ERASMUS UNIVERSITY ROTTERDAM

ERASMUS SCHOOL OF ECONOMICS

THE PD-LGD RELATION FOR US MORTGAGES

MICHAEL HOGERS

403238

SUPERVISOR:

B.J.L. KEIJSERS

CO-READER:

H.J.W.G. KOLE

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER IN ECONOMETRICS AND MANAGEMENT SCIENCE

FEBRUARY 4, 2018

Abstract

This thesis shows that while macro variables and default rates share common cycles for conventional US mortgages, a unique cycle is observed for loss given default, implying that the relation between the default rate and loss given default is weak for conventional US mortgages. The average loss given default across the US increases from 2002 until the end of the data set in 2014. Similar increases are observed across the U.S. Census Bureau's defined regions, although the West region shows losses that seem impacted by the financial crisis of 2007-2008. The research finds evidence of a unique cycle for loss given default across the US as well as for specific regions, moreover this cycle is also present for uninsured mortgage loans with a high enough loan to value at origination ratio. The research uses a mixed-measurement dynamic factor model that applies a mixture of gaussians to capture the dynamics of a bimodal loss given default distribution. The finding of a unique loss given default cycle for conventional US mortgage loans is innovative, as previous research primarily focuses on corporate loans and finds a shared cycle between defaults and losses.

Keywords: Mixed measurement dynamic factor models, mortgages, loss given default, default rates

Acknowledgement

By finishing this thesis I conclude my time as a student at the Erasmus University Rotterdam. Over the last years I have been taught valuable skills and I deeply appreciate the work of university staff. I would like to thank my thesis advisor Bart Keijsers for his continuous availability and input that steered my thesis into the right direction. Furthermore I would like to thank Burkhard Heppe and Jeroen Batema of Open Source Investor Services for introducing myself to the complexity of mortgages. I would like to thank my parents and Jiawei Xu for their unconditional support during my time as a student. Finally I would like to mention Richard Feynman for being an inspiration and greatly influencing my way of thinking about both everyday and academic subjects,

"Some people say, "How can you live without knowing?" I do not know what they mean. I always live without knowing. That is easy. How you get to know is what I want to know." - Richard Feynman

Contents

1 Introduction

Under the Basel II capital framework it is possible to use an advanced internal rating based approach. This approach requires accurate estimates of parameters that determine the credit risk of a banks' financial assets. The two most important parameters are probability of default and loss given default. While literature on probability of default and loss given default studies various characteristics that affect both probability of default and loss given default, a relatively small amount of research considers residential mortgages. The subprime mortgage crisis that started the US recession of 2007-2009 (Barth, 2009) makes it apparent that residential mortgages require considerate attention. To quantify exposures from residential mortgages it is required to create probability of default and loss given default estimates, motivating a clear understanding of these parameters and factors that drive them.

The main objective of this study is to gain a better understanding of the relation between probability of default and loss given default for US residential mortgages. This research uses a mixed measurement dynamic factor model (MM-DFM) as it connects observations from different families of parametric distributions. Earlier research by Koopman (2010), Creal et al. (2014) and Keijsers et al. (2017) use a similar approach by utilizing the MM-DFM for bonds and bank loans. This study is most similar to the latter and extends the application of MM-DFM by using it to analyze US residential mortgage loans, which is shown to have a unique bimodal loss distribution and different exposures from corporate loans.

The research is interesting for several reasons. First, research on losses from loans is extensive. However, earlier work mainly focuses on defaults only or defaults and macro risk (Creal et al., 2014). This research focuses on defaults, losses, macro and housing risk together. Second, the data set allows investigating the relation between probability of default and loss given default over time across a long period with plenty of default observations. Third, to my knowledge literature that uses a similar data set with plentiful default and loss observations is scarce with recent research by An & Cordell (2017) being a notable exception. Furthermore, this thesis shows how loss given default for US mortgages follow its own dynamic, unrelated to macroeconomic or default events. This is a new finding in the credit risk literature, as previous research into underlying factors focuses on corporate loans where losses are shown to be dependent on factors common to macro and default.

The research is based on residential mortgage data from Fannie Mae. The data is publicly available to support Fannie Mae's goal to increase insights in credit risk. The research studies 818,463 quarterly default and 431,200 quarterly loss observations over the period 2000-2015. The total amount of unique loans equals 23,957,641. These loans are a subset of Fannie Mae's 30-year, fixed-rate amortizing loans. Existing latent factor research often uses either samples or time periods that are smaller. Keijsers et al. (2017) uses approximately 22,000 bank loan defaults over the period 2003-2010. Creal et al. (2014) use 1,342 defaults for Moody's rated US firms over the period 1982-2010. Qi & Yang (2009) use collateralized mortgage loans with 241,293 default observations, studying only loss given default. Most similar however is An & Cordell (2017) who study loss given default for the same data set.

Evidence from corporate loans suggests that loss given default is volatile and that it tends to rise when the number of defaults increases. Altman et al. (2004) provide a survey of this phenomenon for corporate loans and discusses a reversal in the traditional approach of assuming defaults and losses independent. Schuermann (2004) discusses causes of changes in loss given default. Schuermann states that the presence and quality of collateral are one of the main causes of significant differences in loss given default. Mortgage loan data alternatively suggests that average loss given default rises steadily from 10% in 2002 to 50% post-crisis. This steady increase before the financial crisis seems unrelated to macroeconomic causes as the housing market only started declining in 2007. An & Cordell (2017) state that observed loss severities in this pre-crisis period are likely problem loans with property located in regions that did not have a similar recovery of housing prices as the rest of the US. Interestingly, a similar increase from 2002 to 2007 is not observed for the default rate. Post 2007 an explanation for the increase of average loss given default is the extreme downturn of housing prices as well as an extreme increase in defaulted loans which requires significant mortgage servicing resources and court dockets, leading to increased costs in case of a foreclosure. Nonetheless, these arguments do not explain why for the entire sample mortgage losses persist around 50% after the crisis. This thesis identifies a unique factor underlying loss given default for US mortgages, which is in line with An & Cordell (2017), who explain the persistent increase in losses by identifying a regime shift, which can be seen as the cause of this unique factor.

Due to the variation in size and other characteristics of the US housing market it is important to investigate regional differences. Wheaton & Nechayev (2008) show that the significant increase in prices between 1998 and 2005 differs notably between Metropolitan Statistical Areas (MSA), where larger MSAs have a larger unexplained price growth, possibly due to the presence of speculative buying that is positively correlated with the size of a MSA. The Fannie Mae data set contains compelling loss given default differences per state, providing an argument that the findings of significant differences between regional areas of Wheaton & Nechayev (2008) may hold for conventional mortgages on a state level. This research finds support of small differences per state between default rates and larger differences between average loss given default. Across the U.S. Census Bureau's regions results show the existence of unique loss given default dynamics decoupled from macro or default cycles, indicating that this thesis' findings for regions are similar to the findings across the US.

Besides regional characteristics, this thesis also researches loan-to-value categories. Qi & Yang (2009), Calem & LaCour-Little (2004) and Pennington-Cross et al. (2003) discuss how loan-to-value is a significant explanatory variable for loss given default, and I therefore create loan-to-value loan subsets based on stress tests by the Office of Federal Housing Enterprice Oversight to inspect how findings hold across these different risk groups. Loans with a loan-to-value above 80% show a considerably larger default rate, not only during the crisis but also before, whereas loans with a loan-to-value below 60% show a very limited influence of the financial crisis. In accordance with results from L. S. Goodman & Zhu (2015), loans with a loan-to-value above 80% almost all have a private mortgage insurance, resulting in the lowest average loss given default over time for this category. While the financial crisis strongly affected the default rate for these high loan-to-value loans there is no similar effect for the average loss given default. Results of this paper show that the existence of a unique loss given default dynamic, decoupled from macro or default cycles, is most apparent for the category with loan-to-value's above 80%, which is expected given the prevalence of mortgage insurance.

Furthermore, Wheaton & Nechayev (2008) find that house prices over the period 1998-2005 can not be explained by increases in demand fundamentals such as population, income growth and declines in interest rates. Shiller (2006) suggests taking into account additional fundamentals such as rental-price ratios, vacancy rates and interest rates, where he concluded that the divergence of these variables implicated an impending crash of house prices. Across the US we see that the homeowner vacancy rate was at an all time high before both the financial crisis of 2007-2008 and corresponding decline in house prices. In contrast, the 30-year fixed rate mortgage US average had been stable when the financial crisis started. Finally the change in US real disposable income had been decreasing several years before the financial crisis, ultimately resulting in a negative change during and shortly after the financial crisis. This thesis shows that across the US increases in rental and housing vacancy rates are accompanied with increases in loss given default, whereas increases in the default rate are mostly associated with decreases in the change of industrial production and disposable income and increases in UR.

This study provides contributions to both mortgage loan literature and market practice. The mixed measurement dynamic factor model allows to see how macroeconomic, default and loss given default share common dynamics. Across the US, loss given default observations warrant their own dynamic factor and a model with three factors is sufficient to capture the dynamics in the data. The findings motivate that there is no clear connection between the default rate and loss given default for US mortgages; loss given default warrants a latent factor that specifically captures its own dynamics, whereas default rates can be explained well by macroeconomic factors. The findings support An & Cordell (2017), who state that loss given default is influenced by business practice or government intervention and only marginally by default rates. As the estimation results in this thesis show that loss given default follows its own dynamics, unrelated to macro variables or default data, the relation between default rates and loss given default for US mortgages is surprisingly weak. When studying separate US regions that comprise of several states in a comparable geographic location similar results are observed, indicating that the absence of a strong relation between probability of default and loss given default also holds on a regional scale. Moreover, this absence of a strong relation is also observed for uninsured mortgages with a loan to value ratio above 60%.

2 Data

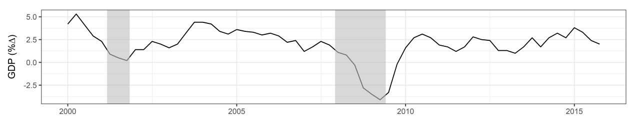
The complete data set consists of two parts: a macroeconomic and a mortgage part. The macroeconomic data set contains time series of US macro variables: gross domestic product (GDP), industrial production (IP), unemployment rate (UR), rental vacancy rate (RVR), homeowner vacancy rate (HVR), disposable income (DI) and the 30-year fixed mortgage rate (FRM). Tsatsaronis & Zhu (2004) describe that the macroeconomic variables GDP and DI influence long-term housing demands, while FRM influences short-term housing demand. Besides GDP, variables such as IP and UR are often used in similar research (Koopman et al. (2012), Creal et al. (2014), Creal (2017), Keijsers et al. (2017)), such that IP and UR are also considered as macro variables for comparing results. Furthermore, Mishkin (2007) discusses how house prices are primarily influenced by supply constraints, the rental and homeowner vacancy rates serve as proxies for these supply constraints.

The default data set contains a subset of Fannie Mae's 30-year fixed-rate US residential mortgages. The mortgages are fully amortizing, fully documented and acquired by Fannie Mae on or after January 1, 2000. The residential mortgages are originated in the period 2000-2015 and amount to a total of 23,957,641 loans. Each individual loan has performance data that reflects the loan status and its characteristics from origination and onwards. Characteristics are static such as the state the property is located in or dynamic such as the first 180 days of delinquency or foreclosure date. In order to reduce the large size of the Fannie Mae data, only static characteristics and specific dynamic characteristics that are relevant for the research are used, such as the location of the collateral property, loan-to-value ratio at origination and origination, inactivity, default and foreclosure dates.

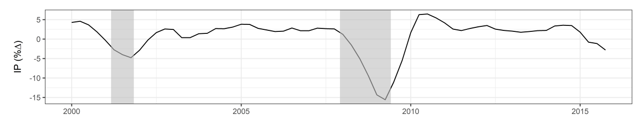
2.1 Macroeconomic data

The time series for GDP and IP show a strong downturn around 2002 and 2008 due to recessions, while UR shows an upturn. Figure 1 contains the time series plots and shows that RVR, HVR, DI and FRM are not similarly affected by recessions. RVR steadily increased after the recession of 2002 and then declines again short after the recession of 2008. As expected, a likewise dynamic is observed for HVR. Moreover, the change in DI preceding the crisis in 2008 was mostly positive but possesses a cyclical behaviour. FRM declined quickly during the 2002 recession, after which it steadily increases, but seems to have started declining already before the 2008 crisis commences. Figure 1 shows the time series for all seven macro variables, note that GDP, IP, UR, DI and FRM are the growth rates with respect to the previous year and all observations are quarterly.

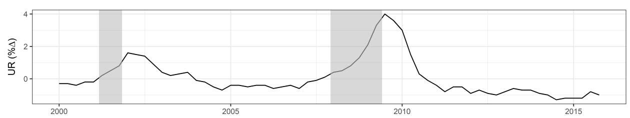
Incorporating all seven macroeconomic time series, that could be interpreted as general macroeconomic time series (GDP, IP, UR) and more housing related (RVR, HVR, DI, FRM), allows investigating to which extent these (two categories) affect residential mortgages, as the information is condensed into a latent factor. Literature shows several examples where these variables are applied to credit risk for residential mortgages. Elul et al. (2010) and Goodman & Smith (2010) consider the unemployment rate for predicting mortgage defaults. Qi & Yang (2009) research residential mortgages and find that distressed housing markets show a significant increase of mortgage loss severity, a signal that can be captured with housing related time series. Guiso et al. (2013) show that the gross domestic product not only influences the loss on a defaulted mortgage, but also the default frequency.



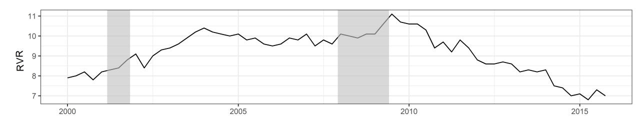
(a)



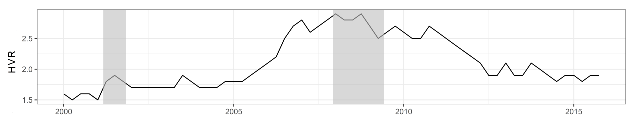
(b)



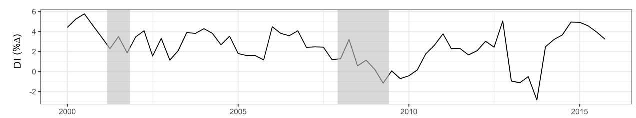
(c)



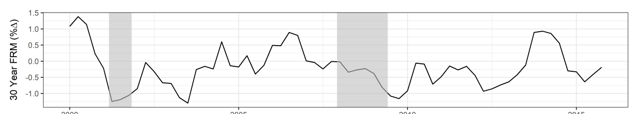
(d)



(e)



(f)



(g)

Figure 1: This figure presents the macroeconomic data that is used to construct the macroeconomic factors. The data source is the St. Louis FED (Federal Reserve Economic Data). The data is transformed as follows. GDP: growth rate compared to the same quarter of the previous year. IP: growth rate compared to the same quarter of the previous year. UR: growth rate compared to the same quarter of the previous year. RVR: no manipulations. HVR: no manipulations. DI: growth rate compared to the same quarter of the previous year. FRM: no manipulations. The subpanels contain the seven time series in the same order.

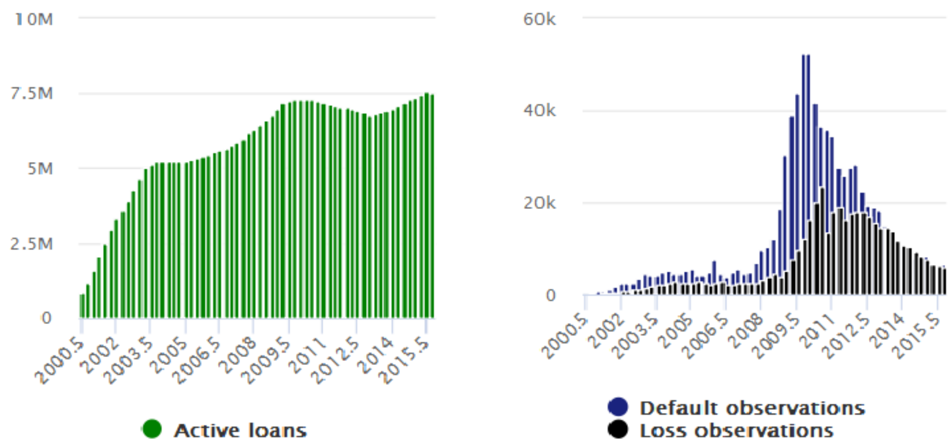
2.2 Mortgage data

Defaults are defined as 180 days of delinquency having occurred, which is in accordance with Basel regulations, in which the IRB approach allows the use of 180 days of delinquency as a default indicator[1]. This results in a total of 818,463 default observations, leading to an average default rate of 3.42%. The default observations are presented in Figure 2. Default observations have a peak after the financial crisis around 2009-2011 with 447,097 defaults observed in this time period, leading to an average default rate of 5.09%, with a reversal to pre-crisis levels lasting until 2015. This research aggregates both default and loss given default observations to a quarterly frequency.

In case of default a lender can incur a loss when the borrower forecloses on its property. Loss given default then represents this loss as a fraction of the unpaid balance at default. The data set contains 486,644 loss given default observations. Loss given default is defined as the total net loss divided by the defaulted unpaid balance at default, with total net loss equal to the sum of the defaulted unpaid balance, accrued interest and total costs

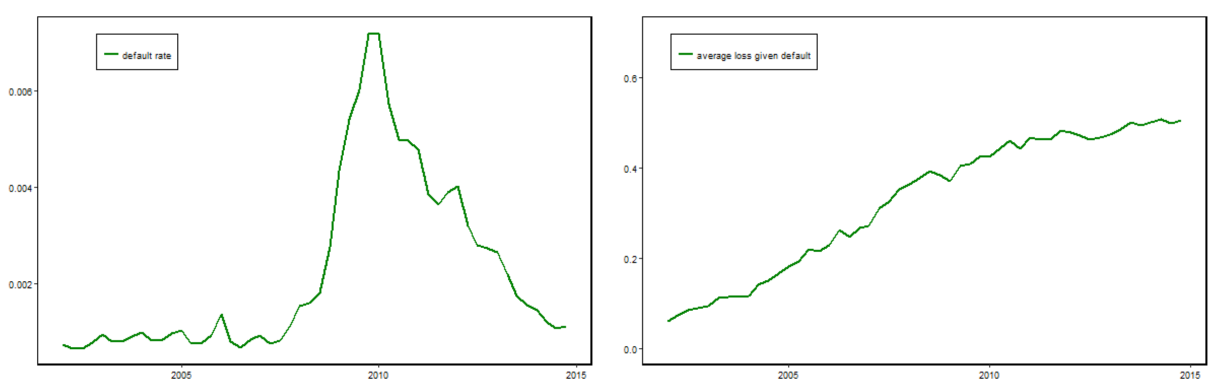
[1]The consultative document by the Basel Committee on Banking Supervision states that the default definition used in the IRB approach allows for the use of a 180 days of delinquency threshold for retail and public sector exposures.

minus total proceeds. Figure 2 shows the amount of active loans, default and loss observations over time. The peak in default observations occurs slightly before the peak in loss observations, which is a result of the workout period and as expected given the workout period observations presented by Figure 19 in the appendix. Figure



(a) (b)

Figure 2: This figure presents the amount of active loans (panel a) and the amount of default and loss observations (panel b) in the data set. The number of active loans are shown per quarter from 2000 to 2015, and the amount of active loans changes are due to new loans entering the mortgage pool and loans being removed. Similarly, the amount of default and loss observations is also shown per quarter from 2000 to 2015. Default observations are registered by means of the 180 days delinquency date and loss observations by the foreclosure date.



(a) (b)

Figure 3: This figure presents the observed default rate and the average loss given default for the period 2002-2014. Panel a contains the observed default rate and panel b the average loss given default over time.

