	Post-imputed	Pre-imputed	Post-imputed	Pre-imputed	Post-imputed	Pre-imputed	Post-imputed
Groceries	0	0	1	1	1	1	1	1
Alcohol	34	37	30	33	31	33	31	34
Cigarettes and tobacco	72	73	73	74	74	75	75	75
Public transport and taxis	72	75	72	73	73	75	72	74
Meals eaten out	16	17	16	18	17	18	17	19
Leisure activities	26	30						
Motor vehicle fuel	9	11	10	12	10	12	9	11
Men's clothing and footwear			35	39	35	39	35	39
Women's clothing and footwear			25	29	24	28	24	29
Children's clothing and footwear			65	67	65	68	65	68
Clothing and footwear	16	19						
Telephone rent and calls	2	3						
Telephone rent and calls, internet charges			3	4	2	3	2	3
Holidays and holiday travel costs	33	36	34	38	32	36	33	37
Private health insurance	48	51	47	51	47	50	45	49
Other insurances			13	17	12	15	11	15
Fees paid to health practitioner			21	25	22	26	21	25

Medicines, prescriptions and pharmaceuticals			9	11	9	11	9	11
Health care	14	17						

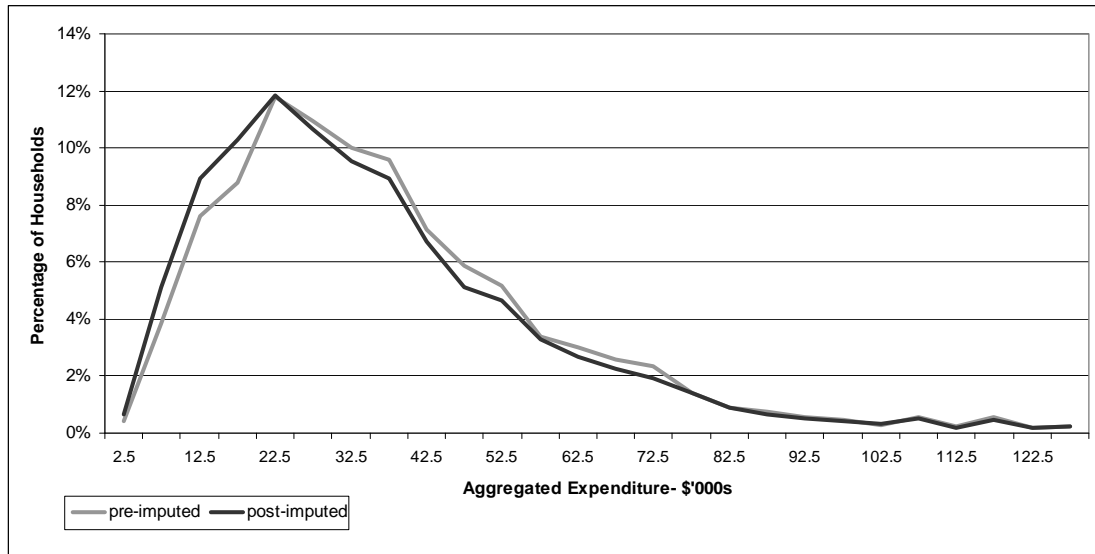
**Table 12: (c'td)**

<i>Variable</i>	<i>Wave 5</i>		<i>Wave 6</i>		<i>Wave 7</i>		<i>Wave 8</i>	
	Pre- imputed	Post- imputed	Pre- imputed	Post- imputed	Pre- imputed	Post- imputed	Pre- imputed	Post- imputed
Electricity bills	2	2						
Gas bills	44	48						
Other heating fuel	87	89						
Electricity, gas bills and other heating fuel	1	2	2	3	2	3	2	3
Repairs, renovation and maintenance to home	37	42	40	44	40	44	39	44
Motor vehicle repairs and maintenance	13	16	15	19	15	18	15	19
Education fees	64	67	66	69	67	69	68	70
Buying brand new vehicles			90	91	90	91	91	92
Buying used vehicles			84	86	85	86	86	87
Computers and related services			50	54	49	51	46	50
Audio visual equipment			61	64	63	66	61	65
Household appliance			64	67	64	67	64	67
Furniture			60	63	60	64	61	65

