18/RT/14

Páipéar Taighde Teicniúil
Research Technical Paper

Modelling default transitions in the UK mortgage market

Fergal McCann

Banc Ceannais na hÉireann
Central Bank of Ireland
Eurosystem

Modelling default transitions in the UK mortgage market

Fergal McCann[a*]
[a] Financial Stability Division,
Central Bank of Ireland

November 2014

Abstract

Using a large panel data set on the population of UK mortgage loans by Irish-headquartered banks, this paper presents a transitions-based model of mortgage default. The estimation departs from cross-sectional methods typically used in mortgage default models, in that the transition both into and out of mortgage default is predicted. Housing equity, regional unemployment and loan interest rates are found to impact the probability of transition into default, while housing equity, interest rates and the time spent in default all significantly affect the probability that a loan transitions out of default ("cures"). The latter finding is indicative of a hysteresis effect in mortgage markets, whereby the longer a mortgage spends in default, the less likely it is that the obligor will begin repayment. This finding provides important impetus for mortgage modification programmes which aim to tackle mortgage arrears at as early a stage as possible. In an extension, the default transitions of owner-occupiers are shown to be far less sensitive to changes in housing equity than those of Buy-to-Let investors. This finding suggests that instances of "strategic default" are uncommon among UK homeowners, but investors may well exercise the "put option" implicit in their contract when house price falls leave them "out of the money".
Keywords: Mortgages, default, credit risk, Markov multi-state model.
JEL: D14; G21.

[*]E-mail: fergal.mccann@centralbank.ie. I thank, in alphabetical order, Edward Gaffney, Robert Kelly, Reamonn Lydon, Gerard O'Reilly and seminar participants in the Central Bank of Ireland Financial Stability Division and the International Finance and Banking Society (IFABS), Lisbon 2014, for helpful discussions. The views presented in this paper are those of the authors alone and do not represent the official views of the Central Bank of Ireland or the ESCB. Any remaining errors are my own.

Non Technical Summary

This paper presents one of a very small number of loan-level models of UK mortgage default. The literature on the determinants of mortgage default has generally been characterised by a dichotomy of views. The purely financial view of mortgage default treats a mortgage in much the same way as any investment, and predicts that when borrowers enter negative equity, i.e. when the value of their outstanding loan exceeds the value of their asset, they will default. The "double-trigger hypothesis", provides the alternative that defaults are explained by a combination of negative equity and affordability shocks.

Using loan-level data on the population of UK mortgages at three Irish-headquartered banks between 2009 and 2013, I model the role of a range of loan-level covariates on the probability of transition both into and out of mortgage default. Previous research has typically focused on models that predict whether or not a loan is in default at a given point in time. The empirical method employed here allows explanatory factors to predict how loans enter default, but also to predict how they impact the probability that a defaulted loan will "cure".

In including the time spent in default (TSD) in the "cure" model, I am able to identify a hysteresis effect of loan delinquency whereby loans that have spent a longer time in default are less likely to transition back to repayment. This hysteresis effect, also identified in the Irish mortgage market by Kelly and O'Malley (2014), is of crucial policy relevance, as it provides strong motivation for programmes which target those in the early stages of mortgage arrears in order to alleviate payment difficulties immediately.

I find support for the hypothesis that both equity (measured by the Loan to Value ratio, LTV) and affordability (measured by regional unemployment and loan-level interest rates) considerations are important in predicting loan default. The results suggest that, in terms of economic magnitude, equity and affordability play a similar and important role in explaining mortgage default transitions, with equity having a marginally larger quantitative effect during the period under study.

I identify differential sensitivities to LTV, interest rates and regional unemployment among owner-occupiers and Buy to Let (BTL) borrowers, with the former being much less likely to enter default in response to changes in these fundamentals. The lower sensitivity of owner-occupiers to LTV suggests that these borrowers are more likely to deploy all available resources to remain current on their mortgages, and that "strategic" or equity-driven default is not prevalent among UK owner-occupiers.

1 Introduction

The central role of the mortgage market in precipitating the global financial crisis has led to a surge in research and policy interest in the area. One particularly pressing question surrounds the conditions under which mortgage borrowers enter default. A purely financial approach to explaining mortgage default adopts an option pricing framework, whereby a borrower "strategically" defaults when entering negative equity (a position where her house value falls below the value of her outstanding debt).[1] While empirical work by Goodman et al. (2010) and Mayer et al. (2009) has pointed to the primacy of negative equity in explaining mortgage default, this view has been challenged by numerous recent studies. Guiso et al. (2013) and Bhutta et al. (2010) have shown using differing methods that equity shortfalls must surpass 50 per cent before strategic default becomes prevalent. Bhutta et al. (2010) also find support for the "double-trigger hypothesis", that defaults are explained by a combination of negative equity and income or employment shocks. Bajari et al. (2008), Gerardi et al. (2013) and Elul et al. (2010), Gyourko and Tracy (2014) for the US, and Lydon and McCarthy (2013) and McCarthy (2014) for Ireland provide further support for the "double-trigger" explanation.[2]

Using a Markov multi-state approach on the population of UK mortgages at three Irish-headquartered banks between 2009 and 2013, I model the role of a range of loan-level covariates on the probability of transition both into and out of mortgage default.[3] In modelling transitions in both directions, I depart from the methodology applied in the literature mentioned above, which generally utilizes discrete choice methods to model the probability that a loan is in default *at a given point in time*, and follow the approach of Kelly (2011) and Kelly and O'Malley (2014). Using this paper's approach, the impact of equity, regional unemployment, interest rates and other factors on loan default can be decomposed into their effect on the probability of loans *entering* default (PD) and the probability of loans *exiting* default ($PCure$).

In including the time spent in default (TSD) in the $PCure$ equation, I am also able to identify a hysteresis effect of loan delinquency. Loans that have been in default for one year have a 50 per cent chance of curing in the baseline model, with this probability falling to 25 per cent for loans in default for 2 years, and to close to zero as TSD passes over 3 years. The identification of this effect is only possible using a model where loans can move out of, as well as into, default. This hysteresis effect, also identified in the Irish mortgage market by Kelly and O'Malley (2014), is of crucial policy relevance,

[1] See Vandell (1995) for a review of early work in this area.

[2] Foote et al. (2008) also cast doubt on the "strategic default" hypothesis by showing that only 10 per cent of negative equity borrowers in Massachusetts in the 1990s defaulted on their mortgages. They motivate this finding using a forward-looking model where it is optimal for negative equity borrowers to remain in their homes as long as there is an expectation of future house price growth.

[3] "Default" is defined in this analysis where the obligor is more than 90 days past due on loans managed in the mortgage book.

as it provides strong motivation for programmes which target those in the early stages of mortgage arrears in order to alleviate payment difficulties immediately.

I find support for the hypothesis that both equity (measured by the Loan to Value ratio) and affordability (measured by regional unemployment and loan-level interest rates) considerations are important in predicting loan default. In explaining transitions into default, a one-standard deviation increase in LTV from its 2009-2013 mean (equating to a move from an LTV of 80 to an LTV of 114) leads to an increase in PD from 1.8 to 2.8 per cent, with the analogous effect for interest rate and unemployment being 1.9 to 2.5 per cent and 1.85 to 2, respectively. In explaining transitions out of default, a one-standard-deviation increase in LTV from its mean leads to a fall in $PCure$ from 56 to 46 per cent, while the analogous effect for interest rates is a decrease from 54 to 48 per cent. These results suggest that, in terms of economic magnitude, equity and affordability play a similar and important role in explaining mortgage default transitions, with equity having a marginally larger quantitative effect during the period under study.

I identify differential sensitivities to LTV, interest rates and regional unemployment among owner-occupiers and Buy to Let (BTL) borrowers, with the former being much less likely to enter default in response to changes in these fundamentals. The lower sensitivity of owner-occupiers to LTV suggests that these borrowers are more likely to deploy all available resources to remain current on their mortgages, and that "strategic" or equity-driven default is not prevalent among UK owner-occupiers. Owner-occupiers' $PCure$ is also much less sensitive to LTV, suggesting that, as equity conditions improve, these borrowers are less readily able to move out of default. Conversely, BTL borrowers move out of default more swiftly in response to improvements in their equity position, suggesting that their default position was more likely to have been "strategic" to begin with. This method of measuring the prevalence of strategic default is only possible in a transitions-based, multi-state framework.

From a financial modelling perspective, the usage of a transitions-based, as opposed to point-in-time discrete choice model approach, brings with it significant advantages. When applying the coefficients of a logit or probit model to macroeconomic scenario inputs for the purposes of estimating likely future losses, it is only possible to arrive at a *lifetime* $\hat{P}D$ for each loan. It is not possible to start at a stock of defaulted loans at $t = 0$, and estimate the probability that loans move both into and out of that stock at $(t = 1, 2, ..n)$. McDonald et al. (2010) highlight a number of these advantages when applying survival analysis to mortgage PD modelling. However, the survival analysis approach can only model transition probabilities in one direction: from performing to default. The advantage of the Markov Multi-State Model (MSM) approach is that, with knowledge of the starting stock of defaulted loans, the coefficients from the model can be applied to any scenario X values to estimate $\hat{P}D$ and $PC\hat{u}re$, which allows the movement into and out of default to be calculated over all t. This results in a significantly more flexible PD-EAD-LGD[4] framework for estimating likely future losses. An application of the *PD* framework here to the Central Bank of Ireland's Loan Loss Forecasting framework is detailed in Gaffney et al. (2014).

