**DRAFT – Do not quote without permission (Comments welcome)**

# HOUSING LEVERAGE IN AUSTRALIA

Luci Ellis*, Jeremy Lawson** and Laura Roberts-Thomson**

Paper presented at the HILDA conference, March 2003

*Economic Analysis Department
**Economic Research Department

Reserve Bank of Australia

The authors thank Gianni La Cava for providing invaluable assistance with the imputation of household income for some respondents in the HILDA (Living in Australia) survey, and Steven Stillman of the New Zealand Department of Labour for providing Stata code implementing our main estimation technique. We have benefited from helpful conversations with Alex Heath and other colleagues at the Reserve Bank. This paper would not have been possible without the hard work of the developers and sponsors of the Living in Australia Survey. Responsibility for any remaining errors rests with the authors. The views expressed in this paper are those of the authors and should not be attributed to the Reserve Bank.

# Abstract

A home is the single largest purchase that most households make, and it is one that usually requires some debt financing. It is important to understand this decision because housing debt is such a large component of households' balance sheets. In this paper, we use unit record data from the HILDA survey to examine households' decisions about their leverage using both graphical and econometric techniques. Because housing leverage is observed only when a household is an owner-occupier and has not already paid off any loans used to fund the home's purchase, we find that we must correct for this selection bias before drawing conclusions about population behaviour. We find evidence of considerable diversity in housing leverage, correlated with households' observable characteristics. The results indicate that leverage is affected both by forces external to the household and households' conscious decisions.

JEL Classification Numbers: D12, G21, R21

Keywords: household survey, housing debt, leverage

# Table of Contents

# HOUSING LEVERAGE IN AUSTRALIA

**Luci Ellis, Jeremy Lawson and Laura Roberts-Thomson**

## 1. Introduction

The effect of monetary policy on household consumption in Australia is most visibly mediated through the effect of movements in mortgage interest rates on household cash flows, and thus consumption and saving decisions. The magnitudes of these effects are largely determined by the amount of debt that households carry, and the sensitivity to policy rates of interest rates paid on that debt. Australia's housing sector has long been characterised by relatively high home ownership rates and a predominance of variable-rate mortgages. Thus it might be expected that fluctuations in housing prices would have a relatively strong effect on consumption, and that housing would have an important role in the transmission of monetary policy (McLennan, Muellbauer and Stephens 1999). When the cash rate rises, variable mortgage rates also rise, impinging on the cash flows of indebted households. This will tend to reduce their consumption, unless they are able to offset this cash-flow effect with further borrowing, or reductions in excess repayments of principal.

It is therefore crucial for policy makers to understand the housing market and how it affects the rest of the economy. Our focus in this paper is on household balance sheets, and one dimension of those balance sheets in particular – households' housing leverage. Although households' debt-income ratios determine the relative effect of a given-sized change in interest rates on their cash flows, we take the view that it is leverage – the debt-assets ratio – that critically determines the level of debt that households are willing to bear, in addition to the burden of repayments.

Leverage might be expected to influence the behaviour of households and intermediaries through a number of channels. For example, households' desired value for leverage might be endogenously affected by the business cycle or uncertainty; in the model in Carroll and Dunn (1997), an increase in the expected probability of unemployment causes some households to assess their precautionary savings as inadequate, and thus reduce their consumption to repair their balance sheets.[1]

Leverage is also relevant to the transmission mechanism because of its likely implications for credit supply. Increases in interest rates might not induce households to reduce consumption if they can borrow additional funds, but intermediaries' willingness to lend more to households might be reduced if leverage is particularly high. Households with higher leverage might therefore be less likely to offset the cash-flow effects of monetary policy with further borrowing. Their ability to smooth their consumption during downturns might thus be constrained by asset-price developments associated with that downturn (Bernanke and Gertler 1995; Carroll and Dunn 1997). In addition, since real estate is widely used as collateral for loans, the level of leverage is a determinant of the balance-sheet risk of financial institutions (Kent and Lowe 1997; Schwartz 2002).

Finally, the interaction of leverage with movements in house prices will determine the prevalence of negative equity, which in turn has implications for labour market flexibility (Henley 1999) and the behaviour of the real estate market (Genesove and Mayer 1997). Although we do not cover these issues in this cross-sectional study, an understanding of housing leverage may help explain housing market features with dynamic or panel aspects, such as pricing inertia and the correlation between falling prices and a larger stock of unsold homes.

In this paper, we focus on cross-sectional, microeconomic aspects of households' housing leverage, with a view to understanding which households are most likely to be affected by changes in interest rates or falls in housing prices. In the next section, we discuss the HILDA data set and our approach to imputing missing income data; the econometric results underlying our imputation methods are presented in Appendix A. After undertaking some preliminary graphical analysis in Section 3, we set up our core econometric model in Section 4 and discuss its implications. A brief conclusion follows in Section 5. Appendix B contains the results for our selection equations.

---

1 Smith, Sterne and Deveraux (1994) found that households in the UK sharply increased lump-sum repayments to mortgage lenders in the early 1990s, in response to the falls in housing prices and increased prevalence of negative equity; this effect was large enough to affect the personal sector savings rate.

# 2. The HILDA Data Set

In this study we use data from the Household Income and Labour Dynamics in Australia (HILDA) Survey. HILDA is a household-based panel or longitudinal survey that aims to track all members of an initial sample of Australian households over time. The survey was commissioned by the Department of Family and Community Services and is directed by the Melbourne Institute; AC Nielsen collected the data. This study uses data from the first wave of the survey, collated from interviews conducted of some 14 000 individuals living in almost 7 700 households over the second half of 2001. The survey covers four broad areas: economic wellbeing, labour market dynamics, family dynamics, and subjective wellbeing.

## 2.1 Constructing the housing leverage variable

Housing leverage is typically expressed as a loan-to-valuation (LTV) ratio, in contrast to the debt-equity ratios commonly used in analysis of corporate finance. We can construct LTV ratios for homeowners' principal residence using information from the household questionnaire in Wave 1 of HILDA. Homeowners were asked to estimate the current value of their home, and to report the amount they currently owed on loans taken out against that home, including institutional mortgages, loans from family, friends and other members of the community, and home equity loans. To calculate the LTV ratio, we total the outstanding amounts of all borrowings against the principal residence and divide it by the estimate of the home's value. Households who rented, or who occupied their home rent-free but did not own it, were not asked these questions. Information on other properties such as investment properties or holiday homes was not included, so our study relates specifically to owner-occupiers' principal residences.

The use of subjective valuations raises the question of their accuracy. Goodman and Ittner (1993), using US survey data, find that there is a small positive bias of about 6 per cent in homeowners' estimates, but that the mean absolute error of estimates tends to be larger at about 15 per cent. This may be due to rounding errors: 42 per cent of households reporting estimated home values in HILDA reported a figure that was a multiple of $50 000. We believe our analysis is unlikely to be significantly biased by homeowners' subjectivity. If the bias in their estimates is small and, as Goodman and Ittner find, unrelated to owners' characteristics, then our point estimates will not be significantly biased even if these rounding errors make them less precise. In any case, households' behaviour presumably depends on their perceptions of their leverage rather than realised leverage, especially if they are not intending to sell their homes in the near future.

## 2.2 Missing data and income imputation

Compared with similar international household surveys, HILDA does not suffer greatly from problems of missing data (Watson and Wooden 2002). However, there is a relatively high incidence of missing data for income-related questions. We can separate the most common reasons for non-response into “item non-response” and “incomplete households”. *Item non-response* occurs when a member of a selected household agrees to be interviewed, but then either refuses, or is unable, to answer some of the questions asked. This is the main source of missing data, accounting for 64 per cent of the missing household income information. Most of the missing income data is due to item non-response for income sourced from business (missing 23.5 per cent of people with business income) and investments (missing 8.1 per cent of recipients of this kind of income). Wages and salaries (missing 7.2 per cent of wage earners) and government benefits and pensions (missing 1.4 per cent of benefit recipients) have lower incidences of missing data.

The other major source of missing data is the 810 *incomplete households,* accounting for 10.5 per cent of the household sample and 36 per cent of the missing household income information; these are households in which not all eligible adult members agreed, or were able, to be interviewed. The HILDA data set as distributed does not include an entry for household income if any of its eligible members was not interviewed, or did not report complete income information; in all, 29 per cent of households have a missing value for household income, which is clearly an unacceptable data loss.

In such circumstances we have two choices. We can drop the 29 per cent of households for which income data is missing from the sample, or impute the income of the individuals with missing data. Our choice to impute income for missing individuals is shaped by two factors. First, because income non-response is not random or uncorrelated with the variable(s) of interest, the missing cases cannot be safely dropped from the sample (Watson and Wooden 2002). For example, men, individuals outside the labour force, individuals living in Tasmania and Perth, people that have been divorced, and people that have a high regard for their leisure time (and generally have low incomes) were more likely to offer complete income information than other individuals. Second, we have a large cross-section of information from the HILDA survey that permits us to do a reasonable job of imputing income for missing individuals.