3 shows the default rate and average loss given default over time for the data set, it presents how the default rate changes significantly around the financial crisis, while the average loss given default increases steadily over time from approximately 0.1 to 0.5 and seems unaffected by the 2007-2008 crisis and other macroeconomic

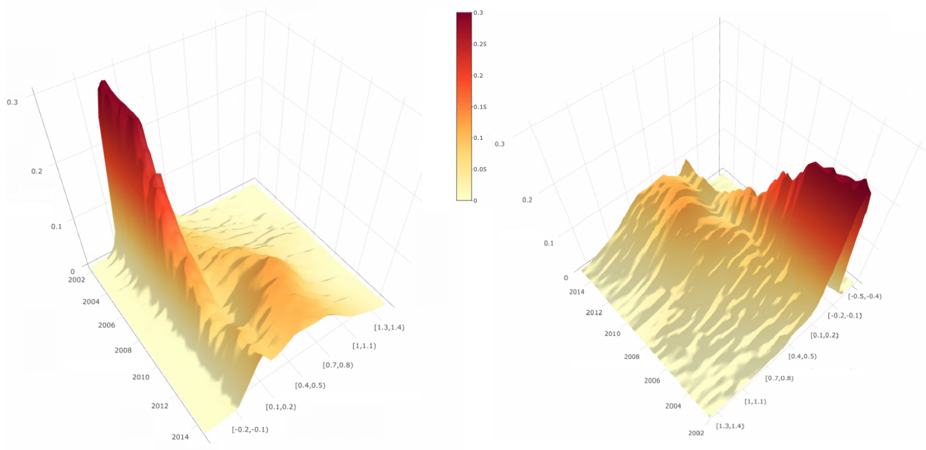


Figure 4: This figure shows US mortgage loss given default over time from two different angles. The color represents the ratio of observations that fall into a certain loss range. The beginning of the period has a large fraction of defaults with close to zero losses, running up to and after the financial crisis of 2007-2008 a significant fraction of loss given default observations shift from zero towards 0.5.

events. Figure 4 shows in detail the loss given default distribution over time. The fraction of loss given default observations close to zero is very high during the first observed years. An & Cordell (2017) states that this may be due to the improving economic conditions after the recession of 2002, causing the housing price and thus the mortgage collateral quality to improve significantly.

3 Models

This section describes the general model specification in subsection 3.1. The other subsections each discuss a model component. Creal et al. (2014) and Keijsers et al. (2017) use a similar model with three components. These components link together macroeconomic variables, default rate and loss given default via latent factors.

3.1 The joint model

A mixed measurement model in the style of Koopman et al. (2012), Creal et al. (2014) and Keijsers et al. (2017) is used, and the notation of the latter paper is followed. The observations can follow different distributions and depend on latent factors. At each point in time the model contains N variables, and the variables may not all be observed at all points in time. The variables are separated in three sets: macro, default and loss given default variables labeled as m, d and l, respectively. y_{it}^c denotes the time t observation of variable i in set c, where $c \in \{m, d, l\}$. N^c denotes the size of the category c, and variables are collected in the vector $\boldsymbol{y} = (\boldsymbol{y}^{\boldsymbol{m}'}, \boldsymbol{y}^{\boldsymbol{d}'}, \boldsymbol{y}^{\boldsymbol{l}'})'$. At time t the following observation densities are considered for the variables,

$$y_t^m \sim N(\mu_{mt}, \Sigma_m),$$

$$y_{it}^l \sim Bernoulli(p_t^l),$$

$$y_{it}^d = \begin{cases} N(\mu_0, \sigma_0^2) & \text{if } s_{it} = 0 \\ N(\mu_{1t}, \sigma_1^2) & \text{if } s_{it} = 1 \end{cases}.$$

The core component of the joint model is a set of K latent factors $\boldsymbol{f_t}$, which link all observed processes. The latent factors can be separated into three categories. K^m macro factors $\boldsymbol{f_t^m}$ capture business and housing related cycles. K^l default factors $\boldsymbol{f_t^l}$ influence the default and loss given default variables. K^d loss given default factors $\boldsymbol{f_t^d}$ affect only the loss given default variables. $\boldsymbol{f_t^l}$ and $\boldsymbol{f_t^d}$ capture the credit cycle dynamics unrelated to a business cycle. This general setup inspired by Creal et al. (2014) and Keijsers et al. (2017) allows investigating whether or not there is a relation between loss given default and probability of default determinants based upon business cycles or specific factors.

3.2 The latent factor model

Following Koopman et al. (2012) the latent factor $\boldsymbol{f_t}$ is assumed to follow a VAR(1) process,

$$\boldsymbol{f_{t+1}} = \Phi \boldsymbol{f_t} + \boldsymbol{\eta_t}, \qquad \boldsymbol{\eta_t} \sim N(\boldsymbol{0}, \Omega), \tag{1}$$

with Φ a diagonal matrix and η_t serially uncorrelated. $\boldsymbol{f_t}$ is stationary, such that $|\phi_{kk}| < 1$. The initial state vector $\boldsymbol{f_1}$ follows the latent process' unconditional distribution $N(0, \Sigma_{\boldsymbol{f}})$, where $\Sigma_{\boldsymbol{f}} = \Phi \Sigma_{\boldsymbol{f}} \Phi^{'} + \Omega$. For identification it is imposed that the unconditional variance is equal to the identity matrix, $\Sigma_{\boldsymbol{f}} = \boldsymbol{I}$.

3.3 The macroeconomic model

The first variable set contains N^m macro variables, depending linearly on the latent macro factor(s),

$$\boldsymbol{y_t^m} = \boldsymbol{\alpha^m} + \boldsymbol{B^m} \boldsymbol{f_t^m} + \boldsymbol{\epsilon_t^m}, \qquad \boldsymbol{\epsilon_t^m} \sim N(\boldsymbol{0}, \Sigma^{\boldsymbol{m}}), \tag{2}$$

where $\boldsymbol{\alpha^m}$ is a vector of size N^m containing intercepts, and $\boldsymbol{B^m}$ a $N^m \times \mathrm{K}^m$ matrix with coefficients. For identification, $\boldsymbol{B^m}$ is lower triangular with sign restriction on the diagonal elements. Macro variables are standardized to have mean zero and unit variance, such that comparisons between factor loadings are simplified.

3.4 The default model

The second variable set contains the default status, $\boldsymbol{y_t^d}$. Loan i at time t can be performing ($y_{it}^d = 0$) or in default ($y_{it}^d = 1$). Conditional on $\boldsymbol{f_t}$, y_{it}^d follows a Bernoulli distribution with default probability p_t^d,

$$y_{it}^d \sim Bernoulli(p_t^d), \tag{3}$$

$$p_t^d = \Lambda(\alpha^d + \boldsymbol{\beta_m^{d'}} \boldsymbol{f_t^m} + \gamma^{\boldsymbol{d'}} \boldsymbol{f_t^d}) \tag{4}$$

where $\Lambda(\cdot)$ is the logistic function.

3.5 The loss given default model

The loss distribution of the mortgages is presented in Figure 4 and shows notions of a bimodal distribution. Therefore a similar approach as Keijsers et al. (2017) is used, who distinguish between loans with a severe loss and a mild loss and use a mixture of normal distributions to model the bimodal loss distribution. In order to examine whether a time-varying location, scale and mixture probability is appropriate, a mixture of normals is fit to the loss distribution for every quarter for the period 2002-2014. Figure 5 shows the results: μ_1 and p^l appear to differ significantly over time, whereas μ_0 and σ_0 stay very stable over time; σ_1 shows a slight increase at first but remains relatively stable over time. Given the empirical evidence in Figure 4 and the estimation

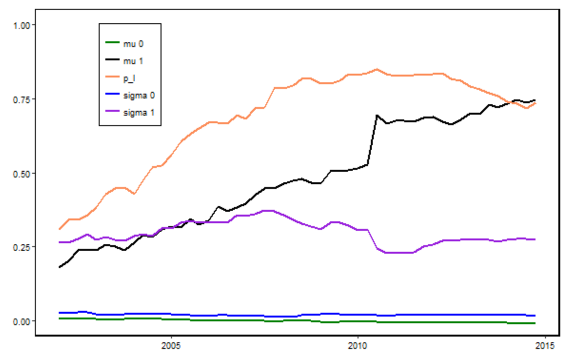


Figure 5: This figure presents the estimation results of a normal mixture as in the loss given default component, estimated for each quarter in the period 2002-2014. The figure shows the location and scale parameters of each mixture component and the mixture probability of component 1.

results in Figure 5, I impose time variation on μ_{1t} and p_t^l and model the loss given default according to the equations below,

$$y_{it}^l = \begin{cases} N(\mu_0, \sigma_0^2) & \text{if } s_{it} = 0, \\ N(\mu_{1t}, \sigma_1^2) & \text{if } s_{it} = 1, \end{cases} \tag{5}$$

$$\mu_{1t} = \alpha_{\mu_1} + \beta_{\mu_1} f_t^m + \gamma_{\mu_1} f_t^d + \delta_{\mu_1} f_t^l, \tag{6}$$

$$s_{it} | f_t \sim Bernoulli(p_t^l), \tag{7}$$

$$p_t^l = \Lambda(\alpha^l + \beta_m^{l'} f_t^m + \gamma^{l'} f_t^d + \delta^{l'} f_t^l), \tag{8}$$

where y_{it}^l is the loss given default observation, modeled as a mixture of two normal distributions. It is assumed that conditional on f_t, y_{it}^l is independent. The mean of the component corresponding with $s_{it} = 1$ depends on the latent factors and varies over time. The latent variable s_{it} indicates to which normal distribution an observation corresponds and is drawn from a Bernoulli distribution with a time-varying probability p_t^l dependent on the latent factors.

4 Methods

The parameters of the joint model are estimated by means of Bayesian inference, and a Gibbs sampler is applied as a Markov chain Monte Carlo (MCMC) algorithm. Bayesian inference allows splitting into parts the estimation of the model parameters and latent factors, by drawing from conditional posteriors. The model consists of four connected parts for which parameters and factors are estimated.

4.1 Bayesian inference

Bayesian inference overcomes the complication of the troublesome likelihood of the MM-DFM, which is not available in closed form. Besides, Markov chain Monte Carlo (MCMC) estimation of parameters and factors can be accompanied with prior information in the Bayesian setting. Also, MCMC is more robust than the alternative of importance sampling in a frequentist setting (Koopman, 2010). For this study the priors are uninformative and specified in the appendix. I create 100,000 draws and treat the first 50,000 as the burn-in period. Increasing the number of draws does not alter the results.

Each parameter is simulated from its conditional posterior distribution by using a Gibbs sampler, and draws from the full posterior distribution of all parameters are obtained. As the joint model consists of several components and the Gibbs sampler allows sampling the parameters block-wise, the inference is described per model part. Note that the appendix contains prior, likelihood and posterior specifications of the following parts.

The first component of the joint model is the macroeconomic component. This macroeconomic component is given in equation (2), where we define $\boldsymbol{A^m}$ as the matrix that collects all intercepts and slopes. Using standard results for multivariate regression models we can sample $\boldsymbol{A^m}$ from a matricvariate normal distribution and $\Sigma^{\boldsymbol{m}}$ from an inverse Wishart distribution. $\boldsymbol{A^m}$ is drawn until it satisfies the identification restrictions, that is, $\boldsymbol{B^m}$ is a lower triangular matrix with sign restrictions on the diagonal. In order to identify the latent macroeconomic factors the diagonal elements of $\boldsymbol{B^m}$ are restricted to be negative.

The second component, the default component, is given in equation (3) and (4). The default indicator y^d_{it} is drawn from a Bernoulli distribution and the corresponding probability of default p^d_t depends on the latent factors in a non-linear way via the logistic function. To draw the parameters of the logistic regression the results of Polson et al. (2013) are used, namely drawing latent variables ω^d_t from a Pólya-gamma distribution, which facilitates the drawing of factor loadings α^d, β^d and γ^d from a multivariate normal distribution. Factor loadings are drawn until $\gamma^d > 0$.

The third component, the loss given default part, which consists of equation (5), (6), (7) and (8), models the bimodal loss distribution. The location and scale parameters of the component, μ_0 and σ_0 corresponding to $s_{it} = 0$, are fixed over time and are drawn from a normal and inverse gamma-2 distribution, respectively. The component corresponding to $s_{it} = 1$ has a time-varying location parameter and a fixed scale parameter. The time-varying parameter μ_{1t} is dependent on the latent factors $(f^m_t, f^d_t, f^l_t)'$ and drawing factor loadings follows standard results for regression models. The scale parameter σ_1 is drawn from an inverse gamma-2 distribution.

To prevent label switching a restriction is imposed such that $\mu_0 < \mu_{1t}$ for all t. Furthermore, the latent factor s_{it} that determines from which component a loss should be drawn is obtained by sampling from a Bernoulli distribution. Lastly, the factor loadings of the logistic regression are once more obtained by drawing latent variables ω_t^l and then drawing the factor loadings α^l, β^l, γ^l and δ^l from a multivariate normal distribution until $\delta^l > 0$.

As the key to combining different parametric distributions, the latent factors are sampled from their conditional density, $\boldsymbol{f} \sim p(\boldsymbol{f}|\boldsymbol{y}, \psi^{(i-1)}, \boldsymbol{f}^{(i-1)})$. The simulation smoothing algorithm of Durbin & Koopman (2002) allows sampling from this posterior in three steps: a Kalman filter step, a Kalman smoother step and a simulation step for the factors. The implementation by Durbin & Koopman (2002) is comparable to Frühwirth-Schnatter (1994) and Carter & Kohn (1994), who first provided ways to use Gibbs sampling for state space models. De Jong & Shephard (1995) provided a more efficient simulation routine, which was further improved by Durbin & Koopman (2002). In order to utilize the simulation smoother algorithm the model components have to be cast into a state space formulation as below,

$$\boldsymbol{y_t} = \boldsymbol{Z}\boldsymbol{f_t} + \boldsymbol{\epsilon_t}, \qquad \epsilon_t \sim N(\boldsymbol{0}, \boldsymbol{H_t}), \tag{9}$$

$$\boldsymbol{f_{t+1}} = \boldsymbol{T}\boldsymbol{\alpha_t} + \boldsymbol{R_t}\boldsymbol{\eta_t}, \qquad \boldsymbol{\eta_t} \sim N(\boldsymbol{0}, \boldsymbol{Q_t}), \tag{10}$$

where equation (9) is the observation equation and equation (10) the state transition equation. Furthermore, $\boldsymbol{y_t}$ contains the macroeconomic time series $\boldsymbol{y_t^m}$ and the ratios κ_t^d/ω_t^d and κ_t^l/ω_t^l, the latter two of which are pseudo data points corresponding to default and loss given default observations. Introduced by Windle et al. (2013), these pseudo data points are distributed normally to simplify sampling from the posterior distribution of $\boldsymbol{f_t}$ by using a data augmentation method, which introduces Gaussian likelihoods conditional on Pólya-gamma variables. It is assumed that $\boldsymbol{f_1} \sim N(\boldsymbol{a_1}, \boldsymbol{P_1})$ with $\boldsymbol{a_1}$ and $\boldsymbol{P_1}$ a vector of zeros and an identity matrix, respectively, imposing that $\boldsymbol{f_1}$ follows its unconditional distribution. Equation (40) and (41) in the appendix show a more detailed version of equation (9) and (10) and allow further understanding the role of the pseudo data points.

5 Results

5.1 Models without loan characteristics

The analysis of estimation results starts by studying models with a different number of latent factors. The model components are as in Section 3 and in this first subsection no loan characteristics are considered. Recall that $\boldsymbol{f_t^m}$ captures business related cycles and the two factors $\boldsymbol{f_t^l}$ and $\boldsymbol{f_t^d}$ pick up credit cycle dynamics unrelated to a business cycle, furthermore $\boldsymbol{f_t^l}$ only influences the probability of default and loss given default and $\boldsymbol{f_t^d}$ only influences loss given default. The analysis of the relation between probability of default, loss given default and also macro variables commences with three different specifications. Each specification has a time-varying mean μ_{1t} and mixture probability p_t^l and fixed variances for the loss given default component of the model, which Figure 5 motivates by showing estimation results of a mixture of normals for every quarter. First the model specifications are discussed and results presented in Table 2 and Figures 6 and 8. Previous research on corporate loans shows that both probability of default and loss given default are driven by similar risk factors (Keijsers et al. (2017), Azizpour et al. (2017), Koopman et al. (2012), Duffie et al. (2007)). These previous results serve as a beginning to investigate different model specifications for mortgage loans. I investigate several model specifications for the full data set using either one, two or three macro factors, and, to test the existence of unique cycles, including default or loss factors.

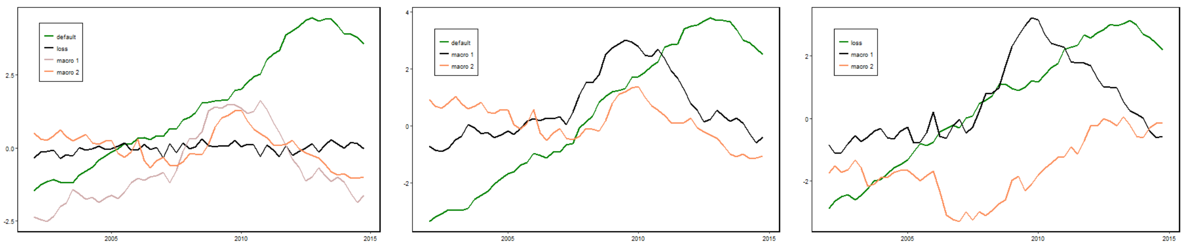
The results for the model with two macroeconomic factors, a default factor and one loss factor show persistent macro, default and loss latent factors, although the posterior range for the second macro factor is relatively wide. The two latent macro factors show explanatory power for all of the macro variables. Figure 7 shows that the model adequately captures the dynamics in the macro variables, similar results are seen for the other two models in Table 2. Models with only one macroeconomic factor do not adequately capture dynamics of the macro variables, as shown by the fits in Figure 21 and as implied by a higher WAIC2 for the one macro factor models (see Table 3), where WAIC2 (Watanabe (2010)) corrects for the number of parameters[2]. The WAIC2 is not improved by a model with three macro factors, showing a higher WAIC2 than a model with two macro factors, as presented by Table 4 in the appendix, with a WAIC2 of 100.935, compared to 57.440 for the identical model with two macro factors. For the model with two macro, one default and one loss factor in Table 2, I observe for the first macro factor negative loadings onto growth rates of the four variables GDP, IP, DI and FRM and positive loadings for UR, RVR and HVR. A change in the first macroeconomic factor mostly affects GDP, HVR and FRM, which have sizable loadings. Similar results are seen for the two macro and one default factor model. The second macro factor mostly influences UR and RVR. The model with two macro and a loss factor is less comparable, but shows a relatively high WAIC2 value for the macro component which indicates a worse fit and therefore explains the different estimation results.

The two macroeconomic factors positively affect the default probability, which is expected given that for the three models an increase in the first macro factor lowers GDP and increases UR and an increase in the second macro factor lowers IP. On average the marginal effect of the first and second macro factor for the two macro, one default and one loss factor model are 0.048% and 0.160%, respectively. Compared to the average marginal effect of the default factor, 0.059%, this is relatively large. All marginal effects are sizable, given that the average default probability are 0.224% per quarter. The marginal effects for the macro factors are comparable or considerably larger, which indicates that no cycle, apart from a macroeconomic related cycle, is present for defaults. Figure 6 shows in panel (a) how the default factor of the main model follows the dynamic of the average loss given default (recall Figure 3, panel (b)). This loss given default dynamic of the default factor indicates that the macro factors already capture most, if not all, default dynamics. For the model with two macro factors and a default factor I observe similar average marginal effects and a default factor that follows average loss given default dynamics.