As a large wealthy country with a highly developed financial sector which underwent substantial turmoil in the years following 2007, the UK is notable for the dearth of micro-level research on the causes of mortgage default. This paper provides the first research to use loan-level data provided by a range of financial institutions in modelling UK mortgage defaults. Lanot and Leece (2014) use data from a securitization from a single US originator in the UK, focussing on the determinants of default and prepayment among sub-prime UK mortgages only. Their findings suggest that self-certification and unobserved heterogeneity in contract choices play an important role in determining default probabilities. Aron and Muellbauer (2010) and Aron and Muellbauer (2011) use aggregate data to model and forecast UK mortgage arrears and possessions, finding a role for debt service ratios, negative equity and unemployment.

Previous work by Gathergood (2009) has used British Household Panel Survey (BHPS) data which allows for the usage of a wealth of detailed borrower level proxies for repayment risk and affordability. However, house price data are self-reported by survey respondants, and mortgage default, as generally defined in banking by the Basel Committee is not observed; rather, a self-reported measure of mortgage repayment difficulty is used as a proxy for default. Boheim and Taylor (2000) also use the BHPS to show that both negative equity and income shocks explain mortgage difficulties in the 1990s in the UK.

The paper proceeds with a discussion of the loan level data (Section 2), summary statistics (Section 3), empirical results (Section 4) and a conclusion (Section 5).

2 Data

The data required to model loan transitions are submitted to the Central Bank of Ireland by financial institutions participating in the Financial Measures Programme (FMP), March 2011.[5] The data delivery process began in March 2011 with data sets pertaining to December 2010 provided. The majority of the fields in the data template relate to point-in-time information at the date of the data drop. However, importantly for the purposes of this study, information on arrears balances is reported for each month going back 12 months from the date of data drop. In the case of the December 2010

[4]Probability of Default, Exposure at Default, Loss Given Default. These model components generally combine to arrive at estimates of expected loss.

[5]The Financial Measures Programme was the capital and liquidity review of six domestic Irish banks, leading to the injection of state support into the banking sector as part of the EC-ECB-IMF package of support to the Irish state. `http://www.centralbank.ie/regulation/industry-sectors/credit-institutions/pages/financialmeasuresprogramme.aspx`

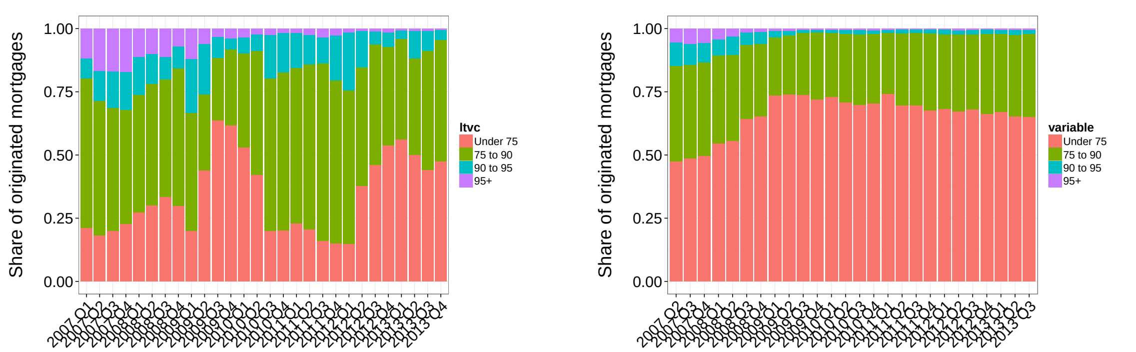
drop, this means that arrears history information for UK mortgage holders goes back to December 2009. The data has been received by the Central Bank of Ireland on an annual, and more recently semi-annual basis, with the most recent data drop being December 2013. The total mortgage lending volume accounted for by Irish-headquartered banks represents 2 per cent of the total UK mortgage market.

The lending patterns of the Irish banks are not necessarily perfectly representative of activity in the UK mortgage market as a whole. Figure 1 shows that, since 2007 (the date at which CML data on lending by LTV category begins), Irish banks have systematically lent at higher LTVs than the UK market in its totality. Since 2009, the volume of lending at above an LTV or 90 has been around 2 per cent in the UK market, while the percentage of Irish banks' mortgages in this category has been between 10 and 20 per cent. The pattern presented suggests that the loan book analysed here is originated with more relaxed credit standards than the population of UK mortgages.

Figure 1: Share of all originated loans per quarter falling into four LTV categories, 2007 to 2013

(a) Irish banks, loan level data

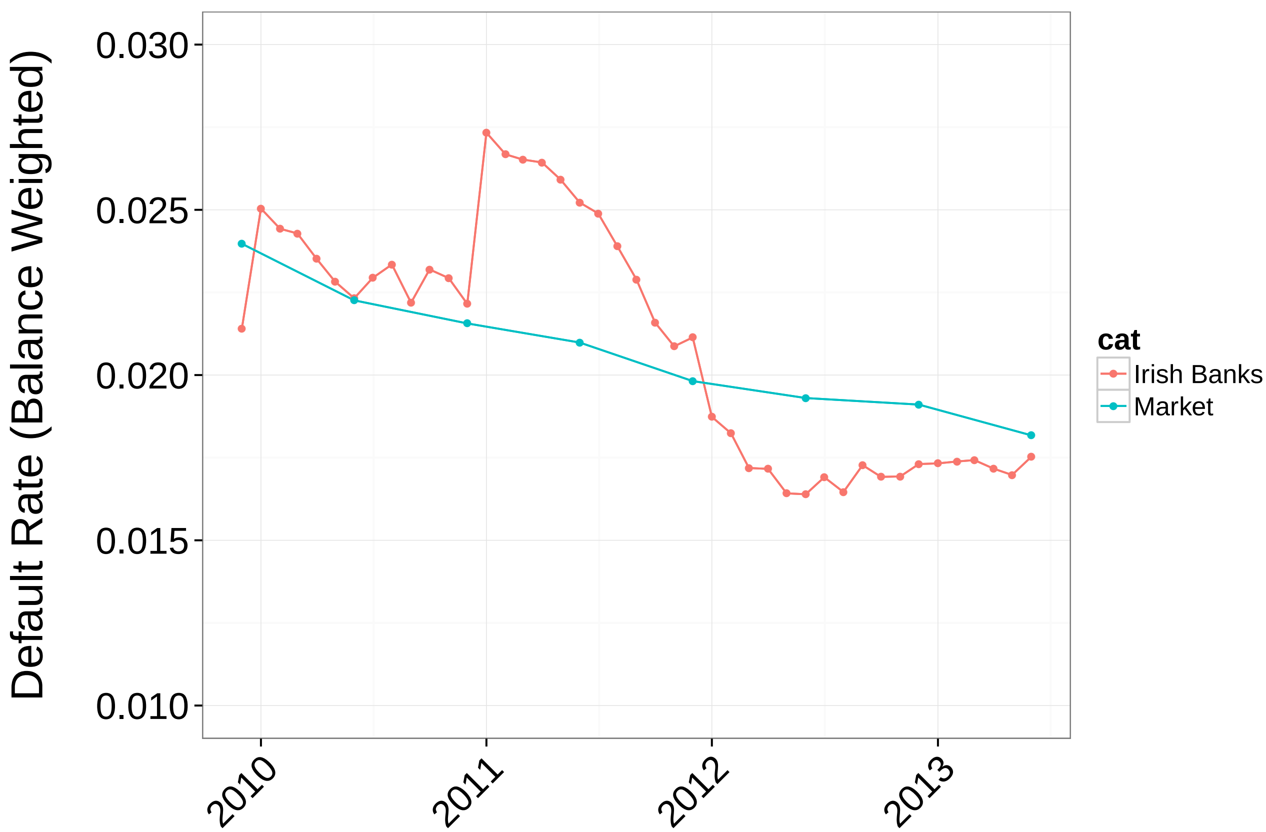
(b) Total UK mortgage market



Data sources: Central Bank of Ireland Loan Level Data; Council of Mortgage Lenders aggregate statistics.

The representativeness of the Irish lenders' mortgage books can also be ascertained by comparing default rates in our data set to those in the market. Figure 2 plots the default rate from the beginning of the LLD sample period on a monthly basis, to that from the CML aggregate series, which is reported on a six-monthly basis. The comparison reveals a common downward trend over the period, with start and end-of-period default rates within one fifth of one per cent of the total market numbers. This comparison provides a level of comfort regarding the generalisability of the results of this paper.

Figure 2: Default rates, 2009 to 2013



Data sources: Central Bank of Ireland Loan Level Data; Council of Mortgage Lenders aggregate statistics.

The combination of four data drops allows the construction of a 49-month panel of loan delinquency status spanning December 2009 - December 2013. Information such as Bank ID, geographic location, Interest Rate Type, Buy To Let flag and Interest Rate are taken from the December 2013 data set. Loans from the December 2013 data set are then linked back to each of the three previous data drops using a unique facility ID provided by each financial institution.

The model is estimated at the property, rather than the loan level due to the fact that, while data for December 2013, 2012 and 2011 are submitted at the loan level with property-level identifiers included, data for December 2010 was submitted at the property level only. To allow information on the months December 2009 - December 2010 to be used in the model, all subsequent data must therefore be aggregated to the property level. A number of crucial fields are reported at the loan level, and cannot be aggregated to the property level without a data-cleaning rule being applied. Table 1 outlines the cleaning rules that are applied in moving the relevant variables to the property level. This results in the collapsing of the data set from 291,630 loans to 244,611 properties. The frequency of loan count per property is documented in Table A.1.