### *2.2.1 Imputation methodology*

Following the recommendations of the HILDA survey team and methods adopted in the British Household Panel Survey (BHPS) we impute income using the "predictive mean matching" method (Little 1988; ISER 2002; Watson and Wooden 2002). This is a stochastic imputation technique that has the advantage of maintaining the underlying distribution of the data by allowing the imputation of error around the mean. Appendix A outlines the method in detail and shows the regression results for the three models.

The nature of the missing data leaves us with the need to impute income for three separate types of missing cases:

1. Individuals that did not complete a person questionnaire and therefore did not report any income information (Type I) (n = 1158)

2. Individuals that completed a person questionnaire but did not provide information on wage income (Type II) (n = 673)

3. Individuals that completed a person questionnaire but did not provide information on non-wage income (Type III) (n = 1621)

Three separate models are estimated to impute income for each type of missing case. For Type I respondents we have information on the characteristics of their household (e.g. value of the dwelling, geographic location, the number of bedrooms) and a limited range of personal information from the household questionnaire. We also have personal information collected about other respondents in the household. These "family variables" include the income, labour force status and occupation of other household members. Both the household and family variables are likely to be correlated with both personal and household income and hence act as useful explanatory variables in the model. We impute total gross financial year income for these individuals.

We have the same information for Types II and III respondents, but also additional personal information obtained from items that they did complete during the interview – labour force status, age, gender, English-speaking background – including information about the sources of their income. This allows us to predict wage and non-wage income in the final two models, and add the income that individuals report from other sources to our estimates. For example, for

Type III individuals we add their imputed non-wage income to any actual reported wage and salary income.

### *2.2.2 Imputation results*

In the regression model for Type I households our model explains nearly 32 per cent of the variation in total gross household income. The root mean square error (RMSE) is about $26 000. In the regression model for Type II households our model explains about 46 per cent of the variation in individuals' wage and salary income and the RMSE is nearly $19 000. In the regression model for Type III households our model explains nearly 21 per cent of the variation in individuals' non-wage income and the RMSE is about $20 500. Although these errors are quite large, we regard the imputation as being relatively successful, not least because it allows us to use reported income for other income and household members that would otherwise be lost.

Our income imputation strategy allows us to recover household income estimates for all but 201 households (about 3 per cent of the sample), ensuring that any bias introduced by dropping missing observations from the sample is minimised. However, because our imputed household income estimates are likely to diverge from the true income that households did not report, we also construct a dummy variable for those households with imputed household income. This dummy was significant in the mortgage equation, but not in either the tenure or the leverage equation.

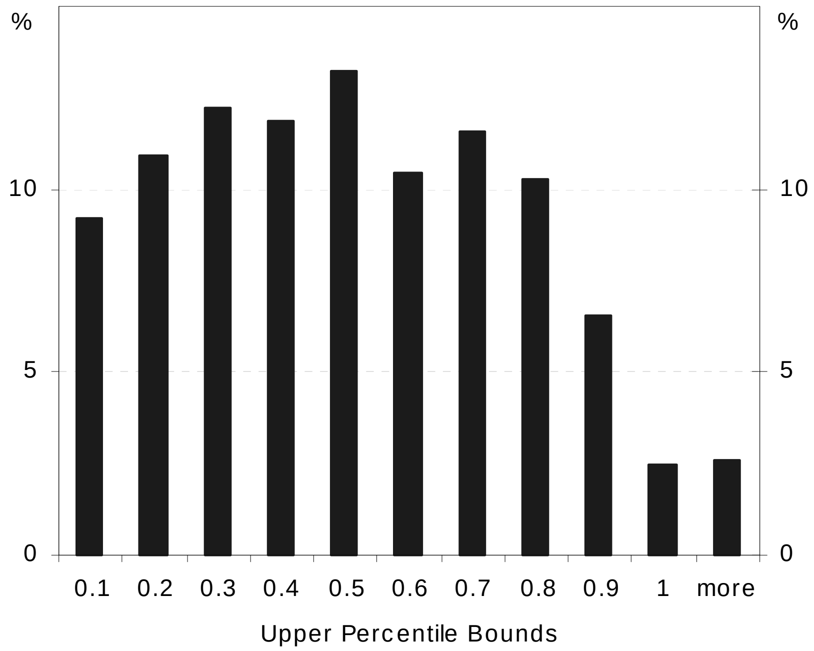
## 3. Preliminary Analysis

Before we outline an estimated model of household housing leverage in Section 4, we first present some stylised facts about housing leverage in Australia, identify some of the household characteristics likely to influence households' housing leverage, and undertake a brief graphical analysis of how three key variables – age of the household head, household income, and household housing wealth – are related to leverage.

According to the HILDA survey the home ownership rate in 2001 was 68 per cent, broadly consistent with both the 2001 Census and the 1998 Household Expenditure Survey (HES). Of these owner-occupying households 42 per cent were still paying off their home, such that households with mortgages represented 29 per cent of the households surveyed. The average level of household leverage implied by the HILDA data is a little below the level implied by the aggregate credit data – despite including loans from friends and family that are not in the aggregate data – at around 15 per cent for all owner-occupying households and 48 per cent for those households with outstanding housing debt, which seems fairly moderate.

Although these figures confirm that on average housing debt is much lower than the value of housing assets, we are also interested in how this leverage is distributed across Australian households. Figure 1 shows that few Australian households have high leverage (defined here as a household with an LTV greater than 80 per cent).[2] Only 11.4 per cent of households with housing loans (less than 4 per cent of all households) had an LTV that exceeded 80 per cent, and less than 3 per cent had negative equity in their homes.[3] There is also considerable diversity in leverage holdings, with similar proportions of households having leverage in each leverage decile below 80 per cent.

**Figure 1: Distribution of Household LVRs**
**Percentage of Households with Housing Loans**




---

2 Our use of an 80 per cent threshold is consistent with evidence in Genesove and Mayer (1997) that household behaviour alters when leverage exceeds this threshold, and it allows us to consider how the incidence of negative equity might change if house prices were to fall by a significant amount.

3 Note that there may be other households with high leverage against their properties not captured here. For example some households that have low leverage on their principal home could be highly leveraged against an investment property. Unfortunately this possibility cannot be explored until Wave 2 HILDA data, which will incorporate questions about households' wealth, are released next year.

### 3.1 Leverage and household characteristics

We can use the wealth of cross-sectional information about Australian households contained in the HILDA dataset to investigate whether the diversity of household leverage holdings across Australian households is related to their observable characteristics.

There are a number of household characteristics that can be expected to influence housing leverage. First, we would expect leverage to be affected by a household's stage in the life cycle. Leverage should initially be low for young households, because the completion of studies, the establishment of careers, and saving to purchase a home, mean that few will be homeowners. Indeed, younger people may not have formed households at all for these reasons. Leverage should then rise as households move into their peak family formation years. Because they have been in the workforce for a relatively short time, such households have usually accumulated little wealth and so must borrow to purchase large assets such as the family home. This debt is then steadily reduced over their working lives so that they are relatively debt-free when their incomes drop sharply upon retirement.

Variables that capture a household's position in the life cycle, such as the household head's *age*, *labour force status*, *marital status*, and whether *children* are present in the home should therefore be important in explaining housing leverage. We also expect that the time since the mortgage was taken out, which we proxy using *time the household has lived in the home*, will be an important determinant of household leverage.

A household's means, and in particular its *income*, should also affect its leverage by influencing its willingness, need and ability to take on housing debt. We expect housing leverage to increase with income for three main reasons. First, higher-income households are more likely to be homeowners in the first place, because they are more likely to have been able to save the necessary downpayment. Second, we expect households that have paid off their mortgages to have lower incomes than households with extant mortgages, because retired lower-income households prefer to have already paid off their debt. Finally, housing leverage may be expected to increase with income if financial institutions apply easier lending criteria to high-income households. The expected importance of household means in predicting leverage also leads us to anticipate that alternative indicators of household means, such as self-assessed *income adequacy*, the past occurrence of *family breakdown*, and receipt of *government income support*, should also influence leverage. Offsetting this, high-income households have more scope to pay off their loan quickly and therefore own their home outright.

Finally, we would expect households' attitude to home ownership and housing debt to influence their housing leverage. For example, households that are *uncomfortable about taking on debt* may accumulate more savings before purchasing a home than households that feel more comfortable taking on debt. They may also choose to pay their loans off faster. Other variables that may pick up different attitudes to home ownership or housing debt across households include households' *ethnic background*, *aversion to risk*, and *credit card usage*.