The loss given default component is most extensive as it depends on all previously mentioned factors as

[2]Gelman et al. (2014) recommend WAIC2 as it has a closer resemblance to leave-one-out cross validation than WAIC. Furthermore, WAIC2 is preferred above AIC and DIC as WAIC2 averages over the posterior distribution instead of conditioning on a point estimate.

well as its own factor. First, the location parameters are discussed. The location of the fixed mixture component is estimated as 0.001 for all three models. Similarly, the (time-varying) location of the other mixture component is on average 0.406. For the model with two macro, one default and one loss factor we see a fixed value of 0.382



(a) (b) (c)

Figure 6: This figure presents the mean of the posterior distributions of the latent factors that accompany the estimated factor loadings and autoregressive parameters of Table 2. The latent factors of the model with a macroeconomic, default and loss factor are displayed in subpanel (a). The latent factors of the model with two macroeconomic and a default factor are displayed in subpanel (b). The latent factors of the model with two macroeconomic factors and a loss factor are presented in subpanel (c).

for this location. With a positive loading on the first macro factor of 0.050 and a negative loading of -0.055 on the second macro factor, the macro factors have similar influence on the location but in opposite ways. An increase in the first macro factor is accompanied with an increase in UR, RVR and HVR and a decrease in GDP, IP, DI and FRM, providing evidence that on average a positive change to UR, RVR, HVR and a negative change to GDP, IP, DI and FRM is accompanied with larger losses for a fraction of US mortgages. I observe the same relation for the mixture probability, with loadings of 0.119 and -0.093 for the first and second macro factor, respectively. This indicates that for the first macro factor a positive change to UR, RVR, HVR and a negative change to GDP, IP, DI and FRM not only creates larger losses for a fraction of US mortgages, but also increases this fraction of loans with larger losses. For the second macro factor I observe a similar effect, however here the second macro factor increases UR and RVR and decreases IP, HVR, DI and FRM and results in a lowering of the loss of a fraction of loans as well as a decrease in the fraction of loans itself. These findings are in line with those of Caselli et al. (2008), for Italian mortgage loans a change in GDP is negatively related to loss given default, whereas a change in UR is positively related.

Moreover for the loss given default component, we see that for the mixture probability the loading onto the default factor is 0.039 and the loading onto the loss factor 0.005. This can be explained as the default factor captures the dynamics for the loss given default, which leaves no information for the loss factor, as seen in Figure 6 subpanel (a). This is further confirmed by the model with two macro and one default factor, which shows similar factors and loadings. Furthermore, the model with two macro and one loss factor shows similar results as the previous two models and the loss factor is similar to the default factor of the other two models. This may be interpreted as defaults being explainable by two macro factors, while loss given default needs its own factor. This would indicate that defaults and macro variables are related, while there is no relation between loss given default and macro variables. Keijsers et al. (2017) present results of two models for their

	2 Macro, default, loss		2 Macro, default		2 Macro, loss	
Panel A: Factor						
ϕ_{m_1}	0.900	*(0.691, 0.989)*	0.893	*(0.652, 0.989)*	0.896	*(0.672, 0.989)*
ϕ_{m_2}	0.729	*(0.049, 0.989)*	0.775	*(0.225, 0.987)*	0.898	*(0.626, 0.978)*
ϕ_d	0.944	*(0.821, 0.989)*	0.945	*(0.845, 0.985)*	-	-
ϕ_l	0.936	*(0.768, 0.992)*	-	-	0.930	*(0.790, 0.990)*
Panel B: Macro						
$\beta_{GDP,1}$	-0.402	*(-0.603, -0.201)*	-0.512	*(-0.705, -0.253)*	-0.393	*(-0.559, -0.097)*
$\beta_{IP,1}$	-0.172	*(-0.474, -0.032)*	-0.316	*(0.573, -0.080)*	-0.250	*(-0.410, -0.017)*
$\beta_{UR,1}$	0.040	*(-0.014, 0.095)*	0.312	*(0.153, 0.481)*	0.388	*(0.166, 0.530)*
$\beta_{RVR,1}$	0.178	*(0.124, 0.213)*	0.261	*(0.162, 0.402)*	0.362	*(0.069, 0.536)*
$\beta_{HVR,1}$	0.802	*(0.633, 0.849)*	0.711	*(0.598, 0.897)*	0.516	*(0.290, 0.695)*
$\beta_{DI,1}$	-0.239	*(-0.408, -0.052)*	-0.477	*(-0.595, -0.287)*	-0.359	*(-0.584, -0.150)*
$\beta_{FRM,1}$	-0.496	*(-0.783, -0.203)*	-0.522	*(-0.816, -0.136)*	-0.563	*(-0.702, -0.397)*
$\beta_{GDP,2}$	-	-	-	-	-	-
$\beta_{IP,2}$	-0.458	*(-0.723, -0.053)*	-0.328	*(-0.549, -0.108)*	-0.040	*(-0.125, 0.000)*
$\beta_{UR,2}$	1.286	*(0.932, 1.530)*	0.908	*(0.756, 1.391)*	-0.249	*(-0.490, 0.016)*
$\beta_{RVR,2}$	1.014	*(0.849, 1.233)*	0.830	*(0.693, 1.159)*	-0.792	*(-1.512, -0.391)*
$\beta_{HVR,2}$	-0.450	*(-0.694, -0.229)*	-0.379	*(-0.596, -0.189)*	-0.637	*(-1.158,-0.381)*
$\beta_{DI,2}$	-0.132	*(-0.210, -0.052)*	0.230	*(-0.190, 0.484)*	-0.130	*(-0.354, 0.130)*
$\beta_{FRM,2}$	-0.062	*(-0.102, -0.001)*	0.073	*(-0.597, 0.482)*	-0.168	*(-0.374, 0.039)*
WAIC2	57.440		54.153		58.697	
Panel C: Defaults						
α^d	-6.666	*(-6.744, -6.587)*	-6.733	*(-6.802, -6.681)*	-6.458	*(-6.813, -6.352)*
β_1^d	0.216	*(0.176, 0.263)*	0.185	*(0.136, 0.224)*	0.567	*(0.544, 0.596)*
β_2^d	0.716	*(0.639, 0.821)*	0.611	*(0.521, 0.731)*	0.157	*(0.098, 0.228)*
γ^d	0.263	*(0.234, 0.292)*	0.261	*(0.232, 0.286)*	-	-
av. $p^d(x10^{-2})$	0.224	*(0.224, 0.224)*	0.224	*(0.223, 0.224)*	0.224	*(0.223, 0.224)*
m.e. of f_1^m	0.048		0.041		0.127	
m.e. of f_2^m	0.160		0.137		0.035	
m.e. of f^d	0.059		0.059		-	
WAIC2	25.915		25.157		26.031	
Panel D: Loss given default						
μ_0	0.001	*(0.001, 0.001)*	0.001	*(0.001, 0.001)*	0.001	*(0.001, 0.001)*
av. μ_1	0.406	*(0.196, 0.547)*	0.406	*(0.197, 0.554)*	0.406	*(0.185, 0.544)*
α_{μ_1}	0.382	*(0.348, 0.416)*	0.371	*(0.330, 0.423)*	0.324	*(0.226, 0.440)*
$\beta_{\mu_1 1}$	0.050	*(0.034, 0.066)*	0.044	*(0.020, 0.067)*	0.007	*(-0.011, 0.025)*
$\beta_{\mu_1 2}$	-0.055	*(-0.085, -0.006)*	-0.061	*(-0.099, -0.020)*	-0.029	*(-0.055, -0.004)*
γ_{μ_1}	0.039	*(0.027, 0.052)*	0.031	*(0.015, 0.050)*	-	-
δ_{μ_1}	0.005	*(0.004, 0.008)*	-	-	0.068	*(0.051, 0.086)*
WAIC2	43.120		52.210		45.010	
$\sigma_0(x10^{-2})$	0.025	*(0.024, 0.026)*	0.025	*(0.025, 0.026)*	0.025	*(0.025, 0.026)*
σ_1	0.099	*(0.098, 0.099)*	0.099	*(0.098, 0.099)*	0.099	*(0.098, 0.099)*
α^l	0.725	*(0.528, 0.939)*	1.160	*(0.799, 1.594)*	1.203	*(0.477, 1.653)*
β_1^l	0.119	*(0.034, 0.209)*	0.021	*(-0.093, 0.125)*	0.022	*(-0.085, 0.141)*
β_2^l	-0.093	*(-0.237, 0.050)*	-0.138	*(-0.285, -0.002)*	0.091	*(-0.051, 0.237)*
γ^l	0.463	*(0.392, 0.538)*	0.392	*(0.307, 0.480)*	-	-
δ^l	0.032	*(0.000, 0.095)*	-	-	0.486	*(0.374, 0.598)*
av. p^l	0.751	*(0.462, 0.939)*	0.750	*(0.467, 0.937)*	0.750	*(0.463, 0.939)*
m.e. of f_1^m	2.225		0.394		0.421	
m.e. of f_2^m	-1.731		-2.592		1.701	
m.e. of f^d	8.659		7.351		-	
m.e. of f^l	0.203		-		9.115	
WAIC2	38194.520		45861.760		30170.030	

Table 2: This table shows the mean of the posterior draws and the 95% highest density interval of the joint model between parentheses. The table reports three model specifications. The specification in the first column uses two macro factors, one default factor and one loss factor. The second specification uses two macro factors, one default factor and omits a loss factor. Lastly, the third specification uses two macro factors, one loss factor and omits the default factor. Panel A shows posterior results for the elements of the autoregressive parameters in Φ. Panel B shows the factor sensitivities $\boldsymbol{B^m}$ for the macroeconomic time series. The intercepts in $\boldsymbol{\alpha^m}$ are not reported due to the standardization of $\boldsymbol{Y^m}$. Panel C reports the posterior results for the logistic regression of the loan status, where α^d is the constant effect, β^d loads onto the macroeconomic factor and γ^d loads onto the loan factor. Panel D outlines findings for the loss given default component. This component models the loss given default as a mixture of Gaussians, where μ_0 corresponds with the expected loss of a 'good' loan and av. μ_1 corresponds with the average loss of a 'bad' loan, which varies over time. Furthermore, a linear model for μ_1 is supposed to capture the time-varying characteristic, with α_{μ_1} the intercept, β_{μ_1} the loading onto the macroeconomic factor, γ_{μ_1} the loading onto the loan factor and δ_{μ_1} the loading onto the loss given default factor. The variance of the Gaussian mixture components are fixed over time and denoted as σ_0 and σ_1. The loadings α^l, β^l, γ^l and δ^l are the constant and loadings onto macro, default and loss factors, respectively.

entire sample of corporate loans, and their estimation results are similar to this thesis' results in that they also have a loss factor that follows the average loss given default dynamics, or in the absence of a separate loss factor a macro factor that follows the average loss dynamics. While the average loss given default for corporate loans in Keijsers et al. (2017) is comparable to macro dynamics, this comparable dynamic between losses and macro is invisible for US mortgage loans and I therefore put more weight on a separate loss factor.

Suspecting that losses need their own factor to be explained we continue analyzing the model for p_t^l. We see that the fixed effect is 0.725 for the model we mainly focus on, with loadings onto the two macroeconomic factors of 0.119 and -0.093, similar in size but different in sign as also seen for the loadings that determine μ_{1t}. The loading onto the default factor is 0.463, which is a lot more sizable than the loading in the μ_{1t} model and relatively large compared to the other factor loadings. Similarly, a large loading onto the default factor is seen for the model without a loss factor. Also, the model without a default factor shows a similarly large loading onto the loss factor. We thus see the same for the p_t^l part as for the μ_{1t} part: a default factor explains most and this default factor follows the same dynamic as the average loss given default over time, besides a model without a default factor but with a loss factor shows a loss factor that is comparable to the aforementioned default factors. The average marginal effect shows expectedly similar results, with large effects for the two default factors of the first two models in Table 2 and the loss factor of the third model. For any of the three models the average mixture probability is similar, around 0.750, indicating that on average a large fraction of loans have losses significantly different from zero. Lastly, we see that scale parameters σ_0 and σ_1 are comparable across models: the former very small and the latter around 0.099, which is expected given Figure 4 and 5.

Regarding the models in Table 2, Figure 7 shows the fit for the macro variables and Figure 8 shows the fit for both the default rate and the average loss given default. The two macro models with two macro factors both fit the macro variables quite well. Furthermore, the fit for the default rate is close to perfect because of the plain Bernoulli distribution. The fit for the average loss given default is also adequate, similarly Figure 24 presents the cross-sectional loss given default distributions for 2006 Q4 and 2012 Q2, which are captured satisfactorily by all model specifications. The appendix shows the fit for models with only one macro factor, which is in general inadequate for the macro variables, as seen in Figure 21 and 22.

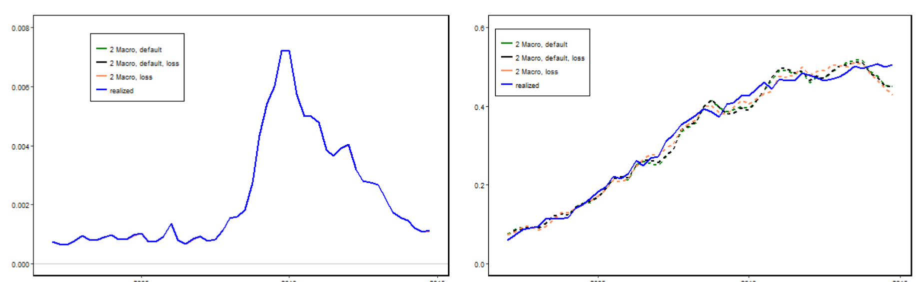
I conclude that a model with two macro factors and one default or loss factor can sufficiently fit macroeconomic, probability of default and loss given default observations. Two macroeconomic factors can explain macro observations and probability of default observations, while loss given default requires a factor that follows a dynamic most similar to the average loss given default. These findings for US mortgages are different from previous research on corporate loans. Altman et al. (2004), Allen & Saunders (2004) and Schuermann (2004) discuss that for corporate loans probability of default and loss given default are driven by the same risk factors. Keijsers et al. (2017) show more recently with the same model how a macro and loan factor accurately capture dynamics for macro variables, probability of default and loss given default for corporate loans. The results in this thesis support the findings of An & Cordell (2017) who find that loss given default for US mortgages follows a new regime that is largely unrelated to the changes in macroeconomic variables or probability of default.

(a) (b) (c)

(d) (e) (f)

(g)

Figure 7: This figure presents the realized time series for GDP, IP, UR, RVR, HVR, DI and FRM. Besides the realized time series it also shows the model fit for each of the four model specifications of Table 2. Each subpanel contains a specific timeseries and model fits. Panel (a), (b), (c), (d), (e), (f), (g) contain GDP, IP, UR, RVR, HVR, DI and FRM, respectively.



(a) (b)

Figure 8: This figure presents the observed default rate and the average loss given default for the period 2002-2014, as well as model fits for the three different model specifications of Table 2. Subpanel (a) contains the observed default rate and subpanel (b) the average loss given default over time.

5.2 Regional characteristics

The analysis so far mainly focuses on the whole sample, with US mortgage loans spread out over the US. In this section a similar analysis is conducted on regions defined by the US Census Bureau. Default rates and loss given default should not be assumed homogeneous across US states, as literature shows that different states have different dynamics for both the default rate and loss given default. Ghent & Kudlyak (2011) show that borrowers from non-recourse states, which are predominantly situated in the West region, such as Arizona, California, Montana, Oregon, and Washington are 30% more likely to default on their loans than lenders in recourse states. In these states a lender is unable to recoup the difference between the mortgage balance and the proceeds from a foreclosure sale. The ability to do is also referred to as a deficiency judgment. Qi & Yang (2009) similarly show that loss given default differs significantly per US division. Figure 9 shows how the default rate and average loss given default vary over time across the US. The default rate for non-recourse states such as California and Arizona increases significantly over time, as compared to other states, with loss given default showing comparable patterns.

This subsection investigates the differences across the US by applying the three main models used in the previous subsection, with two macroeconomic factors and either a default or loss factor or both a default and loss factor. The US Census Bureau recognizes four distinct regions in the US[3]: Northeast, Midwest, South and West[4]. The regional analysis allows examines to what extent the results in the previous subsection hold for different US regions. Figure 9 shows to what extent the different regions are heterogeneous with respect to default rates and losses. In 2002-2005 default rates and losses are relatively comparable across states. The following period 2006-2009 that includes the crisis shows a different picture, with an increase in default rates across the US and significant increases for coastal states. For 2010-2014 the average default rate shows a similar pattern as the previous period with a slight increase in some regions, while the average loss given default further increased across states.

Figure 10 shows the observed default rate and average loss given default over time for each of the four regions. The observed default rates are very similar across the four regions, while the average loss given default has more distinct features per region. The observed default rates all have a peak around 2010. The average loss given default rises steadily from 2002 in each region, while losses in the Midwest region rise fastest. Moreover, losses in the West region seem to decrease around 2010 which seems influenced by the default rate dynamics to a certain extent. The Midwest, Northeast and South region follow similar loss given default patterns as the whole sample, rising from 2002 to 2015. The different pattern for the West region may be explained by a larger presence of non-recourse states. Bhutta et al. (2010) discuss how under-water homeowners in Arizona,

[3] http://www.census.gov/geo/maps-data/maps/pdfs/reference/us_regdiv.pdf

[4] Northeast encompasses New York, New Jersey, Vermont, Massachusetts, Rhode Island, Maine, Connecticut, New Hampshire and Pennsylvania. The Midwest contains all north-central states: Illinois, Indiana, Iowa, Michigan, Minnesota, Missouri, Ohio and Wisconsin. South is made up of Delaware, Florida, Georgia, Maryland, North and South Carolina, (West) Virginia, Alabama, Kentucky, Mississippi, Tennessee, Arkansas, Louisiana, Oklahoma and Texas. The West region consists of Montana, Wyoming, Colorado, New Mexico, Idaho, Utah, Arizona, Nevada, Washington, Oregon, California, Alaska, and Hawaii.

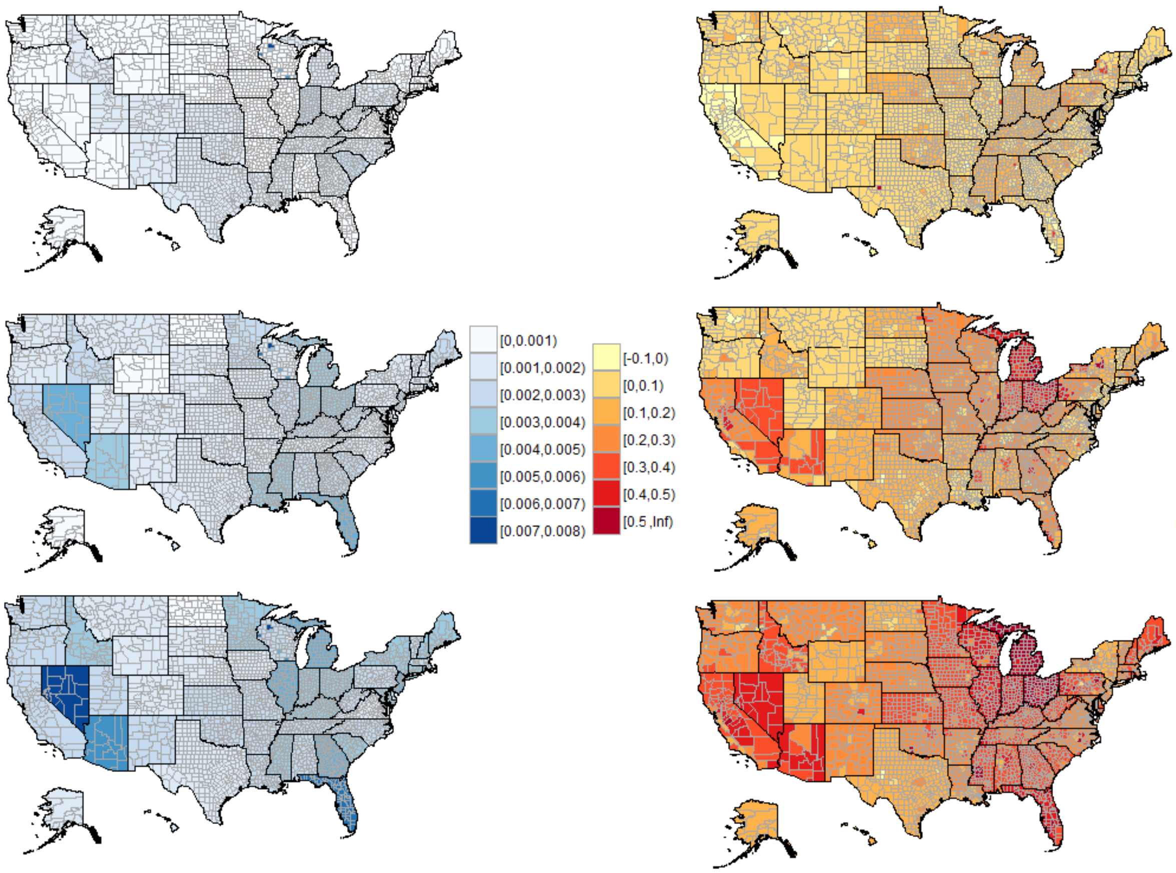
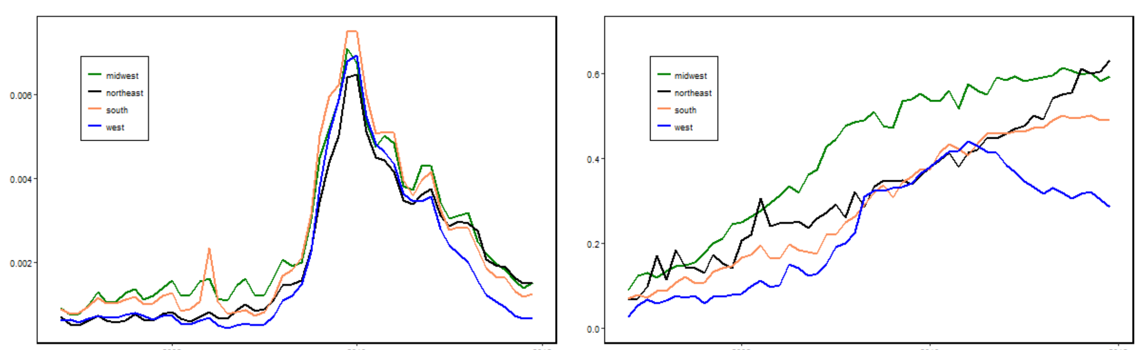


Figure 9: This figure presents the default rate and the average loss over time across the US. The left side geographical chart corresponds with the default rate. The right side charts corresponds with the average loss. The top row shows the respective data for the period 2002-2005, the middle row for 2006-2009 and the bottom row 2010-2014. Most of the loans only report the state they are originated in, the county level values are updated when MSA data is available.