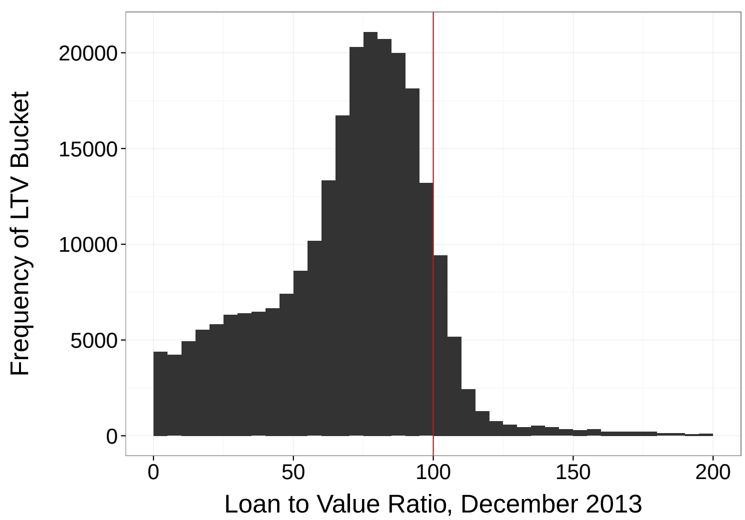
Table 1: Data-cleaning rules applied to transform key variables to property level from loan level

Variable	Rule
Arrears Balance	Summation
Outstanding Balance	Summation
Buy-to-Let Flag	Where multiple BTL flag types, the BTL flag associated with the largest share of the facility's overall outstanding balance.
Installment Amount	Summation
Interest Rate	Balance - weighted average
Interest Rate Type	Where multiple interest rate types, the interest rate type associated with the largest share of the facility's overall outstanding balance.
Maturity Date	Take the most distant (latest) maturity date. Remove observations with negative term remaining.
Origination Date	Take the most recent origination date to allow for potentially higher credit risk of equity releases. Remove observations with negative loan age.
Region	No conficts - already at property level
LTV	No conficts - already at property level. Cap at 300.

To calculate a monthly estimate of the loan to value ratio (LTV), facility balances are rolled backwards from the December 2013 balance with a manipulation of the standard amortization formula using information on term remaining (n) and monthly interest rate (i). This gives a monthly value for outstanding balance of the loan. These estimates are also frozen so that when a loan is in default, the balance does not grow as we move from t to $t-1$. Monthly estimates of a loan's property value are arrived at by indexing the December 2013 bank-provided valuation to each month to December 2009 using the Nationwide regional house price index. Figure A.1 plots the series. The plot shows that Northern Ireland is the only UK region to have undergone a boom and bust cycle in house prices similar to that in the Republic of Ireland. In no other region have house prices continued to fall between 2007 and 2013. For most other regions, the pattern replicated is that prices hit their post-crisis trough in early 2009 and have been stable since. The Outer South East region, and more markedly London, have experienced significant house price appreciation since the crisis, with the London average price now being higher than at any point since 2001.

The distribution of the Loan to Value Ratio (LTV) for all 244,340 properties in the data at December 2013 is plotted in Figure 3. Each bin in the histogram represents a range of 5 in the LTV distribution. The preponderance of positive equity in the UK mortgage market is evident from the histogram, with 90 per cent of the observations being in positive equity (LTV of less than 100 per cent). The majority of the mass of the distribution lies in a range between 60 and 90 per cent LTV.

Figure 3: Distribution of Loan to Value ratio, December 2013



Regional variation in LTVs is a crucial feature of the empirical model utilised in this paper. LTVs will move across time, both as a function of loan amortization and the house price cycle. However, cross-sectional variation is arguably as important a factor in predicting loan delinquency. The significant point-in-time variation in LTV distributions across the UK is highlighted in Figure 4. As of December 2013, there are regions in the UK where there are close to no properties in negative equity (London, Outer Metropolitan, Outer South East). Property values in these regions, in and surrounding the London area, have benefited from significant price appreciation driven by a combination of loose monetary policy and international investor demand since the 2008 financial crisis. On the other hand, in regions such as Northern Ireland, Yorkshire and Humberside, the North and North West, significant portions (27, 20, 17, 15 per cent, respectively) of the properties in the data set remain in negative equity. The larger left tail in the Northern Ireland LTV distribution is explained by that fact that Irish banks' lending to the Northern Irish market is longer established than lending to the rest of the UK, meaning that there is a mass of Northern Irish loans that are much further along the amortization schedule than loans in other regions.

Regional unemployment data are plotted in Figure A.2.[6] The lowest unemployment rates are found in the South East (5.8 per cent at June 2013), South West (6.2) and East (6.7) regions of the UK. The area with the highest unemployment rate throughout the three years of the sample is the North East (10.4 per cent), with the West Midlands (9.8) and Yorkshire & Humberside (8.9) also having relatively high unemployment rates at the sample end-point, December 2013.

All facilities in any period t can take on two statuses: performing and default. Facilities are categorised as defaulted when the following criterion is met in month i:

$$Default \equiv \frac{ArrearsBalance_{it}}{Instalment_{it}} >= 3 \tag{1}$$

This criterion assures that the definition of Default matches the 90 days past due definition of the Basel committee. In order to calculate default status as accurately as possible, the instalment amount that is used is updated at each data drop date. For example,

$$Instalment_{i,Jan2013-Dec2013} = Instalment_{Dec2013} \tag{2}$$

$$Instalment_{i,Jan2012-Dec2012} = Instalment_{Dec2012} \tag{3}$$

$$Instalment_{i,Jan2011-Dec2011} = Instalment_{Dec2011} \tag{4}$$

$$Instalment_{i,Dec2009-Dec2010} = Instalment_{Dec2010} \tag{5}$$

This ensures that where instalment amounts change over time due to modifications to loan terms, these changes are captured at discrete points where point-in-time information on the instalment amount is available. Given that month-to-month instalment amounts are not available in the data set, the above approach represents the most accurate reflection of a borrower's monthly repayment amounts.

2.1 Stratified sampling to generate PD model data set

The sample size of the December 2013 UK mortgage data set is 244,340 properties, leading to a maximum possible sample size for empirical modelling of 11.97 million, in the case of a fully balanced 49-month panel. Such a large sample presents substantial computing challenges, given that the mod-

[6] Data table entitled *Source: Office for National Statistics (ONS), Regional Labour Market: HI00 - Headline LFS Indicators for All Regions, October 2013.* Data are seasonally adjusted.

elling procedure adopted in this paper is computationally intensive, even at small sample sizes. For this reason, a 10 per cent sample of the December 2013 data is drawn, and these loans are linked to the December 2012, 2011, 2010 data drops. This leads to a model-ready data set with delinquency status information for 1,073,152 observations.

To generate the model sample, the data are stratified by bank, region (London, Northern Ireland, other), negative equity dummy, Buy to Let flag and default status, as depicted in Figure 5. Within each of the 48 strata, a random sample is drawn, resulting in a data set in which the share of each categorical variable is within 0.1 per cent of that category's share in the population.

Table A.3 breaks the data down by the categories in Figure 5, reporting average outstanding balance, median outstanding balance and the share of each category in the total number of observations. The table reveals that the loan sizes and number of observations in each category are extremely similar between the December 2013 population and the stratified model data set, confirming that the stratified sample is indeed representative of the population.

The panel structure of the stratified "long" data set is described in table A.2. At both bank 1 and bank 2, the vast majority of properties appear in all 49 months from December 2009 to December 2013. Bank 3, on the other hand, entered the sample at a later date, with the majority of properties appearing 36 times. The table provides evidence that within each bank, the panel structure is highly balanced.

Summary statistics on all variables used in the MSM model are provided in Table 2. The average interest rate in the data set is 3.1 per cent, with the average LTV being 84.2 per cent and the average unemployment rate at 7.9 per cent. The data set is split almost evenly between Principal Dwelling House (PDH) and Buy to Let (BTL) loans, respectively. Tracker mortgages are the most common interest rate type, with only 8 per cent of all mortgages having a fixed interest rate. London and Northern Ireland are the regions with the largest representation in the data set.

Figure 6 plots the evolution of the percentage of loans, by count and balance, in default over time in the stratified model data set. The default rate was decreasing through 2011 into 2012, but rose slightly in 2013. The maximum default rate by balance was 2.8 per cent in January 2011. In all months, the balance-weighted default rate is higher than the default rate by count, indicating that larger loans are on average more likely to default. The discrete jump in default rates between December 2010 and January 2011 is partly explained by an increase in sample size at that point, with improved coverage within existing banks and the addition of one bank to the sample.