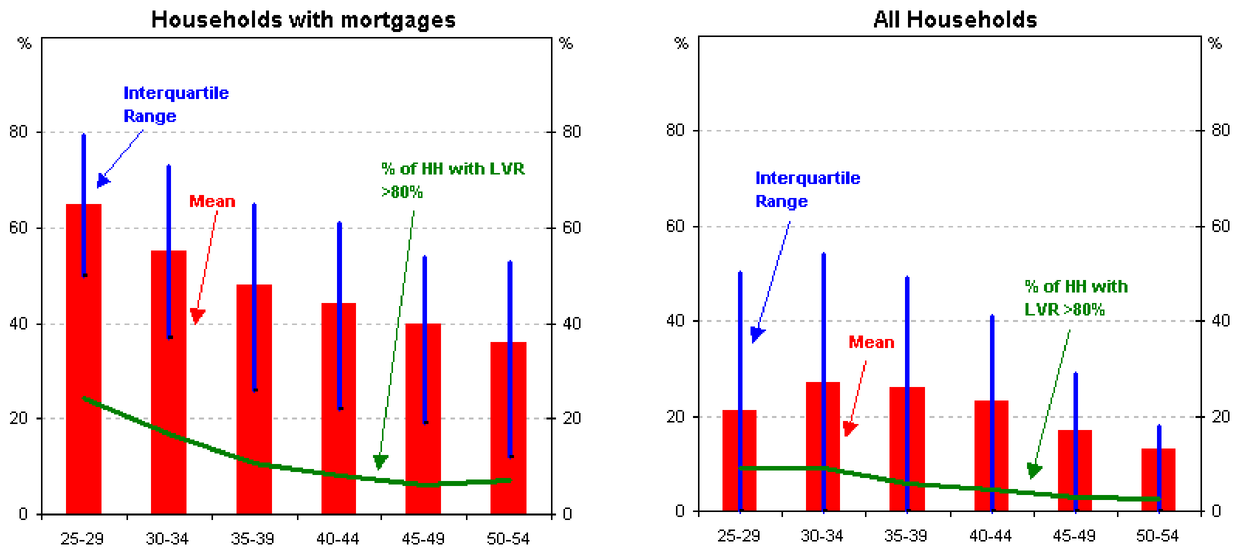
## 3.2 Graphical analysis

In the previous section we identified a number of household characteristics that can be expected to influence a household's housing leverage. Before using this information in an econometric model, it is worth examining graphically how three key variables – age of the household head, household income, and household housing wealth – are related to housing leverage, treating households without mortgages as having zero leverage.

### *3.2.1 Age of the household head*

Figure 2 shows how housing leverage varies across age cohorts in two panels. In addition to mean leverage for each cohort we also include information about the distribution of leverage – the inter-quartile range, and the percentage of households with an LVR greater than 80 per cent. The left-hand panel contains data only for households with mortgages, while the right-hand panel includes data for all households.

**Figure 2: Housing Leverage by Age Group**



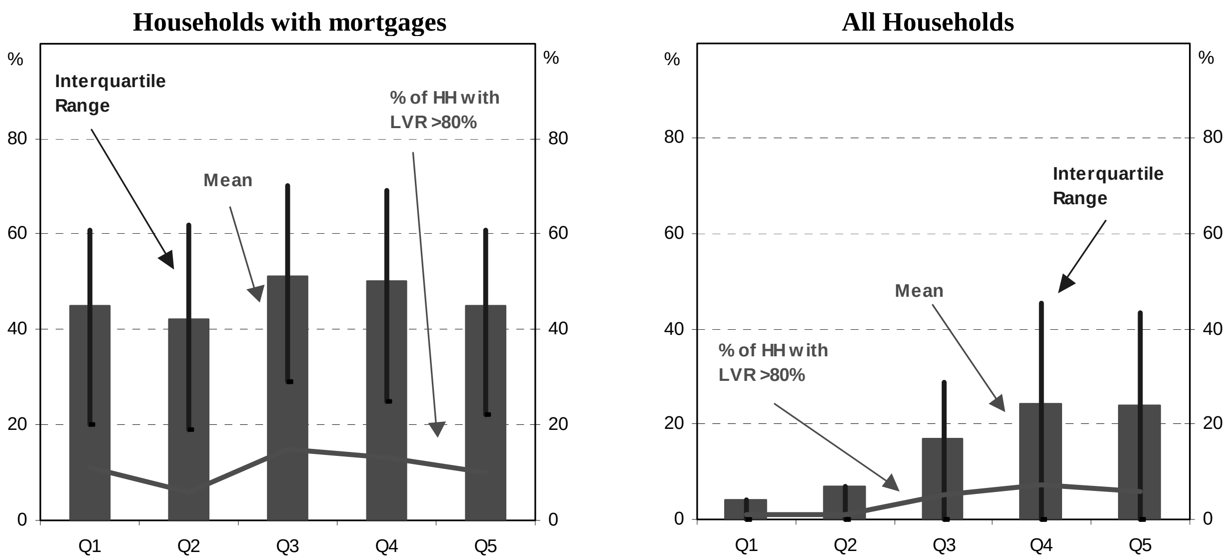

Looking first to the left panel, younger households with mortgages appear to have higher leverage than their older counterparts. Average leverage was almost 70 per cent for households where the household head was aged between 25 and 29, compared with 40 per cent for households with a head aged between 45 and 49. Higher average leverage for younger households is also reflected in a greater incidence of high leverage. About 25 per cent of households with heads aged between 25 and 29 had leverage exceeding 80 per cent, a number that falls to just over 6 per cent for households with heads aged between 45 and 49. Because there is some association between age of the household head and time spent in homeownership, it is difficult to distinguish whether this is a pure age effect, or a reflection of the time the household has had to pay off their initial loan. This issue will be investigated in our econometric analysis in the following sections.

The right panel of Figure 2 shows that when younger households' lower homeownership rate is taken into account, their apparent higher leverage is reduced, and the nature of the relationship between age and leverage alters. Average leverage now increases to a peak when the household head is aged 30 to 34 before falling away as households pay off their housing debt. This observation is important because it shows that if we use a sample that only includes households with mortgages, we may exaggerate the incidence of leverage when households in a category are less likely to have a loan, either because of a lower home ownership rate, or because homeowners in the category are more likely to have paid their loan off. In our graphical analysis we make use both samples so that we can observe pockets of high leverage in categories that otherwise have low debt. However, in our subsequent empirical modelling we must take this sample selection bias into account to draw accurate population inferences about the relationship between household characteristics and housing leverage.

### *3.2.2 Household income*

Figure 3 suggests that there may be a hump-shaped relationship between household income and leverage. Average leverage is greatest among upper-middle income households, and falls away somewhat for both low-income and high-income households. When the lower homeownership rate, and share of households with existing loans in the bottom two income quintiles is taken into account (right panel), their average leverage looks particularly low. This may be an age effect; the average age of household heads in quintiles 1 and 2 is higher (at 55) than in the three highest income quintiles (at about 42).

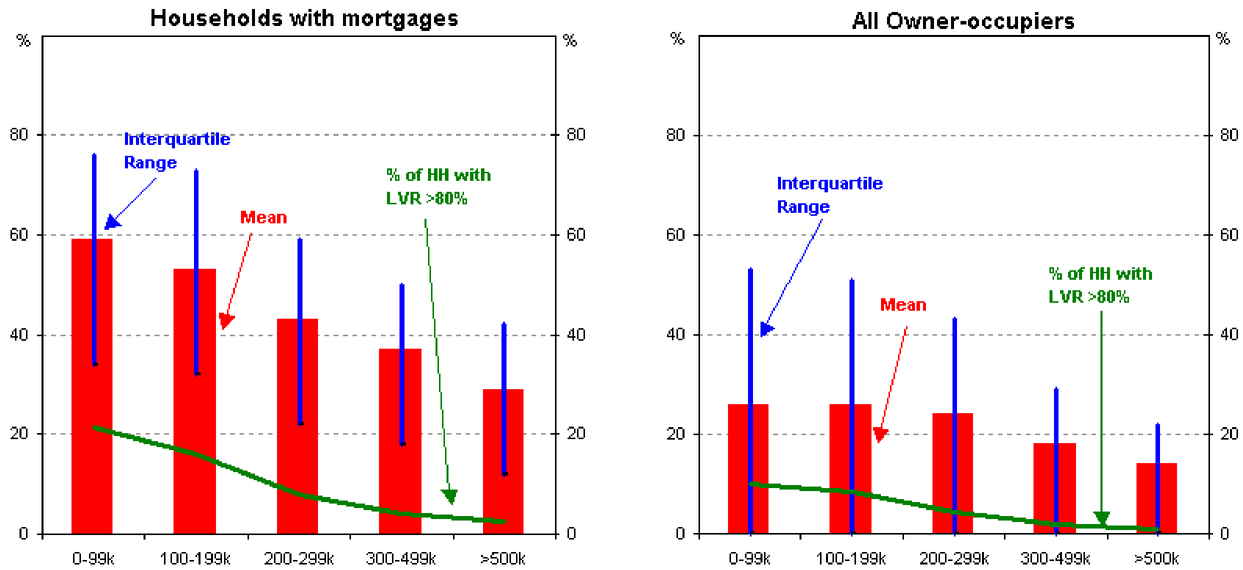
**Figure 3: Housing Leverage by Income Quintile**




### *3.2.3 Housing wealth*

Although the Wave 1 HILDA Survey did not ask households about their non-housing wealth, their estimates of the current market value of their homes should capture a large proportion of their assets. According to Figure 4 there is a strong negative relationship between households' housing leverage and the value of their properties. Households with properties worth less than \$100 000 had an average leverage (at almost 60 per cent) roughly double that of households with properties worth in excess of \$500 000. The low average incidence of leverage among households with expensive properties is reflected in the much lower likelihood (3 per cent) of their being leveraged above 80 per cent, at 3 per cent compared to over 20 per cent for households with homes worth less then $100 000. However, because owner-occupying households that have paid off their mortgages were more likely to have homes worth less than $100 000, their average leverage falls to about the same amount as households with homes worth between $100 000 and $199 000 when this is taken into account (panel 2).