(a) (b)

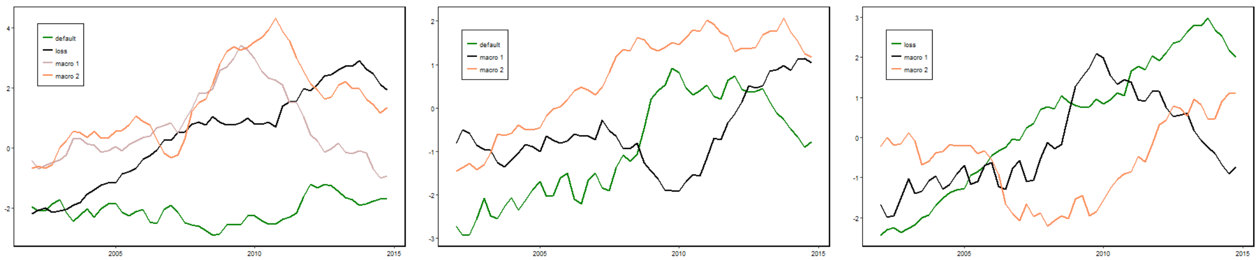
Figure 10: This figure presents the observed default rate in subpanel (a) and the average loss given default in subpanel (b). Each subpanel contains time series for the four regions Midwest, Northeast, South and West over the period 2002-2014.

Californa and Nevada were very likely to abandon their home during the financial crisis, providing a possible explanation for the different average loss pattern in the West region. The Midwest, Northeast, South and West contain 193,119, 123,452, 280,043 and 244,082 default observations, respectively, over the period 2002-2015. Furthermore the regions contain 121,444, 36,171, 151,410 and 137,383 loss given default observations over this period. The total amount of loans per region is 5,228,415, 3,728,378, 7,059,134 and 9,190,706. Resulting in average default rates of 3.69%, 3.31%, 3.97% and 2.66%.

In order to further determine whether defaults and loss given default follow their own dynamics, we use the same models as in the previous subsection. For each region parameter estimation results of the three model specifications are presented in the appendix in Table 5 to 8. The first model uses two macroeconomic factors, a default factor and a loss given default factor. The second uses only two macroeconomic factors and a default factor. The third only uses two macroeconomic factors and a loss given default factor. Recall that for the whole sample these three model specifications show evidence that losses follow their own dynamic.

The Midwest region contains all North-central states, a region for which Figure 9 and 10 show across time a spike in default rates and a slow but steady increase for the average loss given default. Midwest default rates follow similar patterns as the other regions, although the overall level is higher. The average loss given default for the Midwest region is similar to what is observed for the whole sample: it increases from early 2002 and seems unaffected by the financial crisis or the spike in defaults. The average probability of default is 0.249% per quarter, higher than the 0.224% quarterly default probability of the whole sample. Furthermore, the average of μ_{1t} is equal to 0.497 (with an average mixture probability of 0.780), which is also significantly higher than the 0.406 estimated on the whole sample with a slightly lower mixture probability of 0.751. To further investigate whether losses follow their own dynamic we look into the default and loss given default components. The default component shows that for the model with two macro, one default and one loss factor the average marginal effects of the two macro factors and default factor are 0.092, 0.060 and 0.195 respectively, indicating that defaults are partially affected by latent macro factors, but mostly by a separate default factor. For the loss given default component we see that μ_{1t} is mostly explained by the loss factor, with a loading of 0.075. The macro factors have loadings of 0.032 and 0.003 and the default factor a loading of -0.002. For the logistic part that determines the mixture probability we see average marginal effects of -1.742, 3.347, 1.592 and 9.973 for all four factors, respectively. While the macro and default factors influence the mixture probability, a separate loss factor seems most important. Figure 11 shows the mean of the posterior distributions of the latent factors. As for the whole sample, each of the subpanels in Figure 11 shows factors that follow the dynamics of the average loss given default in the Midwest.

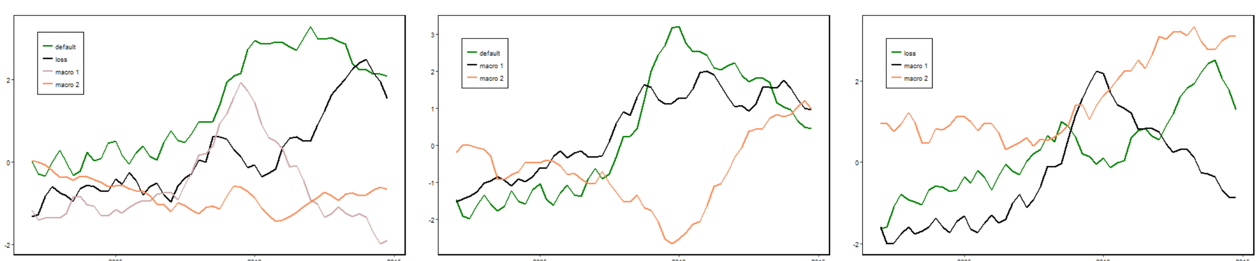
Furthermore, for the Northeast region we observe relatively low default rates and an average loss given default that has been increasing since 2002 as confirmed by both Figure 10 and 11. The average probability of default per quarter is 0.206%, which is the lowest of the entire sample. The average of μ_{1t} is equal to about 0.390 across the three models, with an average mixture probability of about 0.760. The Northeast region has a slightly larger fraction of 'bad' loans than the Midwest region, but with lower losses on average. For the default



(a) (b) (c)

Figure 11: This figure presents the mean of the posterior distributions of the latent factors that accompany the estimated factor loadings and autoregressive parameters of Table 5 in the appendix. The data comprises the Midwest region. The latent factors of the model with a macro, default and loss factor are displayed in subpanel (a). The latent factors of the model with two macro and a default factor are displayed in subpanel (b). The latent factors of the model with two macro factors and a loss factor are presented in subpanel (c).

component we see that the average marginal effect of the default factor is significantly larger than that of the macro factors, indicating that defaults follow their own dynamic, different from the macro variables. The loss given default component shows that a separate loss factor is explanatory for μ_{1t} and the mixture probability, although the macro factors and default factor have factor loadings of similar size. Figure 12 further adds to the findings by showing that all model specifications have factors with unique dynamics. Due to the similar size of the factor loadings for μ_{1t} and the mixture probability, one may find that loss given default in the Northeast region follow their own cycle, while being influenced by latent macro and default factors.



(a) (b) (c)

Figure 12: This figure presents the mean of the posterior distributions of the latent factors that accompany the estimated factor loadings and autoregressive parameters of Table 6 in the appendix. The data comprises the Northeast region. The latent factors of the model with a macro, default and loss factor are displayed in subpanel (a). The latent factors of the model with two macro and a default factor are displayed in subpanel (b). The latent factors of the model with two macro factors and a loss factor are presented in subpanel (c).

The South region shows similar default rates and average loss given default patterns as the Northeast. In 2006 the default rate suddenly increases for southern US mortgages, an effect on the average loss given default is not immediately visible. Similarly to the previous regions the average loss given default in the south increases steadily from 2002. The average probability of default per quarter is equal to 0.241% per quarter, while for the main model specification with two macro, one default and one loss factor the average of μ_{1t} is 0.365

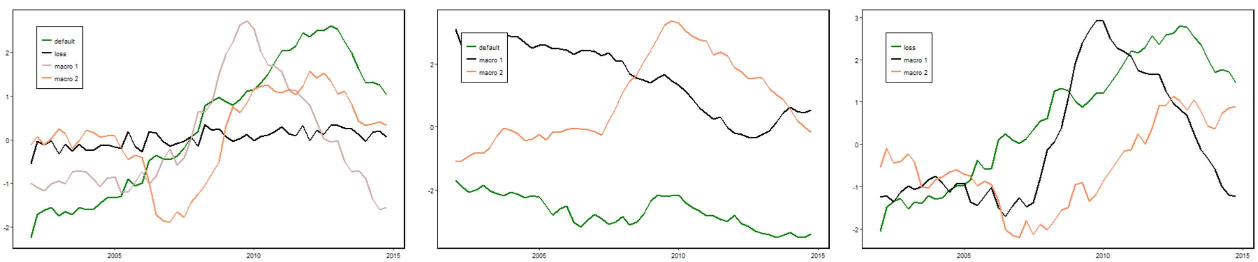
with an average mixture probability of 0.735. The default rate is mostly driven by a separate default factor with a loading of 0.911, which is large compared to the macro loadings of 0.210 and 0.157. For μ_{1t} we see that it is partly driven by the two macro factors with loadings of 0.016 and 0.033 and also by a separate loss factor with a loading of 0.059, the loading onto the default factor is -0.017 with a highest density interval centered around 0 and seems unnecessary. For the mixture probability we see that the separate default factor has by far the largest loading with 0.532, compared to 0.000, 0.157 and -0.052 for the two macro and default factors respectively. For the model with two macro factors and a loss factor a similar result is seen, with an average marginal effect of 10.851 for the loss factor and -0.474 and -1.285 for the two macro factors. Figure 13 show for each model specification a factor that follows the average loss given default dynamic of the South region. The loadings and factors imply a presence of a unique loss given default cycle for the South region.

(a) (b) (c)

Figure 13: This figure presents the mean of the posterior distributions of the latent factors that accompany the estimated factor loadings and autoregressive parameters of Table 7 in the appendix. The data comprises the South region. The latent factors of the model with a macro, default and loss factor are displayed in subpanel (a). The latent factors of the model with two macro and a default factor are displayed in subpanel (b). The latent factors of the model with two macro factors and a loss factor are presented in subpanel (c).

Lastly, the West region is especially interesting as Figure 10 shows that the average loss given default for this region seems to be impacted the most by the crisis: the increase before and the decrease after the financial crisis of the average loss given default is sudden compared to other Regions. The average probability of default per quarter is 0.189%, the lowest of all regions. The average of μ_{1t} is about 0.297 for the three model specifications, with a mixture probability of about 0.723. The West region seems to have default probabilities that can largely be explained by the two macro factors, as average marginal effects for the two macro factors are in general multiples of the default factor; for example for the specification with two macro, one default and one loss factor the average marginal effects are 0.097, 0.066 (for the two macro factors respectively) and 0.013 (for the default factor). When inspecting the model with two macro factors, one default and one loss factor we see that μ_{1t} depends mostly on the first macro factor and the default factor, with loadings of respectively 0.038 and 0.053, significantly larger than the loadings onto the second macro factor and loss factor (being -0.004 and 0.007, respectively). The mixture probability shows the largest loading onto the default factor: 0.528. Figure 14 panel (a) shows how, also for the West region, the default factor largely follows the average loss given default over time, which is expected given the large loadings in the loss given default component onto the default factor.

Furthermore, panel (a) shows how an additional loss factor becomes uninformative. Panel (b) then shows that the specification without a loss factor presents stable dynamics, although the WAIC2 value deteriorates for the loss given default component. Panel (c) shows several factors of which one seems to track the average loss given default.



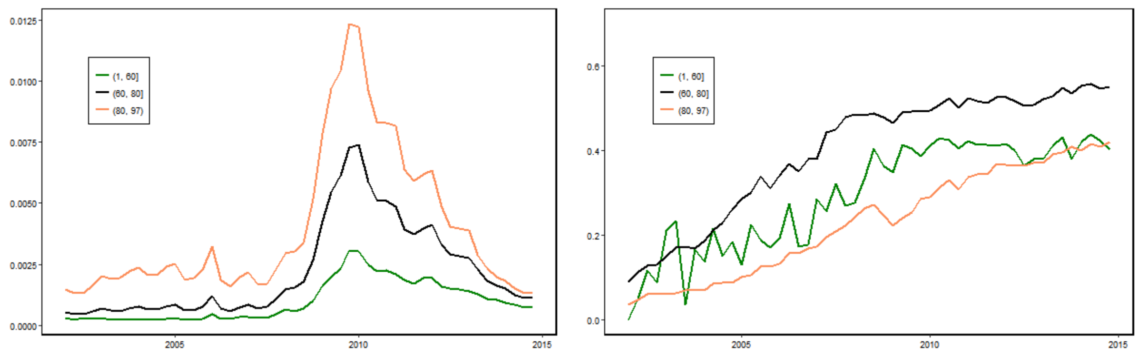
(a) (b) (c)

Figure 14: This figure presents the mean of the posterior distributions of the latent factors that accompany the estimated factor loadings and autoregressive parameters of Table 8 in the appendix. The data comprises the West region. The latent factors of the model with a macro, default and loss factor are displayed in subpanel (a). The latent factors of the model with two macro and a default factor are displayed in subpanel (b). The latent factors of the model with two macro factors and a loss factor are presented in subpanel (c).

5.3 Loan-to-value characteristics

Previous research finds that loan-to-value (LTV) strongly relates to losses: Qi & Yang (2009), Calem & LaCour-Little (2004), Pennington-Cross et al. (2003) all discuss how LTV is a significant explanatory variable for loss given default. In my analysis of LTV characteristics I consider three categories: LTV ≤ 60%, 60 % < LTV ≤ 80%, and LTV > 80%. These categories are based on the categories of the risk-based stress tests for Fannie Mae by the Office of Federal Housing Enterprise Oversight (Frame et al., 2013).

All Fannie Mae mortgages are originated with LTVs below 97%, indicating that at origination the value of the loan is always less than the value of the underlying house. The LTV changes over time due to fluctuating house prices and loan payments, such that the group with a LTV of at least 80% at origination is most susceptible to having a house value unable to cover the loan value in case of economic downturns. Following the data filtration, the data set contains 4,347,671 unique loans with a LTV below 60%, 8,354,780 loans with a LTV between 60% and 80% and 5,088,471 loans with a LTV above 80% (and below 97%). Figure 15 displays the quarterly default rate and average loss given default across the three LTV categories. One can observe that the quarterly default rate increases with LTV, with average default rates of 0.097%, 0.219% and 0.392% for the LTV ≤ 60%, 60 % < LTV ≤ 80%, LTV > 80% categories, respectively. The same figure shows for average loss given default that the LTV category 60% < LTV ≤ 80% is significantly higher than the LTV ≤ 60% category. Interestingly the category with the highest LTV values, LTV > 80%, shows the lowest average given default over time, which is counter-intuitive at first as a high LTV usually relates to a limited ability to cover the outstanding loan in case of a default. The main explanation comes from the mortgage insurance that is present for 234,092 out of 256,027 loss observations in the LTV > 80% category, while the other two LTV categories have no, or a negligible amount of insured mortgages. This finding is in line with L. S. Goodman & Zhu (2015), who research similar mortgage data and find that mortgage insurance plays a significant role in lowering loss given default. I proceed with doing a separate analysis per LTV category, starting with the LTV ≤ 60% category.



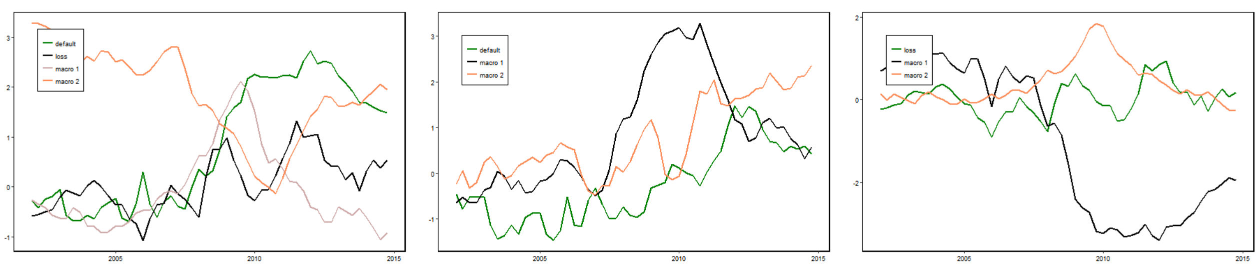
(a) (b)

Figure 15: This figure presents the observed default rate in subpanel (a) and the average loss given default in subpanel (b). Each subpanel contains time series for the three LTV categories: (1, 60], (60, 80] and (80, 97].

The setup is similar as before, comparing and studying results of the mixed measurement dynamic factor model with a combination of two macro factors, one default factor and one loss factor. The means of the posterior distributions of the latent factors are displayed in this subsection, and the estimation results of factor loadings and autoregressive parameters are presented in the appendix.

The category with LTVs below 60% is unique, and shows a relatively mild increase in default rates during the crisis as presented in Figure 15. This is expected given that a low LTV indicates small mortgage payments and thus a higher barrier for a borrower to be unable to cover his or her payments. Moreover, before the financial crisis default rates are comparable between the LTV ≤ 60% and 60% < LTV ≤ 80% category, but the latter category shows significantly more exposure to the financial crisis. The average loss given default in Figure 15 seems rather volatile due to a small number of loss observations (23,050). In general we see that for this category losses slowly increase until the financial crisis, after which the average loss given default flattens. The average quarterly default rate is 0.094%, whereas μ_0 is close to 0 and the average of μ_{1t} equals 0.407, with an average mixture probability of 0.653, implying relatively low losses for this category. Across the three models the latent factors in Figure 16 show unique dynamics. For the model with two macro, one default and one loss factor the average marginal effects of the first and second macro factor onto the default probability is 0.025% and -0.013%, respectively. An increase in the first macro factor is accompanied with a large decrease in change of GDP, IP, DI and UR, and large positive increases in RVR and HVR. The second macro factor mostly positively affects changes in DI and FRM. Most interesting is that on average low LTV loans present a decrease in default as FRM increases, unfortunately explanations are numerous here and can be given based on origination date, repayment type and amortization type, as these factors influence the mortgage interest rate.