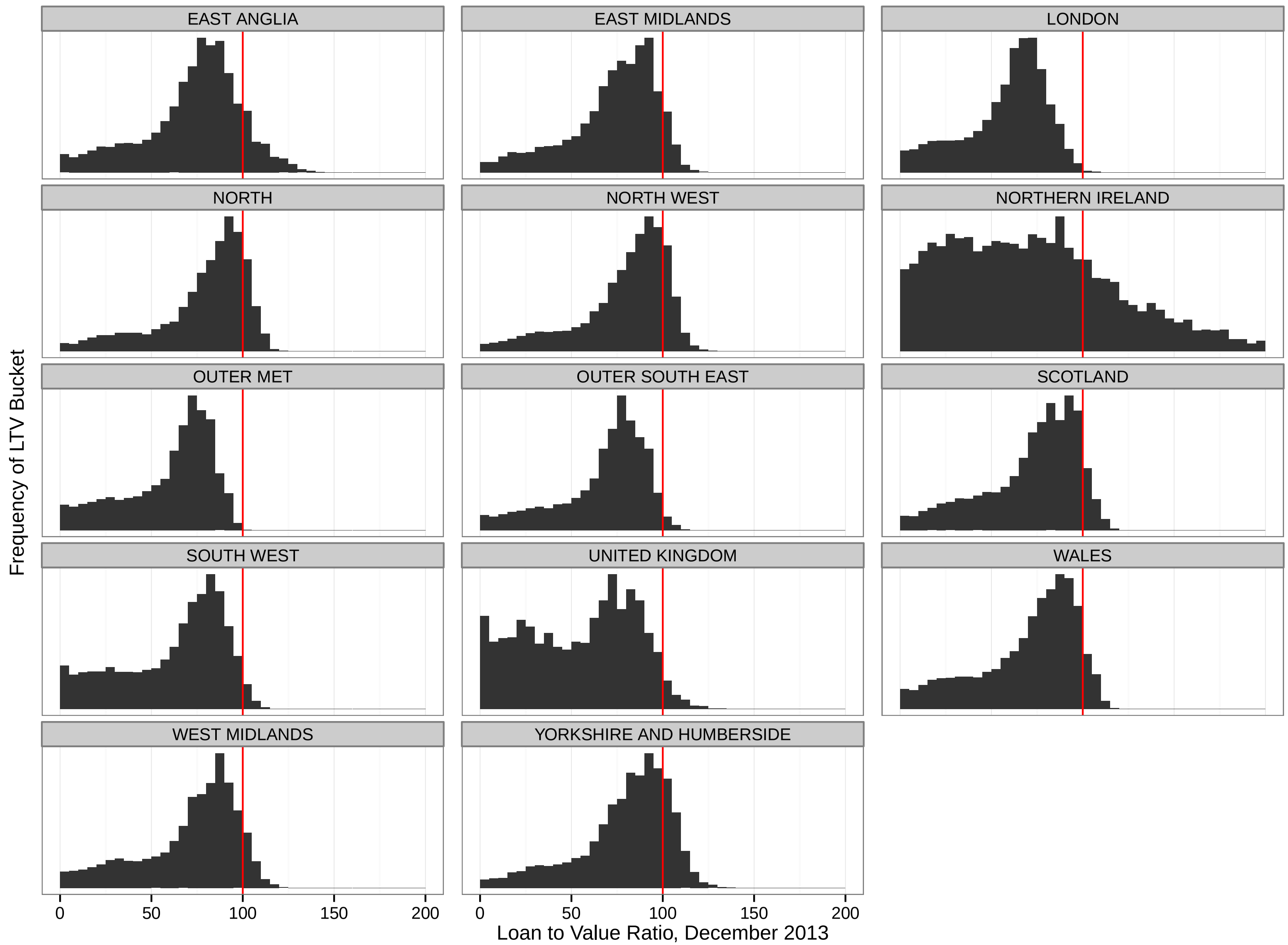
Figure 7 plots the empirical one-year transition probabilities for the stratified data. Annual default transition rates (PD) range between .39 of one percent and 1.31 per cent. There is a pattern to the data whereby in the months of 2010 one-year transition rates are highest, and fall steadily through to mid-2011, with rates rising slightly in the last six months of the sample period. The probability of a

Table 2: Summary statistics, stratified model sample 2009-2013

Continous variables		
	Mean	*Median*
LTV	80.4	84.2
Unemployment	7.9	8.2
Interest Rate	3.1	2.3
Discrete variables		
	Category	*%*
Buy to Let	No	51.1
Buy to Let	Yes	48.8
Interest Rate Type	Fixed	8.0
Interest Rate Type	Tracker	53.3
Interest Rate Type	Variable	38.7
Region	EAST	3.2
Region	EAST MIDLANDS	5.9
Region	LONDON	20.7
Region	NORTH EAST	3.6
Region	NORTH WEST	10.5
Region	NORTHERN IRELAND	14.8
Region	SCOTLAND	3.9
Region	SOUTH EAST	9.8
Region	SOUTH WEST	9.7
Region	Unallocated	1.1
Region	WALES	4.4
Region	WEST MIDLANDS	6.0
Region	Y'SHIRE & H'SIDE	6.2
Default	No	98.4
Default	Yes	1.6

loan transitioning from default to performing ($PCure$), ranges between 16 and 61 per cent (Figure 8). The wide variation in cure rates over time may partly be explained by the small sample of defaulted loans in any starting month. However, it should be acknowledged that the pattern observed in Figure 8 is a relatively smooth one, suggesting that small sample sizes do not drive any statistically noisy kinks in the data series that would pose a concern for estimation.

Figure 4: Regional distribution of Loan to Value ratio, December 2013



“United Kingdom” refers to loans that are not allocated a region in the dataset.

Figure 5: Stratified sampling methodology

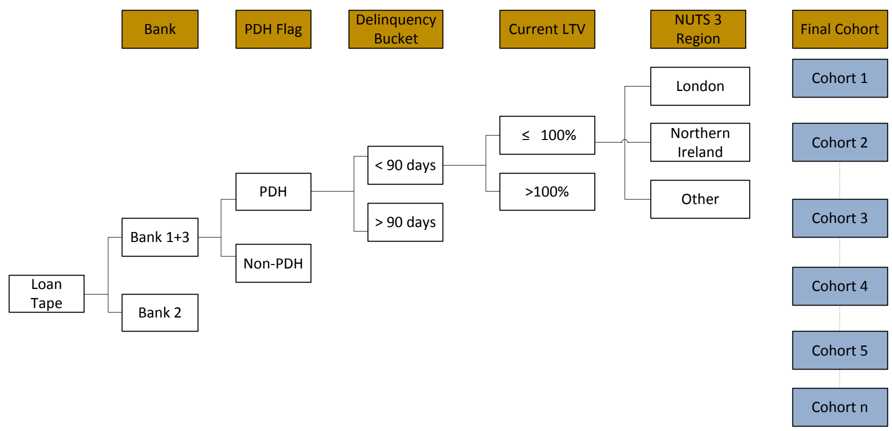


Figure 6: Percentage of loans in default

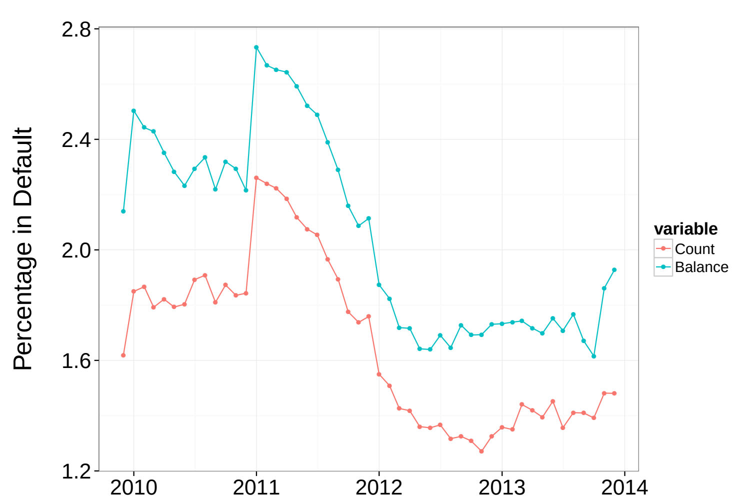


Figure 7: One-year empirical probability of default

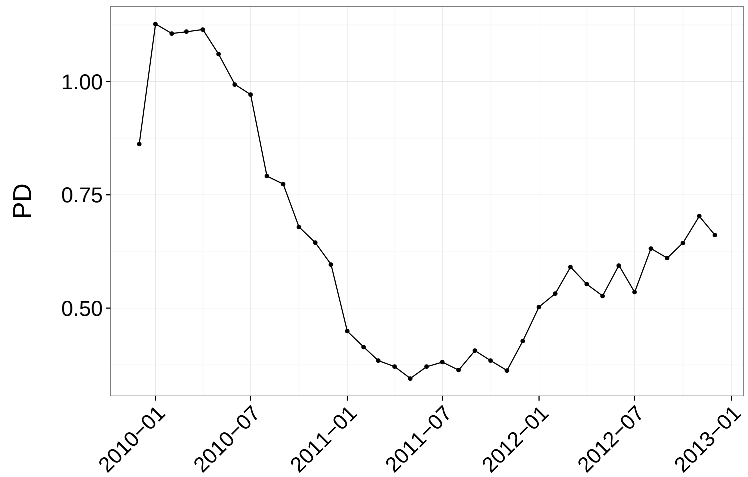
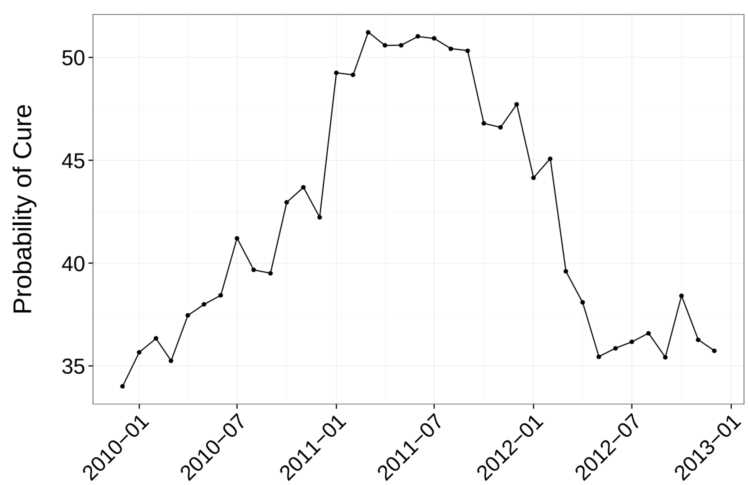


Figure 8: One-year empirical probability of cure



3 Empirical multi-state model

Models of loan default traditionally rely on cross-sectional logistic regression models, which can easily incorporate loan-level covariates but lack a panel data dimension, or transitions-based approaches in the spirit of JP Morgan's CreditMetrics^{TM} and McKinsey's CreditPortfolioReview, which rely on observed empirical loan transitions in discrete time, without the possibility of allowing for an estimated effect of loan-level covariates.