**Figure 4: Housing Leverage by Value of Principal Home**




There are a couple of possible explanations for the negative relationship between housing wealth and housing leverage. First, if a household's housing wealth is positively correlated with its age and tenure length, then the observed relationship may only suggest that younger households, which are more likely to be highly leveraged, are also more likely to own cheaper homes. Second, if more expensive homes have appreciated more rapidly than less expensive homes, households that own more expensive homes will have experienced a larger passive reduction in their leverage, for a given length of tenure.

One of the implications of the negative relationship between leverage and housing wealth is that leverage should likewise be lower in locations with the highest house prices. It is these areas – New South Wales and Victoria, and the inner suburbs of state capitals – that have experienced the strongest rises in house prices recently. In general, the data support this argument. Average leverage is relatively low in New South Wales and Victoria, and particularly in their inner-city suburbs, as is the incidence of high leverage.[4] For example, average leverage in inner Melbourne was around 35 per cent (38 per cent in inner Sydney), compared to over 50 per cent in inner Brisbane and Adelaide. Indeed, only 1 per cent of households living in inner Melbourne had an LTV that exceeded 80 per cent. Provided the average length of tenure (i.e. the households' decision to trade up) is not greatly affected by the rate of house price inflation, this provides some comfort that those areas where concerns about future price falls might be greatest are those where the negative equity consequences are least. This strengthens our general conclusion that the moderate level of aggregate housing leverage in Australia does not seem to be masking important pockets of potential vulnerability.

# 4. Econometric Model And Results

## 4.1 Modelling household leverage

As discussed earlier, few households are able to purchase a house outright. Thus purchasing a home usually requires households to take out a mortgage, which will fund a substantial proportion of the purchase. This loan is then paid off *over time*, at a rate determined by the loan contract, their financial circumstances and whether they decide to refinance at some point. In this way, the amount of leverage held by a household with a mortgage will depend crucially on how long the household has had the loan and its initial size, as well as on the household's ability to make repayments. Leverage for household $i$, $Lv_i$, is thus modelled in a simple linear framework

$$Lv_i = X_i\beta + u_i \tag{1}$$

where $X_i$ is a vector of household characteristics that affect the degree of leverage held by the household, including financial characteristics and variables that proxy the initial loan size and the time since the mortgage began. The error term, $u_i$, is assumed to be normally distributed with a zero mean and a variance of $\sigma_u^2$.

4 Classifying postcodes into inner and outer city groups requires some subjective judgment. We defined suburbs as inner city if they were fairly close to the CBD and considered prestige or fashionable places to live; the specific allocation of postcodes is available from the authors.

### *4.1.1 Selection bias*

Estimating the determinants of household leverage on owner-occupied properties requires examining only those households that satisfy two criteria; they own a home and have a mortgage against it.[5] Households with mortgages select themselves into this group, with all other households' observed leverage being censored at zero. Gronau (1974) and Lewis (1974) showed that "self-selection" may result in samples having unrepresentative characteristics, which will bias estimates of population parameters obtained using standard econometric techniques. This bias is known as sample selection bias, and must be accounted for in the parameter estimation.[6]

There are good reasons to believe that our sample of leveraged households have unrepresentative characteristics; in particular, their age and wealth characteristics are likely to be quite different. Poorer and younger households are less likely to own their own home, since they are less able to satisfy the financial requirements involved. Older and wealthier households, while being more likely to own their home, have greater scope to have repaid their mortgage. The probability of both owning a home and having a mortgage will therefore be greatest for those in the middle age and wealth brackets.

Table 1 illustrates this point. The probability of having a mortgage is greatest for those aged 35–40, irrespective of household income. As income increases, so does the probability of having a mortgage, despite our expectation that wealthy households have more scope to own their homes outright. This pattern presumably reflects wealthier households' greater propensity to own rather than rent, as well as the imperfect relationship between wealth and income. It implies that the subsample of households with a mortgage is likely to have a higher median income than the general population, and an over-representation of middle-aged household heads. If these sample selection rules are correlated with the dependent variable, a specification error is introduced into our leverage equation, causing OLS estimates to be biased (Heckman 1976, 1979).

---

5 Households can also have leverage on other property, but this is not observed in the first wave of the HILDA survey. Wave 2 will contain more information on portfolio decisions.

6 Estimation with a censored sample is traditionally associated with tobit techniques. However these require that the relationships that determine whether or not a household has a mortgage are the same as those that determine the degree of leverage. This is clearly violated in our model, as there are potential factors such as property inheritance that would cause a household to have no mortgage, but are irrelevant in determining the size of mortgages for those households that have them.

**Table 1: Effect of Age and Income on Having a Mortgage**

Per cent of people in each category with a mortgage

| Income Quintile<br>Age Group | Q1 | Q2 | Q3 | Q4 | Q5 |
|---|---|---|---|---|---|
| 20–24 | 8.8 | 8.5 | 16.9 | 29.8 | 36.8 |
| 25–29 | 12.5 | 13.6 | 35.1 | 44.3 | 37.2 |
| 30–34 | 19.5 | 30.0 | 45.2 | 56.1 | 64.7 |
| 35–39 | 26.7 | 34.6 | 49.5 | 65.7 | 68.9 |
| 40–44 | 19.7 | 35.2 | 44.3 | 61.8 | 64.3 |
| 45–49 | 10.4 | 27.8 | 38.3 | 49.2 | 57.7 |
| 50–54 | 20.0 | 29.7 | 34.3 | 40.7 | 47.3 |
| 55+ | 3.4 | 4.5 | 12.2 | 24.3 | 23.3 |

#### *4.1.2 Sample selection rules*

Heckman (1976) formulated a two-step procedure that corrects the bias in the population parameters. This involves modelling the sample selection rules that determine inclusion into the sample. In our model, the two selection rules are whether a household rents or owns – the *tenure decision* – and whether or not the household has a mortgage on its home – the *mortgage decision*.[7] We observe leverage only when a household both owns and has a mortgage.

A household's tenure decision is modelled as a function of financial variables and household characteristics that affect a household's utility from owning. For each *i* household, $T_i^*$ is an unobservable index function that underlies the decision to own. The tenure decision is thus specified as

$$T_i^* = Z_{1i}\gamma_1 + e_{1i}, \;\; T_i = 1 \;\text{ if } T_i^* > 0, \; T_i = 0 \;\text{ if } T_i^* \le 0 \qquad (2)$$

where if $T_i^* > 0$, household *i* is an owner-occupier, and if $T_i^* \le 0$, household *i* rents. $Z_{1i}$ is a vector of financial variables, such as income and labour force status of the household head, and household characteristics.

7 The selection equations used here are not intended as accurate models of the tenure and mortgage decisions underlying a household's leverage position. Rather, they are reduced-form equations that attempt to account for differences in the characteristics of the subsample relative to those of the population. This is partly a result of data limitations; the survey does not contain any detailed wealth information, so more complex portfolio influences are largely ignored.

The mortgage decision is largely a financial one, and thus the decision equation will include variables affecting a household's ability to pay off their loan and the time they have had to pay it. The mortgage decision will thus contain very similar variables to the leverage equation, including financial variables, a proxy for the size of the initial loan and other household characteristics. Consistent with the tenure decision, $M_i$* is an unobservable index function that underlies the mortgage decision

$$M_i^* = Z_{2i}\gamma_2 + e_{2i}\,,\ M_i = 1 \ \text{ if } M_i^* > 0\,,\ M_i = 0 \ \text{ if } M_i^* \le 0 \qquad (3)$$

where $M_i$ = 1 if household $i$ has a mortgage, and $Z_{2i}$ is a vector of variables affecting a household's willingness and ability to pay off a loan. These include income, labour force status of the household head, current value of the property, and attitude to risk.

## 4.2 Two-Step estimation procedure

### *4.2.1 Correlation between the selection equations*

The mean equations of our reduced-form model thus take the form

$$T_i^* = Z_{1i}\gamma_1 + e_{1i} \qquad (4)$$

$$M_i^* = Z_{2i}\gamma_2 + e_{2i} \qquad (5)$$

$$Lv_i = X_i\beta + u_i \qquad (6)$$

There are good reasons to assume that the tenure and mortgage decisions (i.e. $e_{1i}$ and $e_{2i}$) are correlated. Homes are a large, lumpy expenditure that few households can fund without at least some initial resort to debt funding. The importance of finance means that the ability to obtain a mortgage not only affects the mortgage decision, but also affects the tenure decision, since at least some households' decision to rent is driven by their inability to obtain a mortgage. A household's ability to obtain a mortgage is itself dependent on the willingness of financial institutions or other lenders to provide the funds. Some of the determinants of the lender's decision – savings history, past loan defaults and so on – are not recorded in HILDA and so their influence ends up in the residual. This presence of these influences in the residuals of both decision equations may result in error correlation.

