16RT14

Páipéar Taighde Teicniúil
Research Technical Paper

A transitions-based framework for estimating expected credit losses

Edward Gaffney, Robert Kelly, Fergal McCann

Banc Ceannais na hÉireann
Central Bank of Ireland
Eurosystem

A transitions-based framework for estimating expected credit losses

Edward Gaffney, Robert Kelly, Fergal McCann*
Financial Stability Division
Central Bank of Ireland

November 2014

Abstract

This paper presents a framework for estimating losses for residential mortgage loans. At the core is a transitions-based probability of default model which yields directly observable cash-flows at the loan level. The estimated model includes coefficients on unemployment, Loan to Value ratio and interest rates, all of which allow a macroeconomic scenario to be fed through the model and impact loans' probability of default and cure. Other loan-level covariates such as bank, Buy-to-Let status, and vintage also impact loans' transition probabilities. Loss Given Default is also modelled over a three-year horizon combining loan-level collateral information with macroeconomic house price forecasts. The breakout of flows from the stock of defaults allows the impact of loan modifications on recovery rates to be modelled. Unlike other models of mortgage credit risk, this framework allows a hysteresis effect of the time spent in default on the probability of loan cure to be modelled explicitly. In Ireland, an increase in the time spent in default from three months to one year leads to a decrease in the probability of loan cure from 30 to 12 per cent.
Keywords: Mortgages, default, credit risk, Markov multi-state model.

*edward.gaffney@centralbank.ie; robert.kelly@centralbank.ie; fergal.mccann@centralbank.ie. The views presented in this paper are those of the author alone and do not represent the official views of the Central Bank of Ireland or the ESCB. Any remaining errors are our own.

1 Introduction

We propose in this article a framework that explicitly allows the modelling of loan-level transitions both into and out of default which is set up to naturally trace the cash flows of performing and defaulted loans over a set time horizon. The "multi-state model" (MSM, Jackson (2011)) estimation framework uses five years of historical mortgage data for the Republic of Ireland (ROI) and United Kingdom (UK) and allows both loan-level and macroeconomic factors to impact the probability of loan default and cure. A unique feature of the probability of default (PD) estimation model is that the probability of loan cure is negatively influenced by the time that a loan has spent in default ("Time Since Default", TSD). This negative coefficient allows us to model a hysteresis effect of long delinquent spells, which is of particular relevance in countries with large stocks of defaulted loans resulting from the recent financial crisis.

Estimation of the impact of a set of covariates on the probability of both loan default and loan cure provides a natural progression from PD model estimation to an Exposure at Default (EAD) model where defaulted and performing loans at December 2013 move flexibly through each year/quarter/month of a scenario horizon. The EAD model is designed to calculate performing balances, defaulted balances, default flows, cure flows, interest payments, lifetime cures and, crucially, expected losses. All of these values are outputted in each period of the scenario, for example on an annual basis from December 2014 - December 2016 in the case studied in this paper. Scenario-dependent Loss Given Default estimates are also easily arrived at given that collateral valuation data are available at December 2013.

Loan-specific probabilities of default and cure adjust in response to changes in macroeconomic inputs, with house price forecasts entering via the Loan to Value ratio (LTV), unemployment entering at the regional level, and changes in interest rates feeding through to loan-specific interest rate information. At the same time, these probabilities are impacted by loan-level covariates such as vintage, interest rate type, Buy-to-Let (BTL) status, bank and TSD. The explicit link between PD model coefficients and loans' transitions through the EAD model allows the user to avoid the work-arounds, overlays and assumptions that are required in order to generate a link between a static discrete choice PD model such as a logit and a dynamic cash flow model with a specified time horizon.

The covariate-driven transition-based framework outlined in this paper contains a number of advantages compared to the methods through which credit losses are most commonly modelled. These methods are briefly outlined below.

1. "Standardised" models in the parlance of the New Basel Capital Accord ("Basel II", see Basel Committee on Banking Supervision (2001)), which use external ratings such as those provided by external credit assesment institutions (ECAIs) to determine risk-weights for capital charges.

Such models have been commonly used by smaller lenders, and do not require any econometric estimation.

2. Portfolio-level models where the dependent variable is the percentage of loans or balance in default in a given time period. A portfolio in this case could consist of a full asset class, or an asset class such as residential mortgages split into sub-portfolios based on loans' vintage, geographical location or Loan to Value ratio (LTV). A time series of such data is generally modelled as a function of macroeconomic factors such as GDP, unemployment or house prices. Scenario values for these macro factors are then easily applied to starting values for portfolio-level default rates to make a default forecast.

3. Portfolio-level transition models which use the "cohort method", whereby transition probabilities are calculated by directly observing historical migration between ratings or states. Examples include the RiskMetrics^{TM} CreditMetrics model and McKinsey's Credit Portfolio View.

4. At the loan level, a cross-sectional, point-in-time discrete choice model (probit or logit) where the dependent variable is a dummy indicating a loan's default or performing status. Such models are also used by lenders adopting the IRB approach. Scenario inputs for macroeconomic drivers can be used to get an individual loan's fitted probability of default $\hat{PD}$, provided these factors can sensibly predict default status in the discrete choice model.

5. Models such as that of Betancourt (1999) and Grimshaw and Alexander (2011) which model the transition matrix of movement of loans between delinquent states as a Markov chain, but are not capable of allowing loan-level covariates impact the forecast of transition probabilities. These models follow from the Markov chain work of Jarrow et al. (1997) on the term structure of credit risk spreads.

6. The competing risks framework, where loans can move from an initial state, e.g. performing status, into multiple states. Such a model was adopted for the residential model of the US Federal Reserve's Comprehensive Capital Analysis and Review (CCAR, see Board of Governors of the Federal Reserve (2012)) exercise, where the non-performing states modelled were prepayment, impaired (90-180 days past due) and default (greater than 180 days past due) [1].

The model used here to estimate the probability of default is a continuous time maximum-likelihood estimator of a Markov chain which allows for covariates to impact the probability of loan transition between specified states. Lando and Skodeberg (2002) describe the advantages of continuous time estimators of transition matrices over the discrete time, or "cohort" methods of estimating transition

[1]For a list of Frequenty Asked Questions relating to the CCAR exercise, see http://www.federalreserve.gov/newsevents/press/bcreg/bcreg20120420a1.pdf

probabilities that are more commonly used by credit risk practitioners. This framework allows a more accurate estimate of the probability of rare events occuring, can deal easily with data censoring, as well as the modelling of the impact of covariates on transition probabilities. The incorporation of the effect of loan-level covariates into transition probability estimation is the crucial component that allows a bi-directional, loan-level cash flow engine that can respond to changes in a macroeconomic scenario to be developed.

We attempt to briefly provide a list of advantages of the framework presented here over the families of commonly-used models outlined above.

1. Accuracy is increased in a model which incorporates an estimated empirical relationship between **Time Since Default** and the probability of cure at the loan level. Kelly and O'Malley (2014) and McCann (2014) confirm the existence of this relationship for the Ireland and UK residential mortgage markets, respectively. The finding implies that, as loans' duration in default gets longer, the likelihood that the loan will return to repayment decreases. Models without an estimate of such a relationship may only allow for its existence using an overlay. The incorporation of such an effect is always relevent, but particularly crucial in markets with high opening default rates such as Greece, Ireland and Spain, as it provides an in-built distinction between the stock and the flow of defaults.

2. Cross-sectional estimation techniques such as the logit cannot identify the existence of cures or their relationship to loan-level and macro covariates. When cross-sectional data is used, the estimation of cure probabilities must rely on the imposition of an assumed cure rate, or their incorporation into an assumption around the "roll to liquidation" for an estimated default stock.

3. Cross sectional loan-level models such as logits can only provide an estimate of the *lifetime* probability of default for a loan. A cash flow model which incorporates the timing of this default, the cure from default, and the timing of repossession cannot be devised based on PD estimates from a logit without extensive use of mechanical overlays and assumptions.

4. In an environment where year-on-year default flows are not reliably specified, it is impossible to allow for an impact of the **Time to Repossession** (TTR) to enter the model. Modelling changes in Time to Repossession (the time between default and the repossession of collateral, also referred to as the "Emergence Period") can allow for anticipated policy changes to have an impact on model outcomes and is of heightened importance in an environment such as Ireland where the legal and policy environment have been evolving considerably in response to the financial crisis.