A Markov multi-state model of transitions in continuous time offers an attractive alternative to either of the two aforementioned approaches. A continuous time estimation of a multi-state default model was first proposed by Lando and Skodeberg (2002). The authors point out significant advantages over the discrete time, "cohort method" approach adopted in many industry-standard transition models such as those mentioned above, including, but not limited to, the estimation of non-zero transition probabilities in instances where no loans are observed to have moved from state r to s between time period t and $t+1$. Lando and Skodeberg (2002) provide simple practical examples of this feature of the continuous-time Markov model.

The application of a time-homogenous Markov multi-state model to UK mortgage default data is carried out in the *MSM* package in the R statistical language (Jackson, 2011). In a time-homogenous Markov model, the transition probability from state r to s over time u to $t+u$ $p_{rs}(u;t+u)$ is a function of the *distance* between dates rather than the dates themselves. This removes the need to track the location of two time points in calendar time when estimating these models.

A multi-state model describes how an individual moves between a series of states in continuous time. Suppose an individual is in state $S(t)$ at time t. The movement on the discrete state space $1, ..., R$ is governed by transition intensities:

$$q_{rs}(t; z(t)) : r; s = 1, ..., R \tag{6}$$

These may depend on time t, or, more generally, also on a set of individual-level or time-dependent explanatory variables $z(t)$. The intensity represents the instantaneous risk of moving from state r to state $s \neq r$

$$q_{rs}(t; z(t)) = lim_{\delta t \to 0} P(S(t + \delta t) = s | S(t) = r)/\delta t. \tag{7}$$

The q_{rs} form a RxR matrix Q whose rows sum to zero, so that the diagonal entries are defined by $q_{rr} = \sum_{s \neq r} q_{rs}$. In the case of the default model used in this paper, where 0 represents performing loans and 1 indicates loans in default, the matrix Q is represented by

$$Q = \begin{pmatrix} -q_{01} & q_{01} \\ q_{10} & -q_{10} \end{pmatrix}$$

The generator matrix Λ is a RxR matrix for which

$$Q(t) = exp(\Lambda t) \tag{8}$$

where the exponential function is the matrix exponential, and Λt is the matrix Λ multiplied at every entry by t. The elements of the generator are estimated as

$$\lambda_{rs} = \frac{N_{rs}(T)}{\int_0^T Y_r(s)ds} \tag{9}$$

where $N_{rs}(T)$ is the total number of transitions from r to s in the sample period and $\int_0^T Y_r(h)dh$ is the total number of "loan-month" combinations that are spent in state h. All information on time spent in all states is now being used in estimation.

One can obtain maximum-likelihood estimates of the transition probability matrices by first obtaining the maximum-likelihood estimate of the generator and then applying the matrix exponential function and scaling by the time horizon Lando and Skodeberg (2002). Note that the results of the MSM model are always time-dependent; a time horizon for the transition probabilities, and the effect of covariates, must always be specified. The full likelihood is then the product of probabilities of transition between observed states,over all individuals i and observation times j:

$$L(Q) = \prod_i L_i = \prod_{i,j} L_{i,j} \tag{10}$$

Each component $L_{i,j}$ is the entry of the transition matrix $P(t)$ at the $S(t_{ij})$th row and $S(t_{i;j+1})$th column, evaluated at $t = t_{i;j+1} - t_{ij}$. The likelihood $L(Q)$ is maximized in terms of $log(q_{rs})$ to compute the estimates of $_q rs$, using standard optimization algorithms.

One powerful feature of this class of model is the potential for the inclusion of covariates which explain the transitions betweeen r and s. The effect of a vector of explanatory variables z_{ij} on the transition intensity for individual i at time j is modelled using proportional intensities, replacing q_{rs} with

$$q_{rs}(z_{ij}) = q_{rs}^{(0)} exp(\beta_{rs}^t z_{ij}) \tag{11}$$

where $q_{rs}^{(0)}$ is the rs entry of Q, the matrix of baseline transition intensities. Recall the relationship between Q and the time-independent generator matrix Λ: $Q(t) = exp(\Lambda t)$. The likelihood is then maximized over the $q^{(0)}_{rs} and \beta_{rs}$. The interpretation of coefficients from such a model is that, for a one-unit increase in z, the increase in the risk of transitioning from r to s increases by β.

4 Results

Table 3 provides the coefficient results from the baseline Markov multi-state model. The coefficient in the right hand panel on Time Since Default (TSD) indicates that the risk of a loan transitioning from default back to performing status ($PCure$) decreases by a 7.98 per cent for each additional month that a loan is in default. This hysteresis effect on loan default is statistically significant using the estimated 95 per cent confidence intervals.

The effect of LTV on the probability of default is that a one percentage point increase in LTV leads to a one per cent increase in the risk of a transition from performing to default. Similarly, a one per centage point increase in the unemployment rate leads to a 9.48 per cent increase in the risk of such a transition, while a 100 bps increase in the mortgage interest rate leads to a 19 per cent increase. Given that these two latter factors both relate to the affordability of the mortgage, it is clear that these results provide confirmation that both components of the "double trigger" are at play in explaining the movement of loans into default.

The negative and statistically significant coefficients on LTV and interest rates in the $PCure$ equation indicate that both equity and affordability also play role in reducing the likelihood of loans curing from default. This suggests that the "double trigger" findings identified in cross-sectional studies are likely composed of effects on both the inflow and outflow of mortgage defaults. These bi-directional impacts can only be identified using a methodology similar to that used in this paper.

But to Let mortgages are found to be more likely to default, and less likely to cure than Principal Dwelling House (PDH) mortgages. A similar effect is found for variable rate versus fixed mortgages. Finally, tracker mortgages are found to be less likely to cure than fixed mortgages, but do not enter the PD equation with a significant coefficient.

The economic significance of the coefficient estimates in Table 3 is more easily explained by plotting the model-estimated PD and $PCure$ for varying values of each X variable, holding other variables at their mean.

The impact of Time Since Default (TSD) on the probability of cure ($PCure$) is quantified in Figure 9. The effect is such that increase from one to two years in default leads to a halving in the probability that a loan will recover. At the mean value among those in default ($T\bar{S}D$), $PCure$ is equal to 57 per cent. Increasing TSD by one standard deviation ($SD_{TSD} = 13 months$) leads to a fall in $PCure$ to 29 per cent, which represents an 49 per cent decrease in the probability that a mortgage will recover from default status. Crucial policy implications emerge from this result: measures such as

Table 3: Coefficients from Multi State model, all mortgages

	PD		$PCure$	
	Coeff	95% CI	Coeff	95% CI
Time Since Default			-0.0798	(-0.08794,-0.07197)
Bank 2	0.1562	(-0.0601,0.3727)	-0.2551	(-0.4908,-0.0195)
Bank 3	-0.5328	(-0.8056,-0.2599)	-0.5043	(-0.8672,-0.1414)
LTV	0.0098	(0.0085,0.0111)	-0.0080	(-0.0096,-0.0063)
Unemployment	0.0948	(0.0547,0.1348)	0.0172	(-0.0263,0.0608)
Interest Rate	0.1948	(0.1469,0.2428)	-0.1178	(-0.1787,-0.0568)
Buy to Let Dummy	0.357	(0.2337,0.4804)	-0.2928	(-0.4444,-0.1412)
IRT Variable	0.2083	(0.0098,0.4068)	-0.2123	(-0.4076,-0.0171)
IRT Tracker	-0.0712	(-0.2692,0.1266)	-0.2417	(-0.4329,-0.0506)

the Home Affordable Modification Program (HAMP) in the USA, and the Central Bank of Ireland's Mortgage Arrears Resolution Targets (MART) framework, where lenders have been either incentivised or obliged to engage with distressed borrowers in order to arrive at a sustainable modified mortgage, can mitigate the risk of borrowers ending up "trapped" towards the right hand side of the curve in Figure 9, with little chance of recovery.

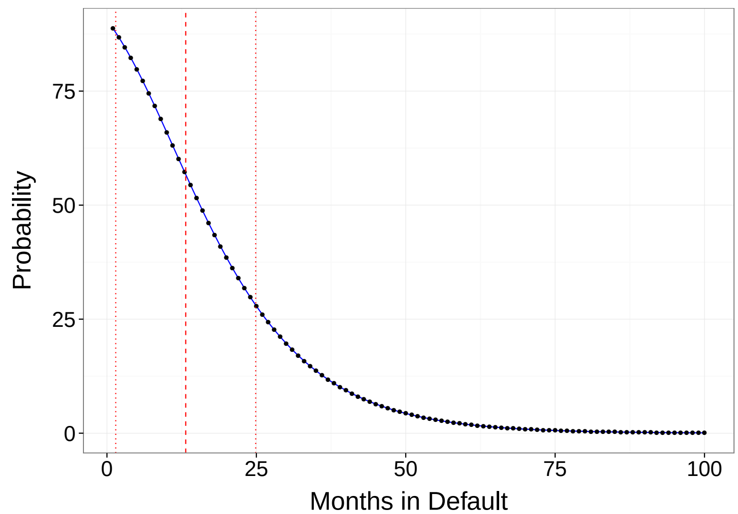
While Time Since Default is restricted to only enter the $PCure$ equation, all other covariates in Table 3 enter both the PD and $PCure$ equation. Figure 10 plots the relationship between interest rates and the transition probabilities. The model predicts that moving from $ir\bar{a}te$ to $(ir\bar{a}te + SD_{irate})$ leads to an increase in PD from 1.9 to 2.5 per cent. The analogous effect in the $PCure$ equation is a fall from 54 to 48 per cent in the transition probability.