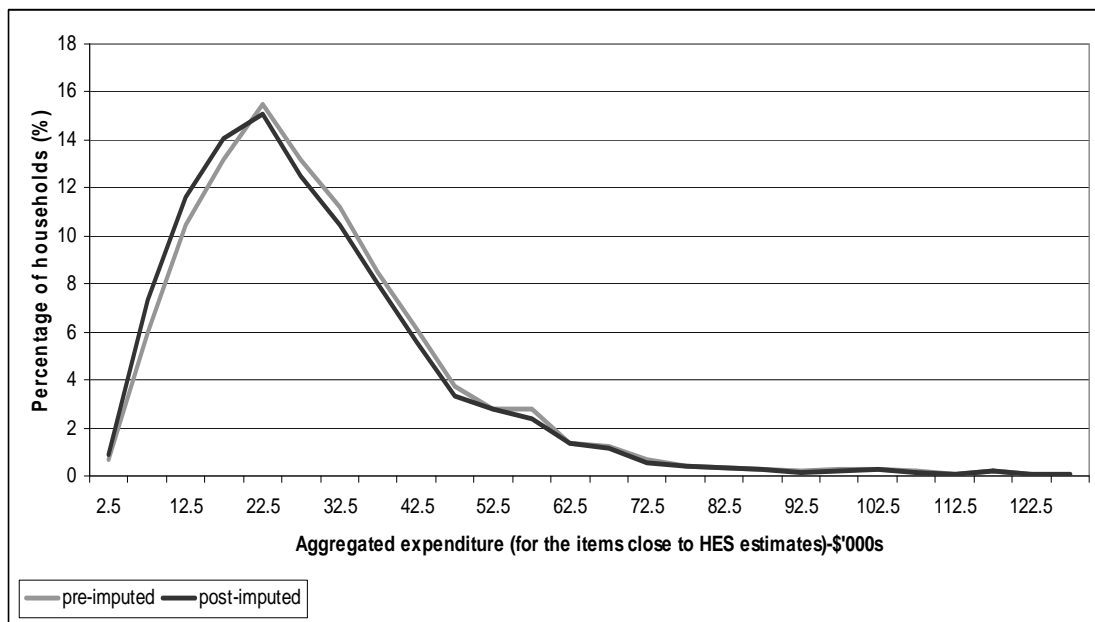
We then looked at the distribution of pre- and post-imputed expenditure data graphically. The aggregated expenditure, which is the sum of all the expenditure items collected, is examined here.

Figure 1 shows the weighted distribution of the aggregated expenditure collected in wave 8 before imputation and after imputation. Note that the mid-point of the expenditure band defined is provided on the graph. For example, the aggregated expenditure \$2,500 on the graph actually represents the households with aggregated expenditure of \$0 to \$5,000. The grey line illustrates the weighted distribution of the aggregated expenditure before the imputation, and the black line shows the weighted distribution of the aggregated expenditure after imputation. When comparing the two lines, we can see that imputation has pulled up the percentage of households with aggregated expenditure less than \$25,000 a small amount. The percentage of households with aggregated expenditure between \$25,000 and \$80,000 has been pulled down slightly. Figure 2 illustrates the weighted distribution of the aggregated household expenditure for these items close to HES estimates (i.e., excluding the items listed in the previous section that were not close to the HES estimates). Similar conclusions can be drawn from Figure 2 as from Figure 1, where the imputation increases the proportion of households with low aggregated expenditure slightly, but pulls down the proportion of households with high aggregated expenditure. Since similar patterns are observed in other waves, though the graphs on other waves are not provided in this paper.

