RESERVE BANK
OF AUSTRALIA

# Research Discussion Paper

# Mortgage-related Financial Difficulties: Evidence from Australian Micro-level Data

Matthew Read, Chris Stewart and
Gianni La Cava

RDP 2014-13

The Discussion Paper series is intended to make the results of the current economic research within the Reserve Bank available to other economists. Its aim is to present preliminary results of research so as to encourage discussion and comment. Views expressed in this paper are those of the authors and not necessarily those of the Reserve Bank. Use of any results from this paper should clearly attribute the work to the authors and not to the Reserve Bank of Australia.



ISSN 1320-7229 (Print)

ISSN 1448-5109 (Online)

# Mortgage-related Financial Difficulties: Evidence from Australian Micro-level Data

Matthew Read*, Chris Stewart** and Gianni La Cava*

Research Discussion Paper
2014-13

November 2014

*Economic Research Department
**Payments Policy Department
Reserve Bank of Australia

The authors would like to thank Dr Bill Measday and MARQ Services for their generosity in answering questions about the loan-level data used in this paper. They would also like to thank Luci Ellis, Rochelle Guttmann, Paul Hutchinson, Greg Kaplan, Josef Manalo, David Rodgers, Kylie Stewart, Puay Sim, Grant Turner and seminar participants at the Reserve Bank of Australia for useful discussions and feedback. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Reserve Bank of Australia. The authors are solely responsible for any errors.

Authors: readm and lacavag at domain rba.gov.au

Media Office: rbainfo@rba.gov.au

# Abstract

We investigate the factors associated with the incidence of mortgage-related financial difficulties in Australia. We use two complementary micro-level datasets: loan-level data on residential mortgages from two Australian banks, which we use to analyse the factors associated with entering 90+ day housing loan arrears; and household-level data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey, which we use to explore the factors associated with households missing mortgage payments.

The loan-level analysis indicates that the probability of entering arrears increases with the loan-to-valuation ratio (LVR) at origination, and is particularly high for loans with an LVR above 90 per cent. In contrast, the probability of entering arrears is lower for loans that are repaid relatively quickly. Additionally, the probability of entering arrears varies across different loan types; for example, low-documentation loans are more likely to enter arrears, even after controlling for whether the borrower was self-employed. The likelihood of entering arrears increases with the contract interest rate, which is consistent with lenders setting higher interest rates for riskier borrowers. The household-level analysis suggests that the probability of missing a mortgage payment is particularly high for households with relatively high debt-servicing ratios. Households that have previously missed a payment are also much more likely to miss subsequent payments than households with unblemished payment histories.



JEL Classification Numbers: G21, R29, R31

Keywords: household surveys, loan-level data, mortgage default

# Table of Contents

# Mortgage-related Financial Difficulties: Evidence from Australian Micro-level Data

**Matthew Read, Chris Stewart and Gianni La Cava**

## 1. Introduction

Housing loans account for a large proportion of both households' and lenders' balance sheets. The incidence of mortgage-related financial difficulties is, therefore, an important indicator of the financial health of households and lenders.

The pronounced deterioration of housing loan performance in the United States in the mid to late 2000s, and its role in contributing to the global financial crisis, has stimulated research into the determinants of mortgage default in the United States. In contrast, there appears to be little publicly available research on this topic in Australia.[1] This may be because the economic downturn in Australia was relatively mild and the associated deterioration in housing loan performance was, by international standards, benign (Figure 1). There has also been a paucity of adequate data; only a relatively short span of aggregate data on loan performance and some key explanatory variables, such as lending standards, has been available. These factors have made it difficult to examine the determinants of mortgage-related financial difficulties in Australia using aggregate data on housing loan performance. Instead, micro-level data are needed.

In this paper, we investigate the factors associated with mortgage-related financial difficulties in Australia using two separate, but complementary, micro-level datasets: loan-level securitised mortgage data from two Australian banks and household-level data from the HILDA Survey. The datasets are complementary for two reasons:

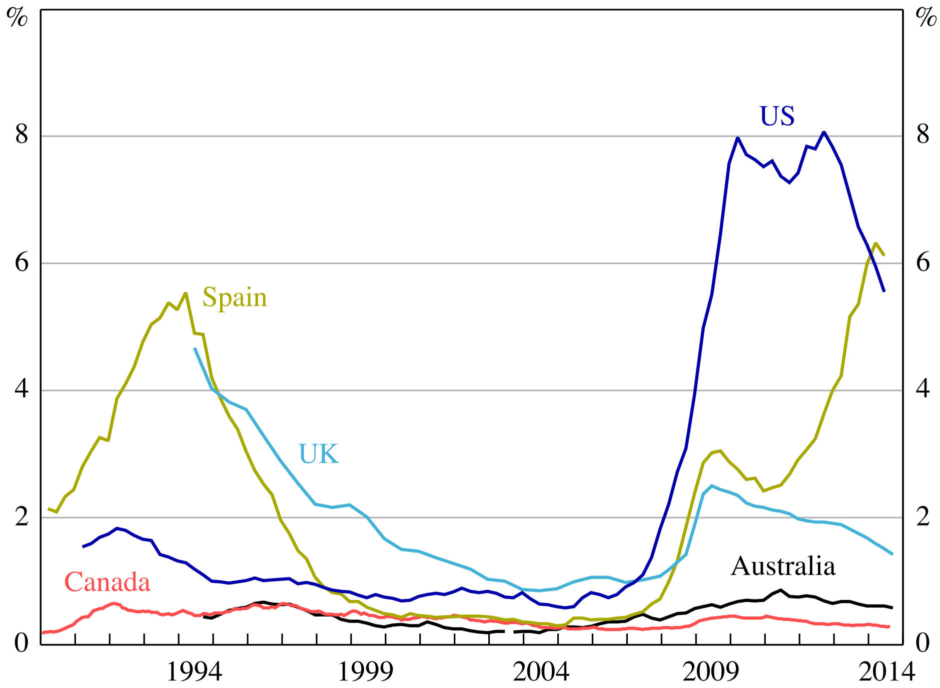
1. They include different types of information on loan, borrower and collateral characteristics. For instance, the loan-level dataset contains detailed information on loan characteristics, such as the LVR at origination and the actual interest rate charged on the loan, while the household-level dataset

1 Berry, Dalton and Nelson (2010) describe the drivers and impacts of mortgage default based on interviews with Australian households that had defaulted.

provides rich information on borrower characteristics, such as income and labour force status.

2. They provide different perspectives on mortgage-related financial difficulties. The loan-level data provide information on how many days a loan is in arrears (i.e. behind schedule on its required payments). This allows us to analyse the factors associated with 90+ day arrears, which are a precursor to default and possible loan losses for lenders. In contrast, the household-level data identify whether a household has missed at least one mortgage payment in the past year. While less severe than falling into 90+ day arrears, this measure provides insight into the early stages of mortgage-related financial difficulties.

**Figure 1: Banks' Non-performing Housing Loans**
Share of housing loans




Notes: There are differences in definitions and institutional coverage across countries; Australian data prior to September 2003 are past-due loans
Sources: APRA; Banco de España; Canadian Bankers Association; Council of Mortgage Lenders; Federal Deposit Insurance Corporation; RBA

The literature on mortgage-related financial difficulties typically focuses on two broad theories of mortgage default. Under both theories, households default in order to best smooth consumption in the face of unexpected shocks to their housing wealth, income or required expenditure. While this paper does not attempt to formally test the two theories, they provide a useful framework for considering the factors that drive mortgage-related financial difficulties.

So-called 'equity' theories of default assume that the decision to default is based on a rational comparison of the financial costs and benefits of continuing to make mortgage payments. Under these theories, default is analogous to a borrower exercising a put option when the value of their mortgaged property falls sufficiently relative to their outstanding mortgage debt (i.e. when the option is 'in the money'). These theories therefore emphasise the role of dwelling prices and amortisation (the extent of principal repayment) in explaining mortgage default. However, empirical studies commonly find that borrowers do not default as soon as they enter negative equity (e.g. Fuster and Willen 2013; Gerardi *et al* 2013). This may be due to the costs associated with default, including reputational costs and the associated negative effects on future access to credit (Elul 2006).

In contrast, 'ability-to-pay' theories maintain that borrowers do not strategically default based on their equity position, but only default when their incomes no longer cover their minimum loan payments and some subsistence level of expenditure. These theories focus on the role of liquidity constraints and credit market imperfections in explaining mortgage default.

These two theories are sometimes combined into so-called 'double-trigger' theories of default. Under these theories, borrowers only default if they experience a shock that makes them unable to pay their mortgage *and* they have negative housing equity.[2] An ability-to-pay shock, such as a negative shock to income, should not be sufficient on its own for a borrower to default. This is because a borrower with positive housing equity can sell the mortgaged property to pay back the loan or reduce their payment size by refinancing. However, these options are not typically available when the borrower has negative equity. Furthermore, negative equity should not be sufficient for a borrower to default; if the borrower expects housing prices to recover and default is not costless, it may be optimal for the borrower to continue to service the loan. Additionally, borrowers may delay default if they expect further significant dwelling price falls, as the value of the default option increases with falling dwelling prices (Kau, Keenan and Kim 1994).

2 For a more detailed description of double-trigger theories of default, see Aron and Muellbauer (2010).

Empirical approaches to testing theories of mortgage default have been eclectic, with studies simultaneously investigating both equity and ability-to-pay factors using a variety of models and data (e.g. aggregate data, loan-level data and household-level surveys). These studies generally find that ability-to-pay and equity factors are both important in determining whether a borrower defaults. Some US studies find evidence of strategic default (e.g. Ghent and Kudlyak 2010), although the level of negative equity at which this occurs has been estimated to be quite high (Bhutta, Dokko and Shan 2010). Some studies find that ability-to-pay factors, such as unemployment and the mortgage interest rate, play a large role in mortgage default behaviour (e.g. Fuster and Willen 2013). Other studies find that both factors, and sometimes their interaction, are important (e.g. Elul *et al* 2010; Gerardi *et al* 2013). Appendix A summarises some recent international studies of mortgage default.

Our paper makes two key contributions to the literature on mortgage-related financial difficulties. First, to the best of our knowledge, this is the first paper to use micro-level data to quantitatively analyse mortgage-related financial difficulties in Australia. Second, we find evidence to suggest that both ability-to-pay and equity factors have significant correlations with the incidence of mortgage-related financial difficulties.