In Figure 16 the two macro factors both pick up a signal around the financial crisis, whereas the default factor shows a steep increase and the loss factor cycles over time. The other two models in panel (b) and (c) show similar factor dynamics. For the two macro, one default, one loss factor the default factor shows an average marginal effect of 0.055%, which is expected given the factor's shape. The factors evenly affect μ_{1t}, with similar orders of magnitude for each factor's loading. However, for the mixture probability, the default and loss factor have sizable average marginal effects of 13.601% and 4.601%, respectively, whereas marginal effects for the macro factors are -0.367% and -1.059%, indicating that the mixture probability for loss given default is mostly related to latent factors irrelevant to macro, however the default factor affects both the default rate and mixture probability quite significantly.



(a) (b) (c)

Figure 16: This figure presents the mean of the posterior distributions of the latent factors that belong to the estimated factor loadings and autoregressive parameters of Table 9 in the appendix. The latent factors of the model with a macro, default and loss factor are displayed in subpanel (a). The latent factors of the model with two macro and a default factor are displayed in subpanel (b). The latent factors of the model with two macro factors and a loss factor are presented in subpanel (c).

The category with LTVs between 60% and 80% shows the highest average loss over time and a relatively moderate increase in default rates around the financial crisis, as shown in Figure 15. This LTV subset of mortgage loans shows the highest average loss over time, which is explainable by mortgage insurance for the 60 to 80 percent LTV category. This subset of the mortgage loans shows an average quarterly default rate p^d of 0.218%, whereas μ_0 is equal to -0.001 and the average of μ_{1t} equals 0.444, with an average mixture probability of 0.827. Here the analysis is based on the 2 macro, 1 default, 1 loss factor model, and conclusions of the other models can be derived analogously. The first macro factor shows negative loadings for the change in GDP, IP, DI and FRM and positive loadings for UR, RVR and HVR, whereas the second macro factor has negative loadings on all macro variables except for UR. The first and second macro factor show average marginal effects for the default rate of 0.066% and 0.030%, respectively, whereas a separate default factor has a marginal effect of 0.147%. In general decreases in the changes in GDP, IP and DI are accompanied with increases in the default rate, whereas a decrease in UR, RVR or HVR relate to a decrease in the default rate. An additional default factor has the largest explanatory power for the default rate, with an average marginal effect of 0.147%. μ_{1t} seems to have comparable loadings across latent factors, although only the loss factor shows a positive loading. For the mixture probability the loss factor shows a large average marginal effect of 6.991%, which is sizable especially when compared to the first and second macro factor, which are with -0.184% and -0.974% as average marginal effects, respectively. The default factor shows an average marginal effect of 2.056%. In general the factors all show unique dynamics in Figure 17, with either default or loss factors following an average loss given default pattern over time.

(a) (b) (c)

Figure 17: This figure presents the mean of the posterior distributions of the latent factors that belong to the estimated factor loadings and autoregressive parameters of Table 10 in the appendix. The latent factors of the model with a macro, default and loss factor are displayed in subpanel (a). The latent factors of the model with two macro and a default factor are displayed in subpanel (b). The latent factors of the model with two macro factors and a loss factor are presented in subpanel (c).

Furthermore the category with LTVs above 80%, where the default rate is highest, is with the lowest average loss given default due to private mortgage insurances. The average default rate is 0.392%, while μ_0 is estimated as 0.004 and μ_{1t} averages to 0.314, with an average mixture probability of 0.681. The default rate is significantly higher than that of the other two categories, not only during but also before the financial crisis. The average loss given default for the above 80 % category is mostly comparable to that of the below 60 % category, reaching a similar average loss at the end of the data set. I discuss again the two macro, one default, one loss factor model in detail. The average marginal effect of the first and second macro factor is 0.111% and 0.225%, respectively. The first factor shows negative loadings for GDP, UR, RVR, DI and FRM, and positive loadings for IP and HVR, whereas the second factor has negative loadings for IP and DI, and positive loadings for UR, RVR, HVR and FRM. The second factor shows significantly larger absolute coefficients for almost all macro variables, explaining the higher average marginal effect of this factor onto p^d of 0.225%. The two macro, one default, one loss model shows the lowest WAIC2 value, 53.440 as compared to 65.317 and 59.415 for the other two models. Moreover, the default factor shows an average marginal effect of 0.110%, while capturing a pattern related to the average loss given default. The WAIC2 for the default component is the lowest for the two macro, one default, one loss model, with a value of 22.441, whereas the other two models show values of 25.83 and 26.023. The factor loading onto μ_{1t} is highest for the default factor with a value of 0.057, which is expected as it follows dynamics of the average loss given default, whereas the other factors show negligible loadings. Similarly, for the mixture probability p_t^d One observes that the average marginal effect of the default factor is highest, at 12.334%, although the first and second macro factors have a relatively large affect with 1.756% and -5.062%.

(a) (b) (c)

Figure 18: This figure presents the mean of the posterior distributions of the latent factors that belong to the estimated factor loadings and autoregressive parameters of Table 11 in the appendix. The latent factors of the model with a macro, default and loss factor are displayed in subpanel (a). The latent factors of the model with two macro and a default factor are displayed in subpanel (b). The latent factors of the model with two macro factors and a loss factor are presented in subpanel (c).

6 Conclusion

The probability of default and loss given default of conventional US mortgages both follow different cycles. Probability of default can largely be explained by macro factors, which indicates that probability of default for conventional US mortgages is mostly related to macro events. Interestingly, loss given default follows its own cycle that is unrelated to either macro or default observations, presenting a weak relation between probability of default and loss given default for US mortgages. These results not only hold across the US, but also for regions in the US. Moreover, the results also persist for uninsured mortgages with a loan to value ratio above 60%. The weak relation implies that probability of default or loss given default of conventional US mortgage loans may be modeled accurately without accounting for their dependence.

References

Allen, L., & Saunders, A. (2004). Incorporating systemic influences into risk measurements: A survey of the literature. *Journal of financial services research*, *26*(2), 161–191.

Altman, E., Resti, A., & Sironi, A. (2004). Default recovery rates in credit risk modelling: a review of the literature and empirical evidence. *Economic Notes*, *33*(2), 183–208.

An, X., & Cordell, L. R. (2017). Regime shift and the post-crisis world of mortgage loss severities. *FRB of Philadelphia Working Paper No. 17-8*.

Azizpour, S., Giesecke, K., & Schwenkler, G. (2017). Exploring the sources of default clustering. *Journal of Financial Economics (JFE), Forthcoming*.

Barth, J. (2009). *The rise and fall of the us mortgage and credit markets: A comprehensive analysis of the market meltdown*. John Wiley & Sons.

Bhutta, N., Shan, H., & Dokko, J. (2010). The depth of negative equity and mortgage default decisions.

Calem, P. S., & LaCour-Little, M. (2004). Risk-based capital requirements for mortgage loans. *Journal of Banking & Finance*, *28*(3), 647–672.

Carter, C. K., & Kohn, R. (1994). On gibbs sampling for state space models. *Biometrika*, *81*(3), 541–553.

Caselli, S., Gatti, S., & Querci, F. (2008). The sensitivity of the loss given default rate to systematic risk: new empirical evidence on bank loans. *Journal of Financial Services Research*, *34*(1), 1–34.

Creal, D. (2017). A class of non-gaussian state space models with exact likelihood inference. *Journal of Business & Economic Statistics*, *35*(4).

Creal, D., Schwaab, B., Koopman, S. J., & Lucas, A. (2014). Observation-driven mixed-measurement dynamic factor models with an application to credit risk. *Review of Economics and Statistics*, *96*(5), 898–915.

De Jong, P., & Shephard, N. (1995). The simulation smoother for time series models. *Biometrika*, *82*(2), 339–350.

Duffie, D., Saita, L., & Wang, K. (2007). Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, *83*(3), 635–665.

Durbin, J., & Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, *89*(3), 603–616.

Elul, R., Souleles, N. S., Chomsisengphet, S., Glennon, D., & Hunt, R. M. (2010). What 'triggers' mortgage default? *FRB of Philadelphia Working Paper No. 10-13*.

Frame, W. S., Gerardi, K., & Willen, P. S. (2013). Supervisory stress tests, model risk, and model disclosure: lessons from ofheo. In *Working paper, fed. reserve bank atlanta*.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of time series analysis*, *15*(2), 183–202.

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for bayesian models. *Statistics and Computing*, *24*(6), 997–1016.

Ghent, A. C., & Kudlyak, M. (2011). Recourse and residential mortgage default: evidence from us states. *The Review of Financial Studies*, *24*(9), 3139–3186.

Goodman, & Smith. (2010). Residential mortgage default: Theory works and so does policy. *Journal of Housing Economics*, *19*(4), 280–294.

Goodman, L. S., & Zhu, J. (2015). Loss severity on residential mortgages: Evidence from freddie mac's newest data. *The Journal of Fixed Income*, *25*(2), 48–57.

Guiso, L., Sapienza, P., & Zingales, L. (2013). The determinants of attitudes toward strategic default on mortgages. *The Journal of Finance*, *68*(4), 1473–1515.

Hartmann-Wendels, T., Miller, P., & Töws, E. (2014). Loss given default for leasing: Parametric and nonparametric estimations. *Journal of Banking & Finance*, *40*, 364–375.

Höcht, S., & Zagst, R. (2007). Generalized maximum expected utility models for default risk-a comparison of models with different dependence structures. *Journal of Credit Risk*, *3*(3), 3–24.

Keijsers, B., Diris, B. F., & Kole, E. (2017). Cyclicality in losses on bank loans. *Journal of Applied Econometrics, Forthcoming*.

Koopman, S. J. (2010). *Mixed measurement dynamic factor models* (Unpublished doctoral dissertation). VU University Amsterdam.

Koopman, S. J., Lucas, A., & Schwaab, B. (2012). Dynamic factor models with macro, frailty, and industry effects for us default counts: the credit crisis of 2008. *Journal of Business & Economic Statistics*, *30*(4), 521–532.

Mishkin, F. S. (2007). *Housing and the monetary transmission mechanism* (Tech. Rep.). National Bureau of Economic Research.

Pennington-Cross, A., et al. (2003). *Subprime and prime mortgages–loss distributions* (Tech. Rep.). Federal Housing Finance Agency.

Polson, N. G., Scott, J. G., & Windle, J. (2013). Bayesian inference for logistic models using pólya–gamma latent variables. *Journal of the American statistical Association*, *108*(504), 1339–1349.

Qi, M., & Yang, X. (2009). Loss given default of high loan-to-value residential mortgages. *Journal of Banking & Finance*, *33*(5), 788–799.

Schuermann, T. (2004). *What do we know about loss given default?" ch. 9 in david shimko (ed.) credit risk: Models and management.* London, UK: Risk Books.

Shiller, R. J. (2006). Long-term perspectives on the current boom in home prices. *The Economists' Voice*, *3*(4).

Tsatsaronis, K., & Zhu, H. (2004). What drives housing price dynamics: cross-country evidence. *BIS Quarterly Review, March 2004*.

Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*(Dec), 3571–3594.

Wheaton, W., & Nechayev, G. (2008). The 1998-2005 housing "bubble" and the current "correction": What's different this time? *Journal of real Estate research*, *30*(1), 1–26.

Windle, J., Carvalho, C. M., Scott, J. G., & Sun, L. (2013). Efficient data augmentation in dynamic models for binary and count data. *arXiv preprint arXiv:1308.0774*.

Appendices

A Figures

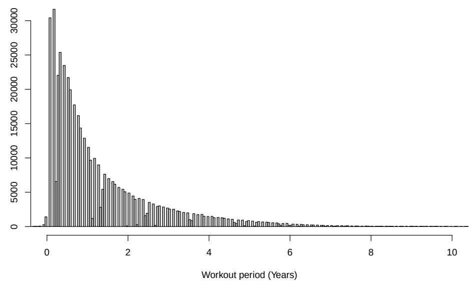
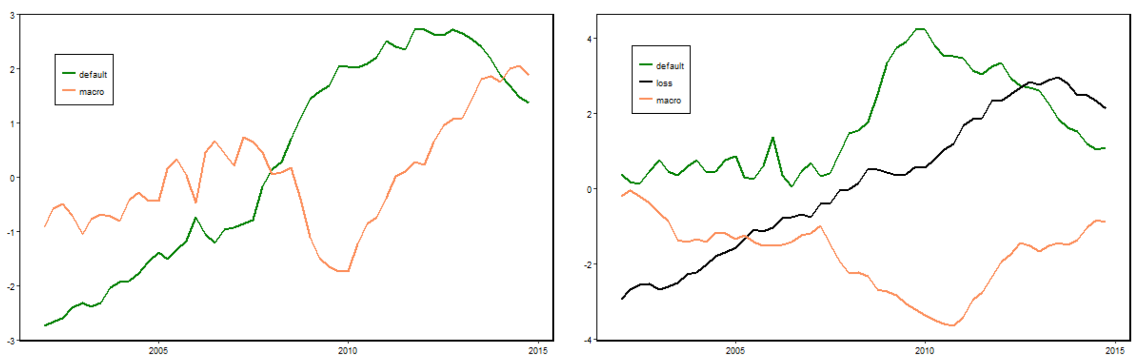


Figure 19: This figure presents the empirical workout period of the mortgage loans over the period 2000-2015. The x axis represents the number of years the workout period lasts and the y axis represents the amount of observations.

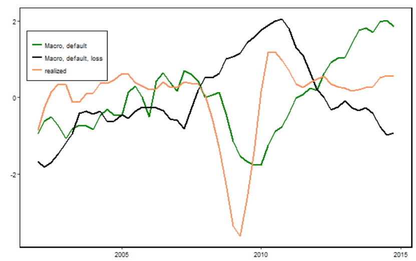


(a)

(b)

Figure 20: This figure presents the mean of the posterior distributions of the latent factors that accompany the estimated factor loadings and autoregressive parameters of three model specifications with one macroeconomic factor. The latent factors of the model with only a macroeconomic and default factor is shown in subpanel (a). The latent factors of the model with a macroeconomic, default and loss factor are displayed in subpanel (b).

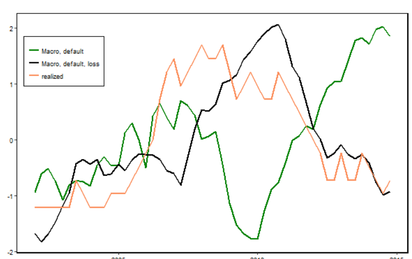
(a)



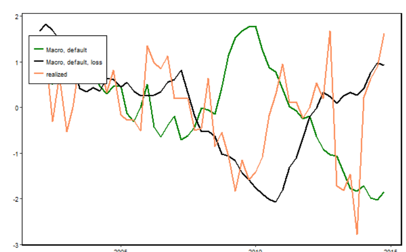
(b)

(c)

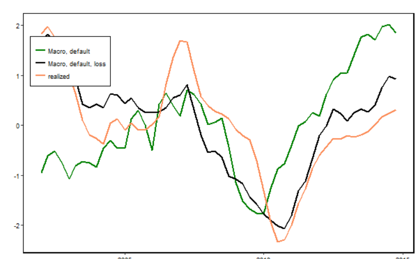
(d)



(e)

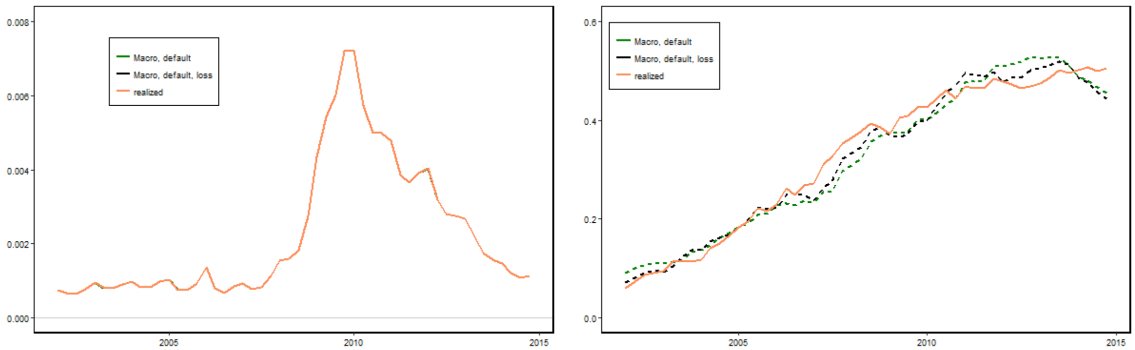


(f)



(g)

Figure 21: This figure presents the realized time series for GDP, IP, UR, RVR, HVR, DI and FRM. Besides the realized time series it also shows the model fit for each of the two model specifications of table 20. Each subpanel contains a specific timeseries and model fits. Panel (a), (b), (c), (d), (e), (f), (g) contain GDP, IP, UR, RVR, HVR, DI and FRM respectively.



(a)

(b)

Figure 22: This figure presents the observed default rate and the average loss given default for the period 2002-2014, as well as model fits for the two different model specifications of table 3. Panel (a) contains the observed default rate and panel (b) the average loss given default over time.

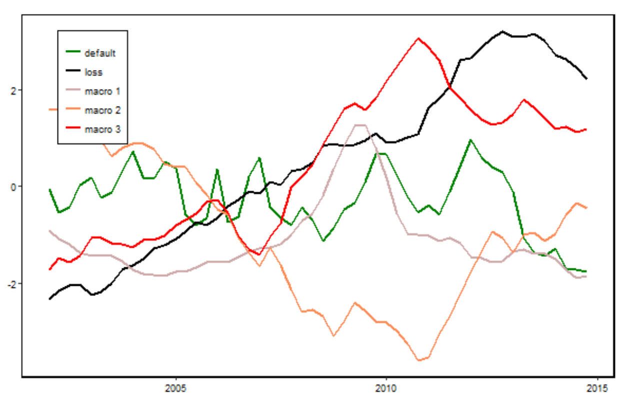
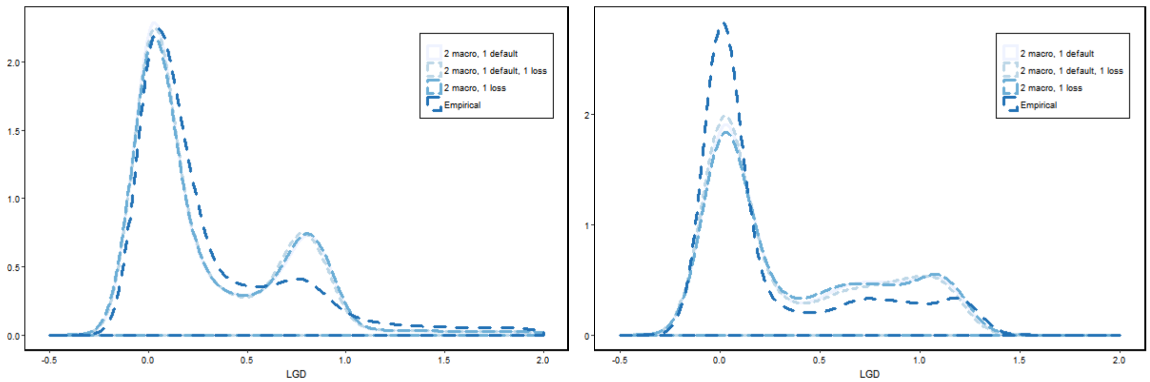


Figure 23: This figure presents the mean of the posterior distributions of the latent factors that accompany the estimated factor loadings and autoregressive parameters of the three macro factor model specification with one default and loss factor.

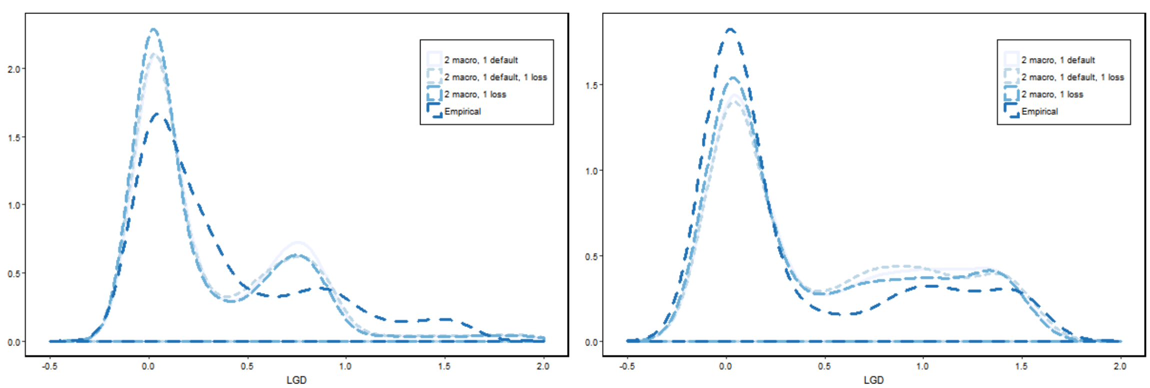
LGD Fits



(a) (b)

Figure 24: This figure presents the cross-sectional loss given default distribution fits for the quarters 2006 Q4 (subpanel (a)) and 2012 Q2 (subpanel (b)). The fits are given for three models: with two macro factors and one default and loss factor, or two macro factors and either one default or loss factor.