**Figure 1: Weighted distribution of aggregated household expenditure, wave 8**



**Figure 2: Weighted distribution of aggregated household expenditure (for items closed to HES estimates), wave 8**

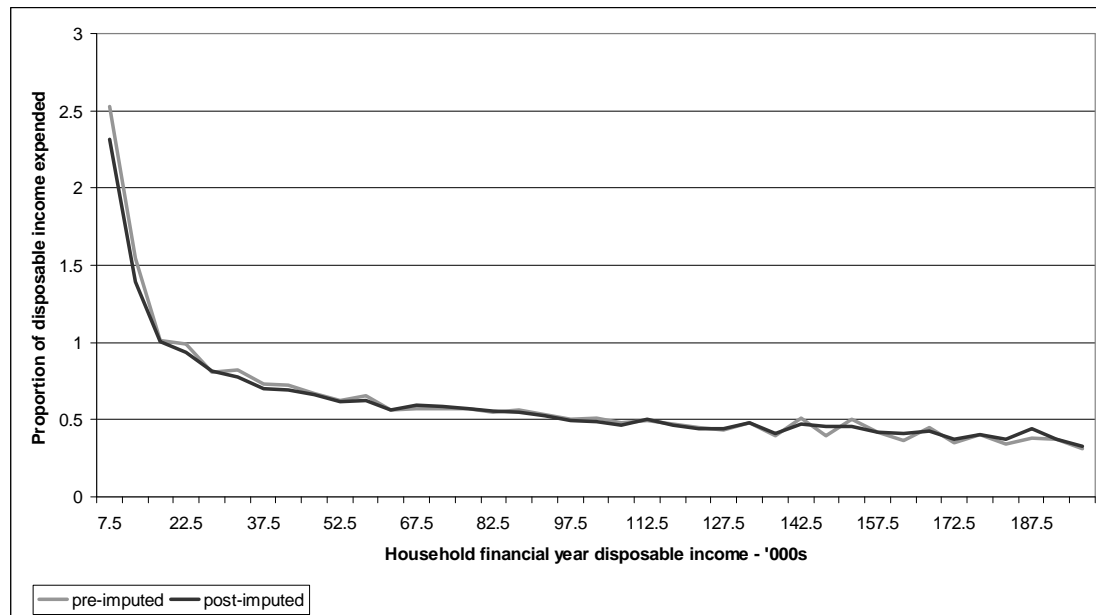


***Effects of Imputation on the Relationship Between Income and Expenditure***

Second, we look at the relationship between income and expenditure and whether the relationship is preserved after imputation. Figure 3 plots household disposable income versus the proportion of income spent on all the expenditure items collected for wave 8. Again, the mid-point of the income band is plotted on the graph. The grey line presents the proportion of financial year disposable income spent before imputation, and the dark line shows this information after imputation. Note that we do not attempt to collect all household expenditure items in HILDA. The aggregated expenditure used in the analysis is the sum of all the expenditure items collected, hence this does

not show a complete picture of how household expenditure changes relative to income. As we can see from Figure 3, low income households, especially these households with financial year disposable incomes less than \$30,000, typically spend more than they earn, which indicates that they are shifting resources to make a living (e.g. use savings, borrow money). For households with income more than \$100,000, they use less than half of what they earn on the expenditure items we collect. The grey line and dark line are very close, indicating that the imputation preserves the relationship between income and expenditure.

**Figure 3: Relationship between household disposable income and aggregated expenditure (wave 8)**



### ***Weighted Average and Proportion of Mean Expenditure Imputed***

Table 13 shows the weighted mean value and the amount that imputation contributes to the expenditure components for selected items. For monthly rent payments, about 1 per cent of the mean rent payments was contributed by imputed cases in wave 8. In wave 1, 7.4 per cent of the monthly mortgage repayments were imputed, the percentage of imputed of mortgage repayments drops to 4.9 in wave 8. For the expenditure variables collected in the SCQ, the imputed values contributed more to the mean value calculated. More than 18 per cent of the annualised household expenditure on groceries was imputed in wave 8.

**Table 13: Mean Expenditure (including imputed values) and proportion of mean expenditure imputed for selected items, wave 1-8 (weighted)**

Variable	Wave							
	1	2	3	4	5	6	7	8
<b>Household</b>								
<b><i>Rent payments (Collected in wave 1-8 HQ)</i></b>								
Mean	198	203	206	215	224	242	258	278
Proportion imputed	0.9	0.8	1.0	0.7	1.3	0.8	0.9	1.3
<b><i>Mortgage repayments (Collected in wave 1-8 HQ)</i></b>								
Mean	320	332	370	398	457	515	577	696
Proportion imputed	7.7	5.8	6.0	4.6	5.4	3.1	4.1	4.7



*Weekly household expenditure on grocery (Collected in wave 1,3,4 and 5 HQ)*

Mean	131	.	135	141	147
Proportion imputed	1.5	.	1.1	0.9	1.2

**Table 13: (c'td)**

Variable	Wave							
	1	2	3	4	5	6	7	8
<b>Household</b>								
<i>Annualised household expenditure on grocery (Collected in the SCQ from wave 5)</i>								
Mean					7,623	8,091	8,534	9,113
Proportion imputed					14.7	14.8	16.8	18.5
<i>Annualised household expenditure on alcohol (Collected in the SCQ from wave 5)</i>								
Mean					1,130	1,289	1,292	1,370
Proportion imputed					13.9	16.3	15.7	17.9
<i>Annualised household expenditure on motor vehicle fuel (Collected in the SCQ from wave 5)</i>								
Mean					1,827	2,322	2,182	2,676
Proportion imputed					15.0	15.0	16.5	17.7