The relationship between LTV and transition probabilities is plotted in Figure 11. Moving from $L\bar{T}V$ to $(L\bar{T}V + SD_{LTV})$ leads to an increase in PD from 1.8 to 2.8 per cent. An identical one-standard-deviation increase from $L\bar{T}V$ leads to a fall in $PCure$ from 56 to 46 per cent.

Given that the regional unemployment rate is statistically significant in the PD equation alone, the impact on PD only is plotted in Figure 12. The model predicts that an increase from 10 to 20 per cent in a region's unemployment rate will lead to an increase in PD from 2.2 to 5.2 per cent, while an increase from $\bar{ue}$ to $(\bar{ue} + SD_{ue})$ leads to a relatively small increase in PD from 1.85 to 2 per cent.

The relative economic importance of the continuous covariates mentioned above is summarised in Table 4.

Figure 9: Impact of Time Since Default on $PCure$

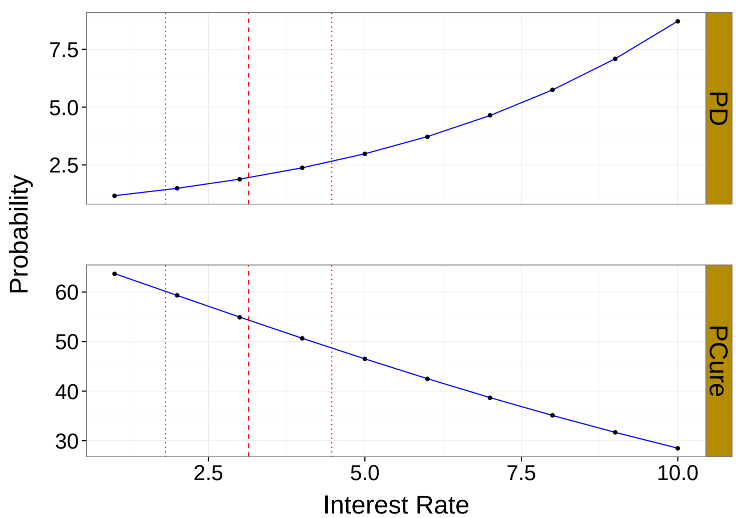


Note: Estimated $PCure$ at values from 0 to 100; Model with LTV, Interest Rate, Unemployment set to their mean, Bank=2, BTL=Yes, Interest Rate Type=Variable. Mean TSD among those in default plotted in dashed red line; Mean Interest Rate +- standard deviation of Interest Rate plotted in dotted red line.

Table 4: Impact of one-standard deviation increase in continuous covariates

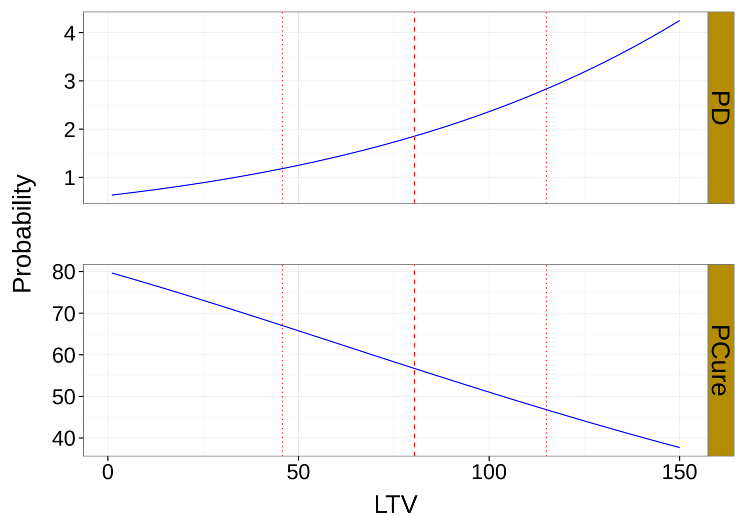
	Probability of Default			Probability of Cure		
Variable	$\bar{X}$	$\bar{X} + SD_X$	% Change	$\bar{X}$	$\bar{X} + SD_X$	% Change
Time Since Default				57	29	-49%
Interest Rate	1.9	2.5	31.5%	54	48	-11%
LTV	1.8	2.8	55.6%	56	46	-18%
Unemployment	1.85	2	8.1%			

Figure 10: Impact of Interest Rate on PD and $PCure$



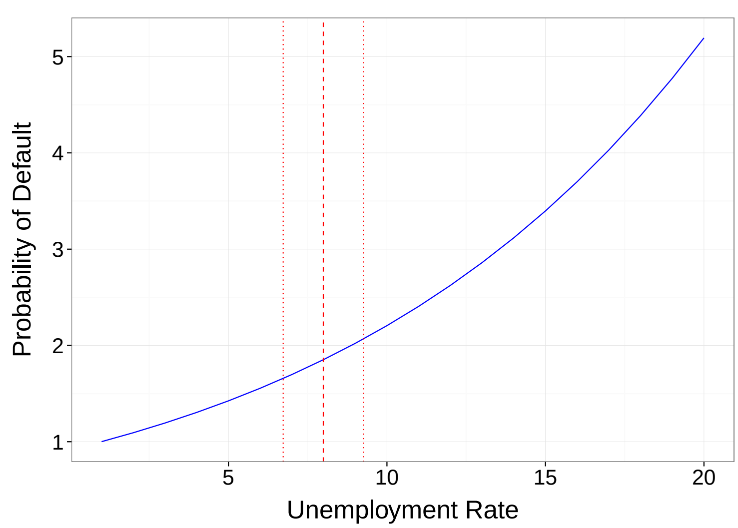
Note: Estimated PD and $PCure$ at Interest Rate values from 0 to 10; Model with Time Since Default, LTV and Unemployment at their meanBank=2, BTL=Yes, Interest Rate Type=Variable. Mean Interest Rate plotted in dashed red line; Mean Interest Rate +- standard deviation of Interest Rate plotted in dotted red line.

Figure 11: Impact of LTV on PD and $PCure$



Note: Estimated PD and $PCure$ at LTV values from 0 to 300; Model with Interest Rate, Time Since Default and Unemployment at their mean, Bank=2, BTL=Yes, Interest Rate Type=Variable. Mean LTV plotted in dashed red line; Mean LTV +- standard deviation of LTV plotted in dotted red line.

Figure 12: Impact of Unemployment on PD



Note: Estimated PD and $PCure$ at Unemployment values from 0 to 30; Model with Interest Rate, Time Since Default, LTV set to their mean, Bank=2, BTL=Yes, Interest Rate Type=Variable.

4.1 Sensitivity of owner-occupiers

The empirical analysis up to this point has modelled Principal Dwelling Houses (PDH) and Buy to Let properties (BTL) jointly. In this section I restrict the sample to PDH properties only, given that the incentives of these borrowers are likely to be different to those of BTL investors, and policy interest in the mortgage market is predominantly focused on alleviating payment distress among those who have purchased to occupy their own home. Table 5 reports results of the same specification as presented in Table 3 for PDH mortgages only, with the Buy to Let dummy removed from the model. Coefficient signs and statistical significance levels remain similar between the pooled model and the PDH-only model, suggesting that the relationships identified in the pooled model also hold in the PDH segment.

The sensitivity of PDH borrowers to changes in fundamentals such as Interest Rate, LTV and Unemployment is of immense policy and academic interest. Figure 13 shows that there is a substantial difference between the effect of LTV in the pooled model, which includes both PDH and BTL loans, when compared to the PDH-only model. The slope of the blue line, which plots the LTV effect in the pooled model and is identical to that plotted in Figure 11, is steeper than the PDH-only slope, plotted in red. The difference in slopes is particularly apparent in the $PCure$ equation.

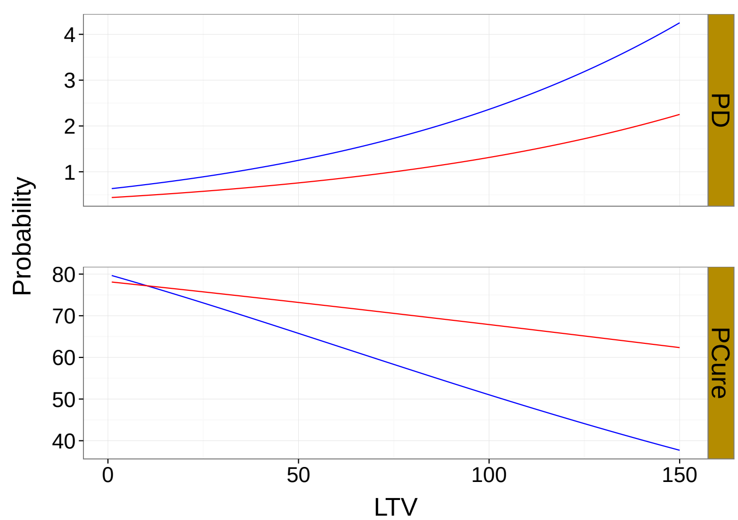
Figure 14 reports that, for interest rates, the slope of the pooled and PDH-only lines are close to identical for both the PD and $PCure$ equations, suggesting that owner-occupiers are as sensitive to interest rates as the pool of owner-occupiers and investors. Figure 15 shows a marked distinction between owner-occupiers and the pooled model, with the pooled PD model (blue line) having a much steeper unemployment slope than the PDH-only model.