5. Assumptions around the share of a defaulted loan that rolls to repossession, which would be

necessary in a cross-sectional model where repossessions were not explicitly modelled, can be avoided in the framework proposed here. Rather, for a loan that defaults at $t = 1$, we can model losses by allowing an estimated component of that default to cure at years $t = 2...t = (1+TTR)$ and allowing the lender to repossess that component of the defaulted loan which has not been estimated to cure by the year $t = (1 + TTR)$.

6. Relative to portfolio-level models, the framework proposed here allows for granular estimation of the impact of loan- and macro-level covariates on the movement between performing and default status.

7. From an accounting perspective, the estimation of year-on-year flows into both default and performing status, combined with an assumption around the Time to Repossession, allows the timing of expected losses to be explicitly modelled. In cross-sectional models, in order to turn model estimates into three-year expected losses, an assumption must be made around the share of the lifetime losses that occur in the desired horizon.

An obvious drawback of this framework is the reliance on loan-level data, which can be time-consuming to collect, remediate and understand. Further, if practitioners wish to run a multi-state PD model such as that used in this paper, the process involves the use of a computationally complex, time-intensive program.

The paper proceeds as follows: Section 1.1 provides an overview of the mechanics of the model; Section 2 provides a description of the data sources used; Section 3 describes the PD estimation method and results; Section 4 describes the cash flow engine used to generate estimates of Exposure at Default; Section 5 describes the Loss Given Default methodology; Section 6 concludes.

1.1 Model overview

Figure 1 provides an overview of the entire model environment. A range of historical input data is required in order to run the model, including the loan-level data used to create the estimation data set and macroeconomic time series that provide covariates for estimation.

The Loan Loss Forecasting model (LLF) estimates expected annual cash flows at the loan level. Cash flows, in this context, include flows between performing and defaulted states, as well as repayments of principal and write-offs of defaulted balances. The model combines information on each loan's characteristics, the macroeconomic scenario inputs, coefficients on transitions to default and cure and information on starting collateral to estimate loss given default (LGD). The model described by this paper estimates cash flows for each loan at the end of 2014, 2015 and 2016, using "scoring" inputs as of December 2013. The PD model is run in the MSM package (Jackson, 2011) in the R software. The

LLF cash flow engine, in which annual transition probabilities, cash flows and losses are calculated, is hosted in a custom-built Graphical User Interface (GUI) designed by the authors in the gWidgets package of the R software.

The model provides a range of outputs, as reported in the green box of Figure 1: performing balance, defaulted balance, annual default and cure flows, interest payments, life time cures and expected losses.

Figure 2 visualises the evolution of a hypothetical performing loan in the LLF. Certain aspects are simplified in this illustration. For instance, the probabilities of default (PD) and cure (PC) are fixed at 5 and 10 per cent, respectively. In practice, these probabilities change according to their time-variant determinants, such as the macroeconomic scenario and the loan's age. The key advantage of the LLF framework over stock delinquency frameworks, in which lifetime PD is estimated using a logit, is that the coefficients of the model can be directly transformed into probabilities of transition both into and out of default, over a specified time horizon. The red arrows in each case indicate that 5 per cent of period t's performing balance will transition to the defaulted state at $(t + 1)$, while 90 per cent of period t's defaulted balance will remain in default at $(t+1)$. The green arrows indicate that 10 per cent of the default balance at t will move to performing at $(t + 1)$, while 95 per cent of the period t performing balance will remain performing at $(t + 1)$. Performing balances are then amortised at a rate of 2 per cent in this illustration.

Figure 3 provides an illustration of how the LLF uses the default flows in Figure 2 to estimate expected loss. Time to repossession has been input as two years. At $t = 1$ in this hypothetical example, €5,000 is expected to be in default. The two green arrows with cure rates of 10 and 8 per cent of the balance indicate that the portion of the loan that defaulted at $(t + 1)$ will be allowed to cure for two years, and that the probability of cure will fall as time elapses. If time to repossession were set to three years, there would be a third year of cure from the $(t + 1)$ default flow, with a probability lower than 8 per cent. This pattern is generated automatically in the model by the inclusion of an estimated coefficient on Time Since Default in the cure equation.

The LLF cash flow engine models losses in a very granular fashion. The LLF treats flows into default in different periods separately, even though they may emanate from the same initial $t = 0$ performing loan. Therefore, "lifetime cure" and expected loss are calculated on the distinct default pools originating from each loan. In the hypothetical example in Figure 3, the €93,100 that does not default in period 1 has a PD applied to it in period 2 (in this case 5 per cent). This €4,655 is a $t = 2$ default flow, and from this pool, there will be $t = 3$ and $t = 4$ cure flows of 10 and 8 per cent, respectively. At $t = 3$ in this example, there are two cure flows occuring: a one-period cure from the $t = 2$ default flow, and a two-period cure from the $t = 1$ default flow. Due to the negative coefficient on Time Since Default in the $PCure$ equation, it will always be the case that the cure rate in the second period will be lower than that in the first period.

Standard Financial Statement (SFS) data are only used when an enhanced Loss Given Default model, (Gaffney and Dunne, 2014)), which allows banks to provide mortgage modifications to borrowers in default, is active. This satellite model is described in more detail in Section 5.2.