**Cross-wave Comparison**

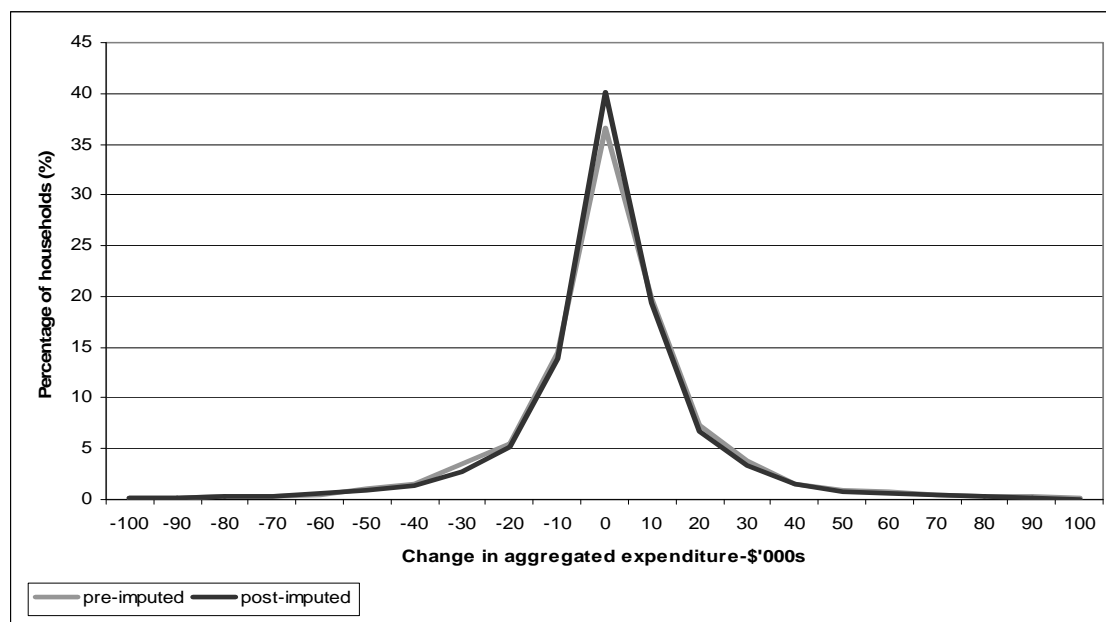
The cross-sectional examination of the imputed expenditure data has not raised any alarming concerns. We now examine the longitudinal component of the imputation.

Friedman’s permanent income hypothesis (Friedman, 1957) suggests that short-term changes in income have little effect on consumer spending behaviour, hence, consumption should be highly correlated over time. However, in HILDA we do not collect total expenditure and there may be some shift in what items households spend money on over time, resulting in lower correlations when we consider a smaller group of expenditure items. Further, correlations are subject to outliers. A small number of outliers can result in big change of the correlation ratio. Hence, instead of looking at correlation between waves, we assess the distribution of expenditure changes over time for the households can be longitudinally linked (i.e. households with no change in the membership between waves).

Figure 4 illustrates the change in the aggregated expenditure of all the items collected in HILDA between wave 7 and wave 8 for the households which can be longitudinally linked between the two waves. For the X axis, the difference in aggregated expenditure between wave 7 and 8 (i.e., wave 8 aggregated expenditure-wave 7 aggregated expenditure) was plotted and the mid-points of the aggregated expenditure bands are shown on the graph. The grey line shows how the aggregated expenditure change for the households before expenditure imputation, and the dark line presents the same information after imputation. There are 5,342 households which can be linked between wave 7 and 8 and therefore included in the post-imputed estimates in this graph. Since aggregated expenditure is used here, the pre-imputed data only includes households that answer every expenditure item in waves 7 and 8, which is 3,048 households. We can see after imputation, the distribution of change in aggregate expenditure is more peaked. This suggests that after imputation, we have more households with the change of aggregate expenditure between the waves within the range of -\$5,000 to \$5,000. We also looked at the change of the aggregate expenditure between wave 6 and wave 7 and wave 5 and wave 8 and observed a very similar pattern – the post-imputed distribution is slightly more peaked. This suggests that after imputation, we have more households with fewer changes in aggregate

expenditure between waves. Nevertheless, the differences between the pre- and post-imputed data are really small, and no serious concerns have been raised about the imputation.

**Figure 4: Distribution of changes in the aggregated expenditure between wave 7 and 8**



### Childcare Expenditure

One expenditure item which is collected in the HILDA Survey, but has not been included in this paper is child care costs. HILDA collects child care costs in HQ from wave 1. The child care questions have changed a few times across the waves:

- In wave 1, only information about child care used while the parents were working was collected. The questions about child care used while parents could undertake non-employment related activities are included from wave 2.
- The level of details collected is different across waves:
  1. For wave 2 to 4, summary information for non-employment related child care cost was collected. From wave 5, a similar level of detail to employment related child care was collected.
  2. In wave 1, the cost of each type of child care for each child was collected. From wave 2 onwards, we collect the total cost for each type of child care for school aged children and those not yet at school.

Expenditure on child care has not been considered in the current expenditure imputation work, and is excluded from all the analysis in this paper. For Release 9, we will incorporate the imputation for the child care costs. Table 14 gives an idea of the extent of the missingness for child care costs. There are not many cases with missing child care expenditure, however, the percentage of cases with missing child care expenditure varies a lot across the waves for some variables. For example, we have more cases with missing total child care costs for all school age children that is not employment related in wave 2 than other waves. How to best impute the childcare costs will be considered before the next release.