This paper provides a useful input into the analysis of housing finance in Australia for a few reasons. First, the micro-level analysis provides a new 'bottom-up' assessment of the risks associated with housing lending. Second, the information could be used as an input into stress tests of the housing lending exposures of authorised deposit-taking institutions and mortgage insurers. Third, it could be useful in informing decisions about the design of the prudential policy framework. More broadly, the information could help to inform decisions about the level of risk that lenders, their investors and regulators are willing to accept.

The remainder of the paper is organised as follows. In Section 2, we analyse entry into 90+ day housing loan arrears using newly available loan-level data and a competing risks regression framework. In Section 3, we analyse missed mortgage payments using the HILDA Survey and a discrete choice modelling framework. Finally, Section 4 concludes.

# 2. Loan-level Determinants of Housing Loan Arrears

In this section, we use data on individual residential mortgages to explore the factors associated with entering 90+ day housing loan arrears. Borrowers in arrears by 90+ days are behind on their payments by at least three monthly contractual payments. We focus on loans that are at least 90 days in arrears, as these should correspond to borrowers that are experiencing serious financial difficulties rather than short-term liquidity problems. Additionally, arrears of this duration are consistent with the definition of default in the Basel II regulations.

## 2.1 Data

The loan-level dataset used in this paper is provided by MARQ Services (a firm that provides investors with information on the collateral pools backing residential mortgage-backed securities (RMBS)).[3] It contains monthly observations on housing loans that were originated between 1994 and 2013 and were securitised by two non-major banks. During the sample period from late 2009 to early 2014, the loan pool contained around 72 000 loans, with an average of around 25 monthly observations per loan.[4] Around 1 300 of these loans (1.8 per cent) were in arrears by more than 90 days at some point in the sample.

To investigate the representativeness of the sample, we compare its composition against that of broader samples at a particular point in time. Table 1 compares the sample against on-balance sheet and securitised housing loans. In the first case this information comes from APRA, while in the second case it is from Perpetual (the trustee for the majority of RMBS in Australia). Table 2 presents further selected descriptive statistics for the MARQ sample.

3 Details on how the dataset is cleaned and constructed are available from the authors on request.

4 The average number of observations per loan is substantially smaller than the length of the sample period (which spans around 50 months). This is because some loans enter the loan pool after the beginning of the sample period in 2009 and some loans are repaid early (and thus drop out of the sample).

**Table 1: Loan Pool Characteristics**
Share of loans outstanding, by value, December 2011

| Characteristic | Data source | | |
|---|---|---|---|
| | MARQ | APRA | Perpetual |
| 90+ day arrears rate | 0.6 | 0.7 | 0.5 |
| Fixed rate | 7.5 | 12.6 | 11.0 |
| Interest only | 23.2 | 32.7 | 21.5 |
| Investor | 26.0 | 32.9 | 26.9 |
| Low doc | 8.3 | 5.4 | 6.2 |
| Loan purpose | | | |
| Home improvement | 3.5 | na | 1.9 |
| Property purchase | 49.6 | na | 53.7 |
| Refinance | 36.3 | na | 24.4 |
| Other | 10.6 | na | 20.0 |
| State | | | |
| NSW | 32.8 | na | 31.6 |
| QLD | 39.7 | na | 23.7 |
| VIC | 15.1 | na | 23.3 |

Notes: Institutional coverage differs across data sources and across loan characteristics within data sources; in the APRA data, loans that are 90+ days in arrears include impaired loans that are not past due; 'other' includes construction, home equity loans and loans where the purpose is unknown

Sources: APRA; Authors' calculations; MARQ Services; Perpetual

**Table 2: Selected Descriptive Statistics for MARQ Loan Sample**
December 2011

| Characteristic | Percentiles | | | Mean |
|---|---|---|---|---|
| | 25th | 50th | 75th | |
| LVR at origination (%) | 47.4 | 69.7 | 80.0 | 64.1 |
| Interest rate (%) | 6.8 | 7.0 | 7.3 | 7.1 |
| Local unemployment rate (%) | 4.0 | 5.1 | 5.9 | 5.0 |
| Required payment ($'000) | 0.7 | 1.3 | 1.9 | 1.4 |
| Loan age (years) | 4.8 | 6.0 | 7.6 | 6.6 |

Note: 'Local unemployment rate' is the unemployment rate in the Statistical Area Level 4 (SA4) region in which the mortgaged property is located (there are around 90 SA4 regions in Australia and the median number of postcodes per region is around 30)

Sources: ABS; Authors' calculations; MARQ Services

As at December 2011, the sample contained 43 800 loans worth around $8.5 billion (equivalent to about 0.7 per cent of housing credit). The sample appears broadly similar in composition to the Perpetual loan pool across a number of loan characteristics. Notably, the 90+ day arrears rate for the sample is similar to arrears rates calculated using the other two data sources. While there are some differences in composition between the sample and the broader loan pools, this does not necessarily imply that the results from our analysis will be biased. We are interested in the *relationship between* certain variables and housing loan arrears. As long as the performance of the loans in this sample responded to these variables in the same way as loans in the broader loan pool, then our results will generalise to the population of housing loans.

Table 3 presents 90+ day arrears rates in the sample across a range of loan characteristics. Broadly speaking, the patterns in these arrears rates are consistent with aggregate data sources. For example: the arrears rate on low-doc loans is much higher than on full-doc loans; the arrears rates on investor and owner-occupier loans are broadly similar; the arrears rate on fixed-rate loans is lower than on variable-rate loans; and the arrears rate tends to increase with the LVR at origination.[5] Overall, however, we suggest caution in making inferences about the broader population of housing loans based on this sample; the sample contains loans from a small subset of lenders and lending practices may differ across lenders.

One potentially important variable that is not available in the dataset is the minimum required mortgage payment. For amortising loans, we estimate this using a credit-foncier model, which assumes that borrowers make constant payments over the life of the loan so that the loan principal is paid down to zero at loan maturity (based on the prevailing interest rate). For interest-only loans, the required payment is estimated as the product of the remaining loan balance and the interest rate.

The valuation for the mortgaged property available in the dataset is the value at loan origination and is not updated over time. We use hedonically adjusted price series to estimate dwelling price growth. For Sydney, Melbourne and Brisbane we use postcode-level indices we have estimated from unit-record data provided by Australian Property Monitors (APM) (see Appendix B for details on construction of the hedonic dwelling price indices). For all other areas, we use capital city or rest-of-state hedonic indices provided by RP Data-Rismark, as postcode-level

5 The arrears rate on loans with an LVR greater than 100 per cent at origination is lower than on loans with an LVR between 60 and 100 per cent at origination. However, the dataset contains information on only the first property securing each loan, implying that loans with multiple properties as security will have an LVR at origination that is overestimated.

**Table 3: 90+ Day Arrears Rates by Loan Characteristic**
Share of loans outstanding, by number

| | | | |
|---|---|---|---|
| LVR at origination | | Loan documentation | |
| $0 \leq \text{LVR} < 60$ | 0.18 | Full doc | 0.33 |
| $60 \leq \text{LVR} < 80$ | 0.52 | Low doc | 1.70 |
| $80 \leq \text{LVR} < 90$ | 0.59 | Loan purpose | |
| $90 \leq \text{LVR} < 100$ | 0.72 | Home improvement | 0.21 |
| $\text{LVR} \geq 100$ | 0.51 | Property purchase | 0.35 |
| Employment type | | Refinance | 0.51 |
| Self-employed | 1.13 | Other | 0.44 |
| Wage earner | 0.32 | Property purpose | |
| Interest rate type | | Investor | 0.49 |
| Fixed | 0.12 | Owner-occupier | 0.42 |
| Variable | 0.46 | Payment type | |
| | | Amortising | 0.43 |
| | | Interest only | 0.44 |

Note: Arrears rates calculated over entire sample (October 2009 to January 2014)
Sources: Authors' calculations; MARQ Services

dwelling price data were not available outside Sydney, Melbourne and Brisbane. While both types of indices are only estimates, they should be a reasonable proxy for borrowers' beliefs about property values to the extent that borrowers adjust their beliefs based on observing sales prices of nearby properties. Gerardi *et al* (2013) argue that perceived valuations are more relevant to mortgage default decisions than actual values, as households take into account their own valuation of their property when choosing whether to default.

## 2.2 Modelling Framework

Duration analysis provides a framework for modelling 'time-to-event' data. In our case, the time-to-event is the time between loan origination and a loan falling into arrears. Importantly, duration models can provide estimates of the effects of covariates on the probability of entering arrears. The advantages of using duration analysis to model housing loan arrears are that it accounts for 'right-censoring' (where the ultimate outcome of the loan is not observed) and can parsimoniously account for time dependence (where the probability of entering arrears is a function of the time since loan origination).

Duration analysis of arrears data is complicated by the fact that most housing loans are paid down in full before or when the loan matures. Application of standard duration analysis techniques is inappropriate in the presence of 'competing risks' – that is, events that prevent observational units from ever experiencing the event of interest. In this case, the competing risk is the loan being paid down in full. Standard duration analysis techniques would treat loans that have been paid down in full as being censored. They would also assume that these loans could still fall into arrears, which is clearly inappropriate, as a loan that has been paid down no longer exists and thus has zero probability of entering arrears. In cases where the probability of experiencing the competing risk is correlated with the covariates of interest, standard duration analysis techniques can yield misleading estimates of the effects of these covariates on the probability that the event of interest occurs.

Competing risks regression models provide a framework for analysing time-to-event data in the presence of competing risks. Competing risks frameworks have previously been used to model default for housing, commercial and personal loans. For example, for the United States, Deng, Quigley and Van Order (2000) estimate a competing risks model for residential mortgage default and prepayment, while Ciochetti *et al* (2002) estimate a similar model for commercial mortgages. Watkins, Vasnev and Gerlach (2014) estimate a competing risks model using data on personal loans made by an Australian bank.

In standard duration analysis, the hazard function, $h(t)$, approximates the instantaneous probability of an event occurring at time $t$ conditional on it having not occurred before time $t$.[6] In a competing risks framework, Fine and Gray (1999) propose a model for the hazard function of the subdistribution of the event of interest, which they call the 'subhazard' (for technical details, see Appendix C). When incorporating covariates, the model for the subhazard of entering arrears (denoted by the subscript $a$) takes a proportional hazards form:

$$h_a\left(t|\mathbf{z}_{it}\right) = h_a^0(t)\exp\left(\mathbf{z}_{it}'\gamma\right), \tag{1}$$

where $\mathbf{z}_{it}$ is a vector of explanatory variables corresponding to loan $i$, $\gamma$ is a vector of coefficients and $h_a^0(t)$ is the baseline subhazard, which accounts for time dependence (outside of the effects of time-varying covariates). The model is semi-parametric, since the shape of the baseline subhazard is left unspecified.