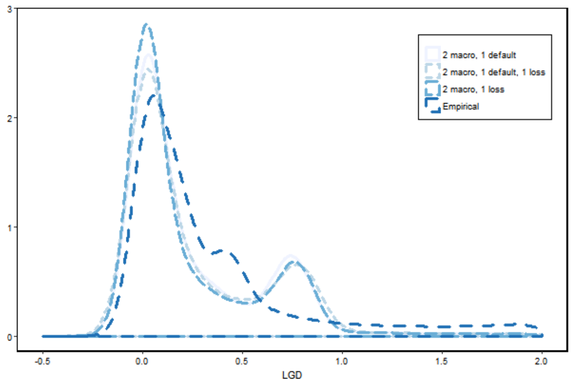
Midwest



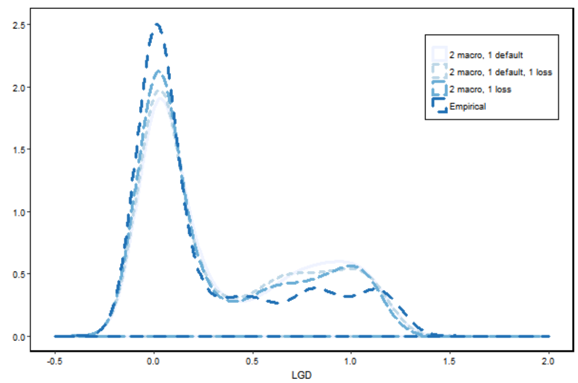
(a) (b)

Figure 25: This figure presents the cross-sectional loss given default distribution fits for the quarters 2006 Q4 (subpanel (a)) and 2012 Q2 (subpanel (b)). The fits are given for three models: with two macro factors and one default and loss factor, or two macro factors and either one default or loss factor. The data sub-sample is the Midwest region.

Northeast



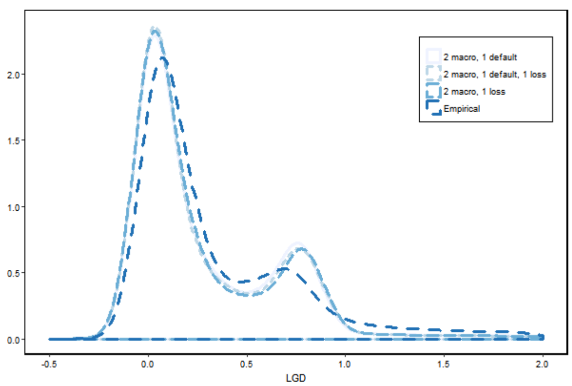
(a)



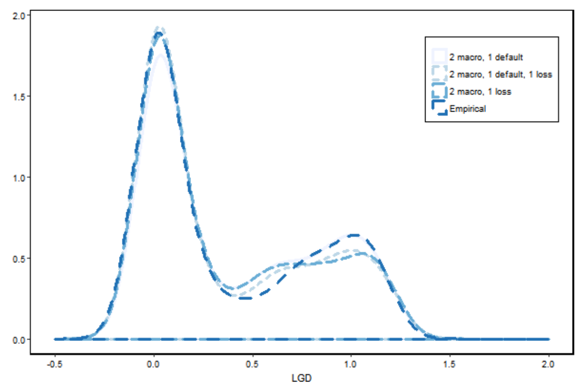
(b)

Figure 26: This figure presents the cross-sectional loss given default distribution fits for the quarters 2006 Q4 (subpanel (a)) and 2012 Q2 (subpanel (b)). The fits are given for three models: with two macro factors and one default and loss factor, or two macro factors and either one default or loss factor. The data sub-sample is the Northeast region.

South



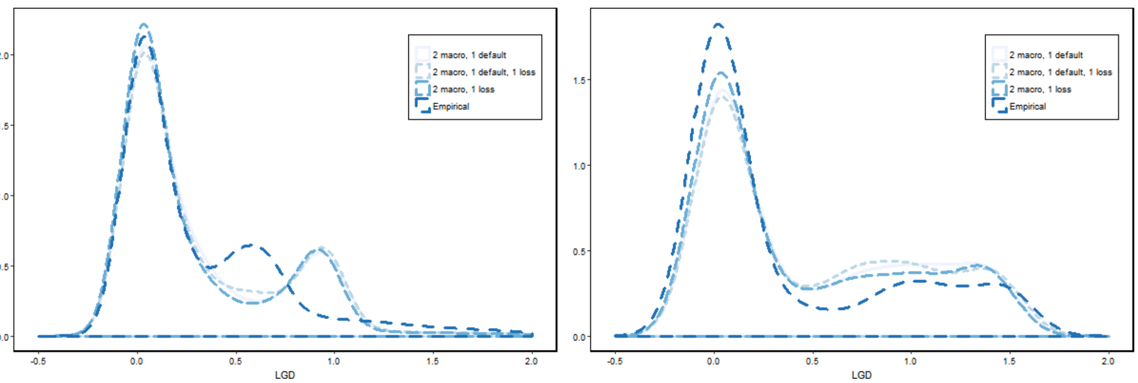
(a)



(b)

Figure 27: This figure presents the cross-sectional loss given default distribution fits for the quarters 2006 Q4 (subpanel (a)) and 2012 Q2 (subpanel (b)). The fits are given for three models: with two macro factors and one default and loss factor, or two macro factors and either one default or loss factor. The data sub-sample is the South region.

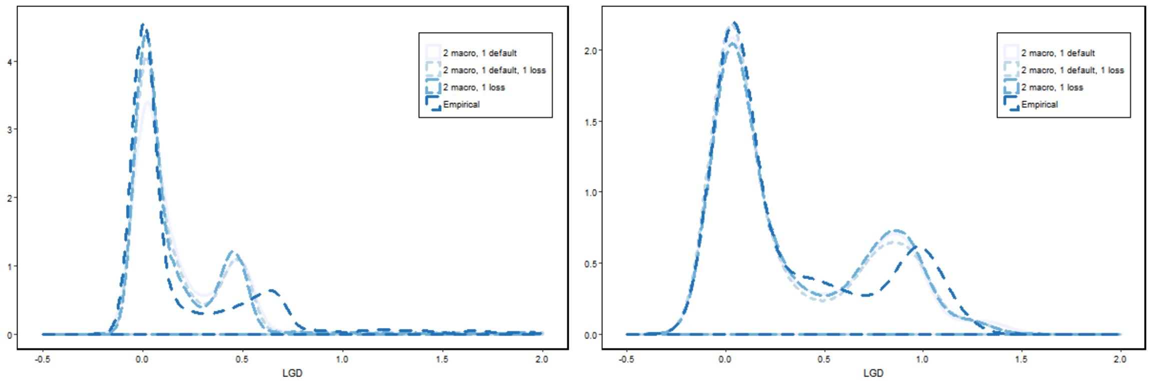
West



(a) (b)

Figure 28: This figure presents the cross-sectional loss given default distribution fits for the quarters 2006 Q4 (subpanel (a)) and 2012 Q2 (subpanel (b)). The fits are given for three models: with two macro factors and one default and loss factor, or two macro factors and either one default or loss factor. The data sub-sample is the West region.

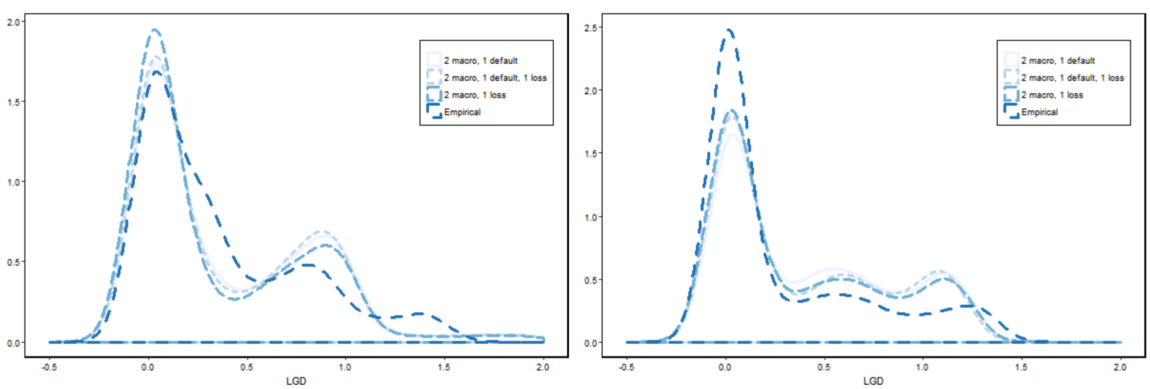
LTV below 60%



(a) (b)

Figure 29: This figure presents the cross-sectional loss given default distribution fits for the quarters 2006 Q4 (subpanel (a)) and 2012 Q2 (subpanel (b)). The fits are given for three models: with two macro factors and one default and loss factor, or two macro factors and either one default or loss factor. The data sub-sample is the below 60% LTV category.

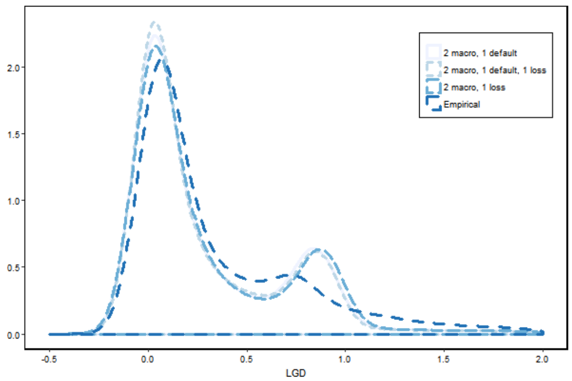
LTV between 60% and 80 %



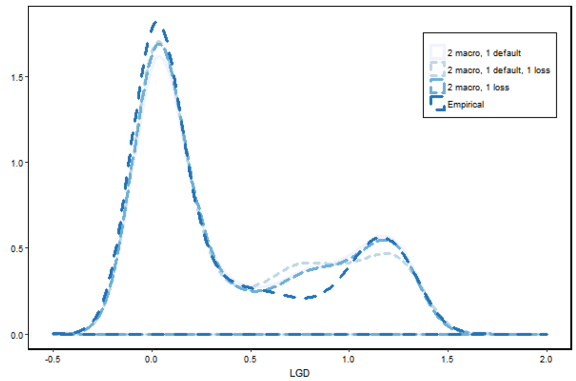
(a) (b)

Figure 30: This figure presents the cross-sectional loss given default distribution fits for the quarters 2006 Q4 (subpanel (a)) and 2012 Q2 (subpanel (b)). The fits are given for three models: with two macro factors and one default and loss factor, or two macro factors and either one default or loss factor. The data sub-sample is the between 60% and 80% LTV category.

LTV above 80%



(a)



(b)

Figure 31: This figure presents the cross-sectional loss given default distribution fits for the quarters 2006 Q4 (subpanel (a)) and 2012 Q2 (subpanel (b)). The fits are given for three models: with two macro factors and one default and loss factor, or two macro factors and either one default or loss factor. The data sub-sample is the above 80% LTV category.

B Tables

	Macro, default		Macro, default, loss	
Panel A: Factor				
ϕ_{m_1}	0.840	*(0.443, 0.985)*	0.929	*(0.770, 0.990)*
ϕ_d	0.928	*(0.785, 0.978)*	0.924	*(0.804, 0.989)*
ϕ_l			0.936	*(0.768, 0.992)*
Panel B: Macro				
$\beta_{GDP,1}$	-0.056	*(-0.162, 0.000)*	-0.050	*(-0.129, -0.001)*
$\beta_{IP,1}$	0.115	*(-0.053, 0.272)*	-0.169	*(-0.374, 0.032)*
$\beta_{UR,1}$	-0.593	*(-0.775, -0.408)*	-0.133	*(-0.568, 0.177)*
$\beta_{RVR,1}$	-0.777	*(-1.014, -0.560)*	-0.538	*(-1.059, -0.190)*
$\beta_{HVR,1}$	0.002	*(-0.279, 0.269)*	-0.622	*(-0.992, -0.322)*
$\beta_{DI,1}$	-0.060	*(-0.344, 0.212)*	-0.165	*(-0.110, 0.472)*
$\beta_{FRM,1}$	0.194	*(-0.191, -0.585)*	-0.963	*(-1.101, -0.783)*
WAIC2	65.867		68.421	
Panel C: Defaults				
α^d	-6.502	*(-6.532, -6.466)*	-7.423	*(-7.589, -7.322)*
β_1^d	-0.442	*(-0.479, -0.396)*	-0.068	*(-0.212, 0.030)*
γ^d	0.400	*(0.370, 0.434)*	0.533	*(0.467, 0.616)*
av. $p^d(x10^-2)$	0.224	*(0.223, 0.224)*	0.224	*(0.223, 0.224)*
m.e. of f_1^m	-0.099		-0.015	
m.e. of f^d	0.090		0.119	
WAIC2	25.682		25.761	
Panel D: Loss given default				
μ_0	0.001		0.001	
av. μ_1	0.406	*(0.229, 0.564)*	0.406	*(0.197, 0.553)*
α_{μ_1}	0.388	*(0.345, 0.426)*	0.338	*(0.265, 0.402)*
$\beta_{\mu_1 1}$	0.019	*(0.000, 0.037)*	-0.049	*(-0.086, -0.006)*
γ_{μ_1}	0.058	*(0.045, 0.073)*	-0.016	*(-0.049, 0.018)*
δ_{μ_1}			0.005	*(0.041, 0.080)*
$\sigma_0(x10^-2)$	0.025	*(0.024, 0.026)*	0.025	*(0.024, 0.026)*
σ_1	0.099	*(0.098, 0.099)*	0.099	*(0.098, 0.099)*
α^l	1.180	*(0.821, 1.50)*	1.099	*(0.577, 1.562)*
β_1^l	0.241	*(0.116, 0.355)*	-0.055	*(-0.226, 0.120)*
γ^l	0.466	*(0.354, 0.588)*	0.037	*(-0.092, 0.166)*
δ^l			0.498	*(0.392, 0.604)*
av. p^l	0.750	*(0.467, 0.938)*	0.750	*(0.462, 0.939)*
m.e. of f_1^m	4.522		-1.030	
m.e. of f^d	8.723		0.693	
m.e. of f^l			9.333	
WAIC2				

Table 3: This table presents the estimation results accompanying figure 20, displaying two model specifications with either three unique factors or a specification with no loss factor.

			3 Macro, default, loss			
			Panel A: Factor			
ϕ_{m_1}	0.880				*(0.565, 0.990)*	
ϕ_{m_2}	0.926				*(0.782, 0.989)*	
ϕ_{m_3}	0.905				*(0.708, 0.988)*	
ϕ_d	0.764				*(0.533, 0.988)*	
ϕ_l	0.924				*(0.761, 0.990)*	
			Panel B: Macro			
	factor 1		*factor 2*		*factor 3*	
β_{GDP}	-1.216	*(-1.412, -1.021)*	-	-	-	-
β_{IP}	-1.479	*(-1.812, -0.984)*	-0.240	*(-0.444, -0.050)*	-	-
β_{UR}	1.359	*(0.842, 1.683)*	0.291	*(0.050, 0.409)*	-0.044	*(-0.089, -0.010)*
β_{RVR}	0.423	*(0.209, 0.681)*	-0.522	*(-0.894, -0.301)*	-0.713	*(-0.909, -0.503)*
β_{HVR}	0.143	*(0.071, 0.259)*	-0.891	*(-0.955, 0.801)*	-0.482	*(-0.645, -0.259)*
β_{DI}	-0.780	*(-0.944, -0.029)*	-0.218	*(-0.394, 0.039)*	-0.447	*(-0.591, -0.305)*
β_{FRM}	-0.359	*(-0.500, -0.109)*	0.084	*(0.020, 0.145)*	-0.466	*(-0.674, -0.201)*
WAIC2			100.935			
			Panel C: Defaults			
α^d	-6.168				*(-6.256, -6.087*	
β_1^d	0.270				*(0.216, 0.320)*	
β_2^d	0.019				*(-0.106, 0.054))*	
β_3^d	0.457				*(0.417, 0.491)*	
γ^d	0.376				*(0.291, 0.431)*	
av. $p^d(x10^-2)$	0.224				*(0.223, 0.225*	
m.e. of f_1^m			0.060			
m.e. of f_2^m			0.004			
m.e. of f_3^m			0.102			
m.e. of f^d			0.084			
WAIC2			25.607			
			Panel D: Loss given default			
μ_0	0.000				*(-0.001,0.000)*	
av. μ_1	0.404				*(0.204, 0.540)*	
α_{μ_1}	0.335				*(0.279, 0.390)*	
$\beta_{\mu_1 1}$	-0.004				*(-0.026, 0.016)*	
$\beta_{\mu_1 2}$	-0.036				*(-0.053, -0.020)*	
$\beta_{\mu_1 3}$	-0.002				*(-0.026, 0.020))*	
γ_{μ_1}	-0.017				*(-0.034, 0.001)*	
δ_{μ_1}	0.044				*(0.034, 0.058)*	
$\sigma_0(x10^-2)$	0.025				*(0.025, 0.025)*	
σ_1	0.099				*(0.098, 0.099)*	
α^c	0.928				*(0.432, 1.403)*	
β_1^c	-0.067				*(-0.233, 0.071)*	
β_2^c	-0.027				*(-0.141, 0.092)*	
β_3^c	0.146				*(-0.007, 0.302)*	
γ^c	-0.012				*(-0.105, 0.076)*	
δ^c	0.448				*(0.345, 0.560)*	
av. p^c	0.750				*(0.463, 0.939*	
m.e. of f_1^m			-1.263			
m.e. of f_2^m			-0.498			
m.e. of f_3^m			2.740			
m.e. of f^d			-0.230			
m.e. of f^c			8.403			
WAIC2			39865.440			

Table 4: This table presents the estimation results accompanying figure 23, displaying a model specification with three macro factors and one default and one loss factor.