Figure 1: Overview of LLF

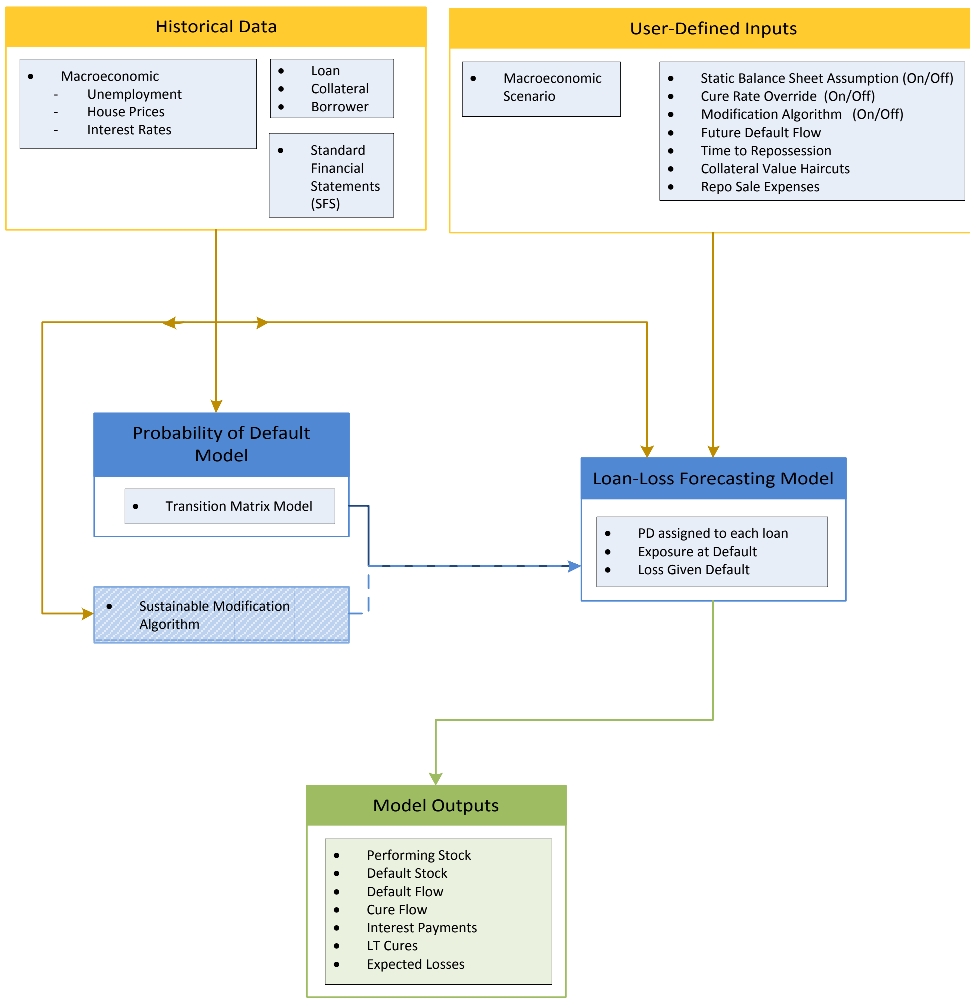
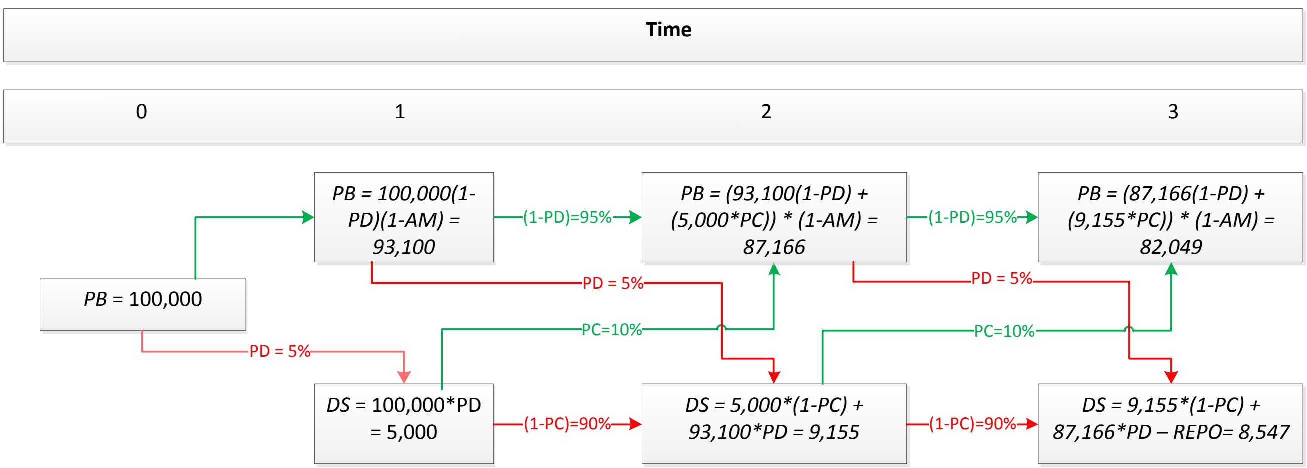
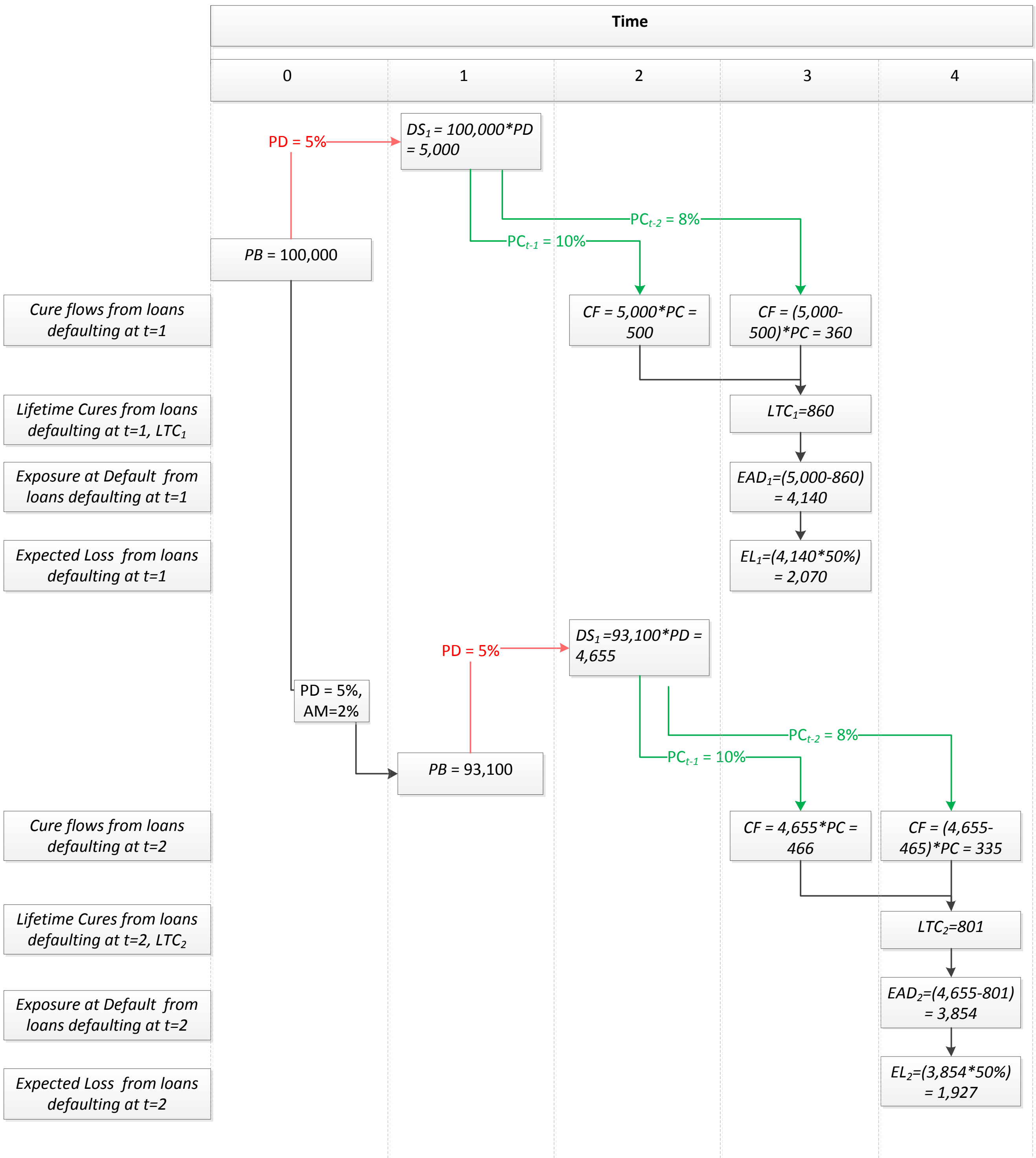


Figure 2: Default and cure flows for 3-year horizon of a hypothetical performing loan



Hypothetical loan with a $t = 0$ balance of €100,000, a constant set of parameters: PD of 5%, $PCure$ of 10% and an amortisation rate (AM) of 2%. PB refers to performing balance, DS to default stock in each year. PD and $PCure$ will vary at the loan level and will derive from the loan-level multi-state model's coefficients. $REPO$ refers to the €4,050 of $t = 1$ default stock that has not cured by $t = 3$, and is therefore repossessed.

Figure 3: Derivation of expected loss for portion of loan defaulting at $t = 1$ and $t = 2$



Hypothetical loan with a $t = 0$ balance of €100,000, and a constant set of parameters: PD of 5%; $PCure$ of 10% and 8% one year and two years after default, respectively, reflecting the negative impact of time in default in the model; Loss Given Default of 50%. PB refers to performing balance, DS refers to default stock. PD and $PCure$ will vary at the loan level and will derive from the loan-level multi-state model's coefficients. PD does not vary over time in this example due to the simplifying assumption of an unchanging macroeconomic environment for the purposes of this display.

2 Data

Since 2011, the Central Bank of Ireland has received loan-level data submissions from banks participating in the Financial Measures Programme[2] (FMP). The Irish mortgage data set covers €90,539 million at June 2013, or 66 per cent of the Irish mortgage market. The UK mortgage data cover €29,814 million at June 2013, or 2.01 per cent of the outstanding mortgage credit in the UK.[3] These data cover the population of loans at the participating institutions, with information provided on the loan's balance, arrears performance, origination, maturity, interest rate, loan to value ratio and a range of other loan-specific variables of interest.

A monthly history of loan performance can be created for an individual loan by linking across the 5 data submissions that have been received at June 2013 using unique loan-level identifiers. At each data submission date, a 12-month history of arrears balances is submitted. Table 1 reports the arrears information used from each data submission, with the historical data set totalling 64 months in Ireland and 43 months in the UK.