6 For the random time-to-event variable $T$, $h(t) = \lim_{\delta\to 0}\{\Pr(t \le T < t+\delta|T \ge t)/\delta\}$.

The estimation results reported in Section 2.3 are exponentiated coefficients (i.e. $\exp(\gamma_k)$ for variable $k$), which are known as 'subhazard ratios' (SHRs). An SHR of $\exp(\gamma_k)$ means that a one unit increase in variable $k$ results in the subhazard being $\exp(\gamma_k)$ times its original value. Therefore, an SHR greater than one implies that an increase in the covariate results in the subhazard increasing. The significance levels reported in Section 2.3 correspond to the null hypothesis that the coefficient on that variable is equal to zero, which is equivalent to an SHR of one.

We estimate a competing risks regression model for mortgage arrears, where the competing risk is full payment (either before or at loan maturity). A loan is classified as having been paid down in full if it drops out of the loan pool before the latest report date. Loans that are outstanding on the latest report date but are not in arrears are considered right-censored. A loan is classified as being in arrears if it is in arrears by more than 90 days. Once a loan has entered arrears, it is removed from the set of loans 'at risk' of entering arrears – that is, we do not allow loans that are in arrears to 'cure' (i.e. return to performing status without refinancing).

Of loans that entered 90+ day arrears in December 2011, around 40 per cent of these loans had returned to performing status three months later, while around 45 per cent remained in 90+ day arrears.[7] The remaining loans had exited from the loan sample, probably because the borrower paid the loan down by selling the property or they refinanced (although a very small number of borrowers may have had their property repossessed). Given the relatively small sample of loans that cure, it is unlikely that our sample would provide much information on the factors associated with curing. However, this could be an interesting avenue for further research when more loan-level data become available, because curing rates will affect the stock of loans that are in arrears at a given time.

7 These transition rates should not be taken as indicative of typical rates of transition out of arrears for the population of housing loans. Transition rates may vary across lenders based on their processes for collections and their procedures for dealing with customers experiencing financial hardship. They may also vary across time due to changes in these processes and procedures or as a result of macroeconomic factors.

## 2.3 Results

Table 4 presents results for a competing risks regression model for mortgage arrears.[8] As explanatory variables, the model includes the LVR at origination (as a sequence of dummy variables to capture potential nonlinearities), the percentage of the loan balance that has been paid down since origination (i.e. amortisation), the cumulative percentage growth of dwelling prices since origination, the local unemployment rate, the current mortgage interest rate for each loan and a number of other loan characteristics.

**Table 4: Housing Loan Arrears – Competing Risks Model**

| Explanatory variable | SHR | Explanatory variable | SHR |
|---|---|---|---|
| Amortisation | 0.99*** | Investor | 0.91 |
| LVR at origination | | Loan purpose | |
| $60 \leq$ LVR$< 80$ | 1.75*** | Home improvement | 0.47** |
| $80 \leq$ LVR$< 90$ | 1.93*** | Refinance | 1.77*** |
| $90 \leq$ LVR$< 100$ | 3.47*** | Other | 1.16 |
| LVR $\geq 100$ | 2.77*** | Local unemployment rate | 1.03* |
| Dwelling price growth | 1.00 | Low doc | 1.76*** |
| Fixed rate | 0.39*** | Minimum required payment | 1.31*** |
| Interest only | 0.56*** | Self-employed | 1.19 |
| Interest rate | 1.35*** | | |
| Number of observations | | 1 612 645 | |
| Number of loans | | 63 468 | |
| Number entered arrears | | 1 056 | |
| Number paid in full | | 25 841 | |
| Number censored | | 36 571 | |

Notes: ***, ** and * denote statistical significance at the 1, 5 and 10 per cent levels, respectively; standard errors are clustered by loan; 'amortisation' is the percentage decrease in the loan balance since origination; 'dwelling price growth' is the cumulative percentage growth of dwelling prices since origination; 'minimum required payment' is measured in thousands of dollars

Sources: ABS; APM; Authors' calculations; MARQ Services; RP Data-Rismark

8 Results from an alternative model that accounts for the discrete timing of observations in the dataset and for unobserved heterogeneity (but ignores the presence of the competing risk) are presented in Appendix D. These results are broadly similar to the results from the competing risks model.

### *2.3.1 Equity factors*

The model provides evidence to suggest that equity factors are associated with the probability of falling into arrears; both the LVR at origination and the amount of amortisation since origination have statistically (and economically) significant SHRs. The subhazard of entering arrears tends to increase with the LVR at origination. For example, a loan with an LVR at origination of between 90 and 100 per cent has an estimated subhazard of entering arrears that is about 3½ times that of a loan with an LVR less than 60 per cent (to put these results into perspective, in Section 2.3.4 we select a 'base' loan and examine how the probability of entering arrears before a certain loan age varies with loan characteristics). Additionally, the subhazard of entering arrears appears to increase nonlinearly, and is particularly high for loans with an LVR between 90 and 100 per cent; a loan with an LVR at origination of between 80 and 90 per cent has a subhazard of entering arrears that is about 1.1 times higher than that of a loan with an LVR between 60 and 80 per cent, but a loan with an LVR of between 90 and 100 per cent has a subhazard of entering arrears that is almost twice that of a loan with an LVR between 80 and 90 per cent.

Somewhat counterintuitively, the results suggest that loans with an LVR at origination greater than 100 per cent are less likely to fall into arrears than loans with an LVR between 90 and 100 per cent. However, as mentioned previously, this is likely to reflect measurement error; the dataset contains information on only the first property securing each loan, implying that loans with multiple properties as security will have an LVR at origination that is overestimated (because the value of the collateral is underestimated).

The amortisation variable has an estimated SHR that is statistically smaller than one, indicating that an increase in cumulative amortisation is associated with a decrease in the subhazard of entering arrears. The magnitude of the effect appears fairly small, at 0.99. However, the effect of an $x$ percentage point increase in cumulative amortisation will be associated with a subhazard that is about $0.99^x$ times lower. For example, a 10 percentage point increase in cumulative amortisation is associated with a subhazard that is around 0.9 times lower, while a 50 percentage point increase in cumulative amortisation is associated with a subhazard that is around 0.6 times lower. Of course, income is also likely to play a role in this relationship; borrowers with higher incomes can pay down their loans faster than other borrowers and will be less likely to enter arrears for other reasons related to their higher income.

One caveat with these results (and the results from the model more generally) is that, given that we do not have data on borrower incomes, we cannot construct a meaningful measure of borrowers' debt-servicing burdens. Therefore, the estimated relationship between the LVR at origination (or the amount of amortisation) and the incidence of arrears may be biased due to the unobserved effect of debt-servicing burdens. Although the estimated required payment should partly control for this, ideally the required payment should be scaled by the borrower's income, since borrowers with higher incomes should be able to meet higher payments.

The SHR for dwelling price growth since loan origination is not statistically significant, suggesting that changes in dwelling prices have not been associated with changes in the incidence of arrears in this sample. This may reflect a lack of sufficient variability in dwelling prices in the sample period. It could also reflect measurement error, since borrowers who entered arrears in this sample may have experienced changes in dwelling prices that were different to the path of dwelling prices implied by the indices that we have used.

#### *2.3.2 Ability-to-pay factors*

The results suggest that borrowers with higher mortgage interest rates have a higher subhazard of entering arrears; a loan with an interest rate 1 percentage point higher than that of an otherwise identical loan is estimated to have a subhazard of entering arrears that is around 1.4 times higher. The mechanism through which this effect might be expected to work is that the higher interest rate increases the required payment, making it more likely that the borrower's income is insufficient to cover their loan payments and subsistence-level expenditure. However, our model controls for the estimated required payment, suggesting that the effect of interest rates on arrears in the model is not just due to such a 'debt-servicing channel'. Instead, the estimated effect may reflect the fact that lenders charge higher interest rates on loans that are more likely to fall into arrears (i.e. higher-risk loans) as compensation for this risk. In our model, we are able to control for some observable loan risk characteristics, such as the loan documentation type. However, when negotiating a borrower's interest rate, lenders may also take into account variables that do not appear in this dataset, such as the borrower's income or wealth; additionally, the lender's existing relationship with the borrower is likely to be an important factor.[9] Overall, the estimated relationship between the mortgage interest rate and the probability of entering arrears is consistent with lenders using risk-based pricing.

The results suggest that ability-to-pay shocks, proxied by the local unemployment rate, have a small but statistically significant (at the 10 per cent level) correlation with the probability of entering arrears. This estimate almost certainly understates the effect of a borrower actually becoming unemployed on the probability that they enter arrears. Indeed, Gyourko and Tracy (2013) show that using unemployment rates to proxy for borrowers' actual (unobserved) employment statuses can result in a severe attenuation bias. This is supported to some extent by our analysis in Section 3 using the separate household-level dataset in which we observe each borrower's labour force status directly.

### *2.3.3 Loan characteristics*

In terms of loan characteristics, fixed-rate and interest-only loans are estimated to have lower subhazards of entering arrears than variable-rate and amortising loans, respectively. While borrowers on fixed-rate loans are insulated against changes in lending rates during their fixed-rate period, our model includes the mortgage interest rate and the estimated required payment, so it is unclear why fixed-rate borrowers should be less likely to fall into arrears. The estimated subhazards for fixed-rate and interest-only loans are possibly biased to the extent that the take-up of fixed-rate and interest-only loans is correlated with income (and potentially with other omitted variables, such as financial sophistication). Another possibility is that these loans tend to enter arrears only after they 'reset' to variable rates (in the case of fixed-rate loans) or amortising payments (in the case of interest-only loans). However, estimating a version of the model that only uses the characteristics of the loan as at loan origination yields very similar results.

9 Based on a linear regression, the relevant loan characteristics available in the sample (e.g. loan size, loan documentation type, interest rate type and LVR at origination) explain only around 35 per cent of the variation in the difference between the interest rate for each loan and the advertised standard variable rate for the corresponding lender.

Despite the results suggesting that interest-only and fixed-rate loans are less likely to enter arrears than other loans, it is important to remember that these results are conditional on cumulative amortisation. To the extent that these loans amortise more slowly than other loans, increases in these types of loans can represent increasing risk, as the results suggest that slower rates of amortisation are associated with a higher probability of entering arrears. Loans that amortise more slowly may also generate greater loan losses for lenders if those loans default.