Midwest

	2 Macro, default, loss		2 Macro, default		2 Macro, loss	
Panel A: Factor						
ϕ_{m_1}	0.893	*(0.643, 0.979)*	0.843	*(0.439, 0.989)*	0.869	*(0.618, 0.991)*
ϕ_{m_2}	0.931	*(0.787, 0.989)*	0.880	*(0.595, 0.987)*	0.872	*(0.567, 0.989)*
ϕ_d	0.935	*(0.800, 0.988)*	0.906	*(0.743, 0.985)*	-	-
ϕ_l	0.914	*(0.736, 0.999)*	-	-	0.919	*(0.750, 0.990)*
Panel B: Macro						
$\beta_{GDP,1}$	-0.431	*(-0.613, -0.363)*	-0.104	*(-0.108, -0.075)*	-0.432	*(-0.659, -0.204)*
$\beta_{IP,1}$	-0.354	*(-0.518, -0.295)*	0.000	*(-0.018, -0.160)*	-0.269	*(-0.494, -0.003)*
$\beta_{UR,1}$	0.696	*(-0.649, 0.791)*	-0.487	*(-0.577, -0.473)*	0.288	*(0.030, 0.562)*
$\beta_{RVR,1}$	1.048	*(0.829, 1.150)*	-1.137	*(-1.169, -1.126)*	0.086	*(-0.262, 0.444)*
$\beta_{HVR,1}$	0.824	*(0.772, 0.975)*	-0.630	*(-0.643, -0.597)*	0.354	*(0.009, 0.809)*
$\beta_{DI,1}$	0.017	*(-0.208, 0.124)*	-0.222	*(-0.244, 0.028)*	-0.435	*(-0.672, -0.189)*
$\beta_{FRM,1}$	0.149	*(0.040, 0.371)*	-0.666	*(0.569, 0.676)*	-0.663	*(-0.848, -0.471)*
$\beta_{GDP,2}$	-	-	-	-	-	-
$\beta_{IP,2}$	-0.037	*(-0.128, -0.012)*	-0.018	*(-0.135, -0.006)*	-0.051	*(-0.150, 0.000)*
$\beta_{UR,2}$	-0.188	*(-0.238, -0.015)*	0.014	*(-0.049, 0.032)*	-0.231	*(-0.543, 0.083)*
$\beta_{RVR,2}$	-0.511	*(-0.607, -0.277)*	0.051	*(0.035, 0.071)*	-0.698	*(-1.082, -0.285)*
$\beta_{HVR,2}$	-0.200	*(-0.441, -0.142)*	0.671	*(0.669,0.690)*	-0.732	*(-1.039,-0.445)*
$\beta_{DI,2}$	-0.458	*(-0.563, -0.257)*	-2.415	*(-0.333, -0.229)*	-0.074	*(-0.388, 0.205)*
$\beta_{FRM,2}$	-0.713	*(-0.872, -0.641)*	-0.725	*(-0.735, -0.592)*	-0.085	*(-0.442, 0.263)*
WAIC2	94.104		62.554		77.089	
Panel C: Defaults						
α^d	-5.177	*(-5.638, -4.825)*	-5.832	*(-6.099, -5.580)*	-6.071	*(-6.179, -5.954)*
β^d_1	0.369	*(0.291, 0.445)*	-0.235	*(-0.293, -0.176)*	0.551	*(0.483, 0.642)*
β^d_2	0.243	*(0.139, 0.358)*	-0.015	*(-0.110, 0.090)*	0.014	*(0.108, 0.178)*
γ^d	0.786	*(0.614, 0.937)*	0.518	*(0.417, 0.593)*	-	-
av. $p^d(x10^-2)$	0.249	*(0.248, 0.250)*	0.249	*(0.246, 0.252)*	0.249	*(0.246, 0.252)*
m.e. of f^m_1	0.092		-0.058		0.137	
m.e. of f^m_2	0.060		-0.004		0.004	
m.e. of f^d	0.195		0.129		-	
WAIC2	24.778		26.205		26.598	
Panel D: Loss given default						
μ_0	0.001	*(0.001, 0.001)*	0.003	*(0.002, 0.004)*	0.001	*(0.001, 0.001)*
av. μ_1	0.464	*(0.408, 0.541)*	0.497	*(0.253, 0.650)*	0.497	*(0.251, 0.637)*
α_{μ_1}	0.433	*(0.300, 0.568)*	0.463	*(0.396, 0.535)*	0.454	*(0.373, 0.531)*
$\beta_{\mu_1 1}$	0.032	*(0.000, 0.064)*	0.015	*(-0.007, 0.037)*	0.020	*(-0.010, 0.051)*
$\beta_{\mu_1 2}$	0.003	*(-0.028, 0.034)*	0.095	*(0.060, 0.130)*	-0.027	*(-0.057, 0.002)*
γ_{μ_1}	-0.002	*(-0.053, 0.050)*	0.025	*(-0.006, 0.056)*	-	-
δ_{μ_1}	0.075	*(0.056, 0.097)*	-	-	0.073	*(0.054, 0.094)*
$\sigma_0(x10^-2)$	0.046	*(0.043, 0.048)*	0.046	*(0.043, 0.048)*	0.046	*(0.043, 0.048)*
σ_1	0.119	*(0.118, 0.120)*	0.119	*(0.118, 0.120)*	0.119	*(0.118, 0.120)*
α^l	1.266	*(0.404, 2.133)*	1.831	*(1.341, 2.302)*	1.367	*(0.696, 2.002)*
β^l_1	-0.100	*(-0.308, 0.110)*	0.474	*(0.337, 0.606)*	0.080	*(-0.092, 0.258)*
β^l_2	0.191	*(-0.024, 0.407)*	0.476	*(0.280, 0.672)*	0.079	*(-0.126, 0.268)*
γ^l	0.091	*(-0.199, 0.363)*	0.345	*(0.168, 0.530)*	-	-
δ^l	0.570	*(0.432, 0.716)*	-	-	0.586	*(0.454, 0.725)*
av. p^l	0.774	*(0.619, 0.911)*	0.780	*(0.482, 0.960)*	0.780	*(0.470, 0.958)*
m.e. of f^m_1	-1.742		8.131		1.371	
m.e. of f^m_2	3.347		8.175		1.360	
m.e. of f^d	1.592		5.916		-	
m.e. of f^l	9.973		-		10.084	
WAIC2	20051.430		19249.190		18833.450	

Table 5: This table presents the estimation results accompanying figure 11, displaying three model specifications with either three unique factors or a specification with no default or loss factor.

Northeast

	2 Macro, default, loss		2 Macro, default		2 Macro, loss	
Panel A: Factor						
ϕ_{m_1}	0.850	*(0.448, 0.989)*	0.875	*(0.607, 0.989)*	0.885	*(0.629, 0.988)*
ϕ_{m_2}	0.800	*(0.220, 0.988)*	0.854	*(0.492, 0.987)*	0.894	*(0.606, 0.990)*
ϕ_d	0.922	*(0.733, 0.989)*	0.916	*(0.747, 0.985)*	-	-
ϕ_l	0.871	*(0.661, 0.992)*	-	-	0.862	*(0.661, 0.990)*
Panel B: Macro						
$\beta_{GDP,1}$	-0.717	*(-0.903, -0.401)*	-0.351	*(-0.706, -0.103)*	-0.430	*(-0.687, -0.226)*
$\beta_{IP,1}$	-0.749	*(-0.924, -0.532)*	-0.131	*(-0.501, 0.119)*	-0.262	*(-0.469, 0.014)*
$\beta_{UR,1}$	1.071	*(-0.314, 1.295)*	-0.114	*(-0.675, 0.157)*	0.522	*(0.162, 0.718)*
$\beta_{RVR,1}$	0.772	*(0.424, 1.213)*	-0.306	*(-0.893, 0.095)*	0.523	*(-0.017, 0.833)*
$\beta_{HVR,1}$	0.479	*(0.233, 0.819)*	0.401	*(0.160, 0.602)*	0.686	*(0.376, 0.867)*
$\beta_{DI,1}$	-0.436	*(-0.808, -0.052)*	-0.395	*(-0.766, -0.032)*	-0.414	*(-0.703, -0.188)*
$\beta_{FRM,1}$	-0.343	*(-0.714, -0.103)*	-0.517	*(-0.893, -0.276)*	-0.565	*(-0.734, -0.387)*
$\beta_{GDP,2}$	-	-	-	-	-	-
$\beta_{IP,2}$	-1.415	*(-1.510, -1.353)*	-0.047	*(-0.014, 0.000)*	-0.035	*(-0.132, 0.000)*
$\beta_{UR,2}$	1.714	*(1.932, 1.412)*	-0.458	*(-0.749, -0.245)*	-0.445	*(-0.655, -0.092)*
$\beta_{RVR,2}$	0.347	*(0.149, 0.532)*	-0.889	*(-1.253, -0.613)*	-0.984	*(-1.582, -0.694)*
$\beta_{HVR,2}$	-1.764	*(-1.943, -1.423)*	-0.628	*(-0.941, -0.380)*	-0.645	*(-1.224,-0.302)*
$\beta_{DI,2}$	-0.379	*(-0.510, -0.039)*	0.011	*(-0.289, 0.315)*	-0.018	*(-0.334, 0.237)*
$\beta_{FRM,2}$	0.471	*(0.222, 0.654)*	0.434	*(0.107, 0.825)*	-0.007	*(-0.214, 0.273)*
WAIC2	57.440		54.153		58.697	
Panel C: Defaults						
α^d	-6.989	*(-7.231, -6.799)*	-6.695	*(-6.957, -6.505)*	-6.597	*(-6.791, -6.441)*
β_1^d	0.308	*(0.217, 0.379)*	-0.029	*(-0.156, 0.098)*	0.567	*(0.526, 0.642)*
β_2^d	0.131	*(-0.145, 0.329)*	0.019	*(-0.199, 0.182)*	0.575	*(0.094, 0.255)*
γ^d	0.548	*(0.501, 0.628)*	0.495	*(0.442, 0.553)*	-	-
av. $p^d(x10^-2)$	0.206	*(0.204, 0.207)*	0.206	*(0.204, 0.207)*	0.206	*(0.203, 0.204)*
m.e. of f_1^m	0.063		-0.006		0.118	
m.e. of f_2^m	0.027		0.004		0.039	
m.e. of f^d	0.112		0.102		-	
WAIC2	25.915		25.157		26.002	
Panel D: Loss given default						
μ_0	0.000	*(0.000, 0.001)*	0.001	*(0.001, 0.001)*	0.001	*(0.001, 0.001)*
av. μ_1	0.385	*(0.277, 0.471)*	0.398	*(0.197, 0.554)*	0.398	*(0.245, 0.602)*
α_{μ_1}	0.263	*(0.138, 0.392)*	0.379	*(0.311, 0.443)*	0.317	*(0.171, 0.452)*
$\beta_{\mu_1 1}$	-0.016	*(-0.057, 0.026)*	0.092	*(0.030, 0.161)*	0.011	*(-0.027, 0.051)*
$\beta_{\mu_1 2}$	-0.094	*(-0.191, 0.001)*	0.051	*(0.008, 0.096)*	0.034	*(-0.033, 0.095)*
γ_{μ_1}	0.029	*(-0.025, 0.077)*	0.007	*(-0.033, 0.046)*	-	-
δ_{μ_1}	0.066	*(0.027, 0.107)*	-	-	0.458	*(0.029, 0.118)*
$\sigma_0(x10^-2)$	0.001	*(0.001, 0.001)*	0.001	*(0.001, 0.001)*	0.001	*(0.025, 0.026)*
σ_1	0.105	*(0.103, 0.107)*	0.105	*(0.103, 0.107)*	0.105	*(0.104, 0.107)*
α^l	0.660	*(-0.100, 1.433)*	1.412	*(1.080, 1.786)*	0.716	*(-0.234, 1.577)*
β_1^l	-0.055	*(-0.304, 0.203)*	0.392	*(0.046, 0.731)*	0.160	*(-0.066, 0.397)*
β_2^l	-0.186	*(-0.703, 0.314)*	0.522	*(0.281, 0.812)*	0.398	*(0.044, 0.735)*
γ^l	0.377	*(0.061, 0.667)*	0.307	*(0.115, 0.491)*	-	-
δ^l	0.499	*(0.323, 0.689)*	-	-	0.458	*(0.260, 0.668)*
av. p^l	0.777	*(0.594, 0.944)*	0.777	*(0.595, 0.955)*	0.774	*(0.548, 0.954)*
m.e. of f_1^m	-0.960		6.791		2.802	
m.e. of f_2^m	-3.220		9.042		6.970	
m.e. of f^d	6.540		5.330		-	
m.e. of f^l	8.655		-		8.012	
WAIC2	48055.820		46532.010		42982.310	

Table 6: This table presents the estimation results accompanying figure 12, displaying three model specifications with either three unique factors or a specification with no default or loss factor.

South

	2 Macro, default, loss		2 Macro, default		2 Macro, loss	
Panel A: Factor						
ϕ_{m_1}	0.906	*(0.707, 0.986)*	0.920	*(0.789, 0.989)*	0.855	*(0.672, 0.989)*
ϕ_{m_2}	0.890	*(0.608, 0.990)*	0.813	*(0.310, 0.973)*	0.829	*(0.626, 0.978)*
ϕ_d	0.921	*(0.739, 0.978)*	0.865	*(0.504, 0.980)*	-	-
ϕ_l	0.922	*(0.752, 0.995)*	-	-	0.928	*(0.790, 0.990)*
Panel B: Macro						
$\beta_{GDP,1}$	-0.367	*(-0.422, -0.325)*	-0.263	*(-0.439, -0.071)*	-0.656	*(-0.981, -0.378)*
$\beta_{IP,1}$	-0.145	*(-0.259, -0.036)*	-0.113	*(-0.281, -0.081)*	-0.526	*(-0.922, -0.295)*
$\beta_{UR,1}$	0.357	*(0.240, 0.577)*	0.238	*(0.015, 0.464)*	0.669	*(0.310, 1.007)*
$\beta_{RVR,1}$	0.441	*(0.328, 0.568)*	0.364	*(0.102, 0.691)*	0.511	*(0.257, 0.0.902)*
$\beta_{HVR,1}$	0.833	*(0.713, 0.873)*	0.628	*(0.436, 0.917)*	0.679	*(0.349, 0.957)*
$\beta_{DI,1}$	-0.234	*(-0.370, -0.121)*	/0.232	*(-0.414, -0.026)*	-0.496	*(-0.771, -0.285)*
$\beta_{FRM,1}$	-0.263	*(-0.312, -0.181)*	-0.377	*(-0.591, -0.158)*	-0.625	*(-0.844, -0.440)*
$\beta_{GDP,2}$	-	-	-	-	-	-
$\beta_{IP,2}$	-0.133	*(-0.263, -0.011)*	-0.087	*(-0.257, 0.000)*	-0.364	*(-0.690, -0.039)*
$\beta_{UR,2}$	0.071	*(-0.323, 0.226)*	-0.405	*(-0.935, 0.154)*	-0.315	*(0.007, 0.854)*
$\beta_{RVR,2}$	-0.399	*(-0.669, -0.300)*	-1.081	*(-1.835, -0.334)*	0.436	*(0.108, 0.934)*
$\beta_{HVR,2}$	-0.931	*(-0.995, -0.641)*	-0.892	*(-1.381, -0.468)*	0.224	*(-0.195,0.603)*
$\beta_{DI,2}$	-0.336	*(-0.479, -0.185)*	-0.076	*(-0.505, 0.313)*	0.138	*(-0.190, 0.440)*
$\beta_{FRM,2}$	-0.875	*(-0.925, -0.819)*	0.001	*(-0.471, 0.620)*	0.821	*(0.560, 1.071)*
WAIC2	86.808		54.153		58.697	
Panel C: Defaults						
α^d	-5.069	*(-5.523, -4.722)*	-6.114	*(-6.179, -6.038)*	-6.483	*(-6.553, -6.392)*
β_1^d	0.210	*(0.136, 0.264)*	0.294	*(0.245, 0.338)*	0.708	*(0.658, 0.760)*
β_2^d	0.157	*(0.014, 0.356)*	0.533	*(0.478, 0.607)*	-0.512	*(-0.687, -0.351)*
γ^d	0.911	*(0.742, 1.042)*	0.406	*(0.349, 0.447)*	-	-
av. $p^d(x10^-2)$	0.241	*(0.240, 0.242)*	0.239	*(0.238, 0.241)*	0.240	*(0.239, 0.241)*
m.e. of f_1^m	0.050		0.070		0.170	
m.e. of f_2^m	0.038		0.128		-0.123	
m.e. of f^d	0.219		0.098		-	
WAIC2	25.614		25.157		26.001	
Panel D: Loss given default						
μ_0	0.002	*(0.001, 0.002)*	0.001	*(0.001, 0.001)*	0.001	*(0.001, 0.001)*
av. μ_1	0.365	*(0.315, 0.430)*	0.378	*(0.180, 0.534)*	0.378	*(0.186, 0.538)*
α_{μ_1}	0.283	*(0.172, 0.388)*	0.373	*(0.324, 0.423)*	0.0.329	*(0.260, 0.405)*
$\beta_{\mu_1 1}$	0.016	*(-0.002, 0.035)*	0.041	*(0.017, 0.064)*	0.020	*(-0.001, 0.040)*
$\beta_{\mu_1 2}$	0.033	*(0.004, 0.065)*	0.031	*(-0.006, 0.066)*	-0.017	*(-0.042, -0.008)*
γ_{μ_1}	-0.017	*(-0.056, 0.023)*	-0.040	*(-0.062, -0.018)*	-	-
δ_{μ_1}	0.059	*(0.046, 0.076)*	-	-	0.0.061	*(0.051, 0.086)*
$\sigma_0(x10^-2)$	0.003	*(0.003, 0.003)*	0.003	*(0.003, 0.003)*	0.003	*(0.003, 0.003)*
σ_1	0.097	*(0.098, 0.099)*	0.096	*(0.095, 0.097)*	0.096	*(0.096, 0.096)*
α^l	0.758	*(0.234, 1.423)*	1.086	*(0.680, 1.548)*	0.780	*(0.181, 1.432)*
β_1^l	0.000	*(-0.135, 0.154)*	0.257	*(0.048, 0.463)*	-0.024	*(/0.194, 0.136)*
β_2^l	0.157	*(-0.059, 0.366)*	0.255	*(-0.018, 0.557)*	-0.064	*(-0.229, 0.106)*
γ^l	-0.052	*(-0.265, 0.161)*	-0.436	*(-0.619, -0.265)*	-	-
δ^l	0.532	*(0.426, 0.660)*	-	-	0.542	*(0.374, 0.598)*
av. p^l	0.735	*(0.581, 0.839)*	0.723	*(0.416, 0.929)*	0.723	*(0.412, 0.929)*
m.e. of f_1^m	-0.013		5.154		-0.474	
m.e. of f_2^m	3.060		5.098		-1.285	
m.e. of f^d	-1.007		-8.733		-	
m.e. of f^l	10.363		-		10.851	
WAIC2	30051.430		27053.280		26098.370	

Table 7: This table presents the estimation results accompanying figure 13, displaying three model specifications with either three unique factors or a specification with no default or loss factor.