Table 1: Arrears history information in Central Bank of Ireland loan-level data

Submission Date	Arrears information provided	
	Ireland	UK
June 2013	Jan - Jun 2013	Jan - Jun 2013
December 2012	Jan - Dec 2012	Jan - Dec 2012
December 2011	Jan - Dec 2011	Jan - Dec 2011
December 2010	Jan 2008- Dec 2010	Dec 2009 - Dec 2010
Number of months	64	43

The data sets for PD model estimation are constructed as a quarterly loan-level panel in the case of ROI and a monthly property-level panel in the case of the UK. The UK data set must be constructed at the property, rather than loan level due to the fact that the December 2010 data was submitted at the property level only. In all subsequent data drops, loan-level data must be aggregated to the property level (using unique identifiers) to allow property-level data sets to be merged to create the arrears time series.

[2] http://www.centralbank.ie/regulation/industry-sectors/credit-institutions/pages/financialmeasuresprogramme.aspx

[3] Mortgage market data are available in Table A5.3, "Lending secured on dwellings", of the Bank of England's Monetary and Financial Statistics, Section A, "Money and Lending". http://www.bankofengland.co.uk/statistics/Pages/bankstats/2013/jun13/default.aspx

2.1 Scenario design

In order to arrive at a loan loss forecast, scenario inputs must be provided. As an example of the type of information required to run the LLF, we provide details on the Irish baseline and adverse scenarios that are used in the Comprehensive Assessment (CA) of European banks that is being carried out throughout 2014 by the European Banking Authority (EBA) and Single Supervisory Mechanism (SSM). Figure 2 reports the numbers for GDP growth, house price inflation, the unemployment rate, 10-year government bond yields and interest rates for Ireland, while Figure 3 provides the analogue for the UK. The numbers in the adverse scenario are reported as deviations from the baseline. The baseline scenario for Ireland is for a gradual improvement in economic conditions from end-2013, with GDP growing by around 2 per cent per annum, unemployment falling to 11.4 per cent and interest rates remaining unchanged. The way in which the model is developed in this paper means that only unemployment, house prices and interest rates have an effect on EL forecasts.

Table 2: EBA/SSM baseline and adverse scenarios for Ireland

	Baseline			Adverse		
Variable	2014	2015	2016	2014	2015	2016
GDP	1.8	2.9	2.4	-3	-3.6	-1.9
HICP Inflation	0.8	1.1	1.2	-0.1	-0.7	-1.1
House Price Inflation	3.6	6.9	7.9	-7.1	-7.9	-6.5
Unemployment Rate	11.9	11.2	11.4	0.5	1.7	2.6
10-year bond yield	3.4	3.7	3.8	148bps	109bps	109bps
Interest Rates	0	0	0	80bps	0	0

Table 3: EBA/SSM baseline and adverse scenarios for UK

	Baseline			Adverse		
GDP	2.5	2.4	1.6	-3.3	-3.7	-1
HICP Inflation	2	2	2.1	-0.3	-1.9	-3.3
House Price Inflation	4.9	5.1	3.5	-12.4	-13.8	-7.8
Unemployment Rate	6.8	6.5	6.4	1.9	4.4	5.1
10-year bond yield	2.8	3.2	3.3	+139bps	+103bps	+103bps
Interest Rates	0	0	0	80bps	0	0

3 Probability of Default Estimation 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 $t+1$. Kelly (2011) develops such a model for the Irish mortgage market, with Kelly and O'Malley (2014) adding the effect of loan-level covariates, while McCann (2014) carries out a similar analysis for the UK mortgage exposures of Irish banks.

The application of a time-homogenous Markov multi-state model to ROI and 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 $p_{rs}(u; t+u)$ 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:

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

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 $s \neq r$

$$q_{rs}(t; z(t)) = lim_{\delta t \to 0} P(S(t + \delta t) = s | S(t) = r)/\delta t. \tag{2}$$

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

$$Q(t) = exp(\Lambda t) \tag{3}$$

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

$$\lambda_{rs} = \frac{N_{rs}(T)}{\int_0^T Y_r(s)ds} \tag{4}$$

where $N_{rs}(T)$ is the total number of transitions from r to s in the sample period and $\int_0^T Y_r(s)ds$ is the total number of "loan-month" or "loan-quarter" combinations that are spent in state s. All information on time spent in all states is now being used in estimation (Lando and Skodeberg, 2002).

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

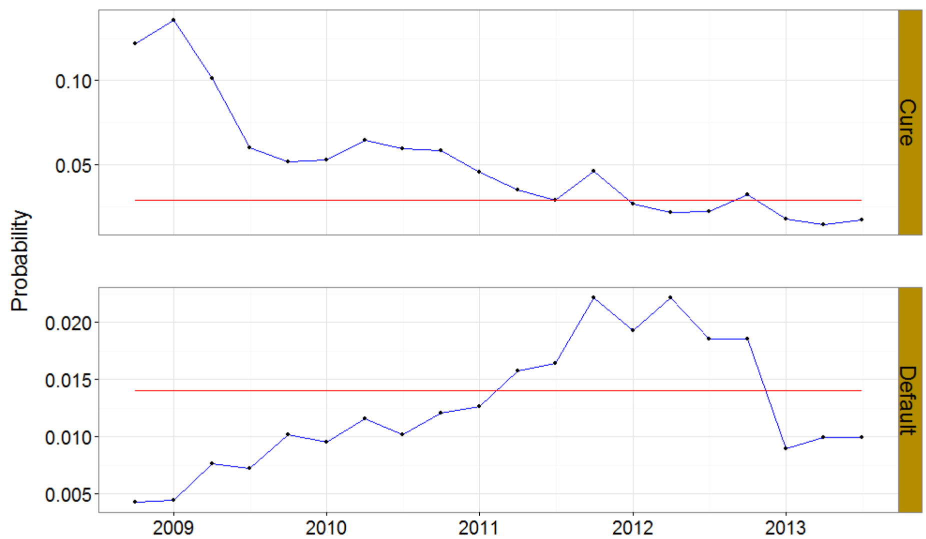
$$q_{rs}(z_{ij}) = q_{rs}^{(0)} exp(\beta_{rs}^t z_{ij}) \tag{5}$$

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 $q_{rs}^{(0)}$ and β_{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 the hazard ratio $exp(\beta_{rs})$.

The quarterly empirical transitions in the ROI data are plotted in Figure 4, along with the horizontal red line which depicts the transition probability for an MSM model with no covariates. The plots report a sharp quarter-on-quarter increase in PD transition rates from 2009 through to late 2011, with quarterly PD remaining close to this peak for four quarters before falling to a lower level in 2013, reflecting the improving conditions in the Irish economy. $PCure$ on the other hand was consistently above the model-estimated overall level from 2008 through to mid-2011, and has been stable since early 2012.

The covariates included in the PD models are outlined in Table 4. The model specification is

Figure 4: Quarterly empirical transition rates and model-estimated transition probability, ROI mortgages



almost identical in the ROI model of Kelly and O'Malley (2014) and the UK model of McCann (2014), with the only exception being that the UK model does not include loan age among the covariates. The coefficients of the ROI model are reported in Kelly and O'Malley (2014) and McCann (2014). We focus in the rest of this section on the economic magnitude of the effect of selected coefficients on interest.