Also relating to loan characteristics, the results indicate that low-doc loans (that is, loans where the borrower's income has not been documented, assessed and verified, such as by checking pay slips or business activity statements) have a subhazard of entering arrears that is around 1.8 times greater than that of full-doc loans, after controlling for other factors. This does not simply reflect the tendency for low-doc loans to be extended to self-employed borrowers, who tend to have more volatile incomes, as we control for whether the borrower was self-employed at the time of loan origination.[10] The estimated SHR for low-doc loans could reflect a correlation with the level of borrower income, but could also reflect higher-risk borrowers self-selecting into this product category. The results also suggest that refinanced loans have a subhazard of entering arrears that is 1.8 times that of loans for property purchase. This could reflect the fact that some borrowers refinance because they are having difficulty making their payments, implying that there is also self-selection of some riskier borrowers into this loan type.

### *2.3.4 Economic significance*

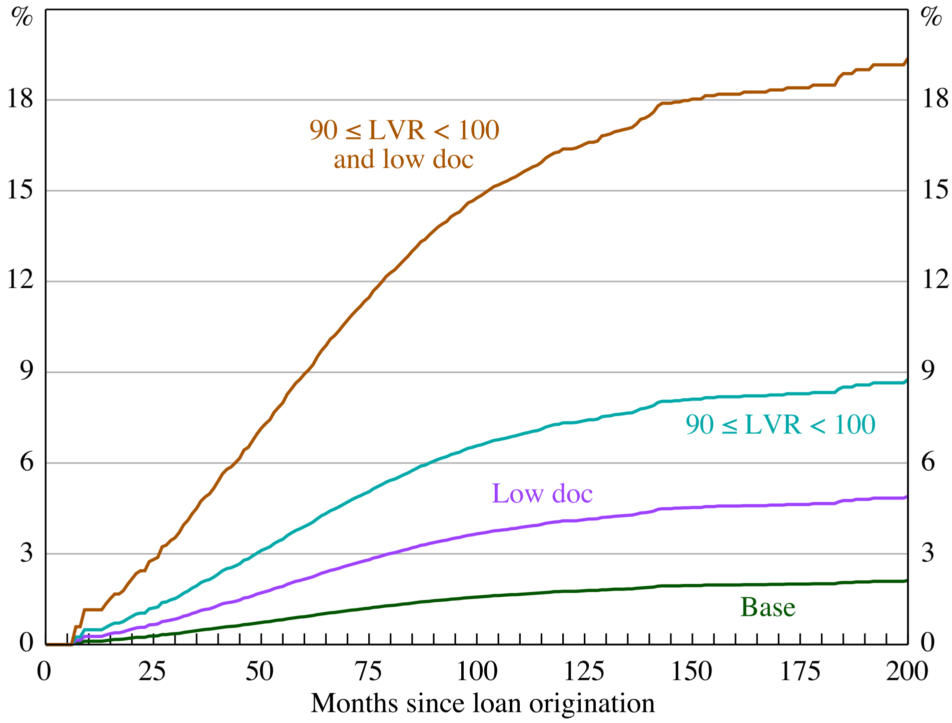
A potentially useful way to consider the economic significance of these results is by examining the cumulative incidence function (CIF), which gives the probability of a loan entering arrears before time $t$ (for technical details, see Appendix C). Figure 2 plots the CIF for hypothetical loans with different characteristics.[11] The characteristics of the 'base' loan are the modes of the categorical variables and the means of the continuous variables (see note to Figure 2 for details), while the

10 Around 80 per cent of low-documentation loans in the sample were to borrowers that were self-employed when the loan was approved.

11 The CIFs are calculated based on the results of an alternative competing risk model that only uses information from the time of loan origination. This model excludes cumulative amortisation, dwelling price growth and the local unemployment rate.

other series show how the probability of entering arrears changes as certain loan characteristics vary from the base case.

**Figure 2: Cumulative Incidence of Mortgage Arrears**




Note: 'Base' is a loan to a wage-earning borrower for the purchase of owner-occupied property, with full income documentation, principal and interest payments, a variable interest rate of 6.9 per cent, an LVR less than 60 per cent and a required payment of $1 680 per month

Sources: Authors' calculations; MARQ Services

The probability that the base loan enters arrears within the first five years is 0.9 per cent. A low-doc loan that is otherwise identical to the base loan has a probability of entering arrears in the first five years of 2.2 per cent, while a loan with an LVR between 90 and 100 per cent at origination has a probability of around 3.9 per cent. A loan that is low doc and has a high LVR at origination is much more likely to enter arrears than a loan with just one of these characteristics; 8.9 per cent of these loans would be expected to enter arrears in the first five years.

## 3. Household-level Determinants of Missed Mortgage Payments

In this section, we identify household characteristics associated with missing a mortgage payment using household-level survey data. This complements the analysis in the previous section by allowing us to use a range of variables that are not available in the loan-level dataset, including the borrower's labour force status and income.

## 3.1 Data

As part of the 2006 and 2010 wealth modules, the HILDA Survey – an annual household-based longitudinal study – asked respondents if they had been unable to meet a payment by the due date on any housing or property loan in the previous 12 months because of financial difficulties. The share of households with owner-occupier mortgage debt that reported missing a mortgage payment was around 5½ per cent in 2006 and 6 per cent in 2010.[12]

While missing a mortgage payment does not necessarily correspond to the borrower defaulting, it represents an early stage of the default process and provides a signal of financial difficulties. For example, around 14 per cent of households that missed a mortgage payment in 2010 reported being behind schedule on their mortgage payments at the time of the 2010 survey, compared with 2½ per cent of households that did not miss a payment. Additionally, around 5 per cent of households that missed a mortgage payment in 2006 reported selling their home due to financial difficulties at some point in the following four years, compared with 2 per cent of households that did not miss a payment. These statistics imply that mortgage-related financial difficulties are often temporary; only a small proportion of households that report missing a mortgage payment go on to report experiencing more serious financial difficulties.

### *3.1.1 Ability-to-pay factors*

A common measure of a borrower's ability to comfortably make their mortgage payments is the debt-servicing ratio (DSR), defined as the percentage of household disposable income used to service mortgage debt. The measure of mortgage payments available in the HILDA Survey is based on households' reported 'usual payments' on owner-occupier housing debt, which may include regular and excess repayments of principal, as well as interest payments. The DSR may help to

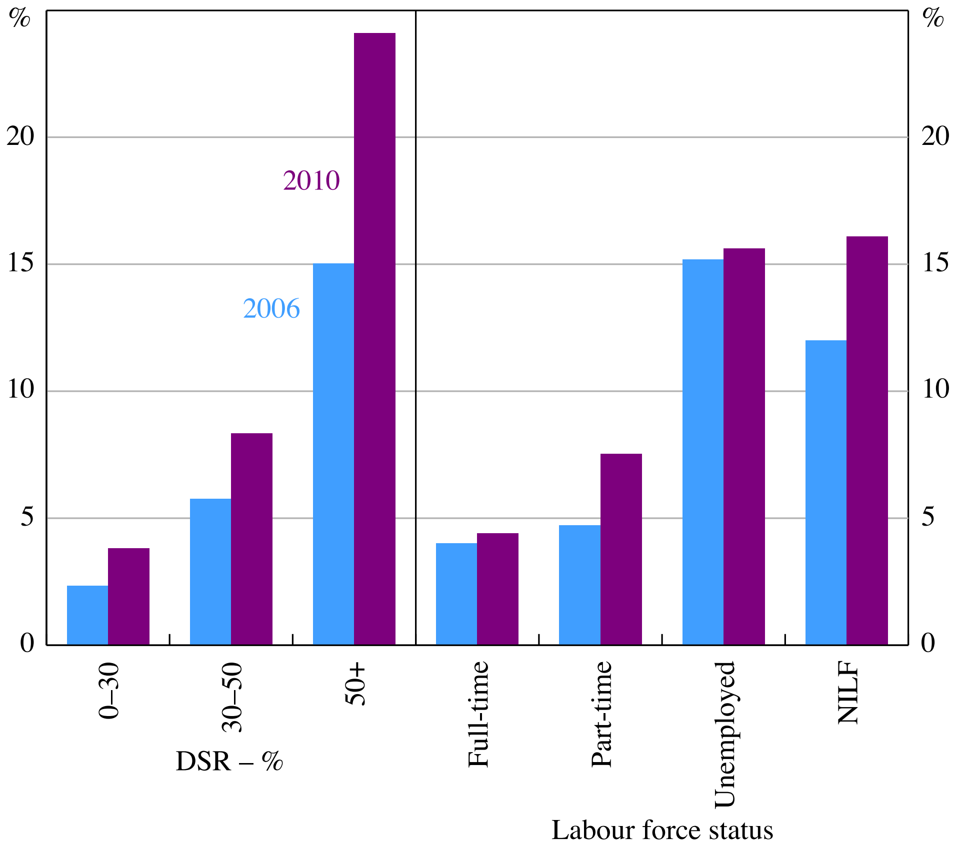
12 Here, 'households with owner-occupier mortgage debt' are defined as those that had an owner-occupier mortgage at the time of the 'current' or 'previous' survey (e.g. in 2009 or 2010). The reason for using this definition is outlined in Appendix E.

identify households that are particularly vulnerable to income or expenditure shocks.

The share of households that reported missing a mortgage payment tends to increase with the DSR (Figure 3). Borrowers that were unemployed or not in the labour force (NILF) were also more likely to miss a mortgage payment relative to those in employment. This may largely reflect correlation with income and DSRs; individuals that are unemployed or NILF tend to have lower incomes and higher DSRs.

**Figure 3: Missed Mortgage Payments by DSR and Labour Force Status**

Share of households by category




Note: DSR and labour force status of household head at previous survey

Sources: Authors' calculations; HILDA Release 12.0

Previous behaviour in servicing mortgage and other debt may also provide useful insight into the propensity for a household to miss a mortgage payment, potentially by capturing a household's 'willingness to pay'. Households that missed a mortgage payment in 2006 were substantially more likely to miss a payment in 2010 than households that did not miss a mortgage payment, despite the substantial period of time between the two observations (Table 5). May and Tudela (2005) suggest three potential explanations for this:

1. The conditions relevant to a borrower meeting their debt obligations may be altered if they have previously missed a mortgage payment; this is sometimes referred to as state dependence. For example, missing a mortgage payment may make it more difficult to access credit in the future. If borrowers cannot costlessly renegotiate their mortgage terms (such as through refinancing) when facing payment difficulties, then payment problems may persist. Furthermore, the borrower may be less averse to missing payments if the associated stigma is lessened by their previous experience.