Similarly, some households come to own homes through other means such as inheritance, which do not require saving and taking out mortgages. The determinants of inheritance are not readily modelled, at least using explanators available in the HILDA data set. These households would therefore generate a positive residual in the ownership selection equation and a negative one in the mortgage selection equation, creating another source of correlation between the two selection equations, opposite to the source mentioned above.

Other factors also point to interdependency between the two decisions, particularly the dual consumption and investment properties of housing. When a household's housing investment demand is sufficiently close to its consumption demand, tax advantages and positive externalities to owner-occupation increase the probability that the owner of a single property will occupy it.[8] On the other hand, many of the unobserved factors explaining past mortgage repayments, and thus whether a mortgage has now been paid off, will be unrelated to the past tenure decision.

### *4.2.2 Sequential versus joint decisions*

The error structure of the model depends crucially on whether the two "decisions" are made sequentially or jointly. The distinction between sequential and joint decisions relates to the interdependency of the two decisions, rather than their timing, although these may be related. We use a decision matrix to illustrate this difference (Figure 5).

**Figure 5: Selection Rule Matrix**

| | | Selection equation 2 | |
|---|---|---|---|
| | | Yes | No |
| Selection equation 1 | Yes | 1 | 2 |
| | No | 3 | 4 |

8 These largely derive from a landlord's inability to fully extract the cost of occupation from a tenant (Henderson and Ioannides 1983).

If the two decisions are defined over the entire set of observations, so that all four cells are logically possible, then they are made jointly, regardless of whether the outcomes represented by the cells are actually observed. Suppose Centrelink are interested in the determinants of a person's earnings after a period of unemployment. Earnings are only observed if the person obtains work, so the two underlying selection rules are whether a person registered for benefits when unemployed and whether the person became employed at a later date. These decisions are made jointly since the rules apply to all unemployed people; any unemployed person can both register for benefits and subsequently obtain work. Importantly, a person can obtain work regardless of whether they had previously registered for benefits, meaning that if Centrelink only observes those who have registered, it misses the wages for unregistered people. Failure to observe this "decision" to register for unemployment benefits is known as partial observability in the joint decision model (Lee and Maddala 1985; Tunali 1986).

Sequential decisions, on the other hand, are characterised by one decision only being defined given a particular *outcome* of the other decision (Maddala 1986). This is the case with our sample selection rules; a household cannot "decide" to have a mortgage on an owner-occupied property if they are not owner-occupiers. Here, one cell is logically impossible (cell 3), and we have a three-cell model.

### *4.2.3 Estimating the reduced-form model*

The sequential nature of our selection rules and the correlation between them implies that we have a reduced-form model comprised of (7), (8) and (9).

$$T_i^* = Z_{1i}\gamma_1 + e_{1i} \tag{7}$$

$$M_i^* = Z_{2i}\gamma_2 + e_{2i} \tag{8}$$

$$Lv_i = X_i\beta + u_i \tag{9}$$

In this system, $e_{1i}$ is distributed as a standard normal, $e_{2i}$ is also standard normal, but defined only on the subpopulation $T_i^*>0$ and $u_i$ is normally distributed on the subpopulation $T_i^*>0$, $M_i^*>0$ with variance $\sigma_u^2$ (Lee and Maddala 1985). The three errors may be correlated; the technique used to produce the estimates in the next section allows for this.

The selection bias in the leverage equation is corrected using two Heckman corrections. First, the mortgage equation is estimated using maximum-likelihood probit techniques, adjusted for the selection bias resulting from the tenure decision. Consistent estimates of $\beta$ are then obtained using a Heckman-Lee type correction. An adjusted leverage equation (11) is estimated by OLS

$$Lv_i = X_i\beta + \alpha_1\lambda_1 + \alpha_2\lambda_2 + v_i \tag{10}$$

where $\alpha_1\lambda_1 + \alpha_2\lambda_2 = E\left(u_i \middle| T_i^* > 0, M_i^* > 0\right)$. These variables are approximated using the estimated parameters in the selection equations; because we have a double-selection model, these $\lambda$s are not identical to the inverse Mills ratio used in standard Heckman-Lee single-selection models.

One important implication from the use of the Heckman-Lee two-step correction is that the OLS estimate of the error variance, $\sigma_u^2$, is biased, as are the estimates of the covariances between equations and the parameter variances (Heckman 1979; Tunali 1986). These are corrected using approximations to the true distribution using the delta method (Tunali 1986), although this only affected assessments of statistical significance at conventional levels in one or two cases.

Despite the obvious benefits of the Heckman-Lee correction technique, the parameter estimates are quite sensitive to violations of the underlying normality assumptions. Greater generality has been obtained through the use non-parametric estimation techniques, however this is beyond the scope of our paper.[9] Our preferred specifications, presented in the next section and Appendix B, preserve normality, but possibly at the cost of some efficiency.

### 4.3 Estimation results

Table 2 shows results from the estimation of the Heckman-Lee selection model for household housing leverage, as well as the results from the model when selection bias is not taken into account. The coefficients can be interpreted as respresenting percentage points of leverage. The results for the tenure and mortgage equations appear in Appendix B. Overall the results are reasonable with our model explaining about 30 per cent of the variation in households' housing leverage, and an *F*-test of the null hypothesis that all the coefficients are equal to zero rejected at all conventional levels of significance. The selection equations themselves also provide a reasonable fit of the decision processes, despite necessarily excluding important portfolio variables.

9 For examples and test of these techniques see Heckman et al. (1998).

For ease of interpretation, we split variables into three categories of household characteristics – the stage of the household in the life cycle, household means, and characteristics relating to finance and housing – and a group of location dummies to control for region-specific effects and the effect of unexpected changes in house prices on housing leverage. This classification also ensures that we retain a focus on economically important variables. The HILDA survey contains a large number of potentially relevant variables, many of which are highly correlated with each other. In addition to the multicollinearity problem this presents for single-equation estimation, we also need to ensure sufficient identification of each of the three equations in our selection model.

It turns out that identification is the weakest aspect of our model. The factors influencing tenure and mortgage overlap significantly. Thus, despite our best efforts to use different regressors in the two selection equations, we only account for a portion of the selection bias in the mortgage equation resulting from the tenure selection. As a result, the fitted values for leverage implied by the bias-corrected equation seem unreasonably high for households that do not in fact have leverage, even though the estimated parameters ostensibly provide information about the whole population, not just the subpopulation with leverage. This leads us to suspect that the bias-corrected results presented in Table 2 fall some way short of the true population estimates. They are nonetheless sufficiently different from the coefficients in the unadjusted conditional leverage equation in some key dimensions, that we can regard them as an improvement on the conditional OLS results when assessing population behaviour.

**Table 2: Household Housing Leverage Regression Results**

| | Bias Corrected | Not Bias Corrected | | Bias Corrected | Not Bias Corrected |
|---|---|---|---|---|---|
| **Personal Characteristics of Household Head and Demographic Variables** | | | | | |
| Age | –0.8*** | –0.3** | Age squared | –0.002 | 0.0005 |
| Retired | –16.1 | –0.6 | Married | 4.3* | 2.2 |
| De facto | 3.9 | 3.6 | Divorced | –2.3 | –2.3 |
| Separated | 13.2*** | 12.4** | Recently separated | –5.1 | –5.1 |
| Student | 10.5 | 10.3 | Intends to retire early | –2.6 | –3.1 |
| Recent migrant | 8.6 | 7.2 | University educated | –2.1 | –1.1 |
| Born in UK | –4.2 | –4.9 | Born in Europe | 0.9 | 3.6 |
| Father born in UK | 6.0** | 4.2* | Father born in Europe | –6.5** | –6.5** |
| **Household Means Variables** | | | | | |
| Household income | 1.0** | 0.3 | Income squared | –0.2 | –0.1 |
| Socio-economic status | –0.01* | –0.01* | Satisfaction with pay | 0.4 | 0.4 |
| Casual job | –12.0*** | –7.8*** | Fixed-term contract | –3.8 | –3.7 |
| Likelihood of losing job in next 12 months | –0.04 | –0.04 | Had difficulty paying mortgage on time | 5.9** | 6.0** |
| Income adequacy | –2.0** | 0.1 | | | |
| **Financial and Housing Variables** | | | | | |
| Time lived at address | –1.0*** | –0.5*** | Planned pay off date | 1.2*** | 1.3*** |
| Ahead of mortgage repayment schedule | –2.1* | –2.0 | Satisfaction with neighbourhood | –0.9*** | –0.9*** |
| Long saving horizon | 3.3** | 3.4** | Short saving horizon | 2.0 | 0.2 |
| Attitude to borrowing | 0.8 | 0.4 | Attitude to risk | –0.5 | –1.5** |
| Took out institutional loan to buy house | 11.9*** | 11.7*** | Recently moved to smaller home | –10.3* | –10.7* |
| Moves in last 10 years | 2.1** | 1.2** | No recent moves | –6.4*** | –7.6*** |
| Condition of home | 3.3*** | 3.3*** | Not a first-home buyer | 5.4*** | 5.7*** |
| Semi-detached dwelling | 6.3** | 6.0** | Pays credit card on time | –1.6*** | –0.6 |
| **Location Variables** | | | | | |
| Inner Sydney | 0.3 | 1.0 | Outer Sydney | 9.4*** | 9.6*** |
| Non-metro NSW | 7.3** | 9.7*** | Outer Melbourne | 9.34*** | 9.7*** |
| Non-metro Victoria | 13.3*** | 12.7*** | Inner Brisbane | 15.8*** | 20.2*** |
| Outer Brisbane | 18.6*** | 18.2*** | Non-metro Queensland | 17.6*** | 17.2*** |
| Inner Adelaide | 15.3** | 15.6** | Outer Adelaide | 15.1*** | 14.8*** |
| Non-metro SA | 21.6*** | 23.8*** | Inner Perth | 9.2 | 4.6 |
| Outer Perth | 17.2*** | 14.7*** | Non-metro WA | 13.4** | 13.6*** |
| Inner Hobart | 2.3 | 1.2 | Outer-city Tasmania | 17.8*** | 17.7*** |
| Non-metro Tasmania | 37.0*** | 36.6*** | Canberra | 18.2*** | 19.2*** |
| Northern Territory | 23.0*** | 28.2*** | House price growth | –0.1 | –0.1 |
| Tenure selection term | –0.7 | | Mortgage selection term | 31.3*** | |
| Adjusted $R^2$ | 0.281 | 0.271 | | | |

Note: ***, **, and * represent significance at 1, 5 and 10 per cent levels.