Figures 13 and 15 suggest that investors appear more sensitive to equity and affordability considerations than owner-occupiers. Given that owner-occupiers directly consume a housing service through their mortgage, this finding appears intuitive. Further, in a system of full recourse such as the UK's personal bankruptcy regime, it is likely that owner-occupiers will be more hesitant in deciding to enter default than investors. Further, the preference for home-ownership revealed by those holding owner-occupier mortgages should lead one to predict that these borrowers will deploy all available resources, such as household savings and lifestyle changes, in order to avoid entering default. The higher loan cure sensitivity of investors to improving LTVs can be interpreted as suggesting that "strategic" or equity-driven default is more prevalent in this market segment: when house prices move into positive equity, investors are in a position to begin repayments and move away from exercising the "put option" implicit in their mortgage contract. Owner-occupiers, on the other hand, are less likely to have entered default for strategic reasons, and are therefore less likely to be in a position to re-start full payment solely as a result of improvements in their equity position.

Table 5: Coefficients from Multi State model, Principal Dwelling Houses only

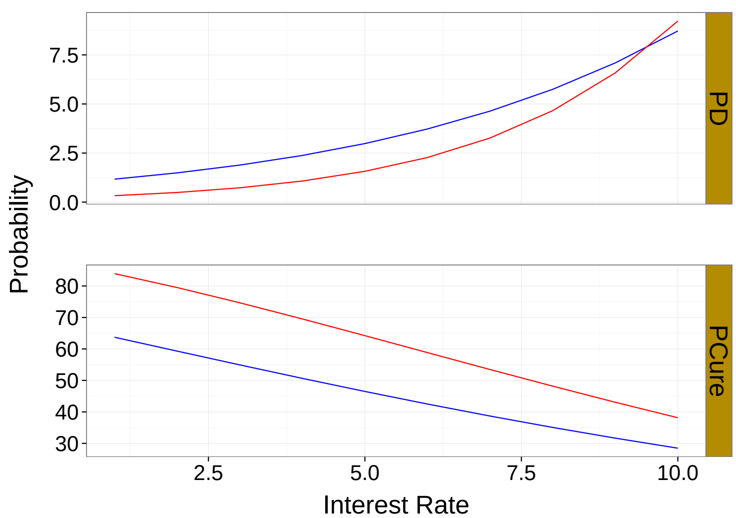
	PD		$PCure$	
	Coeff	95% CI	Coeff	95% CI
Time Since Default				
			-0.0562	(-0.06548,-0.047)
Bank 2	-0.2682	(-0.5298,-0.0065)		
			-0.636	(-0.8944,-0.3776)
Bank 3	1.162	(0.7718,1.551)		
			-0.395	(-0.8663,0.0763)
LTV	0.0096	(0.0079,0.0113)		
			-0.0028	(-0.0048,-0.0008)
Unemployment	0.1077	(0.0503,0.165)		
			0.0481	(-0.0116,0.1079)
Interest Rate	0.3173	(0.2381,0.3965)		
			-0.1412	(-0.2187,-0.0636)
IRT Variable	-0.0919	(-0.3538,0.1699)		
			-0.4431	(-0.6984,-0.1878)
IRT Tracker	-0.142	(-0.3983,0.1143)		
			-0.6015	(-0.8556,-0.3473)

Figure 13: Impact of LTV on PD and $PCure$ across pooled and PDH models



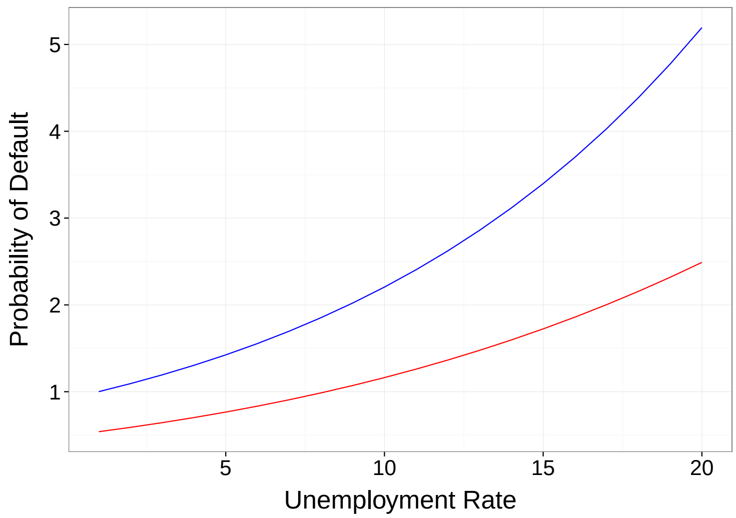
Note: Blue line: pooled model, all mortgages. Red line: PDH model. Estimated PD and $PCure$ at LTV values from 0 to 150; Model with Interest Rate, Time Since Default and Unemployment at their mean, Bank=2, BTL=Yes, Interest Rate Type=Variable.

Figure 14: Impact of Interest Rates on PD and $PCure$ across pooled and PDH models



Note: Blue line: pooled model, all mortgages. Red line: PDH model. Estimated PD and $PCure$ at Interest Rate values from 0 to 10; Model with LTV, Time Since Default and Unemployment at their mean, Bank=2, BTL=Yes, Interest Rate Type=Variable.

Figure 15: Impact of Unemployment Rates on PD and across pooled and PDH models



Note: Blue line: pooled model, all mortgages. Red line: PDH model. Estimated PD and $PCure$ at Unemployment values from 0 to 20; Model with LTV, Time Since Default and Interest Rates at their mean, Bank=2, BTL=Yes, Interest Rate Type=Variable.

5 Conclusion

Using data on the population of UK residential mortgages from three Irish-headquartered banks, this paper provides for the first time a micro-level model of transitions between mortgage delinquency states in the UK. The model implemented is a Markov multi-state model in which loans can transition in both directions between performing and default status, with default status defined as ninety days past due.

We highlight an economically significant hysteresis effect of the time spent in default: a mortgage that has been in default for one year is twice as likely to "cure" back to performing status as a loan that has been in default for one year. This finding provides importance impetus to policies that aim to target those that enter mortgage payment difficulty as early as possible, and find a solution that allows them to continue repayment on modified terms.

The model finds that both equity and affordability play an important role in determining the probability of default (PD) and cure ($PCure$). The magnitude impacts of changes in Loan-to-Value ratios and interest rates are similar in both the PD and $PCure$ samples. Buy-to-Let and variable rate mortgages are more likely to default and less likely to cure than owner-occupier and fixed rate mortgages, respectively.

In running a separate model for owner-occupier mortgages only, I identify that the role of housing equity, as measure by the Loan-to-Value ratio, is much less important among owner-occupiers than Buy-to-Let (BTL) investors. This suggests that "strategic" default has minimal prevalence among UK owner-occupiers, which is likely explained by the full recourse nature of the UK's personal bankruptcy code. The multi-state nature of the model employed in this paper allows me to identify that owner-occupiers are also less likely to *recover* from mortgage default as a response to improvements in equity alone. This finding suggests that affordability must also improve before house price improvements lead to owner-occupiers' loan cure, leading to the conclusion that UK owner-occupiers' defaults were most likely not "strategic" to begin with.

The converse is true among BTL investors, where improvements in equity, independent of changes in affordability, have a much stronger relationship with $PCure$. The fact that BTL mortgages move readily out of default in response to changes in equity provides suggestive evidence that many of these defaults were "strategic" in nature. This method of identifying strategic or equity-driven default is only possible in a modelling framework where loans can move both into and out of default, such as the Multi-State Model used in this paper.

References

Aron, Janine and John Muellbauer, "Modelling and Forecasting UK Mortgage Arrears and Possessions," SERC Discussion Papers, Spatial Economics Research Centre, LSE 0052, Spatial Economics Research Centre, LSE August 2010.

_ and _ , "Modelling and Forecasting with County Court Data: Regional Mortgage Possession Claims and Orders in England and Wales," SERC Discussion Papers, Spatial Economics Research Centre, LSE 0070, Spatial Economics Research Centre, LSE February 2011.

Bajari, Patrick, Chenghuan Sean Chu, and Minjung Park, "An Empirical Model of Subprime Mortgage Default From 2000 to 2007," NBER Working Papers 14625, National Bureau of Economic Research, Inc December 2008.

Bhutta, Neil, Jane Dokko, and Hui Shan, "The Depth of Negative Equity and Mortgage Default Decisions," Finance and Economics Discussion Series, Board of Governors of the Federal Reserve System (U.S.) 2010-35, Board of Governors of the Federal Reserve System (U.S.) May 2010.

Boheim, Rene and Mark P. Taylor, "My Home Was My Castle: Evictions and Repossessions in Britain," *Journal of Housing Economics, Elsevier*, December 2000, *9* (4), 287–319.

Elul, Ronel, Nicholas S. Souleles, Souphala Chomsisengphet, Dennis Glennon, and Robert Hunt, "What "Triggers" Mortgage Default?," *American Economic Review, American Economic Association*, May 2010, *100* (2), 490–94.

Foote, Christopher L., Kristopher Gerardi, and Paul S. Willen, "Negative equity and foreclosure: Theory and evidence," *Journal of Urban Economics*, September 2008, *64* (2), 234–245.

Gaffney, Edward, Robert Kelly, and Fergal McCann, "A transitions-based framework for estimating expected credit losses," Research Technical Paper 16RT14, Central Bank of Ireland 2014.