2. Characteristics that increase the propensity to miss a mortgage payment may be persistent (or invariant) over time. These may include observed characteristics, such as the DSR, as well as unobserved characteristics, such as financial literacy.

3. Persistence in mortgage-related financial difficulties may be observed if a single spell of mortgage-related financial difficulties tends to be long in duration. This explanation seems less applicable here, as our observations are four years apart.

**Table 5: Missed Mortgage Payments – Transition Rates**
Share of households by 2006 category

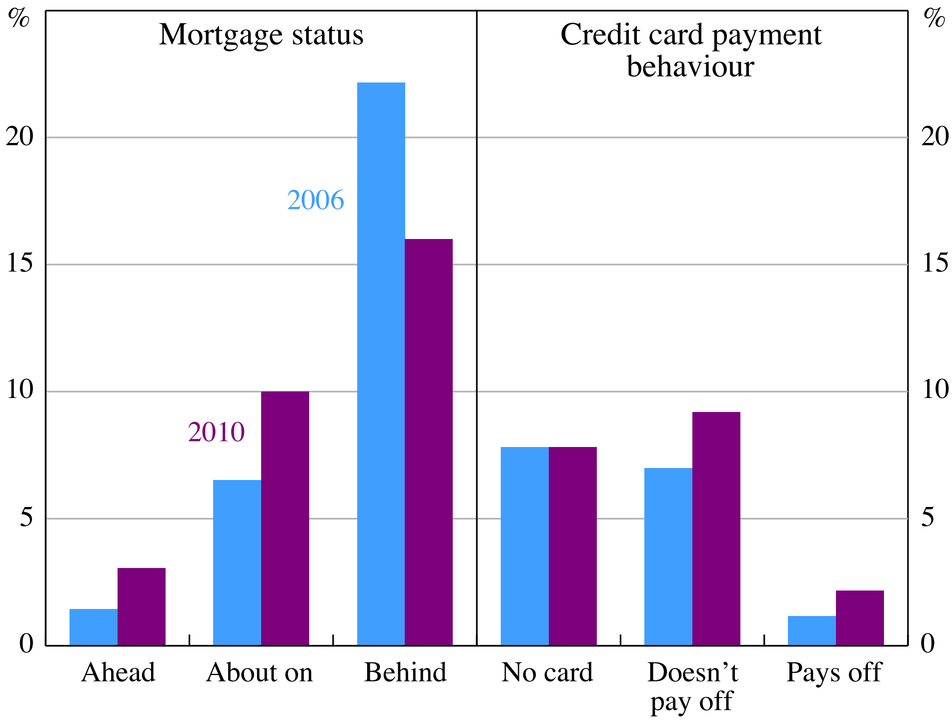
| | **2010** | |
|---|---|---|
| | Missed a payment | Did not miss a payment |
| Missed a payment | 36 | 64 |
| Did not miss a payment | 5 | 95 |

Sources: Authors' calculations; HILDA Release 12.0

The ongoing persistence of mortgage-related financial difficulties may also be observed through the relationship between a household's mortgage status – whether the household reports being ahead, behind or on schedule on their mortgage payments – and whether they miss a mortgage payment (Figure 4). Of households that reported being behind schedule on their mortgage payments in 2009, 16 per cent missed a mortgage payment in 2010; this is likely to reflect factors similar to those that cause persistence in missing mortgage payments

**Figure 4: Missed Mortgage Payments by Previous Payment Behaviour**

Share of households by category




Note: Reported payment behaviour at previous survey

Sources: Authors' calculations; HILDA Release 12.0

(discussed above). By comparison, only 3 per cent of households that were ahead of schedule in 2009 missed a payment in 2010.

Credit card payment behaviour may also provide some information about mortgage payment behaviour. Households that reported always (or almost always) paying off the entire balance on their credit cards each month were less likely to miss mortgage payments than households that did not have a credit card or did not always pay off the entire balance of their credit card each month. This may be because these households are more financially literate or conscious of actively managing their finances; they could also have less variable income or expenditures.

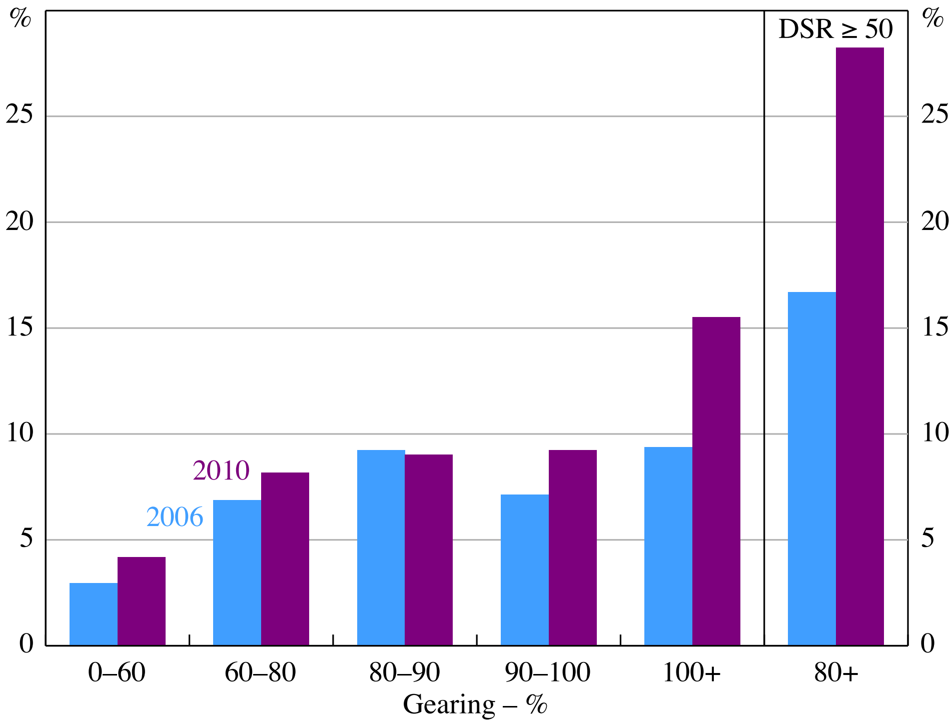
### *3.1.2 Equity factors*

The HILDA Survey data allow us to calculate a household's level of housing gearing using the value of their outstanding mortgage debt and their self-assessed home value. While Windsor, La Cava and Hansen (2014) show that there is considerable dispersion in the difference between home price beliefs and observed (hedonically adjusted) prices, home price beliefs appear to be unbiased on average and (as noted in Section 2.1) households' valuations are likely to be of greater relevance to mortgage default decisions than actual prices (Gerardi *et al* 2013).

There does not appear to be a particularly strong (or stable) relationship between housing gearing and missed payments (Figure 5). Taken at face value, this suggests that equity factors are less important than ability-to-pay factors as a determinant of missing mortgage payments; this is not particularly surprising given that housing lending in Australia is 'full recourse', meaning that, in the event of default, lenders have a claim on some assets other than the mortgaged property. On the other hand, looking at gearing and DSRs together provides some evidence for the importance of equity factors in missing mortgage payments; the incidence of missed mortgage payments among households that have both high gearing and high DSRs is greater than among households that have either high gearing or high DSRs, but not both. This suggests that double-trigger effects (described in Section 1) may play a role in households missing mortgage payments.

**Figure 5: Missed Mortgage Payments by Gearing**

Share of households by category




Note: Gearing and DSR as at previous survey

Sources: Authors' calculations; HILDA Release 12.0

## 3.2 Modelling Framework

The preceding analysis of missed mortgage payments in the HILDA Survey dataset describes only unconditional correlations between missed payments and particular variables. In order to account for cross-correlations between these variables, and thus isolate their direct effects on the probability of missing a mortgage payment, we turn to regression methods. Since the dependent variable is binary – a household either missed a payment or did not – we employ a probit model:

$$Y_i = \begin{cases} 1 & \text{if} \quad Y_i^* \geq 0 \\ 0 & \text{if} \quad Y_i^* < 0 \end{cases}$$

$$Y_i^* = \mathbf{x}_i'\beta + u_i, \quad u_i \sim N(0,1)$$

$$\Pr(Y_i = 1|\mathbf{x}_i) = \Phi(\mathbf{x}_i'\beta).$$

Here, $Y_i = 1$ if household $i$ missed a mortgage payment and $Y_i = 0$ if the household did not miss a payment, $\mathbf{x}_i$ is a vector of explanatory variables for household $i$, $\beta$ is a vector of coefficients and $\Phi$ is the standard normal cumulative distribution function. Under this model, a household misses a mortgage payment when the continuous latent random variable $Y_i^*$ exceeds some threshold (normalised to 0). Notable features of the model are that:

- We estimate the model for the 2010 cross-section and include a lag of the dependent variable to capture possible state dependence in missing mortgage payments. The lagged missed payments variable is a categorical variable (represented by a set of dummy variables), where one of the categories is for non-response.[13] We also include other variables related to the mortgage status (i.e. whether the household is ahead, behind or on schedule) and credit card payment behaviour.[14]

---

13 Households that responded 'no' to missing a payment in 2006 but did not appear to have a property loan in the 2005 or 2006 surveys are treated as non-responders. For more information about this issue, see Appendix E.

14 The mortgage status variable also includes a category for households that only have a 'second mortgage' (e.g. a home equity loan), as these households are not always asked about their mortgage payment status.

- Following May and Tudela (2005), we use lagged values of the explanatory variables (i.e. from the 2009 survey) instead of their contemporaneous values to capture the household's characteristics prior to missing a payment. This should help to minimise endogeneity problems related to reverse causality and allows us to better identify meaningful lead-lag relationships. We exclude households that bought their residence in 2010, as the DSR and gearing recorded in the 2009 survey would not correspond to the dwelling for which a mortgage payment was missed.
- DSRs and gearing are included as categorical variables to capture potential nonlinear relationships between these variables and missed payments.
- The household head's labour force status is included as an explanatory variable. In defining the labour force status, we differentiate between full-time wage earners, part-time wage earners and the self-employed. Self-employed households are likely to have relatively volatile incomes, which may affect their ability to repay loans.
- The state in which the property is located is used to control for geographic factors, possibly related to local conditions in the labour and housing markets.