**Table 14: Number and percentage of households with missing childcare expenditure, wave 1-8**

<i>Variable</i>	<i>Wave</i>							
	1	2	3	4	5	6	7	8
<b>Number of households with missing childcare costs</b>								
Total costs for all school age children during term while parents work	21	21	7	11	4	4	10	7
Total costs for all school age children during school holiday while parents work	25	11	9	6	5	3	8	3
Total costs for all not yet at school children while parents work	10	8	3	5	3	0	6	3
Total costs for all school age children not employment related	-	29	3	11	5	2	10	9
Total costs for all not yet at school children not employment related	-	7	2	7	6	2	11	12
<b>Percentage of households with missing childcare costs (non-zero cases only)</b>								
Childcare total costs for all school age children during term while parents work	9.5	8.0	3.0	5.1	1.8	1.8	4.7	3.6
Childcare total costs for all school age children during school holiday while parents work	12.1	5.4	4.4	3.6	3.0	1.7	5.2	1.8
Total costs for all not yet at school children while parents work	3.0	2.4	1.0	1.6	0.9	0.0	2.0	0.9
Total costs for all school age children not employment related	-	29.0	5.4	17.7	10.6	3.5	22.2	20.9
Total costs for all not yet at school children not employment related	-	3.4	1.0	3.6	3.5	1.4	7.1	8.8

### Concluding Remarks

This paper detailed the imputation methods applied to the expenditure data collected in the HILDA Survey and examined the quality of the imputed data.

The expenditure data in the HILDA Survey are mainly collected in the SCQ from wave 5 onwards, and a few variables are collected in the HQ every wave. Expenditure items collected in the SCQ are more prone to missingness, since SCQs are not always obtained from the person responsible for the household bills. The percentage of missing expenditure items collected in the HQ (like mortgage repayment and rent payment) is less than 2 per cent every wave, whereas for the items collected in the SCQ, more than 15 per cent of households have missing expenditure items. The imputation is done at the household level. The methodology used for expenditure imputation is similar to that used for the income and wealth imputation, where the Little and Su method, the nearest neighbour regression method and the population carry-over method, are used. The biggest challenges for the expenditure imputation were determining whether the household has the expenditure or not and how to impute the lumpy items in order to preserve the irregular nature of these items. The population carryover method was used first to carry over the zeros from abutting waves for non-lumpy items, and the nearest neighbour regression method was used to as a filter process to update the zeros for all cases. When imputing the lumpy items, we did not use the carryover method, as carryover methods are more likely to overstate correlation between waves (Starick and Watson, 2007) and when using the

Little and Su method, the donors were selected from those with the same zero pattern as the recipient's.

The quality of imputation was evaluated. We first looked at the distribution before and after imputation. The mean and median before and after imputation were very close for most variables. For some variables, the mean was slightly lower after imputation. This was partially because we had slightly more zeros in the post-imputed data. The imputation slightly increased the proportion of households with aggregated expenditure less than \$25,000, and slightly pushed down the proportion of households with high aggregated expenditure. The study also showed that the imputation did not change the relationship between income and aggregated expenditure. To explore the longitudinal component of the imputation, the change in the aggregated expenditure between waves was examined for households that can be longitudinally linked. After imputation, the distribution of the change of aggregated expenditure between waves appears to be more peaked. These investigations suggest that the quality of imputed data appears to be good.

We also compared the HILDA mean estimates of expenditure with the HES estimates. Although the comparability between the two surveys is limited, it still gives some indications on how well the expenditure items were measured in HILDA. It appears that the HILDA Survey successfully measured 13 items out of 23 items it collects from wave 6. Expenditure on motor vehicle fuel, clothing and footwear, holiday and holiday travel costs, private health insurance, medicines, prescriptions and pharmaceuticals, education fees, and expenditure on household durables like vehicles, household appliance and furniture appears to be not so well-measured, where the differences between the HILDA and HES estimates are more than 10 per cent. Nevertheless, these items collected in the HILDA Survey may still have longitudinal and distributional value, which is not available from other surveys. Users should be careful when using the items.

## References

- Barrett, G.F., Crossley, T.F., & Worswick, C.(1999). 'Consumption and income inequality in Australia', *Economics Record*, Vol. 76, no. 233, p 116-138.
- Browning, M., Crossley, T.F., & Weber, G. (2003), ' Asking Consumption Questions in General Purpose Surveys', *Economics Journal*, 113 (November), P540-567.
- Crossley, T. and Pendakur, K. (2002), 'Consumption Inequality', D. Green (ed) *Dimension of Inequality in Canada*.
- Friedman, M (1957). '*A Theory of the Consumption Function*'. Princeton University Press. [ISBN 13 978-0691041827](#)
- Hayes, C and Watson, N (2009), 'HILDA Imputation', HILDA Project Technical Paper Series No. 2/09, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.
- Headey, B (2008), 'Poverty Is Low Consumption and Low Wealth, Not Just Low Income', *Social Indicators Research* Oct 2008, Vol 89 Issue 1, p23-39.
- Ringen, S. (1988). 'Direct and Indirect Measures of Poverty'. *Journal of Social Policy*, Vol 17, p 351-365
- Starick, R, and Watson, N. (2007), 'Evaluation of Alternative Income Imputation Methods for the HILDA Survey', HILDA Project Discussion Paper Series No. 1/07,