Gathergood, John, "Income Shocks, Mortgage Repayment Risk and Financial Distress Among UK Households," Discussion Papers 09/03, University of Nottingham, Centre for Finance, Credit and Macroeconomics (CFCM) May 2009.

Gerardi, Kristopher, Kyle F. Herkenhoff, Lee E. Ohanian, and Paul S. Willen, "Unemployment, Negative Equity, and Strategic Default," Working Paper, Federal Reserve Bank of Atlanta 2013-04, Federal Reserve Bank of Atlanta August 2013.

Goodman, Laurie S, Roger Ashworth, Brian Landy, and Ke Yin, "Negative Equity Trumps Unemployment in Predicting Defaults," *Journal of Fixed Income*, Spring 2010, *19* (4), 67–72.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, "The Determinants of Attitudes toward Strategic Default on Mortgages," *Journal of Finance, American Finance Association*, 08 2013, *68* (4), 1473–1515.

Gyourko, Joseph and Joseph Tracy, "Reconciling theory and empirics on the role of unemployment in mortgage default," *Journal of Urban Economics*, 2014, *80* (C), 87–96.

Jackson, Christopher S., "The MSM Package for R," *Journal of Statistical Software*, January 2011, *38* (8), 67–72.

Kelly, Robert, "The Good, The Bad and The Impaired - A Credit Risk Model of the Irish Mortgage Market," Research Technical Papers 16/RT/09, Central Bank of Ireland 2011.

— **and Terence O'Malley**, "A transitions-based model of default for Irish mortgages," Research Technical Paper 17RT14, Central Bank of Ireland 2014.

Lando, David and Torben M. Skodeberg, "Analyzing rating transitions and rating drift with continuous observations," *Journal of Banking & Finance, Elsevier*, March 2002, *26* (2-3), 423–444.

Lanot, Gauthier and David Leece, "Mortgage Loan Characteristics , Unobserved Heterogeneity and the Performance of United Kingdom Securitised Sub-Prime Loans," Umea Economic Studies 876, Umea universitet Department of Economics 2014.

Lydon, Reamonn and Yvonne McCarthy, "What Lies Beneath? Understanding Recent Trends in Irish Mortgage Arrears," *The Economic and Social Review, Economic and Social Studies*, 2013, *44* (1), 117–150.

Mayer, Christopher, Karen Pence, and Shane M. Sherlund, "The Rise in Mortgage Defaults," *Journal of Economic Perspectives*, 2009, *23* (1), 27–50.

McCarthy, Yvonne, "Dis-entangling the mortgage arrears crisis: The role of the labour market, income volatility and housing equity," Research Technical Papers 02/RT/14, Central Bank of Ireland 2014.

McDonald, Ross A., Ania Matuszyk, and Lyn C. Thomas, "Application of Survival Analysis to Cash Flow Modelling for Mortgage Products," *OR Insight*, 2010, *23* (1), 1–14.

Vandell, Kerry D., "How Ruthless is Mortgage Default? A Review and Synthesis of the Evidence," *Journal of Housing Research*, 1995, *6* (2), 245–264.

A Appendix

Table A.1: Number of loans per property, December 2013

Loans	Properties
1	20,8935
2	27,650
3	5,796
4	1,551
5	454
6	136
7	55
8	15
9	5
10	7
11	2
13	1
14	2
16	1
18	1
	244,611

Table A.2: Frequency of unique property appearances in the PD model sample

N	Freq Bank 1	Freq Bank 2	Freq Bank 3
1	7	75	5
2	14	19	4
3	8	23	1
4	2	14	3
5	6	20	4
6	9	314	80
7	1	0	1
8	3	0	0
10	3	0	0
11	4	0	1
12	7	239	3
24	21	444	10
31	0	1	0
36	29	3,483	4,142
37	103	106	0
38	0	55	0
39	0	111	0
40	2	77	0
41	4	80	0
42	1	60	0
43	0	36	0
44	0	24	0
45	1	25	0
46	3	28	0
47	0	20	0
48	0	22	0
49	2,283	13,016	0
Total	2,511	18,292	4,254

Table A.3: Cross tabulations in population and stratified MSM samples

Field	Category	Mean OB	Median OB	% Count	Mean OB	Median OB	% Count
Bank	1	104620	78666	6.98	106149	78140	7.28
Bank	2	112841	98311	75.76	112735	98744	75.22
Bank	3	139295	111089	17.26	139401	112174	17.49
BTL Flag	NO	120073	102135	49.81	119773	102500	49.99
BTL Flag	YES	113620	97602	50.19	114068	98080	50.01
Default	0	116306	99373	98.50	116356	100000	98.51
Default	1	151389	118808	1.50	154175	121240	1.49
Negative Equity	0	116121	98349	90.51	115921	98655	90.23
Negative Equity	1	123635	107624	9.49	126156	108849	9.77
Region	E. ANGLIA	113236	101506	3.12	110778	102409	3.21
Region	E. MIDLANDS	96244	88755	6.36	98687	89112	6.09
Region	LONDON	184818	165461	14.09	184176	167963	14.16
Region	NORTH	85106	74899	3.77	83761	76163	3.76
Region	NORTH WEST	95438	86015	10.96	95557	86275	10.96
Region	N. IRELAND	94583	75641	11.69	97468	75653	12.10
Region	OUTER MET	142710	128564	7.58	143068	129506	7.44
Region	OUTER S.E.	132197	120000	10.31	130928	119098	10.43
Region	SCOTLAND	93156	82399	3.88	93872	84215	3.88
Region	SOUTH WEST	108899	102690	9.59	106885	102778	9.64
Region	Unallocated	115824	91350	1.23	116675	92994	1.18
Region	WALES	90111	82396	4.51	90002	82215	4.37
Region	WEST MIDLANDS	97893	89942	6.51	98475	91142	6.28
Region	Y'SHIRE AND H'SIDE	94481	84135	6.40	94089	83408	6.48

Figure A.1: Regional house prices, 2001 - 2013

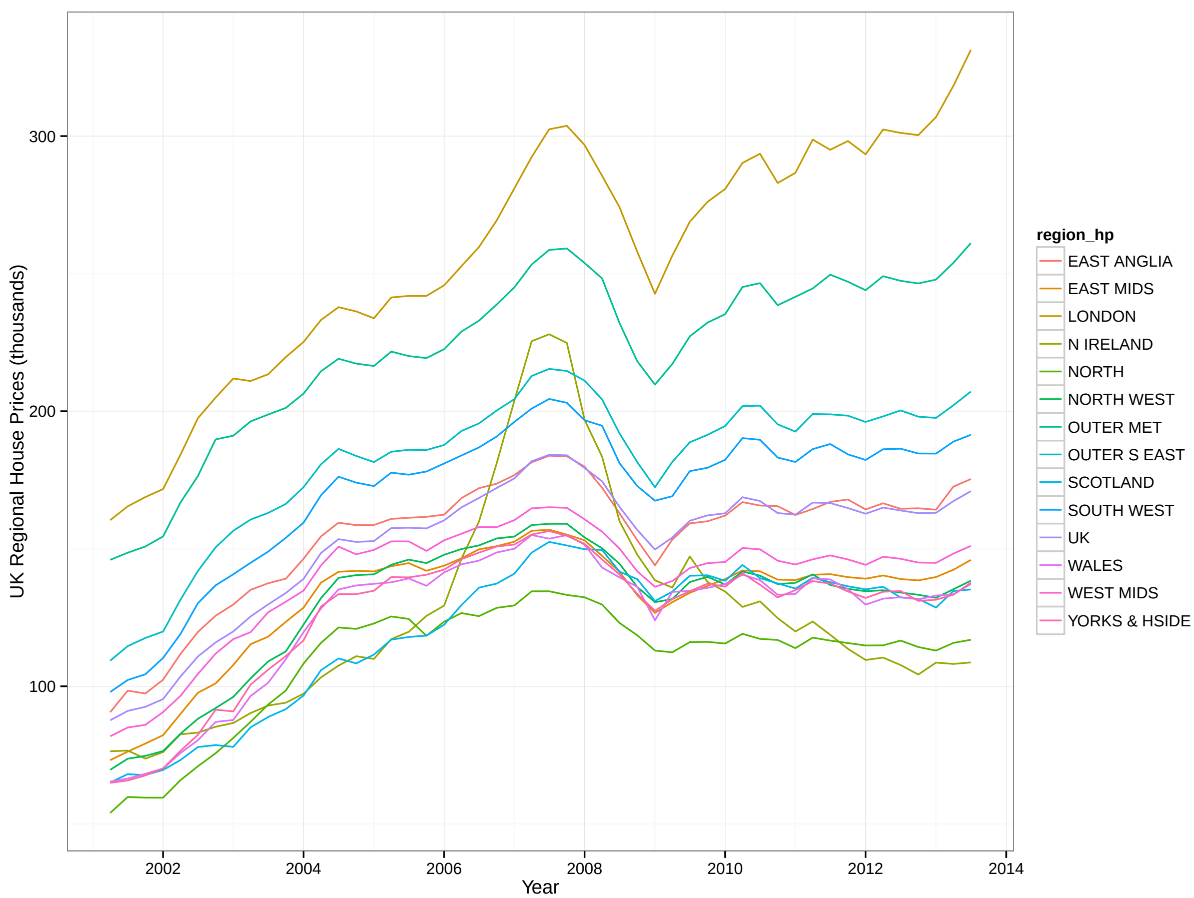


Figure A.2: Regional unemployment rates, 2009 - 2013