### 3.3 Results

The results from our preferred probit model are shown in Table 6. Overall, the results indicate that ability-to-pay factors are strongly positively associated with the probability of missing mortgage payments. Additionally, there is also some evidence to suggest that borrowers with negative equity are more likely to miss a payment.

**Table 6: Missed Mortgage Payments – Probit Model Estimation Results**

| Variable | Coefficient | Marginal effect ppt |
|---|---|---|
| DSR | | |
| 30 ≤DSR<50 | 0.35*** | 3.57** |
| *DSR* ≥50 | 0.76*** | 10.00*** |
| Labour force status | | |
| Full-time self-employed | 0.53*** | 6.14** |
| Part-time employed | 0.28 | 2.73 |
| Unemployed | 0.67 | 8.43 |
| NILF | 0.64*** | 7.92*** |
| Previously missed a payment | | |
| Missed mortgage payment in 2006 | 1.24*** | 22.20*** |
| Non-response to question in 2006 | 0.19 | 1.82 |
| Mortgage payment status | | |
| Ahead of schedule | –0.37*** | –3.65*** |
| Behind schedule | 0.03 | 0.43 |
| Second mortgage only | –0.33 | –3.30 |
| Credit card payment behaviour | | |
| Does not pay off credit card | 0.24* | 2.71* |
| Pays off credit card | –0.27* | –2.19* |
| Current gearing | | |
| 60 ≤ Gearing <80 | 0.24* | 2.38 |
| 80 ≤ Gearing <90 | 0.02 | 0.15 |
| 90 ≤ Gearing <100 | 0.24 | 2.37 |
| Gearing ≥100 | 0.52** | 6.20* |
| Constant | –2.24*** | |
| Number of observations | 1 745 | |
| Pseudo-$R^2$ | 0.21 | |
| Likelihood ratio ($\chi^2_{23}$) | 170.80*** | |

Notes: Marginal effects and corresponding standard errors are calculated for each household based on the observed values of the explanatory variables for that household and are averaged across all households to yield average marginal effects and associated standard errors; ***, ** and * denote statistical significance at the 1, 5 and 10 per cent levels, respectively; results for geographic controls not reported; for further details on model specification see Section 3.2

Sources: Authors' calculations; HILDA Release 12.0

### *3.3.1 Ability-to-pay factors*

The results suggest that having a DSR over 50 per cent is associated with a probability of missing a mortgage payment that is, on average, 10 percentage points higher than for a DSR under 30 per cent. Even after controlling for the

DSR, we find evidence that the labour force status of the household head is correlated with the probability of missing a mortgage payment. Households with a household head that is NILF are 8 percentage points more likely, on average, to miss a mortgage payment than those with a household head that is a full-time wage earner. Full-time self-employed workers are also around 6 percentage points more likely to miss a mortgage payment, possibly reflecting these households' more volatile cash flows. The marginal effect of being unemployed is large in magnitude, at around 8 percentage points, but statistically insignificant, possibly reflecting the small sample of unemployed households (less than 1 per cent of the estimation sample). Replacing the labour force status variables with variables representing the *change* in labour force status yields qualitatively similar results.

Households that had missed a mortgage payment in 2006 are estimated to be around 22 percentage points more likely to miss a payment in 2010, on average. This effect is broadly consistent with aggregate data on non-conforming housing loans (many of which are to borrowers with blemished credit histories); arrears rates on non-conforming loans tend to be far greater than arrears rates on 'prime' lending. This result, however, should be interpreted with some caution, as it may reflect an endogeneity problem. In particular, the lagged dependent variable may be correlated with the latent error term ($u_i$) if there are persistent omitted factors that influence the probability of missing a mortgage payment, such as household wealth and financial literacy.[15] However, we have potentially controlled for such factors by including variables relating to the mortgage status and credit card payment behaviour (discussed below). Even if the effect of state dependence on missing a mortgage payment is overstated by the coefficient on the lagged missed mortgage payment variable, these results still indicate that having previously missed a mortgage payment is a good predictor of subsequently missing another payment. This result supports the practice of lenders using credit scores and other information on previous debt payment behaviour in their credit assessment processes.

15 Data on household wealth are only available every four years in the HILDA Survey's wealth modules. Including variables in the model related to household wealth in 2010 results in statistically significant marginal effects with the expected signs, and has little effect on the estimated marginal effects of the other explanatory variables. However, because these variables are observed after a mortgage payment has been missed, we are unable to disentangle the causal relationship between wealth and missed payments.

Also in relation to previous debt-servicing behaviour, we find that households that are ahead of schedule on their mortgage payments are, on average, 4 percentage points less likely to miss a payment than households that are on schedule. This could be the net result of several factors. First, households that are ahead of schedule on their mortgage payments probably tend to be better at managing their finances than other households. Second, if faced with temporarily lower income, households that are ahead of schedule can comfortably miss a scheduled payment without severe consequences. All else equal, this could make these households *more* willing to miss a payment. Finally, households that are ahead of schedule could avoid missing a payment by drawing down on existing offset account balances or mortgage redraw facilities. These considerations are complicated by uncertainty about whether households would report missing a payment if they are ahead of schedule at the time that they miss the payment. The marginal effect of being behind schedule is not statistically significant, possibly reflecting the small sample of such households (around 3 per cent of the estimation sample).

In terms of the payment of non-mortgage debt, households that do not pay off their entire credit card balance each month are, on average, 3 percentage points more likely to miss a payment than households with no credit card, while households that regularly pay off their credit card are around 2 percentage points less likely to miss a payment.

### *3.3.2 Equity factors*

The estimation results provide weak evidence to suggest that equity factors play a role in missing mortgage payments: the coefficient on the negative equity variable (i.e. gearing greater than 100 per cent) is positive and significant at the 5 per cent level, although the marginal effect of 6 percentage points (relative to having gearing less than 60 per cent) is only significant at the 10 per cent level. The imprecision of this estimate could partly reflect sample size issues, as only around 4 per cent of households in the estimation sample have negative housing equity. It could also suggest that equity factors *by themselves* are not particularly important in driving missed payments; double-trigger theories of default suggest that a household experiencing negative equity would also need to experience an ability-to-pay shock before missing a mortgage payment. However, the very small sample of households with both a high DSR and negative equity, or that are unemployed and have negative equity, makes it difficult to estimate precisely the effects of double-trigger type interactions.

### *3.3.3 Demographics and other life events*

Some studies of mortgage default find that demographic variables, such as age and education, and other life events, such as divorce and illness, play a significant role in mortgage payment behaviour. However, inclusion of variables related to these factors resulted in statistically insignificant effects and their inclusion in the model had little effect on the estimated marginal effects of the other variables. This is also the case when including *changes* in some of these variables, such as marital status and the household head's self-assessed health. Consequently, we have chosen to exclude these variables from our preferred model. The insignificance of these variables, particularly marital status and health, may reflect the wording of the missed payments question; the question asks whether the household had missed a payment due to *financial* difficulties. It is plausible that respondents may differentiate between missing payments due to purely financial difficulties and due to other problems, such as relationship breakdown or illness.

## 4. Conclusion

Our loan-level analysis suggests that loans with high loan-to-valuation ratios (above 90 per cent) are more likely to enter arrears, while loans that are repaid relatively quickly are less likely to enter arrears. Together, these results reinforce the importance of supervisors carefully monitoring changes in lending standards that affect the loan-to-valuation ratio of loans at origination and rates of principal repayment thereafter.

Although interest-only and fixed-rate loans appear less likely to enter arrears, the fact that these loans tend to be repaid relatively slowly (particularly interest-only loans) means that increases in these types of lending can represent an increase in risk. Additionally, low-doc loans appear more likely to enter arrears than other types of loans, even after controlling for whether the borrower was self-employed. This suggests that lenders should maintain sound income documentation and verification policies, and that supervisors should continue to monitor developments in the low-doc lending space.

Borrowers with relatively high mortgage interest rates have a higher probability of entering arrears, even after controlling for the estimated minimum mortgage repayment, which is consistent with riskier borrowers being charged higher interest rates to compensate for their higher risk. We caution, however, that the loan-level results are affected by data limitations, such as a lack of information on borrower income, wealth and labour force status, and a relatively small sample of banks.

Complementary analysis using household-level data suggests that having a high debt-servicing ratio (above 50 per cent) significantly increases the probability of missing a mortgage payment. This highlights the importance of borrowers not overextending themselves by taking out loans of a size that will be difficult to comfortably service. Additionally, it reinforces the importance of lenders maintaining sound debt-serviceability and income-verification policies.

Having previously missed a mortgage payment is also found to be a significant predictor of subsequently missing another mortgage payment. This highlights the heightened risk associated with lending to borrowers with a history of missing payments, and supports the practice of lenders using information on previous debt payment behaviour (such as credit scores) in their credit assessment processes.

Overall, our results reinforce the importance of supervisors carefully monitoring changes in lending standards, as well as the importance of borrowers exercising prudence when taking on mortgage debt.