### *4.3.1 Stage of the household in the life cycle*

A household's stage in the life cycle should be a key explanator of a household's housing leverage. Taking the results in Table 1 at face value, one might expect housing leverage to rise as households purchase homes and acquire debt during their family formation years, and then fall as they repay this debt. Our results are, however, mainly suggestive of leverage falling as households move through the life cycle. Even when we adjust for selection effects, and thus that younger households are less likely to own their homes, leverage is associated negatively with age, with each extra year associated with up to 1 percentage point less leverage, while the age-squared term is negative but insignificant. This contrasts with the positive coefficient on age and negative coefficient on the square of age that would be required to obtain a hump-shaped profile for this relationship.

A number of other variables are suggestive of leverage falling as households move through the life cycle. For a given term for the mortgage, a move-in date that is one year earlier than an otherwise identical household implies that the household's housing leverage will be about 2.2 percentage points lower, since both the move-in date and the expected payoff date are then one year earlier. This suggests an important role for passive paydown of mortgages on schedule; households that are older and have lived in their homes for longer have lower leverage simply because they have had longer to pay their mortgage off. In addition, rising house prices imply that mortgages taken out earlier were likely to be smaller, and thus a smaller ratio to the home's current estimated price. Indeed, the size and precision of the estimated coefficients on these variables are noticeably greater than those for age. Thus it could be concluded that older households generally have lower leverage because they have had longer to pay down their loans, rather than reflecting some pure age effect.

On the other hand, some of the significant explanators of leverage do suggest a more explicit lifecycle interpretation. Married household heads are associated with leverage that is 5 percentage points higher than households with heads that never married. This is suggestive of the importance of family formation in the decision to become an owner-occupier, and thus to take on a mortgage loan. In the conditional equation, though, we see no significant effect of marriage on leverage given that the household does have a mortgage.

### *4.3.2 The means of the household*

Our expectation that leverage should rise with household means is generally supported by the data. For example, our main indicator of household means, household income, is significantly positively related to leverage. However, the magnitude of the effect is small, with each extra ten thousand dollars of income associated with only 1 percentage point higher leverage. Increased income appears to be more closely associated with higher values of both debt and housing assets, than with the ratio of those two variables. We also find evidence that the effect of household income on leverage is non-linear, with the negative sign on the income-squared variable implying that leverage is increasing in income but at a decreasing rate. However, the coefficient on this term is so small that leverage does not begin falling in income until household income reaches about $500 000.

We also included two measures of households' subjective views about their incomes – the ease with which they are making ends meet (income adequacy), and their satisfaction with their pay – in our leverage model to determine whether such measures of income relative to perceived requirements add information above that of measured income. Neither variable is closely correlated with reported actual income. Of the two variables we find that only income adequacy is statistically significant, with households having difficulty making ends meet having higher leverage than households easily making ends meet. This is probably because financially stretched households are more likely to pay their housing debt off more slowly than other households, and because high leverage might be associated with high absolute levels of debt that the household has difficulty servicing. This explanation is supported by the result that households that have had difficulty paying their mortgage on time, a measure of acute financial distress, also have higher leverage than other households.

The means of the household captures more than just its income. Factors such as family breakdown, permanency of employment, negative income shocks and familial support also influence leverage by affecting households' ability to take on and pay off debt. Our results suggest that such factors are important in explaining household housing leverage. For example, households with separated household heads have considerably higher leverage (14 percentage points) than other households. However, the fact that household heads that are divorced do not have higher leverage than other households suggests the impact of family breakdown on leverage may be temporary; this possibility could be confirmed using the longitudinal aspects of the HILDA survey to track households through the process of breakdown.

Households with heads employed casually, and those that more generally regard their current employment as precarious have lower leverage than other households. A household with a casually employed head has about 12 per cent less leverage than a household with a permanently employed head. This is partly because households with more precarious employment, and hence incomes, find it more difficult to save for a deposit and obtain housing finance from intermediaries, and are therefore less likely to own their own homes. However, the biased conditional equation shows that even when these households have a mortgage, it represents a lower leverage ratio on the value of their homes than for other households with mortgages.

#### *4.3.3 Household financial attitudes*

Besides the structural characteristics of households such as age and income, we also expect household attitudes to housing and debt to influence their housing leverage. For example, we may have expected that households that are more comfortable about taking on debt would have higher leverage than other households. However, a variable that summarised households' attitude to borrowing for items such as holidays, cars, and clothes, did not enter significantly into the model, although the point estimate did have the expected sign. We may also have expected households' attitudes to risk would influence their housing leverage. For example, risk-averse households may prefer to pay housing debt off more rapidly. Although the coefficient on this variable is again of the expected sign, its insignificance may suggest that risk-averse households' aversion to debt might be offset by a preference for housing assets over other assets they perceive as riskier; confirming this suspicion will not be possible until Wave 2 data on other kinds of asset and debt are available.

Other attitudinal variables do, however, enter the regression significantly. For example, households that pay their credit card off on time each month have lower leverage than households that do not. This could reflect either such households' preference for paying as little interest as possible on their debt, or perhaps greater financial sophistication and means. Households with long savings horizons have slightly higher leverage than households with shorter savings horizons; this may be picking up life-cycle or other characteristics that are not captured by our variables. Households who have owned more than one home have slightly higher than average leverage, which suggests that households take on more debt when trading up, relative to the asset's value.

*4.3.4 Other household characteristics*

People's cultural background may affect their leverage by influencing attitudes to debt, home ownership and intergenerational transfers. To test this, we included variables representing both the household heads' own country of birth, and their parents' country of birth. Migrants will report both their own and their parents' birthplaces as being outside Australia, while second-generation Australians will report parental birthplaces outside Australia and their own birthplace as Australia.

If the decision to migrate is at least partly motivated by a desire to ensure better economic prospects for one's children, migrant parents might have greater propensity to make gifts or bequests to their children than other parents. We might therefore expect second-generation Australians to have lower leverage than both their parents and other households. This would be captured as a negative coefficient on parental background, with a potentially offsetting positive coefficient for own birthplace. Cultural attitudes would also be best captured by the parental background variable, while the own-birthplace variable would best capture factors specific to the experience of migrants. Including both these variables allows us to distinguish between these competing explanations of household differences in leverage by origin.

We found that only the parental background variables were significant. Households with heads whose fathers were born in continental Europe had leverage 7 percentage points lower than households where the head's father was born in Australia, while households with heads whose fathers were born in the UK had higher leverage than other households.[10] Taken at face value, these results could be interpreted as indicating that a combination of cultural values and intergenerational transfers explains the pattern of slightly lower leverage for European migrants, and more noticeably lower than average leverage for their children. The results for migrants from the UK and their children are more difficult to interpret, but might be suggesting a minor familial tendency to be relatively more comfortable taking on debt, with the difference between migrants and their children perhaps indicating intergenerational transfers from child to parent. Recent migrants from any part of the world were estimated to have higher leverage than the rest of the population, perhaps implying that the cost of establishing themselves in a new country impinges on their ability to fund home purchase from their own resources, but this effect was not statistically significant.

---

[10] Using birthplace of mother or of either parent gave virtually identical results. The number of households with heads that were born in, or children of migrants from, other regions was too small to produce statistically significant estimates, and so these variables were excluded from our preferred specification.

Finally, we also observe that households that took out loans from financial institutions at the time of purchase, and households whose homes are in poor condition have higher leverage than other households. Neither of the estimated coefficients on these variables seem particularly affected by selection bias, implying that these factors impinge on the composition of mortgage funding, but not the decision to own or carry some debt against the home.

Households that did not require a loan from a bank or other financial intermediary presumably acquired the property through inheritance, or used their own resources and bequests, gifts or loans from friends and family to fund their purchase. Although loans from friends and family are included in the measure of leverage used here, the other means of funding the purchase are clearly substitutes for debt that would reduce initial leverage at the time of purchase, and thus remaining leverage given the time elapsed since then. The higher leverage of households with homes in worse condition than average may be an example of reverse causation; households with high leverage may be too financially stretched to pay for renovations, and might be unable to borrow more to do so.