The impact of Time Since Default on the $PCure$ in the ROI and UK models is graphed in Figure 5. In the ROI model, a loan that has been in default for one quarter has a cure probability above 30 per cent. This probability falls to 12 per cent for loans in default one year, and to under 5 per cent for loans in default for two years. The UK model, where cures are more prevalent in the sample period, predicts that a loan that has been in default for 25 months has a 25 per cent chance of curing, while this falls to under 10 per cent for a loan in default for 50 months. These decreasing cure probabilities represent an important hysetesis effect of long durations in default - if loans do not cure in a relatively rapid timeframe, the probability that they will return to performing status diminishes greatly. These findings have important implications for banks' and policy makers' responses to increases in mortgage arrears: the earlier that potentially delinquent mortgages are dealt with through engagement with the

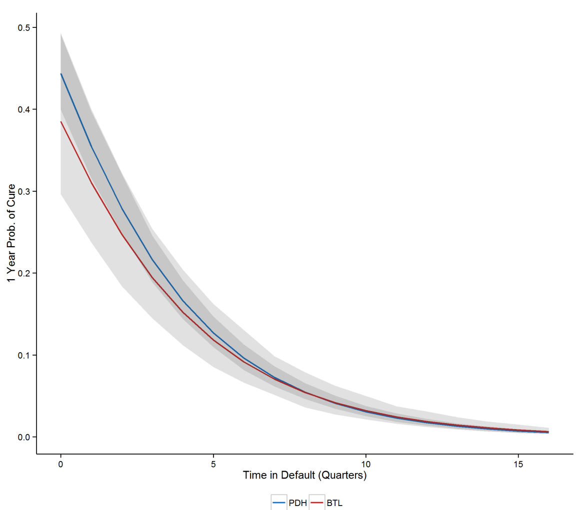
Table 4: Description of covariates included in PD models

Factor	Comments	ROI	UK
Bank ID	Intercept adjustment for bank specific effects for Banks 2, 3 and 4. All coefficients are relative to baseline of Bank 1.	Yes	Yes
Buy-to-Let	Intercept adjustments for buy-to-let mortgages. Baseline is Primary Dwelling Houses.	Yes	Yes
Interest Rate Type	Intercept adjustments for interest rate type effects for Standard Variable Rate and Tracker mortgages. All coefficients are relative to baseline of fixed rate mortgages.	Yes	Yes
Vintage	Vintage (i.e. loan age) is measured in months since the mortgage was issued. Both linear and natural-logged terms enter into the functional form of the model.	Yes	No
Interest Rate	Current interest rates on the mortgage.	Yes	Yes
Time in Default	Time (in months) since loan entered into Default state.	Yes	Yes
Current Loan-to-Value	Current loan-to-value at the property level.	Yes	Yes
Unemployment	National unemployment rate is converted to regional by the model.	Yes	Yes

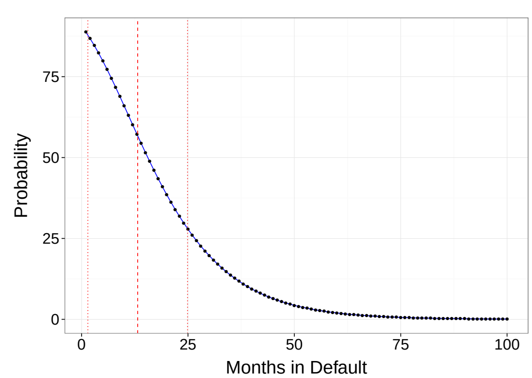
lender and potential loan modification, the more this effect can be mitigated.

Figure 5: Variation in $PCure$ as a function of Time Since Default

(a) ROI model, Kelly and O'Malley (2014)

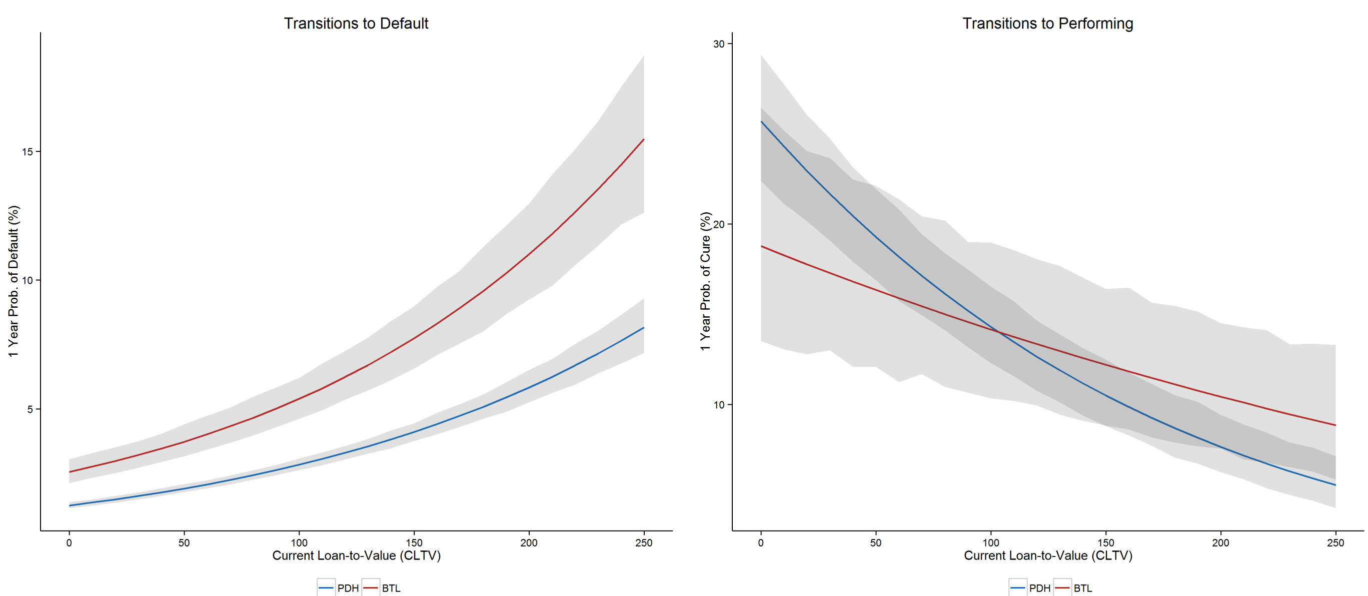


(b) UK model, McCann (2014)



The role played by housing equity in mortgage default is the subject of intense debate. Both Kelly and O'Malley (2014) and McCann (2014) find that higher levels of Loan to Value ratio (LTV) lead to higher PD, but that this effect is more important among buy-to-let (BTL) investor mortgages relative to Principal Dwelling House (PDH) mortgages. Figure 6 reports changes in PD and $PCure$ for varying values of LTV from the Kelly and O'Malley (2014) model. A BTL investor with an LTV of 100 has a PD of 5 per cent, with this probability rising to 12 per cent for an LTV of 200. In the owner-occupier market, the change in PD for commensurate changes in LTV is an increase from 2.5 to 5 per cent. The less important role played by housing equity in explaining owner-occupier defaults lends support empirically to the view that homeowners are more likely to continue to pay their mortgage despite being "out of the money", in order to remain in their homes. Instances of "ruthless" or "strategic" default in response to equity shortfalls appear to be more likely among investors, where the (non-pecuniary) costs of repossession are lower.

Figure 6: The role of housing equity in PD and $PCure$, ROI model, Kelly and O'Malley (2014)



4 Exposure at Default

EAD is calculated at the loan level using information available at December 2013. This is done by modelling expected flows of loan balances between performing and default status, and from default balance to recoveries and cures, over the stress horizon.

All loans are classified as either "amortising" or "non-amortising". It is assumed that all amortising loans will continue to pay down under the amortisation schedule which covers both the principal and interest due on the December 2013 balance, using as inputs the term remaining and interest rate on the loan. In the EAD framework, for non-amortising positions, the amortisation deduction is skipped when calculating balances over the stress horizon.

The interest rate at December 2013, or $t = 0$ is taken from the loan-level tape for each loan. It is specified as a yearly rate. For forecasting purposes, a user-specified index is applied to determine future interest rates, r_t. Interest rates on fixed rate loans remain constant regardless of the future changes in interest rates inputted. For variable and tracker mortgages, the interest rate (r_t path is specified as:

$$r_t = r_{(t-1)} + index_t \tag{6}$$

Where $index_t$ is the interest rate change given by the inputted scenario. In the case of the EBA/SSM adverse scenario given in Table 2, each variable and tracker loan will have r_1, r_2 and r_3 that is 80 bps higher than that at $t = 0$

In the EAD model, the performing loan balance will be prepaid, be paid down or become delinquent. Each year, shares of the performing balance are allocated to each of these three categories by expected value. To determine the amount that is paid down (amortised), a simple re-arrangement of the standard annuity formula is applied. Given that a loan is performing, the balance-independent amortisation rate AM_t is applied to the performing balance:

$$AM_t = \frac{r_t}{((1 + r_t)^{n_t} - 1)} \tag{7}$$

where r_t is the annualised interest rate and n_t is term remaining expressed in years.

The PD and $Pcure$ are assigned to loans over time by a function built into the R code for the LLF framework which uses the coefficients as estimated by the multi-state model reported in Kelly and O'Malley (2014) and McCann (2014). The coefficients for each model covariate are represented by a 2x2 matrix of movement between performing and default status from years t to $t + 1$.