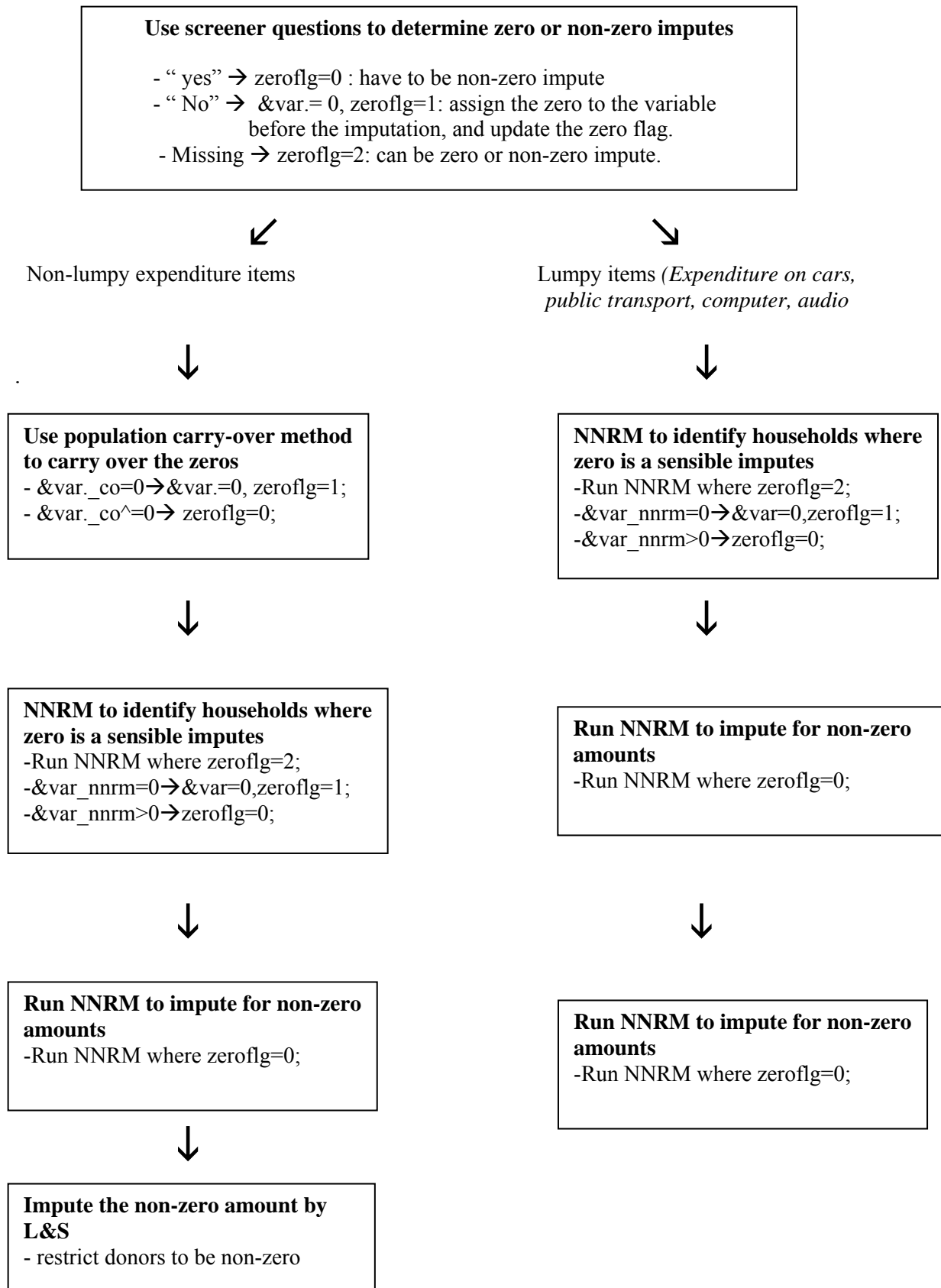
Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Wilkins, R., Warren, D. and Hahn, M.(2009), 'Families, Incomes and Jobs, Volume 4: A Statistical Report on Wave 1 to 6 of the HILDA Survey', Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Wilkins, R. and Sun, C. (2010), 'Assessing the Quality of the Expenditure Data Collected in the Self-completion Questionnaire', HILDA Project Discussion Paper Series No. 1/10.