#### *4.3.5 Location variables*

The final set of variables we considered were a group of location dummies split into three categories for each state – the inner suburbs of the capital city, the outer suburbs of the capital city, and the non-metropolitan regions of the state. These variables are in the model to proxy for different rates of house price growth across these regions over the years, which can be expected to have exogenously influenced households' housing leverage. Indeed, using inner Melbourne as our base category because it is the region that has experienced the most rapid price growth in recent years, we can see that all other regions have higher leverage than inner Melbourne, and for all but two – inner Sydney and inner Perth, it is significantly higher. In general, leverage is higher in non-metropolitan regions than metropolitan regions, and leverage is highest in Tasmania and Queensland. Brisbane's high average leverage is a puzzle, given that its price growth has been rapid in recent years, but this may be partly a base effect.

We found that these regional dummies performed better as proxies for general regional variation in housing price growth than using actual price growth over the past two years, even though the house price data was disaggregated by more narrowly defined regions than the three-way split by state in the dummy variables. Households move at different times and have thus experienced different degrees of inflation of the value of their home since they purchased it. The regional dummies appear to be capturing this variation in holding period in ways that price growth over a fixed period cannot; the estimated coefficient should be interpreted as indicating the relative price inflation over the average holding period in that region. If housing price data had been available at a sufficiently disaggregated level for the whole of Australia and had a long enough history to cover the range of holding periods implied by households' move-in dates, it may have been possible to dispense with the regional dummy variables, but this is not the case with currently available data.

#### *4.3.6 Effect of selection bias*

As expected, the mortgage and tenure adjustment terms are jointly significant in the leverage equation. This is a direct test of the existence of selection bias, and validates our use of this estimation approach. The adjustment term for the tenure decision is, however, insignificant on its own; all the bias seems to be controlled for by the mortgage equation. This result is a puzzle given that the results suggest that we are not capturing the selection from tenure into mortgage very well. We would expect this to come independently through the tenure selection term; its insignificance may thus indicate that the correction of the bias resulting from the tenure selection is imperfect in both the mortgage and leverage equations. Despite this, our results suggest that this double-selection model is superior to a standard single-selection model with Heckman correction. An alternative specification with only one selection equation, distinguishing mortgage-holders from other households, produces different results that appear to fit the data less well than our preferred specification (these results are available from the authors).

Table 2 also highlights that controlling for selection bias is important, particularly for the income variables, and produces superior results to the OLS results from a conditional model. Income significantly affects leverage when we control for selection bias but it has a much smaller and insignificant coefficient in the uncorrected conditional specification. This lines up with the differences in the relationship between income and leverage identified in Figure 3 when the sub-population of households with mortgages is considered, rather than the population as a whole. Conditional on both owning a home and having a mortgage, income is not strongly correlated with housing leverage. Not correcting for this conditionality would lead us to conclude erroneously that low-income households tend to have similar leverage – and thus be as sensitive to moves in interest rates and vulnerable to falls in housing prices – as households with higher income. Correcting for the selection bias leads us to the more accurate conclusion that higher-income households have higher leverage than the population as a whole. This is because these households are more likely to have a mortgage, given that they own their homes.

There are a number of other variables that are associated with greater likelihood of owning a home and having a mortgage, but do not have any relationship with the level of leverage conditional on the household having a mortgage. These variables have significant estimated coefficients in the bias-corrected leverage equation, but not in the uncorrected, conditional equation; their effect occurs entirely through the selection equations. Such variables include having a university education, always paying off one's credit card, feeling that one's income is adequate, being married and having a father who was born in the UK.

Where a variable affects the leverage and selection equations differently, however, the difference between the selection bias corrected and the uncorrected specifications is ambiguous. A household's attitude to risk is one variable that is demonstrably affected by conflicting signs, being insignificant in the corrected specification but significant when we do not correct for the selection bias. The more risk-averse is a household head, the lower is the household's leverage through both the (unadjusted) leverage equation and the tenure equation, since the household is less likely to own. However, having a more risk-averse household head increases the likelihood of a household having a mortgage once they own, offsetting the lower leverage from the other equations when we correct for selection bias.

## 5. Conclusions

Despite our reservations about the adequacy of our results in correcting for selection bias, they clearly show that such adjustments are necessary when examining leverage across the whole household sector. Our double-selection model produces different estimates – and a slightly better fit – than both the conditional OLS equation shown in Table 2 and a single-selection model that treats renters and homeowners without mortgages the same. The unadjusted conditional equation is nonetheless useful. Although parameters estimated only over the subpopulation with mortgages are biased estimates of population behaviour, comparison of the conditional and corrected equations provides a guide to interpreting the source of a particular significant relationship.

Our preferred specification suggests several descriptive conclusions about the pattern of housing leverage in Australia. Overall, the households that are most highly leveraged are those most able to bear the debt – mid-life households with high income – and those least vulnerable to falls in housing prices, living in outer suburbs and non-metropolitan regions that have experienced relatively smaller price gains in recent years. Although there may be pockets of high leverage in any subgroup of the household sector, the general picture accords with aggregate data in suggesting that leverage on the housing stock remains fairly moderate. Young homeowners are likely to have particularly high leverage, but young households in general are less likely to be homeowners. Although there do appear to be differences in leverage between households of different ethnic origins, the reasons for this appear to be a complex mix of attitudes transmitted across generations, and experiences and intergenerational transfers associated with the migration decision.

Age, lifecycle stage and time at address variables all point to a significant role for the passive paydown of debt as scheduled in the household's mortgage contract, in determining current leverage. Similarly, the significance of the locational and regional housing price variables indicates that the inverse relationship between the housing leverage ratio and its denominator might not result in appreciable reactions from households adjusting their debt. These determinants of leverage can all be characterised as largely being beyond the control of individual households, suggesting that at least in the short run, households do not necessarily adjust their balance sheet to maintain a desired leverage ratio.

Against this, however, we find evidence of at least some households making explicit decisions about the end date of their mortgage, which might not be the date specified in the loan contract. If some households did not make conscious decisions about the desired date on which their loan will be fully paid off, move-in date and payoff date would not both be significant in our estimated leverage equation. Amongst these households, at least, we might expect deliberate portfolio reactions to developments in interest rates and housing prices, rather than simply adhering to a predetermined path for their remaining outstanding debt. Observing such reactions would, however, require tracking households through time. The longitudinal nature of the HILDA survey will make it uniquely suited amongst all data sets for Australia, to examination of these household responses. Future waves of the HILDA data set will therefore be essential for further work on understanding households' decisions about their balance sheets, leverage and indebtedness, and in particular their responses to interest rate changes and housing price movements through these channels.

# Appendix A: Income Imputation and Results

As discussed in Section 2.2, the nature of the missing data leaves us with the need to impute income for three separate types of missing cases. For Type 1 individuals we impute total gross financial year income. For Type 2 individuals we impute gross financial year wage and salary income and add this imputed income to their reported gross financial year non-wage and salary income. For Type 3 individuals we impute gross financial year non-wage and salary income and add this imputed income to their reported gross financial year wage and salary income. Table A1 contains all the relevant results.

In all cases missing values are imputed using the predictive mean matching (PMM) method outlined in Little (1988). In the first stage this involves estimating a regression on the variable to be imputed for individuals without missing values – in our case income. Next the model with the highest $R^2$ is used to predict the income of individuals with missing values. For every missing value we find the record with the nearest predicted value. The actual value of this 'donor' is then imputed for the missing value. The advantages of using the PMM method over other single imputation methods, such as simply imputing the conditional mean obtained from a regression, are that it ensures that only feasible values of the variable are imputed, and that a random error component is introduced so that imputed values have a similar variance to the reported values (ISER 2002).

However, because the PMM method (like all single imputation methods) treats imputed values identically to observed values, it fails to represent all the uncertainty about the imputed values (Dillman et al 2002). This is generally overcome in the literature by using a multiple imputation method whereby multiple values for the variable of interest are imputed from a set of M datasets. The results are then combined by taking the average of the estimates from the M datasets, the average of the variances from the M datasets, and a correction for imputation uncertainty (Dilman et al 2002). Despite its limitations, we use the PMM method in our paper because of its ease of implementation and because the implications of multiply imputed data for the error distribution have not been fully explored in the context of a double selection model like the one used in our paper.