The PD and $Pcure$ are represented as a three-dimensional $n * m * p$ array. n is the number of years in the stress horizon; m is the number of loans (or cohorts of loans). Here, p equals one plus the maximum number of years to repossession. This sum is the number of states estimated at each point in time. One state is for performing stock (containing PDs) and the others are for defaulted stock (containing PCures); there is one PCure for each possible value of time since default, which affects the probability of cure.

The coefficients from the multi-state model are combined with values for the time-variant fields for each year, whether user-specified or derived from loan-level data. The EAD model is constructed as a balance flow tree that forecasts the expected performing and defaulted loan balances through time. Each year, PD is used to estimate the proportion of the previous year's performing balance that will be in default in the next period, while $PCure$ is used to estimate the proportion of the previous year's defaulted balance that is expected to cure. After applying default and cure flows, interest is charged on the performing balance. The performing balance is then amortised according to the amortisation rates described above, except for non-amortising contracts such as interest-only loans. The remaining balance is reduced by an amount of expected prepayments determined by the prepayment rate assumption specified in the scenario. Only performing loans incur amortisation and prepayments. Finally, a user-specified share of amortisation and prepayments is added as new lending. Under the "static balance sheet" assumption, this increases the performing balance of each loan by an amount equal to amortisation and prepayments. The Performing (PB) and Delinquent (DB) balances for year t are calculated as follows:

$$PB_t = PB_{t-1} \cdot (1 - PD_t) \cdot (1 - AM_t)(1 - PP) + DB_{t-1} \cdot (PCure_t) \tag{8}$$

$$DB_t = PB_{t-1} \cdot PD_t + DB_{t-1} \cdot (1 - PCure_t) \tag{9}$$

where PD_t is the estimated probability of default, AM is the amortization rate, PP is the user-inputted prepayment rate and $PCure$ is the estimated probability of cure.

Two schematics demonstrate the movements between performing and defaulted balances in the model. These schematics provide a complete view of the life of a performing and defaulted loan, expanding on the illustration in Figure 2. Figure 7 shows the evolution of initial performing balances, while Figure 8 shows an initial defaulted balance. Performing balances are blue, while defaulted balances are grey. On the left-hand side of the schematics, the first node represents the total balance for each loan (or cohort of loans) at $t = 0$. The arrows show how each balance at $t = 1$ is calculated; they are annotated with the probability of reaching the new state from the initial state. The paths for performing balances in each period are calculated using PD and $1 - PD$ (blue arrows), while the paths for defaulted balances are calculated using $PCure$ and $1 - PCure$ (grey arrows).

For example, when the balance is performing at $t = 0$, the probability of default in year 1 is applied to generate the expected defaulted balance. This is represented by movement from nodes A1 to B2 on the diagram. The proportion of the balance that does not default is amortised and prepaid, as shown by movement from nodes A1 to B1. In the second iteration, the same process is repeated for the performing node creating a performing stock at $t = 2$ (C1) and a new defaulted stock at $t = 2$ (C2). The cure rate is then applied to the defaulted balance (B2) at $t = 1$, which generates an additional performing stock (C3). The balance not cured remains in default (C4). Nodes C2 and C4 each represent defaulted balances, but they remain distinct to correctly model the effect of time

Figure 7: Schematic of flows for loans beginning as performing

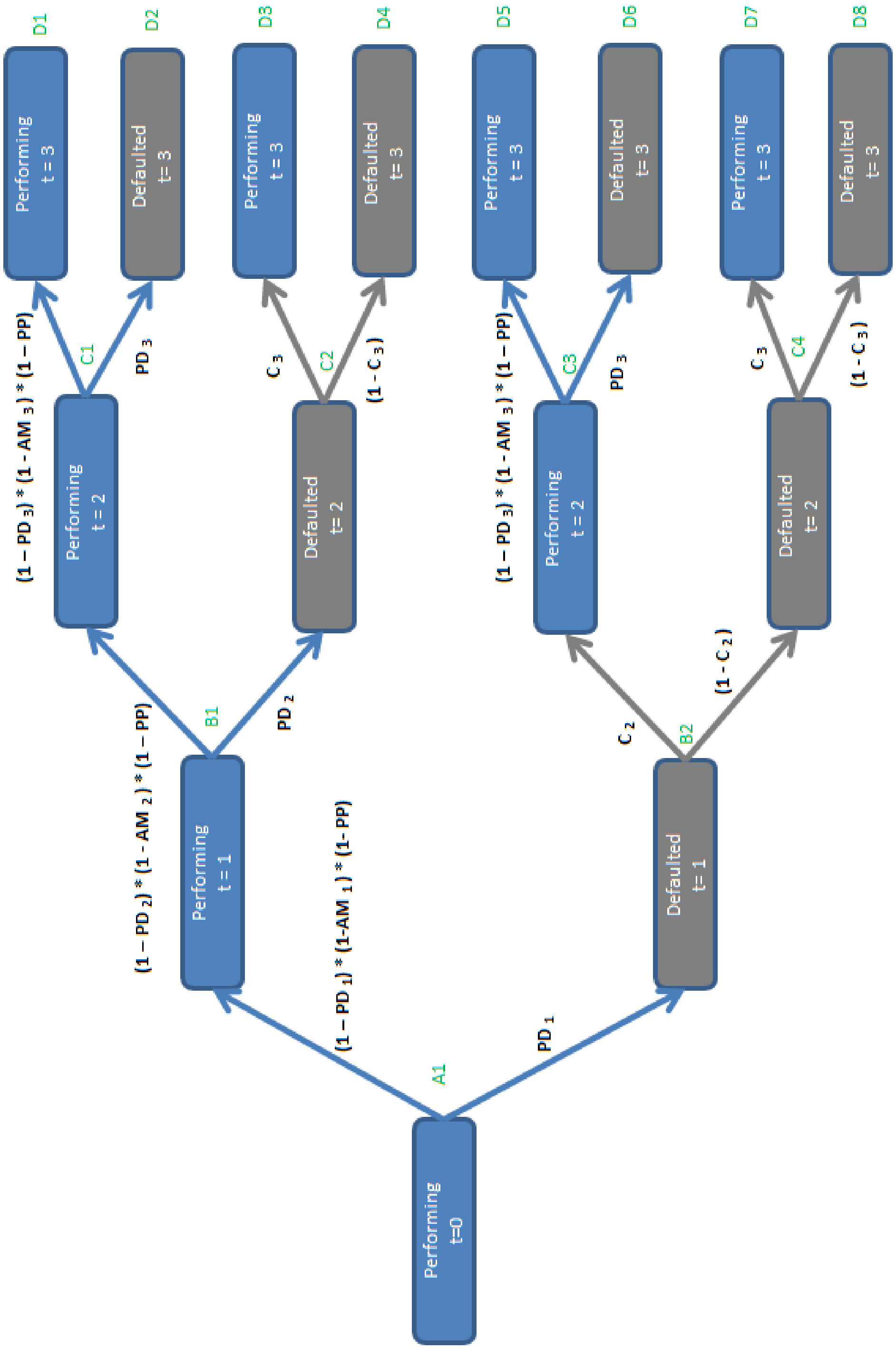
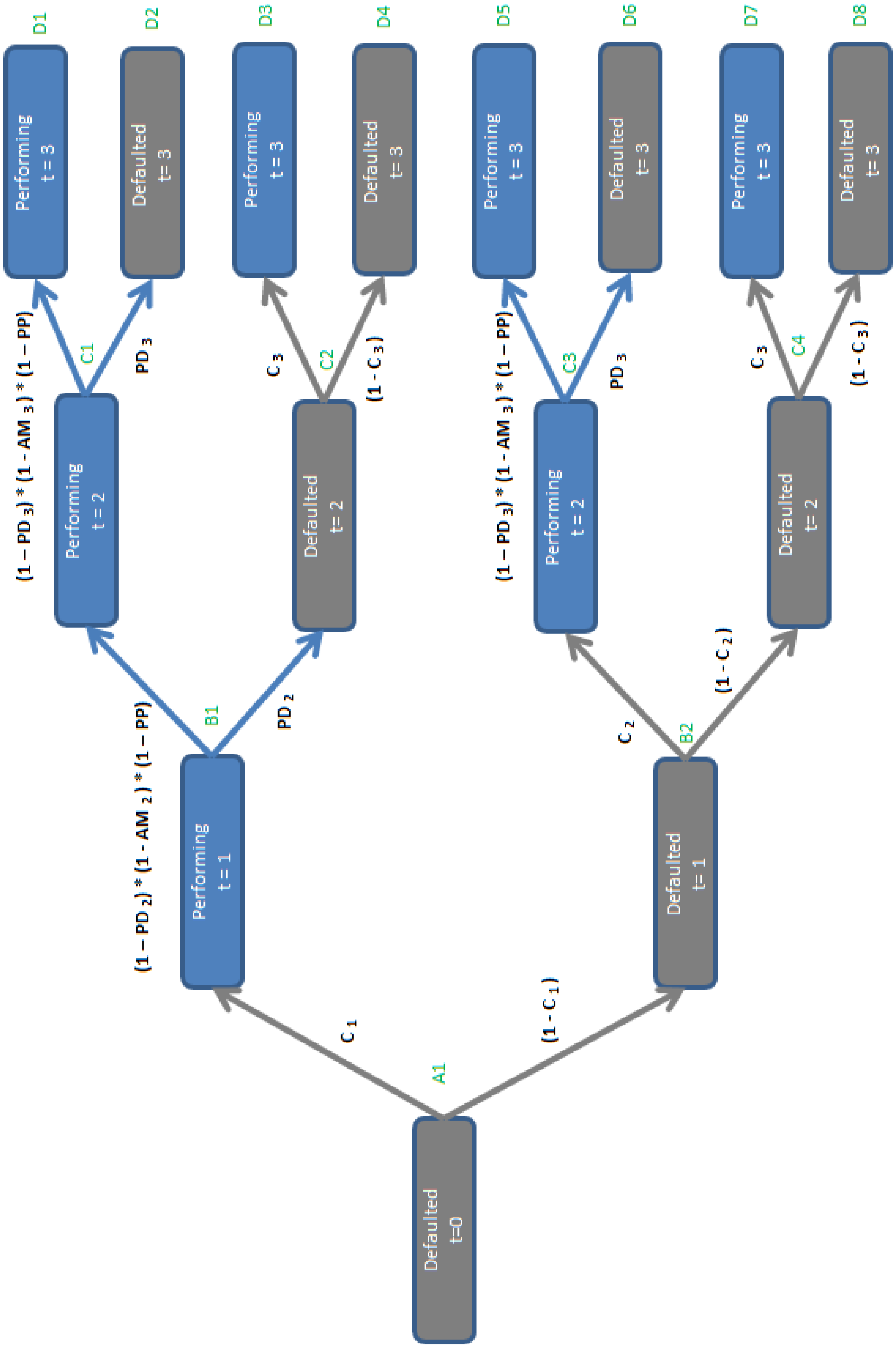