## Appendix 1: Overall Imputation Process for the Expenditure Items

Figure 5: Overall imputation process for expenditure items



## Appendix 2: Variables Used in the Nearest Neighbour Regression Models

---

### Household Level

---

#### Household income

Household financial year wages and salary  
Household financial year government income  
Household financial year business income  
Household financial year total income  
Equivalised disposable household income

#### Household characteristics

Family type  
Number of bedrooms  
Household tenure  
Family size  
SEIFA 2001 Decile of Index of relative socio-economic disadvantage  
SEIFA 2001 Decile of Index of economic resources  
SEIFA 2001 Decile of Index of education and occupation

#### Household summary

Number of adults  
Number of children  
Number of employed people  
Number of unemployed people  
Number of people not in labour force  
Number of males  
Number of females  
Number of people speaking English  
Number of people have long term health condition  
Average adult age  
Average child age  
Age of youngest child  
Age of oldest child  
Age of youngest adult  
Age of oldest person in the household  
Number of people having current benefit income  
Age of person with highest income in the household  
Gender of person with highest income in the household  
Number of people with university degree or higher

#### Household assets

home value

#### Mortgage information

Mortgage amount  
Mortgage schedule  
Other mortgage amount

---

### Appendix 3: Distribution of Expenditure Data Before and After Imputation, Waves 1 to 7

**Table 15: Wave 1 unweighted distribution of expenditure data (non-zero cases) before and after imputation**

Variable	Before Imputation			After Imputation		
	Mean	Median	Standard Deviation	Mean	Median	Standard Deviation
<b>Households (non-zero only)</b>						
Rental payments	647	608	366	647	607	369
Mortgage repayments (first and second)	1,016	850	814	1,024	850	831
Weekly expenditure on grocery	131	120	69	131	120	69
Weekly expenditure on food	100	90	58	100	90	58
Weekly expenditure on meals eaten outside	40	30	42	40	30	43

**Table 16: Wave 2 unweighted distribution of expenditure data (non-zero cases) before and after imputation**

Variable	Before Imputation			After Imputation		
	Mean	Median	Standard Deviation	Mean	Median	Standard Deviation
<b>Households (non-zero only)</b>						
Rental payments	659	608	419	655	607	418
Mortgage repayments (first and second)	1,038	869	783	1,043	869	783

**Table 17: Wave 3 unweighted distribution of expenditure data (non-zero cases) before and after imputation**

Variable	Before Imputation			After Imputation		
	Mean	Median	Standard Deviation	Mean	Median	Standard Deviation
<b>Households (non-zero only)</b>						
Rental payments	689	630	407	687	630	409
Mortgage repayments (first and second)	1,121	934	817	1,126	934	828
Weekly expenditure on grocery	133	120	74	133	120	74
Weekly expenditure on food	103	90	63	103	90	63
Weekly expenditure on meals eaten outside	42	30	41	42	30	41

**Table 18: Wave 4 unweighted distribution of expenditure data (non-zero cases) before and after imputation**

Variable	Before Imputation			After Imputation		
	Mean	Median	Standard Deviation	Mean	Median	Standard Deviation
<b>Households (non-zero only)</b>						
Rental payments	722	652	417	720	652	419
Mortgage repayments (first and second)	1,239	1,043	943	1,234	1,031	940
Weekly expenditure on grocery	137	120	74	137	120	75
Weekly expenditure on food	106	90	63	106	90	63



**Table 18: (c'td)**

<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>
<b>Households (non-zero only)</b>						
Weekly expenditure on meals eaten outside	44	30	42	44	30	42

**Table 19: Wave 5 unweighted distribution of expenditure data (non-zero cases) before and after imputation**

<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>
<b>Households (non-zero only)</b>						
Rental payments	757	695	478	754	695	478
Mortgage repayments (first and second)	1,359	1,130	997	1,362	1,115	1,017
Weekly expenditure on grocery	142	125	79	142	125	79
Weekly expenditure on food	111	100	67	111	100	68
Weekly expenditure on meals eaten outside	47	30	46	47	30	46
Groceries	7,593	6,952	4,224	7,460	6,756	4,204
Alcohol	1,788	1,304	1,727	1,817	1,304	1,777
Cigarettes and tobacco	2,595	2,347	1,870	2,593	2,346	1,880
Public transport and taxis	1,269	926	1,489	1,270	913	1,486
Meals eaten out	1,323	912	1,670	1,330	959	1,661
Leisure activity	1,026	600	1,633	1,045	600	1,888
Motor vehicle fuel	2,056	1,500	1,973	2,041	1,500	1,940
Clothing and footwear	1,091	720	1,251	1,083	720	1,238
Telephone rent and calls	1,454	1,080	1,932	1,461	1,080	1,927
Holidays and holiday travel costs	2,250	1,500	2,888	2,229	1,500	2,825
Private health insurance	1,537	1,400	1,035	1,522	1,400	1,036
Health care	741	500	983	726	500	956
Electricity	899	800	683	884	800	680
Gas	486	400	473	481	400	462
Other heating sources	276	200	297	276	200	296
Utility	1,205	1,050	913	1,166	1,000	902
Home repairs, renovation and maintenance	1,786	750	8,734	1,763	750	8,365
Motor vehicle repairs and maintenance	947	700	1,007	947	700	1,082
Education fees	2,554	830	4,800	2,532	800	4,725