**Table A1: Income Imputation Results**

| Total Income ('000) | | Wage income ('000) | | Non-wage income ('000) | |
|---|---|---|---|---|---|
| Age | 1.6*** | Age | 0.8*** | Age | 0.08*** |
| Age squared | -0.15*** | Age squared | -0.01*** | | |
| Victoria | -0.3 | Victoria | -1.2** | Victoria | 0.6 |
| Queensland | -1.4* | Queensland | 0.0 | Queensland | 0.0 |
| South Australia | -2.4** | South Australia | -1.2* | South Australia | -0.8 |
| Western Australia | -0.9 | Western Australia | -2.4*** | Western Australia | 1.3** |
| ACT | 6.3*** | ACT | 4.8*** | ACT | -1.5 |
| Make ends meet | 3.9*** | Make ends meet | 2.1*** | Make ends meet | 0.9*** |
| Socio-economic | 1.0*** | Socio-economic | 0.1*** | Socio-economic | |
| Has disability | 2.3*** | Business income | -16.0*** | Business income | 20.9*** |
| Lone person | 7.9*** | Gov. benefit | -5.1*** | Gov. benefit | 2.6*** |
| Group household | 6.7*** | Receives interest | 1.2*** | Receives interest | 4.1*** |
| Sole parent, depend. children | 6.8*** | Receives rent | 4.4*** | Receives rent | 4.5*** |
| | | Receives dividends | 2.5*** | Receives dividends | 1.0** |
| Sole parent, no. depend. children | 4.4** | Non-metropolitan | -1.9*** | Age pensioner | -2.3*** |
| | | Inner-city | 1.5*** | Receives royalties | 3.9 |
| Persons in h'hold | -1.4*** | Union member | 5.9*** | Union member | -1.9*** |
| Employed | 13.0*** | Employed | 11.6*** | Employed | -4.2*** |
| Retired | -7.6*** | Retired | -10.6*** | Retired | 4.3*** |
| Home duties | -6.0*** | Spouse's income | 0.0*** | Spouse's income | 0.1*** |
| Multifamily home | -3.7* | Multifamily home | -3.6** | Student | -1.7 |
| | | Household head | 9.9*** | Household head | 6.2*** |
| No. bedrooms | 0.8*** | Has disability | 1.5*** | Health | 0.2 |
| Home's condition | -1.1*** | Home's condition | -0.6*** | Home's condition | -0.3 |
| Home value | 0.03*** | Home value | 0.004*** | Home value | 0.004*** |
| No. of children | 1.4*** | | | No. of children | 0.3* |
| Married | 8.1*** | Education level 2 | -10.9*** | Never married | 3.9*** |
| Separated | 4.7*** | Education level 3 | -11.1*** | Separated | 3.9*** |
| De facto | 10.0*** | Education level 4 | -11.6*** | De facto | 2.3*** |
| Divorced | 4.9*** | Education level 5 | -10.7*** | Divorced | 4.8*** |
| Widowed | 8.5*** | Widowed | 4.4* | Widowed | 1.8* |
| Adjusted $R^2$ | 0.32 | Adjusted $R^2$ | 0.46 | Adjusted $R^2$ | 0.204 |
| RMSE | \$26 000 | RMSE | \$19 000 | RMSE | \$20 500 |

Note: ***, ** and * represent significants at 1, 5 and 10 per cent levels

# Appendix B: The Tenure and Mortgage Equations

**Table B1: Household Tenure and Mortgage Decision Regression Results**

| | Mortgage | Tenure | | Mortgage | Tenure |
|---|---|---|---|---|---|
| **Personal Characteristics of the Household Head and Demographic Variables** | | | | | |
| Age | -3.7*** | 1.1** | Married | | 41.6*** |
| Retired | -47.5*** | | Receives age pension | -25.1** | -16.2* |
| De facto | | 23.9*** | Widowed | | 21.8** |
| Receives youth allowance | -124.0*** | | Home duties | -9.02 | -29.0*** |
| University x age | -0.01 | 0.9** | Speaks English at home | 23.4*** | |
| Born in UK | | -69.9*** | Born in Europe | | -88.0*** |
| Born in NZ | | -97.0*** | Born in Asia | | -135.0*** |
| Father born in UK | 16.3** | | Father born in Europe | | 23.4** |
| Mother born in Asia | | 53.2* | No. children at home | 6.7*** | |
| Number of children | | -6.3*** | Intended no. children | | -40.6*** |
| Int. children x age | | 1.2*** | Long term migrant | | 76.8*** |
| Indigenous | | 35.3** | Non-resident children | | 16.9 |
| **Household Means Variables** | | | | | |
| Income | 0.4*** | 0.5* | Income squared | -0.8*** | -27.4 |
| Socio-economic status of postcode | | 8.1*** | Satisfaction with pay | | -2.3** |
| Casual job | -34.2*** | | Fixed term contract | | -25.9*** |
| Received financial help from friends | | -14.7** | Had difficulty paying mortgage on time | | -29.1*** |
| Income adequacy | -12.3*** | | Satisfaction with finances | | 3.99*** |
| **Financial, Housing and Location Variables** | | | | | |
| Time at address | -1.8*** | 2.9*** | Moved to smaller home | | 43.5*** |
| Attitude to risk | 8.4*** | -5.3** | Short saving time horizon | 9.9* | -8.6 |
| Condition of home | | -20.3*** | Importance of home | | 8.6*** |
| Moves in last 10 years | 11.1*** | -13.6*** | No recent moves | | 47.3*** |
| Small apartment block | | -64.4*** | Tall apartment block | | -97.1*** |
| Semi-detached dwelling | | -66.9*** | Pays off credit card on time | -7.5*** | 4.6*** |
| Has credit card | -47.1*** | -24.2*** | No. of bedrooms | -1.1 | 27.1*** |
| Outer Sydney | | 19.6** | Non-metro NSW | -17.0*** | 42.7*** |
| Outer Melbourne | | 24.6*** | Non-metro Victoria | | 20.3** |
| Inner Brisbane | -33.3** | | Non-metro Queensland | | 18.4*** |
| Non-metro SA | -20.3 | 23.8* | Inner Perth | 48.2* | 54.2*** |
| Outer Perth | 21.9** | 16.2* | Non-metro WA | | 31.3** |
| Non-metro Tasmania | | 43.9*** | House price growth | | 0.7*** |
| Log-likelihood for system | -4462.42 | | | | |

Note: ***, **, and * represent significance at 1, 5 and 10 per cent levels.

## References

**Bernanke B and M Gertler (1995), '**Inside the black box: The credit channel of monetary policy transmission', NBER Working Paper No 5146.

**Caroll CD and WE Dunn (1997),** 'Unemployment expectations, Jumping (S,s) triggers, and household balance sheets', NBER Working Paper No 6081.

**Dillman D, J Eltinge, R Groves and R Little (2002), '**Survey nonresponse in design, data collection, and analysis', Chapter 1 in R Groves, D Dillman, J Eltinge, and R Little (eds), *Survey Nonresponse*, John Wiley & Sons, New York.

**Genesove D and CJ Mayer (1997),** 'Equity and time to sale', *American Economic Review*, 87(3), pp 255–269.

**Goodman J and J Ittner (1993),** 'The accuracy of home owners' estimates of house value', Board of Governors of the Federal Reserve System Working Paper No 131.

**Gronau R (1974),** 'Wage comparisons – a selectivity bias', *Journal of Political Economy*, 82(6), pp 1119–1143.

**Heckman JJ (1976),** 'The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models', *Annals of Economic and Social Measurement*, 5(4), pp 475–492.

**Heckman JJ (1979),** 'Sample selection as a specification error', *Econometrica*, 47(1), pp 153–161.

**Heckman JJ, I Ichimura, J Smith and P Todd (1998),** 'Characterizing Selection Bias using Experimantal Data', NBER Working Paper Series No. 6699.

**Henderson JV and YM Ioannides (1983),** 'A model of housing tenure choice', *American Economic Review*, 73(1), pp 98–113.

**Henley A (1999),** 'The economics of the crazy British housing market', Aberystwyth University Economic Research Paper No 99/8.

**Institute for Social and Economic Research (2002),** 'Weighting imputation and sampling errors', Chapter 5 in British Household Panel Survey (BHPS) User Documentation.

**Kent C and P Lowe (1997),** 'Asset-price bubbles and monetary policy', Reserve Bank of Australia Research Discussion Paper 9709.

**Lee LF and GS Maddala (1985),** 'Sequential selection rules and selectivity in discrete choice econometric models', in GS Maddala (ed), *Econometric Methods and Applications Volume II*, Edward Elgar Publishing Limited, Aldershot, pp 311–329.

**Lewis HG (1974),** 'Comments on selectivity biases in wage comparisons', *Journal of Political Economy*, 82(6), pp 1145–1155.

**Little R (1988),** 'Missing-data adjustments in large surveys', *Journal of Business and Economic Statistics*, 6(3), pp 287–301.

**Maddala GS (1986),** *Limited-Dependent and Qualitative Variables in Econometrics*, (Econometric Society Monographs No. 3), Cambridge University Press, Sydney.

**McLennan D, J Muellbauer and M Stephens (1999)**, 'Asymmetries in housing and financial market institutions and EMU', CEPR Discussion Paper No 2062

**Schwartz AJ (2002),** 'Asset price inflation and monetary policy', NBER Working Paper No 9321.

**Smith J, G Sterne and M Devereux (1994),** 'Personal and corporate sector debt', *Bank of England Quarterly Bulletin*, May, pp 144–155.

**Tunali I (1986),** 'A general structure for models of double-selection and an application to a joint migration/earnings process with remigration', *Research in Labor Economics*, 8(B), pp 235–282.

**Watson N and M Wooden (2002),** 'Toward an imputation strategy for Wave 1 of the HILDA Survey', HILDA Project Discussion Paper Series No. 1/02.