Figure 8: Schematic of flows for loans beginning as defaulted



since default on $PCure$. Therefore, in each period, the total performing and defaulted balance are calculated by summing the appropriate nodes, as follows:

Table 5: Stock of performing and defaulted balances for initial performing loans

Time period	Total Performing Balance	Defaulted Balance
t=0	A1	-
t=1	B1	B2
t=2	C1 + C3	C2 + C4
t=3	D1 + C3 + D5 + D7	D2 + D4 + D6 + D8

Expected losses are calculated based on the flow into default. For initially defaulted loans, a provision is raised against the whole balance, net of recoveries. For each year thereafter, the provision is calculated based on the level of new expected defaults. Using the balance flow tree, it is straightforward to determine this flow directly from the nodes on the tree. The flow into default over a three-year horizon for loans that begin as performing loans is given in Table 6, with reference to the nodes of Figure 7. Similarly, the flow into default over a three-year horizon for loans that begin as defaulted loans is given in Table 7, with reference to the nodes of Figure 8.

Table 6: Flow into default: stock performing at t=0

Time Period	Default Flow
t=0	-
t=1	B2
t=2	C2
t=3	D2 + D6

Table 7: Flow into default - stock defaulted at t=0

Time Period	Default Flow
t=0	A1
t=1	-
t=2	C2
t=3	D2 + D6

5 Loss Given Default

When a loan is deemed to have defaulted and the collateral is repossessed, the proceeds from sale of the property are used to offset the balance of the loan and cover the associated expenses. It is not always the case that the collateral is sufficient to absorb all costs. As a result a loss may be realised.

The LGD is defined as the ratio of this loss to the remaining balance. The factors which will impact the LGD are house prices (which are provided as part of the macroeconomic scenario in Figures 2 and 3) and the legal/administrative and haircut costs associated with the recovery process. The inputs to LGD calculation can be summarised as follows:

- Indexed LTV at time t, calculated from
  - Total loan value - LTV is calculated at the property level so multiple loans secured on one property are aggregated. This total debt balance is then reduced each year in proportion with the amortisation rate.
  - Future expected collateral valuation, which will move in line with the macroeconomic scenario house price forecast.
- A liquidity haircut (often also referred to as a "fire sale discount") which accounts for the fact that distressed properties will likely sell at a discount to those sold in normal conditions.
- Legal and other costs associated with a distressed sale.

The LGD is not estimated by an econometric model but is defined using a formulaic approach. In basic terms, the LGD for a mortgage at the point of default t can be thought of as the following expression:

$$LGD = max(-\frac{CVLH_t - OB_t - C_t}{OB_t}, 0) \tag{10}$$

Where $CVLH_t$ is the collateral value net of liquidity haircut, OB_t is the outstanding loan balance and C_t denotes legal and other costs. The intuition behind the $Max(.)$ statement is that when the collateral value, net of liquidity haircut and legal costs, exceeds the outstanding balance on the loan, LDG is set the floor of zero, i.e. a negative LGD is not plausible.

The recovery rate is defined as 1 minus the LGD. Re-arranging this equation for use in the model results in the following expression used to generate the future recovery.

$$RR = 1 - LGD \tag{11}$$

The model has been calibrated to ensure that the liquidity haircut is applied to the expected value of collateral, while the legal costs are defined as a percentage of the loan balance. Both the liquidity haircut and the impact of legal costs are inputs into the loan loss model. The LGD is always applied to the expected delinquent balance net of cures. The extent to which the loan balance will amortise, be prepaid or provided for will depend on loan characteristics.

5.1 Indexing LTVs and Expected Losses

Property values at $t = 0$ are provided by the subject banks. In order to provide the most granular estimate possible of future house prices, changes in the national house price index, as inputted in the macroeconomic scenario, are converted into national sub-indices. In the UK case, these are the NUTS1[4] regions of the UK. In the Irish case, properties are classified as either being in Dublin or outside Dublin, and being houses or apartments. In both models, for each of the categories mentioned ("sub-index"), an OLS regression using quarterly data from 2006 to 2011 of the following form is run:

$$ln(SubIndex_t) = \alpha + \beta ln(NHP_t) + e_t \tag{12}$$

Where NHP_t is the national house price index. The log-log nature of this specification means that the coefficients can be interpreted as percentage changes in the $SubIndex_t$ for a given 1 per cent increase in the national house price index. This provides for a simple mapping from the inputted macroeconomic house price forecast, which is given in terms of percentage changes from the index at (t=0). Results on these adjustments to the macroeconomic scenario are reported for the ROI in Figure 8 and the UK in Figure 9. The interpretation of the coefficient for Dublin houses in Table 8 is that, for a 1 per cent increase in the national house price index, prices of properties in Dublin will be indexed forward by 1.13 per cent plus the constant term (1.13 - 0.6115).

Table 8: House Price Forecast Adjustments, Republic of Ireland

	Constant	Slope
Dublin Houses	-0.612	1.134
Non-Dublin Houses	0.532	0.889
Dublin Apartments	-1.917	1.388
National Apartments	-1.983	1.401

The final inputs to the LGD calculation are the liquidity haircut and legal/administrative costs parameters. Liquidity haircuts are applied to the estimated property valuation at repossession to reflect the fact that distressed properties are generally sold quickly, potentially at a time when their placement on the market may exert downward pressure on prices, and that buyer preferencs may dictate that demand is lower than normal for such properties. The LLF framework allows for differing liquidity haircuts to be applied to primary dwelling and buy to let properties.