**Table 20: Wave 6 unweighted distribution of expenditure data (non-zero cases) before and after imputation**

<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>
<b>Households (non-zero only)</b>						
Rental payments	807	739	490	805	739	492
Mortgage repayments (first and second)	1,514	1,300	1,214	1,526	1,300	1,235
Weekly expenditure on grocery	8,083	7,821	4,667	7,954	7,561	4,665
Weekly expenditure on food	1,938	1,390	2,098	1,959	1,434	2,138
Weekly expenditure on meals eaten outside	2,602	2,373	1,829	2,610	2,377	1,855

**Table 20: (c'td)**

<i>Variable</i>	<i>Before Imputation</i>			<i>After Imputation</i>		
	<i>Mean</i>	<i>Media n</i>	<i>Standard Deviation n</i>	<i>Mean</i>	<i>Media n</i>	<i>Standard Deviation n</i>
<b>Households (non-zero only)</b>						
Groceries	1,192	834	1,326	1,190	787	1,313
Alcohol	2,554	2,086	2,771	2,577	2,086	2,883
Cigarettes and tobacco	2,532	1,800	3,209	2,554	1,800	3,327
Public transport and taxis	1,808	1,200	2,335	1,809	1,200	2,874
Meals eaten out	1,784	1,200	3,516	1,775	1,200	3,670
Leisure activity	3,610	2,000	5,523	3,579	2,000	5,472
Motor vehicle fuel	1,507	1,300	1,000	1,486	1,298	993
Clothing and footwear	1,273	1,000	1,227	1,259	1,000	1,218
Telephone rent and calls	1,018	500	2,364	1,003	500	2,281
Holidays and holiday travel costs	437	250	1,733	449	241	2,105
Private health insurance	1,190	1,000	1,021	1,185	1,000	1,086
Health care	4,423	1,000	18,911	4,297	1,000	18,306
Electricity	1,010	700	1,150	1,003	700	1,138
Gas	2,557	830	4,822	2,648	900	5,113
Other heating sources	16,05	10,00		16,11	10,00	
Utility	7	0	43,041	2	0	41,495
Home repairs, renovation and maintenance	1,050	600	1,137	1,067	600	1,198
Motor vehicle repairs and maintenance	1,031	500	1,623	1,035	500	1,633
Education fees	1,371	800	2,119	1,377	800	2,128
	1,629	1,000	2,407	1,605	995	2,361

**Table 21: Wave 7 unweighted distribution of expenditure data (non-zero cases) before and after imputation**

<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>
<b>Households (non-zero only)</b>						
Rental payments	865	800	501	862	782	502
Mortgage repayments (first and second)	1,688	1,400	1,443	1,690	1,400	1,472
Weekly expenditure on grocery	8,478	7,821	4,941	8,369	7,821	4,927
Weekly expenditure on food	2,002	1,564	2,469	2,022	1,564	2,411
Weekly expenditure on meals eaten outside	2,747	2,607	2,117	2,768	2,607	2,106
Groceries	1,243	1,043	1,293	1,274	1,043	1,363
Alcohol	2,752	2,086	3,222	2,802	2,086	3,548
Cigarettes and tobacco	2,420	1,800	2,454	2,421	1,800	2,459
Public transport and taxis	1,930	1,200	2,570	1,908	1,200	2,517
Meals eaten out	1,846	1,200	2,338	1,865	1,200	2,482
Leisure activity	3,872	2,000	5,828	3,887	2,000	5,768
Motor vehicle fuel	1,610	1,400	2,074	1,583	1,375	1,977
Clothing and footwear	1,325	1,000	1,350	1,309	1,000	1,303
Telephone rent and calls	1,074	500	2,205	1,052	500	2,118
Holidays and holiday travel costs	458	250	1,199	444	250	1,115
Private health insurance	1,227	1,000	1,013	1,211	1,000	1,020
Health care	4,470	1,000	17,009	4,609	1,000	18,105
Electricity	1,026	738	1,073	1,032	700	1,122
Gas	2,672	1,000	4,938	2,738	1,000	5,016

**Table 21: (c'td)**

<i>Variable</i>	<i>Before Imputation</i>			<i>After Imputation</i>		
	<i>Mean</i>	<i>Media n</i>	<i>Standard Deviation</i>	<i>Mean</i>	<i>Media n</i>	<i>Standard Deviation</i>
<b>Households (non-zero only)</b>	15,60	10,00		15,82	10,00	
Other heating sources	9	0	16,435	3	0	16,850
Utility	1,043	600	1,136	1,056	600	1,189
Home repairs, renovation and maintenance	1,201	500	1,816	1,217	500	1,918
Motor vehicle repairs and maintenance	1,320	800	1,752	1,348	800	1,925
Education fees	1,685	1,000	2,432	1,672	1,000	2,424