# Appendix A: Literature Review

**Table A1: Recent Studies of Home Mortgage Default**

| Study | Equity | Ability to pay (ATP) | Equity or ATP? | Methodology | Unit of analysis | Country |
|---|---|---|---|---|---|---|
| Fuster and Willen (2013) | LVR; Gearing | Interest rate; Unemployment | Both | Competing risks | Loan | US |
| Palmer (2013) | Gearing | DSR | Equity | Discrete-time hazard | Loan | US |
| Krainer and Laderman (2011) | Gearing | Interest rate; Unemployment | ATP | Competing risks | Loan | US |
| Bhutta *et al* (2010) | Gearing | Unemployment | Both | Discrete-time hazard | Loan | US |
| Elul *et al* (2010) | LVR; Gearing | Unemployment | Both | Dynamic logit | Loan | US |
| Bajari, Chu and Park (2008) | LVR; Gearing | Mortgage payments | Both | Competing risks | Loan | US |
| Gerardi *et al* (2008) | Gearing | Interest rate | Equity | Probit | Loan | US |
| Foote, Gerardi and Willen (2008) | LVR | Unemployment; Household income | Both | Cox proportional hazard | Loan | US |
| Deng *et al* (2000) | LVR; Gearing | Unemployment | Both | Competing risks | Loan | US |
| Gerardi *et al* (2013) | Gearing | Unemployment; Liquid assets | Both | Probit and logit | Household | US |
| May and Tudela (2005) | Gearing | DSR; Unemployment | ATP | Random effects probit | Household | UK |
| Lydon and McCarthy (2013) | Gearing | DSR; Unemployment | ATP | Panel regressions | Regional | Ireland |
| Aron and Muellbauer (2010) | Housing debt to equity ratio | DSR; Unemployment | Both | Error correction model | Aggregate | UK |
| Whitely, Windram and Cox (2004) | Net housing wealth ratio | DSR; Unemployment | ATP | Error correction model | Aggregate | UK |

Note: LVR denotes LVR at origination

# Appendix B: Hedonic Dwelling Price Adjustment

The hedonic adjustment method generates estimates of the mean price of dwellings sold for each postcode $j$ and month $t$ conditional on the characteristics of dwellings sold. Following Hansen (2009) and Windsor *et al* (2014), each capital city is treated as a separate market with hedonic models estimated independently for Sydney, Melbourne and Brisbane. The postcode-time dummy hedonic adjustment model is given by Equation B1:

$$\ln(S_{ijt}) = \mathbf{x}'_{it}\beta + \sum_{t=1}^{T}\sum_{j=1}^{J}\lambda_{jt}D_{ijt} + \varepsilon_{ijt} \tag{B1}$$

where $S_{ijt}$ is the sale price of dwelling $i$ in postcode $j$ and month $t$. The vector of explanatory variables, $\mathbf{x}_{it}$, contains dwelling characteristics, including the (log) area in square metres, and dummy variables for the number of bedrooms and the sales mechanism.[16,17] All variables are interacted with the property type, allowing their effects to differ between houses and units. The dummy variable $D_{ijt}$ is equal to one if dwelling $i$ is sold in postcode $j$ and month $t$ and is equal to zero otherwise. From the coefficients on these dummy variables ($\lambda_{jt}$), we calculate the average value of dwellings sold within a postcode in a given month after controlling for observable characteristics. The results of estimating Equation B1 separately for Sydney, Melbourne and Brisbane are shown in Table B1.

16 We do not include the number of bathrooms as an explanatory variable, as this information is largely missing before 2005.

17 Genesove and Hansen (2014) find evidence to suggest that average prices of dwellings sold at auction incorporate information about the underlying trend in prices more quickly than average prices of dwellings sold via private treaty. Inclusion of the sales mechanism as an explanatory variable may also control for other systematic differences in the characteristics of dwellings sold via auction (relative to those sold via private treaty) that are unobserved or otherwise omitted from the model.

**Table B1: Hedonic Model Coefficient Estimates for Logged Sale Price**

| Characteristic | City | | |
|---|---|---|---|
| | Sydney | Melbourne | Brisbane |
| Unit | $-0.72^{***}$ | $0.23^{**}$ | $0.25^{***}$ |
| Bedrooms × property type | | | |
| 2 beds × house | $0.26^{***}$ | $0.47^{***}$ | $0.34^{***}$ |
| 2 beds × unit | $0.38^{***}$ | $0.44^{***}$ | $0.37^{***}$ |
| 3 beds × house | $0.46^{***}$ | $0.69^{***}$ | $0.50^{***}$ |
| 3 beds × unit | $0.79^{***}$ | $0.77^{***}$ | $0.74^{***}$ |
| 4 beds × house | $0.72^{***}$ | $0.91^{***}$ | $0.73^{***}$ |
| 4 beds × unit | $1.19^{***}$ | $1.00^{***}$ | $1.00^{***}$ |
| 5 beds × house | $0.90^{***}$ | $1.08^{***}$ | $0.91^{***}$ |
| 5 beds × unit | $1.25^{***}$ | $0.99^{***}$ | $0.97^{***}$ |
| 6 beds × house | $0.99^{***}$ | $1.16^{***}$ | $0.98^{***}$ |
| 6 beds × unit | $1.41^{***}$ | $1.12^{***}$ | $0.93^{***}$ |
| 7 beds × house | $1.05^{***}$ | $1.10^{***}$ | $1.03^{***}$ |
| 7 beds × unit | $1.63^{***}$ | $0.73^{***}$ | na |
| Sales mechanism × property type | | | |
| Private treaty × house | $-0.06^{***}$ | $-0.06^{***}$ | $-0.10^{***}$ |
| Private treaty × unit | 0.00 | $0.05^{***}$ | 0.02 |
| Area × property type | | | |
| Area × house | $0.02^{**}$ | $0.07^{***}$ | $0.12^{***}$ |
| Area × unit | $0.04^{***}$ | $-0.02$ | 0.01 |
| Constant | $12.69^{***}$ | $11.81^{***}$ | $11.65^{***}$ |
| Number of observations | 601 958 | 669 398 | 219 120 |
| $R^2$ | 0.81 | 0.82 | 0.72 |

Notes: Coefficients of postcode-time dummies omitted; ***, ** and * indicate statistical significance at the 1, 5 and 10 per cent level, respectively; standard errors clustered at the postcode level

Sources: Authors' calculations; APM

## Appendix C: The Competing Risks Regression Model

Let $j$ index events, with $j = a$ corresponding to a loan falling into arrears (i.e. the event of interest) and $j = p$ corresponding to the loan being paid down in full (i.e. the competing risk). The subhazard for entering arrears is then:

$$h_a(t) = \lim_{\delta \to 0} \left\{ \frac{\Pr(t \leq T < t + \delta, j = a | T \geq t \cup (T < t \cap j = p))}{\delta} \right\}. \tag{C1}$$

The subhazard is similar to the hazard function from standard duration analysis, except that it keeps loans that have been paid down in full 'at risk' of entering arrears so that they can be appropriately counted as having zero probability of entering arrears. While the subhazard is somewhat difficult to interpret (in fact, Fine and Gray (1999) describe the 'risk set' associated with this subhazard as being 'unnatural'), it provides a convenient way to model the CIF, which gives the probability of a loan falling into arrears before time $t$ (Rodriguez 2012):

$$\text{CIF}_a(t) = \Pr(T \leq t, j = a) = 1 - \exp\left( - \int_0^t h_a(s) ds \right). \tag{C2}$$

One of the advantages of using this competing risks regression approach is that the covariates included in the model will have effects on the subhazard and the CIF that are in the same direction. This is not necessarily the case when modelling the event of interest using a standard duration model, such as a Cox proportional hazard model, which treats the competing risk as a censoring event; the effect of a covariate on the CIF will depend on how it affects the incidence of the event of interest, but also on how it affects the incidence of the competing risk.

The model is estimated using the stcrreg command in Stata 13. Parameter estimates are obtained by maximising a log-pseudolikelihood function, and standard errors are clustered by loan.

## Appendix D: Discrete-time Duration Model with Unobserved Heterogeneity

One assumption underlying the competing risks model is that the time between loan origination and a loan entering arrears is continuous – that is, that we observe the *exact* time that a loan enters arrears. However, in reality, we only observe the loans at the end of each month and, consequently, the data are discrete. This suggests that a discrete-time duration model may be more appropriate. Another aspect of our competing risks model, which is discussed in Section 2.2, is that we do not allow loans in arrears to cure or enter arrears multiple times. However, if we were to relax this assumption and use information on loans that enter arrears multiple times, we could potentially control for unobserved heterogeneity across loans. We consider these alternative model features as a robustness check of the results from our competing risks regression.

It can be shown that a continuous-time duration model with a proportional hazards representation can be expressed as a complementary log-log regression if the observations are discretised (e.g. Kaplan 2012). The model, which includes a normally distributed random effect to allow for unobserved heterogeneity, can be written as:

$$\log\{-\log[1-h_t(\mathbf{z}_{it})]\} = \mathbf{z}_{it}'\gamma + \theta(t) + \alpha_i, \tag{D1}$$

where $h_t(\mathbf{z}_{it})$ is the hazard of entering arrears during month $t$, $\theta(t)$ is the integrated baseline hazard during month $t$, which we proxy for using a polynomial in loan age, and $\alpha_i$ is a random effect for loan $i$. We estimate this model using the xtcloglog function in Stata 13 and report exponentiated coefficients, which can be interpreted in a similar fashion as the subhazard ratios presented in Section 2.3. For the purposes of this analysis, we abstract from the presence of the competing risk of full payment.

While there are some differences between the results from this model and the results from our competing risks model, overall the key results are qualitatively similar (Table D1). The hazard of entering arrears tends to increase with the LVR at origination, although, like in the competing risks model, it falls for loans with an LVR at origination greater than 100 per cent. Increases in amortisation since origination decrease the hazard of entering arrears. However, in contrast to the results from the competing risks model, increases in dwelling prices significantly decrease the hazard of entering arrears. Investor loans are also significantly less likely to enter arrears than owner-occupier loans.

**Table D1: Housing Loan Arrears – Complementary Log-log Model**

| Explanatory variable | exp($\gamma$) | Explanatory variable | exp($\gamma$) |
|---|---|---|---|
| Amortisation | 0.98*** | Investor | 0.78*** |
| LVR at origination | | Loan purpose | |
| $60 \leq$ LVR$< 80$ | 2.35*** | Home improvement | 0.42*** |
| $80 \leq$ LVR$< 90$ | 2.80*** | Refinance | 2.08*** |
| $90 \leq$ LVR$< 100$ | 5.97*** | Other | 1.15 |
| LVR $\geq 100$ | 5.04*** | Local unemployment rate | 1.02 |
| Dwelling price growth | 0.98*** | Low doc | 3.49*** |
| Fixed rate | 0.36*** | Minimum required payment | 1.07*** |
| Interest only | 0.61*** | Self-employed | 1.36** |
| Interest rate | 1.42*** | | |
| Number of observations | | 1 624 132 | |
| Number of loans | | 63 526 | |

Notes: ***, ** and * denote statistical significance at the 1, 5 and 10 per cent levels, respectively; standard errors are clustered by loan; model includes a random effect for each loan; 'amortisation' is the percentage decrease in the loan balance since origination; 'dwelling price growth' is the cumulative percentage growth of dwelling prices since origination; 'minimum required payment' is measured in thousands of dollars

Sources: ABS; APM; Authors' calculations; MARQ Services; RP Data-Rismark

Although this model provides a potentially useful robustness check for the results from our competing risks model, it is worth noting that the model has two key shortcomings:

1. It ignores the presence of the competing risk of full payment, potentially resulting in misleading estimates.

2. It assumes that the loan-level random effect is uncorrelated with the take-up of particular loan features. This seems unlikely if the random effect captures the time-invariant component of omitted factors such as income and financial sophistication.