The LGD is implemented as a Tobit model to reduce distribution censoring. Z is the LGD as

[4] Nomenclature of Territorial Units for Statistics

Table 9: House Price Forecast Adjustments, United Kingdom

Region	Slope	Constant
East Anglia	0.984***	0.0585*
East Midlands	0.969***	0.158***
London	1.124***	-0.674***
North	0.910***	0.442***
North West	0.915***	0.432***
Northern Ireland	1.029***	0.106
Outer Met	0.998***	-0.0226
Outer South East	1.038***	-0.176***
Scotland	0.828***	0.753***
South West	1.035***	-0.164***
Wales	0.965***	0.187**
West Midlands	0.908***	0.451***
Yorkshire And Humberside	0.941***	0.304***

defined above, σ is the uncertainty other costs and the liquidity discount.

$$LGD(tobit) = N(\frac{z}{\sigma})z + \frac{\sigma}{2\pi}exp(-0.5(\frac{z}{\sigma})^2) \tag{13}$$

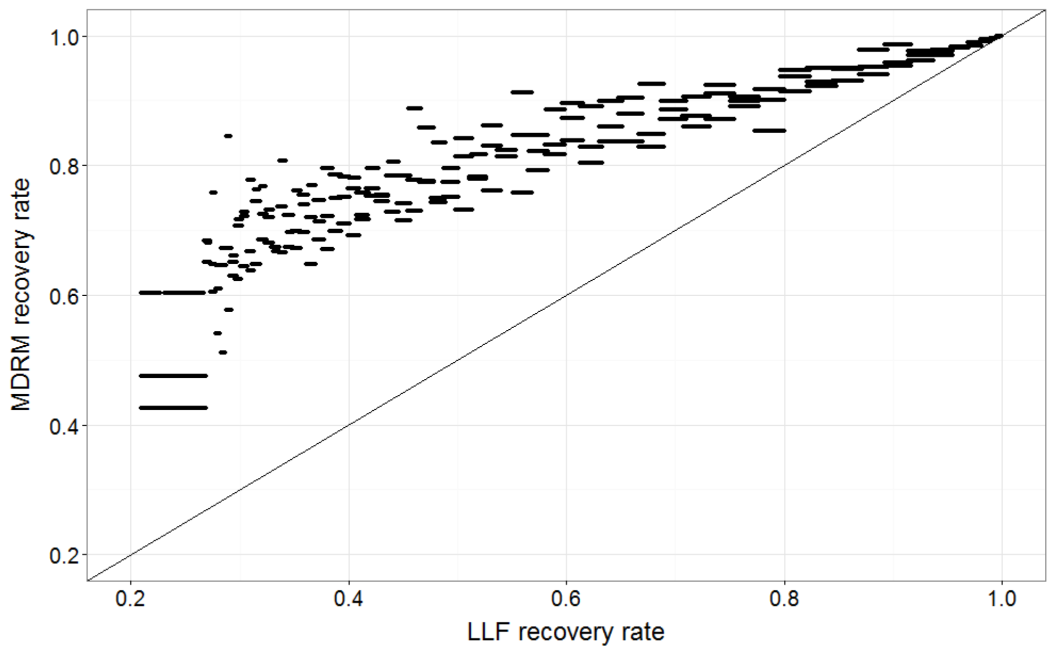
5.2 Satellite Model: Enhanced recovery rates through modification

All estimates of LGD are based on a model in which all defaulted loans, net of the probability of cure, are repossessed once the "time to repossession" window has passed. A model in which these repossession-based LGDs can be altered by allowing banks the choice between a range of mortgage modification options and repossession is proposed by Gaffney and Dunne (2014). The algorithm creates affordable, sustainable modifications for borrowers in mortgage arrears based on Standard Financial Statement (SFS) returns. These include term extensions, interest rate reductions, write-downs and split mortgages. Losses given default are calculated in the same manner as in the loan-loss forecasting tool; the effect of modifications is to prevent repossessions and thereby reduce LGD. The SFS sample is stratified, balance-weighted LGD averages are computed, and each loan's LGD in the loan-loss forecasting model is replaced by the modified LGD from the corresponding SFS stratum.

The algorithm takes SFS data provided by distressed borrowers on their income, expenditure and family size to arrive at a calculation of a sustainable mortgage given the borrower's circumstances. The algorithm allows a bank to sequentially compare the net present value (NPV) of a range of modifications against the return from repossessing the property. The modification procedure is as follows:

1. If the mortgage appears affordable on current terms, no modification is offered.
2. Expand the term of the mortgage by increments of one month, stopping at sustainability or the

Figure 9: Comparison of base LLF and adjusted recovery rates (Gaffney and Dunne, 2014)



maximum of a mortgage term of 40 years or a borrower age of 70 years.

3. Set aside an amount equal to the difference between the sustainable mortgage and the outstanding balance. Depending on circumstances, parts of this amount can be either written off or warehoused.

4. If the modification is worthwhile compared to repossession, it is offered. Otherwise, foreclosure proceeds.

The LGD model is equipped with the option to activate the enhanced LGD model, altering recovery rates and LGDs for the bank. Given that SFS data is not available for all loans in the data, weighted-average adjusted LGDs must be applied to strata of the mortgage book. Figure 9 outlines the relationship between adjusted recovery rates and those from the repossession-based LLF model. By construction, it is not possible for adjusted recovery rates to be lower than those from the LLF.

6 Conclusion

In this paper we have outlined a framework that allows year-on-year flows into and out of default to be traced at the loan level. An illustration is provided for PD models of the Irish and UK mortgage markets. The cash flow engine of the framework allows a direct mapping from the estimated transition coefficients of a Markov Multi-State Model (MSM) with loan and macroeconomic covariates to annualised flows. Loss Given Default can be calculated in each year by allowing loans the opportunity to cure for T years after default, and comparing the remaining exposure at default to loan-level property valuations, indexed to incorporate macroeconomic house price forecasts.

The framework improves upon traditional loan-level default models which generally use logit models to estimate a *lifetime* probability of a loan entering default. In such a setting, the timing of defaults cannot be specified, while cure flows can only enter the model through a mechanical overlay. The framework proposed here, by virtue of the Markov MSM estimation method used, tackles both of these issues directly. When compared to traditionally used transition matrix approaches, the framework also offers advantages, in that loan-level covariates have an estimated impact on the probability of default and cure. Finally, our framework allows the time spent in default to negatively impact cure probabilities, thereby directly factoring into the model the hysteresis effect of long durations in default on expected losses. This characteristic is particularly important in countries experiencing high stocks of defaulted loans, where at the model starting point, loans may already have been in a delinquent state for a long period of time.

References

Basel Committee on Banking Supervision, "Overview of The New Basel Capital Accord," Consultative Document, Bank for International Settlements May 2001.

Betancourt, Luis, " Using Markov Chains to Estimate Losses from a Portfolio of Mortgages," *Review of Quantitative Finance and Accounting*, May 1999, *12* (3), 303–17.

Board of Governors of the Federal Reserve, "Comprehensive Capital Analysis and Review 2012: Methodology and Results for Stress Scenario Projections," mimeo, Board of Governors of the Federal Reserve 2012.

Gaffney, Edward and Peter Dunne, "Enhanced modelling of mortgage Loss Given Default for loan-loss forecasting," Mimeo, Central Bank of Ireland 2014.

Grimshaw, Scott D. and William P. Alexander, "Markov chain models for delinquency: Transition matrix estimation and forecasting.," *Applied Stochastic Models in Business Industry*, 2011, *27* (3), 267 – 279.

Jackson, Christopher, "Multi-State Models for Panel Data: The msm Package for R," *Journal of Statistical Software*, 1 2011, *38* (8), 1–28.

Jarrow, Robert A., David Lando, and Stuart M. Turnbull, "A Markov Model for the Term Structure of Credit Risk Spreads," *Review of Financial Studies*, 1997, *10*, 481–523.

Kelly, Robert, "The Good, The Bad and The Impaired - A Credit Risk Model of the Irish Mortgage Market," Research Technical Papers 13/RT/11, Central Bank of Ireland November 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*, March 2002, *26* (2-3), 423–444.

McCann, Fergal, "Modelling default transitions in the UK mortgage market," Research Technical Paper 18RT14, Central Bank of Ireland 2014.