# Appendix E: Household-level Data

In the 2006 and 2010 HILDA Survey's wealth modules, respondents were asked, 'Do you currently own or have you ever owned a residential property?' If a respondent answered yes, they were then asked, 'During the last 12 months have you (or your household) been unable to meet a mortgage payment by the due date on any housing or property loan because of financial difficulties?' Respondents could answer 'yes', 'no', 'did not own/have a property loan during last 12 months' or 'don't know'. Thus, only respondents that had a property loan in the previous 12 months should have responded 'yes' or 'no' to the second question. However, around half of households that did not have any property loans (including mortgages for non-owner-occupied property) at the time of the survey responded 'no'. Although some of these respondents may have had a property loan during the previous 12 months but repaid it before the interview date, the numbers appear too large to solely reflect this. As a consequence, the share of households that missed a mortgage payment could be understated. To address this problem, in our descriptive analysis we restrict the sample to households that had a mortgage on their residence at the time of the survey or at the time of the previous survey.

This issue should not affect the dependent variable in our probit model, as the inclusion of explanatory variables related to the DSR and LVR implicitly restricts the estimation sample to those households that had owner-occupier housing debt in 2009. However, this issue makes it necessary to recode the missed payments variable from the 2006 survey before including it as a categorical explanatory variable. Households that responded 'no' to missing a mortgage payment in 2006 but did not appear to have a property loan in the 2005 or 2006 surveys are treated as non-responders. The non-response category also includes households that: refused to answer the question; reported not knowing the answer to the question; were part of the survey sample in 2006 but did not respond; or were not part of the survey sample in 2006. Recoding the missed payments variable in this way ensures that the estimation sample is not unnecessarily restricted to households that responded to both 2006 and 2010 surveys. It also avoids the measurement error that would occur when classifying households that did not have a property loan in 2006 as having not missed a payment.

## References

**Aron J and J Muellbauer (2010),** 'Modelling and Forecasting UK Mortgage Arrears and Possessions', Spatial Economics Research Centre, SERC Discussion Paper No 52.

**Bajari P, CS Chu and M Park (2008),** 'An Empirical Model of Subprime Mortgage Default From 2000 to 2007', NBER Working Paper No 14625.

**Berry M, T Dalton and A Nelson (2010),** *Mortgage Default in Australia: Nature, Causes and Social and Economic Impacts*, AHURI Final Report No 145, Australian Housing and Urban Research Institute, Melbourne.

**Bhutta N, J Dokko and H Shan (2010),** 'The Depth of Negative Equity and Mortgage Default Decisions', Board of Governors of the Federal Reserve System Finance and Economics Discussion Series No 2010-35.

**Ciochetti BA, Y Deng, B Gao and R Yao (2002),** 'The Termination of Commercial Mortgage Contracts through Prepayment and Default: A Proportional Hazard Approach with Competing Risks', *Real Estate Economics*, 30(4), pp 595–633.

**Deng Y, JM Quigley and R Van Order (2000),** 'Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options', *Econometrica*, 68(2), pp 275–307.

**Elul R (2006),** 'Residential Mortgage Default', Federal Reserve Bank of Philadelphia *Business Review*, Third Quarter, pp 21–30.

**Elul R, NS Souleles, S Chomsisengphet, D Glennon and R Hunt (2010),** 'What "Triggers" Mortgage Default?', *The American Economic Review*, 100(2), pp 490–494.

**Fine JP and RJ Gray (1999),** 'A Proportional Hazards Model for the Subdistribution of a Competing Risk', *Journal of the American Statistical Association*, 94(446), pp 496–509.

**Foote CL, K Gerardi and PS Willen (2008),** 'Negative Equity and Foreclosure: Theory and Evidence', *Journal of Urban Economics*, 64(2), pp 234–245.

**Fuster A and PS Willen (2013),** 'Payment Size, Negative Equity, and Mortgage Default', NBER Working Paper No 19345.

**Genesove D and J Hansen (2014),** 'Predicting Dwelling Prices with Consideration of the Sales Mechanism', RBA Research Discussion Paper No 2014-09.

**Gerardi K, KF Herkenhoff, LE Ohanian and PS Willen (2013),** 'Unemployment, Negative Equity, and Strategic Default', Federal Reserve Bank of Atlanta Working Paper No 2013-4.

**Gerardi K, A Lehnert, SM Sherlund and PS Willen (2008),** 'Making Sense of the Subprime Crisis', *Brookings Papers on Economic Activity*, Fall, pp 69–159.

**Ghent AC and M Kudlyak (2010),** 'Recourse and Residential Mortgage Default: Theory and Evidence from U.S. States', Federal Reserve Bank of Richmond Working Paper No 09-10R.

**Gyourko J and J Tracy (2013),** 'Unemloyment and Unobserved Credit Risk in the FHA Single Family Mortgage Insurance Fund', NBER Working Paper No 18880.

**Hansen J (2009),** 'Australian House Prices: A Comparison of Hedonic and Repeat-Sales Measures', *Economic Record*, 85(269), pp 132–145.

**Kaplan G (2012),** 'Moving Back Home: Insurance against Labor Market Risk', *Journal of Political Economy*, 120(3), pp 446–512.

**Kau JB, DC Keenan and T Kim (1994),** 'Default Probabilities for Mortgages', *Journal of Urban Economics*, 35(3), pp 278–296.

**Krainer J and E Laderman (2011),** 'Prepayment and Delinquency in the Mortgage Crisis Period', Federal Reserve Bank of San Francisco Working Paper No 2011-25.

**Lydon R and Y McCarthy (2013),** 'What Lies Beneath? Understanding Recent Trends in Irish Mortgage Arrears', *The Economic and Social Review*, 44(1), pp 117–150.

**May O and M Tudela (2005),** 'When is Mortgage Indebtedness a Financial Burden to British Households? A Dynamic Probit Approach', Bank of England Working Paper No 277.

**Palmer C (2013),** 'Why Did so Many Subprime Borrowers Default during the Crisis: Loose Credit or Plummeting Prices?', Massachusetts Institute of Technology, Department of Economics, Job Market Paper.

**Rodriguez G (2012)**, 'Cumulative Incidence', Pop509: Survival Analysis Course Materials, Princeton University, accessed 4 July 2014. Available at <http://data.princeton.edu/pop509/cumulativeIncidence.pdf>.

**Watkins JGT, AL Vasnev and R Gerlach (2014),** 'Multiple Event Incidence and Duration Analysis for Credit Data Incorporating Non-Stochastic Loan Maturity', *Journal of Applied Econometrics*, 29(4), pp 627–648.

**Whitley J, R Windram and P Cox (2004),** 'An Empirical Model of Household Arrears', Bank of England Working Paper No 214.

**Windsor C, G La Cava and J Hansen (2014),** 'Home Price Beliefs in Australia', RBA Research Discussion Paper No 2014-04.

# Copyright and Disclaimer Notices

## APM Disclaimer

The Australian property price data used in this publication are sourced from Australian Property Monitors Pty Limited ACN 061 438 006 of level 5, 1 Darling Island Road Pyrmont NSW 2009 (P: 1 800 817 616).

In providing these data, Australian Property Monitors relies upon information supplied by a number of external sources (including the governmental authorities referred to below). These data are supplied on the basis that while Australian Property Monitors believes all the information provided will be correct at the time of publication, it does not warrant its accuracy or completeness and to the full extent allowed by law excludes liability in contract, tort or otherwise, for any loss or damage sustained by you, or by any other person or body corporate arising from or in connection with the supply or use of the whole or any part of the information in this publication through any cause whatsoever and limits any liability it may have to the amount paid to the Publisher for the supply of such information.

## New South Wales Land and Property Information

Contains property sales information provided under licence from the Department of Finance and Services, Land and Property Information.

## State of Victoria

The State of Victoria owns the copyright in the Property Sales Data and reproduction of that data in any way without the consent of the State of Victoria will constitute a breach of the Copyright Act 1968 (Cth). The State of Victoria does not warrant the accuracy or completeness of the Property Sales Data and any person using or relying upon such information does so on the basis that the State of Victoria accepts no responsibility or liability whatsoever for any errors, faults, defects or omissions in the information supplied.

## State of Queensland

© State of Queensland (Department of Natural Resources and Mines) 2012. In consideration of the State permitting use of this data you acknowledge and agree that the State gives no warranty in relation to the data (including accuracy, reliability, completeness, currency or suitability) and accepts no liability (including without limitation, liability in negligence) for any loss, damage or costs (including consequential damage) relating to any use of the data. Data must not be used for direct marketing or be used in breach of the privacy laws.

## HILDA

The following Disclaimer applies to data obtained from the HILDA Survey and reported in this RDP.

**Disclaimer**

The Household, Income and Labour Dynamics in Australia (HILDA) Survey was initiated and is funded by the Australian Government Department of Social Services (DSS), and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views based on these data should not be attributed to either DSS or the Melbourne Institute.

RESEARCH DISCUSSION PAPERS

These papers can be downloaded from the Bank's website or a hard copy may be obtained by writing to:


Mail Room Supervisor
Information Department
Reserve Bank of Australia
GPO Box 3947
SYDNEY NSW 2001

Enquiries:

Phone: +61 2 9551 9830
Facsimile: +61 2 9551 8033
Email: rbainfo@rba.gov.au
Website: http://www.rba.gov.au


| | | |
|---|---|---|
| 2014-04 | Home Price Beliefs in Australia | Callan Windsor<br>Gianni La Cava<br>James Hansen |
| 2014-05 | The Changing Way We Pay: Trends in Consumer Payments | Crystal Ossolinski<br>Tai Lam<br>David Emery |
| 2014-06 | Is Housing Overvalued? | Ryan Fox<br>Peter Tulip |
| 2014-07 | International Trade Costs, Global Supply Chains and Value-added Trade in Australia | Gerard Kelly<br>Gianni La Cava |
| 2014-08 | The Effect of the Mining Boom on the Australian Economy | Peter Downes<br>Kevin Hanslow<br>Peter Tulip |
| 2014-09 | Predicting Dwelling Prices with Consideration of the Sales Mechanism | David Genesove<br>James Hansen |
| 2014-10 | Financial Reform in Australia and China | Alexander Ballantyne<br>Jonathan Hambur<br>Ivan Roberts<br>Michelle Wright |
| 2014-11 | Exchange Rate Movements and the Australian Economy | Josef Manalo<br>Dilhan Perera<br>Daniel Rees |
| 2014-12 | A State-space Approach to Australian GDP Measurement | Daniel Rees<br>David Lancaster<br>Richard Finlay |

RESERVE BANK OF AUSTRALIA