West

	2 Macro, default, loss		2 Macro, default		2 Macro, loss	
Panel A: Factor						
ϕ_{m_1}	0.871	*(0.555, 0.989)*	0.928	*(0.800, 0.990)*	0.898	*(0.672, 0.989)*
ϕ_{m_2}	0.840	*(0.462, 0.989)*	0.906	*(0.694, 0.990)*	0.865	*(0.626, 0.978)*
ϕ_d	0.907	*(0.731, 0.989)*	0.952	*(0.852, 0.992)*	-	-
ϕ_l	0.234	*(-0.550, 0.992)*	-	-	0.913	*(0.790, 0.990)*
Panel B: Macro						
$\beta_{GDP,1}$	-0.494	*(-0.757, -0.304)*	-0.151	*(-0.453, 0.123)*	-0.338	*(-0.559, -0.097)*
$\beta_{IP,1}$	-0.362	*(-0.636, -0.140)*	-0.144	*(-0.232, 0.051)*	-0.222	*(-0.410, -0.017)*
$\beta_{UR,1}$	0.523	*(0.240, 0.898)*	0.655	*(0.111, 1.012)*	0.383	*(0.166, 0.530)*
$\beta_{RVR,1}$	0.565	*(0.160, 0.999)*	1.109	*(0.721, 1.431)*	0.324	*(0.069, 0.536)*
$\beta_{HVR,1}$	0.738	*(0.505, 1.001)*	0.353	*(0.102, 0.502)*	0.413	*(0.290, 0.695)*
$\beta_{DI,1}$	-0.333	*(-0.592, -0.072)*	0.036	*(-0.012, 0.231)*	-0.303	*(-0.584, -0.150)*
$\beta_{FRM,1}$	-0.368	*(0.576, -0.206)*	-0.058	*(-0.121, 0.031)*	-0.508	*(-0.702, -0.397)*
$\beta_{GDP,2}$	-	-	-	-	-	-
$\beta_{IP,2}$	-0.065	*(-0.191, 0.000)*	-0.259	*(-0.549, 0.000)*	-0.045	*(-0.125, 0.000)*
$\beta_{UR,2}$	-0.060	*(-0.476, 0.404)*	0.647	*(0.321, 0.982)*	-0.287	*(-0.490, 0.016)*
$\beta_{RVR,2}$	-0.506	*(-1.105, 0.120)*	0.784	*(0.493, 1.021)*	-0.802	*(-1.512, -0.391)*
$\beta_{HVR,2}$	-0.651	*(-1.195, -0.313)*	-0.484	*(-0.632, -0.184)*	-0.705	*(-1.158,-0.381)*
$\beta_{DI,2}$	-0.150	*(-0.450, 0.237)*	-0.222	*(-0.385, -0.014)*	-0.029	*(-0.354, 0.130)*
$\beta_{FRM,2}$	-0.405	*(-0.644, -0.169)*	-0.598	*(-0.723, -0.482)*	-0.012	*(-0.374, 0.039)*
WAIC2	57.440		54.153		58.697	
Panel C: Defaults						
α^d	-6.749	*(-6.818, -6.668)*	-4.782	*(-5.061, -4.490)*	-6.663	*(-6.813, -6.352)*
β_1^d	0.514	*(0.472, 0.544)*	-0.304	*(-0.392, -0.220)*	0.598	*(0.544, 0.596)*
β_2^d	0.351	*(0.291, 0.393)*	0.546	*(0.453, 0.614)*	0.049	*(0.098, 0.228)*
γ^d	0.068	*(0.028, 0.102)*	0.708	*(0.644, 0.773)*	-	-
av. $p^d(x10^-2)$	0.189	*(0.188, 0.190)*	0.189	*(0.188, 0.190)*	0.189	*(0.188, 0.190)*
m.e. of f_1^m	0.097		0.041		0.113	
m.e. of f_2^m	0.066		0.137		0.009	
m.e. of f^d	0.013		0.059		-	
WAIC2	25.915		25.157		25.193	
Panel D: Loss given default						
μ_0	-0.001	*(-0.001, 0.000)*	0.001	*(0.001, 0.001)*	0.001	*(0.001, 0.001)*
av. μ_1	0.297	*(0.138, 0.446)*	0.298	*(0.116, 0.460)*	0.298	*(0.116, 0.460)*
α_{μ_1}	0.279	*(0.194, 0.337)*	0.152	*(-0.019, 0.305)*	0.250	*(0.226, 0.440)*
$\beta_{\mu_1 1}$	0.038	*(0.015, 0.059)*	-0.026	*(-0.054, 0.005)*	0.035	*(-0.011, 0.025)*
$\beta_{\mu_1 2}$	-0.004	*(0.036, 0.027)*	0.069	*(0.047, 0.092)*	-0.018	*(-0.055, -0.004)*
γ_{μ_1}	0.053	*(-0.004, 0.086)*	-0.049	*(-0.095, -0.004)*	-	-
δ_{μ_1}	0.007	*(-0.018, 0.055)*	-	-	0.060	*(0.051, 0.086)*
$\sigma_0(x10^-2)$	0.016	*(0.015, 0.017)*	0.016	*(0.015, 0.017)*	0.016	*(0.015, 0.016)*
σ_1	0.063	*(0.061, 0.064)*	0.063	*(0.061, 0.064)*	0.063	*(0.062, 0.064)*
α^l	0.945	*(0.227, 1.533)*	1.240	*(0.327, 2.194)*	0.879	*(0.477, 1.653)*
β_1^l	-0.013	*(-0.193, 0.161)*	-0.605	*(-0.785, -0.430)*	0.095	*(-0.085, 0.141)*
β_2^l	0.128	*(-0.136, 0.384)*	0.236	*(0.077, 0.400)*	0.139	*(-0.051, 0.237)*
γ^l	0.528	*(0.132, 0.753)*	-0.256	*(-0.515, 0.012)*	-	-
δ^l	0.077	*(0.000, 0.461)*	-	-	0.543	*(0.374, 0.598)*
av. p^l	0.723	*(0.470, 0.937)*	0.726	*(0.482, 0.932)*	0.723	*(0.463, 0.939)*
m.e. of f_1^m	-0.260		-0.121		1.902	
m.e. of f_2^m	2.566		4.723		2.784	
m.e. of f^d	10.573		-5.127		-	
m.e. of f^l	1.551		-		10.876	
WAIC2	92414.030		75659.760		85333.810	

Table 8: This table presents the estimation results accompanying figure 14, displaying three model specifications with either three unique factors or a specification with no default or loss factor.

LTV < 60%

	2 Macro, default, loss		2 Macro, default		2 Macro, loss	
Panel A: Factor						
ϕ_{m_1}	0.828	*(0.385, 0.989)*	0.902	*(0.680, 0.990)*	0.930	*(0.777, 0.990)*
ϕ_{m_2}	0.895	*(0.640, 0.989)*	0.843	*(0.388, 0.990)*	0.722	*(-0.056, 0.990)*
ϕ_d	0.897	*(0.642, 0.989)*	0.827	*(0.460, 0.991)*	-	-
ϕ_l	0.677	*(0.144, 0.990)*	-	-	0.520	*(-0.230, 0.990)*
Panel B: Macro						
$\beta_{GDP,1}$	-0.790	*(-1.191, -0.332)*	-0.378	*(-0.532, -0.032)*	-0.073	*(-0.178, -0.004)*
$\beta_{IP,1}$	-0.717	*(-1.198, -0.255)*	-0.298	*(-0.432, 0.051)*	-0.211	*(-0.316, -0.136)*
$\beta_{UR,1}$	0.841	*(0.379, 1.204)*	0.603	*(0.234, 1.012)*	0.319	*(0.231, 0.385)*
$\beta_{RVR,1}$	0.545	*(-0.159, 1.074)*	0.456	*(0.231, 0.713)*	0.443	*(0.328, 0.543)*
$\beta_{HVR,1}$	0.662	*(0.248, 0.956)*	0.524	*(0.321, 0.829)*	0.037	*(-0.064, 0.167)*
$\beta_{DI,1}$	-0.439	*(-0.787, -0.072)*	-0.249	*(-0.423, -0.023)*	-0.013	*(-0.103, 0.126)*
$\beta_{FRM,1}$	-0.003	*(-0.724, 0.562)*	-0.605	*(-0.819, -0.319)*	0.356	*(0.276, 0.515)*
$\beta_{GDP,2}$	-	-	-	-	-	-
$\beta_{IP,2}$	-0.180	*(-0.344, 0.000)*	-0.015	*(-0.040, 0.000)*	-1.383	*(-4.703, -0.726)*
$\beta_{UR,2}$	-0.072	*(-0.536, 0.404)*	-0.696	*(-0.910, -0.423)*	2.002	*(1.238, 5.326)*
$\beta_{RVR,2}$	-0.033	*(-1.054, 1.207)*	-1.088	*(-1.189, -0.819)*	2.066	*(1.174, 4.562)*
$\beta_{HVR,2}$	-0.271	*(-0.879, 0.210)*	-0.408	*(-0.449, -0.342)*	1.654	*(1.123, 3.883)*
$\beta_{DI,2}$	0.396	*(-0.208, 2.965)*	-0.057	*(-0.108, -0.011)*	-1.021	*(-3.199, -0.577)*
$\beta_{FRM,2}$	0.610	*(0.168, 1.109)*	-0.118	*(-0.310, 0.051)*	-0.265	*(-0.644, 1.134)*
WAIC2	56.440		55.269		53.976	
Panel C: Defaults						
α^d	-7.561	*(-9.303, -6.784)*	-7.603	*(-7.936, -7.393)*	-7.976	*(-8.084, -7.862)*
β_1^d	0.269	*(0.158, 0.393)*	0.541	*(0.446, 0.614)*	-0.442	*(-0.484, -0.374)*
β_2^d	-0.134	*(-0.258, -0.012)*	-0.072	*(-0.202, 0.130)*	0.039	*(0.222, 0.595)*
γ^d	0.582	*(0.406, 0.702)*	0.577	*(0.365, 0.809)*	-	-
av. $p^d(x10^{-}2)$	0.094	*(0.093, 0.095)*	0.094	*(0.091, 0.096)*	0.094	*(0.093, 0.094)*
m.e. of f_1^m	0.025		0.051		-0.041	
m.e. of f_2^m	-0.013		-0.007		0.037	
m.e. of f^d	0.055		0.054		-	
WAIC2	27.010		25.999		34.149	
Panel D: Loss given default						
μ_0	-0.002	*(-0.002, 0.000)*	0.001	*(0.001, 0.001)*	0.001	*(0.001, 0.001)*
av. μ_1	0.407	*(0.335, 0.501)*	0.409	*(0.324, 0.502)*	0.410	*(0.333, 0.494)*
α_{μ_1}	0.458	*(-0.003, 1.166)*	0.361	*(0.289, 0.432)*	0.369	*(0.318, 0.428)*
$\beta_{\mu_1 1}$	0.016	*(-0.040, 0.070)*	0.039	*(0.012, 0.067)*	-0.031	*(-0.058, -0.003)*
$\beta_{\mu_1 2}$	-0.028	*(-0.100, 0.045)*	0.022	*(-0.032, 0.078)*	1.899	*(-0.085, 0.117)*
γ_{μ_1}	0.032	*(-0.058, 0.081)*	0.006	*(-0.056, 0.066)*	-	-
δ_{μ_1}	0.007	*(-0.056, 0.072)*	-	-	-0.006	*(-0.062, 0.049)*
$\sigma_0(x10^{-}2)$	0.004	*(0.004, 0.004)*	0.004	*(0.004, 0.004)*	0.004	*(0.004, 0.004)*
σ_1	0.128	*(0.124, 0.131)*	0.128	*(0.125, 0.131)*	0.128	*(0.125, 0.135)*
α^l	0.438	*(-1.420, 1.510)*	0.482	*(0.124, 0.757)*	0.285	*(0.064, 0.509)*
β_1^l	-0.016	*(-0.264, 0.230)*	0.312	*(0.198, 0.412)*	-0.574	*(-0.687, -0.444)*
β_2^l	-0.028	*(-0.299, 0.321)*	0.262	*(0.108, 0.413)*	-0.482	*(-1.563, 0.005)*
γ^l	0.600	*(0.332, 0.861)*	0.655	*(0.410, 0.925)*	-	-
δ^l	0.203	*(0.000, 0.440)*	-	-	0.121	*(0.000, 0.268)*
av. p^l	0.653	*(0.442, 0.866)*	0.651	*(0.404, 0.888)*	0.651	*(0.419, 0.888)*
m.e. of f_1^m	-0.367		7.086		-13.038	
m.e. of f_2^m	-1.059		5.965		-10.947	
m.e. of f^d	13.601		1.488		-	
m.e. of f^l	4.601		-		2.738	
WAIC2	61229.130		64496.350		58199.82	

Table 9: This table presents the estimation results accompanying figure 14, displaying three model specifications with either three unique factors or a specification with no default or loss factor.

60% < LTV ≤ 80%

	2 Macro, default, loss		2 Macro, default		2 Macro, loss	
Panel A: Factor						
ϕ_{m_1}	0.846	*(0.433, 0.989)*	0.831	*(0.396, 0.988)*	0.852	*(0.467, 0.988)*
ϕ_{m_2}	0.966	*(0.911, 0.989)*	0.884	*(0.665, 0.989)*	0.963	*(0.900, 0.978)*
ϕ_d	0.932	*(0.811, 0.989)*	0.909	*(0.717, 0.988)*	-	-
ϕ_l	0.905	*(0.696, 0.972)*	-	-	0.895	*(0.690, 0.990)*
Panel B: Macro						
$\beta_{GDP,1}$	-0.886	*(-1.646, -0.062)*	-0.752	*(-0.959, -0.466)*	-0.187	*(-0.186, -0.080)*
$\beta_{IP,1}$	-1.116	*(-1.732, -0.461)*	-0.667	*(-0.947, -0.402)*	-0.584	*(-0.596, -0.464)*
$\beta_{UR,1}$	1.097	*(0.754, 1.441)*	0.812	*(0.596, 1.050)*	0.638	*(0.639, 0.670)*
$\beta_{RVR,1}$	0.626	*(0.108, 1.253)*	0.740	*(0.458, 0.961)*	0.749	*(0.751, 0.760)*
$\beta_{HVR,1}$	0.323	*(-0.068, 0.658)*	0.835	*(0.626, 1.074)*	0.350	*(0.346, 0.374)*
$\beta_{DI,1}$	-0.642	*(-1.107, 0.073)*	-0.439	*(-0.694, -0.206)*	-0.044	*(-0.036, -0.003)*
$\beta_{FRM,1}$	-0.016	*(-0.481, 0.426)*	-0.284	*(-0.522, -0.050)*	0.129	*(0.110, 0.127)*
$\beta_{GDP,2}$	-	-	-	-	-	-
$\beta_{IP,2}$	-0.610	*(-1.344, -0.015)*	-0.217	*(-0.374, -0.029)*	-0.559	*(-0.567, -0.482)*
$\beta_{UR,2}$	0.381	*(0.041, 0.637)*	0.205	*(0.036, 0.379)*	0.469	*(0.463, 0.478)*
$\beta_{RVR,2}$	-0.162	*(-1.226, 1.302)*	0.294	*(0.134, 0.519)*	0.289	*(0.286, 0.311)*
$\beta_{HVR,2}$	-0.719	*(-1.072, -0.224)*	-0.015	*(-0.209, 0.163)*	0.307	*(0.301, 0.384)*
$\beta_{DI,2}$	-0.268	*(-1.110, 0.996)*	0.167	*(-0.006, 0.406)*	-0.457	*(-0.481, -0.455)*
$\beta_{FRM,2}$	-0.691	*(0.303, 0.963)*	0.635	*(0.454, 0.794)*	-0.733	*(-0.902, -0.741)*
WAIC2	53.440		75.191		54.519	
Panel C: Defaults						
α^d	-4.777	*(-6.764, -3.527)*	-6.458	*(-6.479, -6.436)*	-9.218	*(-9.428, -8.990)*
β_1^d	0.304	*(0.225, 0.406)*	0.560	*(0.490, 0.602)*	-0.058	*(-0.102, -0.003)*
β_2^d	0.138	*(-0.200, 0.383)*	-0.500	*(-0.571, -0.453)*	0.844	*(0.780, 0.890)*
γ^d	0.673	*(0.500, 0.824)*	0.050	*(0.004, 0.104)*	-	-
av. $p^d(x10^-2)$	0.218	*(0.217, 0.220)*	0.218	*(0.216, 0.220)*	0.218	*(0.217, 0.220)*
m.e. of f_1^m	0.066		0.122		-0.013	
m.e. of f_2^m	0.030		-0.109		0.184	
m.e. of f^d	0.147		0.011		-	
WAIC2	22.441		25.573		26.489	
Panel D: Loss given default						
μ_0	-0.001	*(-0.001, 0.000)*	-0.001	*(-0.001, 0.001)*	-0.001	*(-0.001, 0.000)*
av. μ_1	0.444	*(0.370, 0.512)*	0.470	*(0.266, 0.606)*	0.469	*(0.253, 0.603)*
α_{μ_1}	0.087	*(-0.225, 0.356)*	0.447	*(0.370, 0.517)*	0.394	*(0.262, 0.533)*
$\beta_{\mu_1 1}$	-0.003	*(-0.042, 0.035)*	0.016	*(-0.015, 0.046)*	0.009	*(-0.022, 0.041)*
$\beta_{\mu_1 2}$	-0.085	*(-0.142, -0.028)*	0.005	*(-0.028, 0.037)*	0.025	*(-0.011, 0.060)*
γ_{μ_1}	-0.028	*(-0.071, 0.011)*	0.073	*(0.050, 0.097)*	-	-
δ_{μ_1}	0.071	*(0.036, 0.108)*	-	-	0.074	*(0.049, 0.101)*
$\sigma_0(x10^-2)$	0.011	*(0.011, 0.011)*	0.011	*(0.011, 0.011)*	0.011	*(0.010, 0.011)*
σ_1	0.099	*(0.099, 0.100)*	0.099	*(0.099, 0.099)*	0.099	*(0.099, 0.100)*
α^l	1.634	*(0.525, 2.800)*	1.740	*(1.241, 2.164)*	1.019	*(0.294, 1.703)*
β_1^l	-0.013	*(-0.181, 0.173)*	-0.002	*(-0.164, 0.181)*	-0.169	*(-0.313, -0.007)*
β_2^l	-0.068	*(-0.385, 0.243)*	-0.162	*(-0.347, 0.007)*	0.244	*(0.066, 0.405)*
γ^l	0.144	*(-0.035, 0.320)*	0.472	*(0.359, 0.594)*	-	-
δ^l	0.490	*(0.361, 0.640)*	-	-	0.436	*(0.331, 0.548)*
av. p^l	0.827	*(0.774, 0.872)*	0.825	*(0.591, 0.954)*	0.825	*(0.591, 0.954)*
m.e. of f_1^m	-0.184		-0.023		-2.437	
m.e. of f_2^m	-0.974		-2.333		2.526	
m.e. of f^d	2.056		6.821		-	
m.e. of f^l	6.991		-		6.299	
WAIC2	42983.04		37200.870		36773.470	

Table 10: This table presents the estimation results accompanying figure 14, displaying three model specifications with either three unique factors or a specification with no default or loss factor.

LTV > 80%

	2 Macro, default, loss		2 Macro, default		2 Macro, loss	
Panel A: Factor						
ϕ_{m_1}	0.917	*(0.512, 0.989)*	0.806	*(0.432, 0.988)*	0.869	*(0.441, 0.988)*
ϕ_{m_2}	0.809	*(0.321, 0.989)*	0.859	*(0.411, 0.989)*	0.833	*(0.637, 0.978)*
ϕ_d	0.924	*(0.532, 0.989)*	0.919	*(0.417, 0.988)*	-	-
ϕ_l	0.155	*(0.691, 0.972)*	-	-	0.927	*(0.632, 0.990)*
Panel B: Macro						
$\beta_{GDP,1}$	-0.060	*(-0.092, -0.002)*	-0.428	*(-0.859, -0.366)*	-0.271	*(-0.406, -0.080)*
$\beta_{IP,1}$	0.222	*(0.010, 0.321)*	-0.535	*(-0.537, -0.402)*	-0.405	*(-0.532, -0.344)*
$\beta_{UR,1}$	-0.297	*(-0.512, -0.050)*	0.937	*(0.575, 1.442)*	0.486	*(0.139, 0.772)*
$\beta_{RVR,1}$	-0.264	*(-0.431, -0.103)*	1.056	*(0.421, 1.232)*	0.445	*(0.251, 0.750)*
$\beta_{HVR,1}$	0.132	*(-0.045, 0.203)*	0.171	*(0.050, 0.522)*	0.330	*(0.300, 0.374)*
$\beta_{DI,1}$	-0.308	*(-0.509, 0.100)*	-0.169	*(-0.294, -0.106)*	0.049	*(-0.036, 0.073)*
$\beta_{FRM,1}$	-0.715	*(-0.952, -0.513)*	-0.137	*(-0.122, -0.050)*	0.598	*(0.810, 0.127)*
$\beta_{GDP,2}$	-	-	-	-	-	-
$\beta_{IP,2}$	-0.718	*(-1.310, -0.312)*	-0.325	*(-0.374, -0.029)*	-0.507	*(-0.711, -0.212)*
$\beta_{UR,2}$	1.439	*(0.050, 2.039)*	0.351	*(0.036, 0.379)*	0.669	*(0.163, 0.878)*
$\beta_{RVR,2}$	1.389	*(0.596, 2.031)*	0.280	*(0.134, 0.519)*	0.597	*(0.486, 0.811)*
$\beta_{HVR,2}$	0.698	*(0.312, 0.941)*	0.788	*(-0.209, 0.163)*	0.565	*(0.201, 0.934)*
$\beta_{DI,2}$	-0.131	*(-0.812, 0.912)*	-0.331	*(-0.006, 0.406)*	-0.393	*(-0.421, -0.365)*
$\beta_{FRM,2}$	0.698	*(0.343, 0.930)*	-0.760	*(0.454, 0.794)*	-0.577	*(-0.802, -0.311)*
WAIC2	53.440		65.317		59.415	
Panel C: Defaults						
α^d	-6.435	*(-7.434, -5.567)*	-5.002	*(-6.479, -6.436)*	-5.867	*(-6.428, -4.910)*
β_1^d	0.284	*(0.059, 0.432)*	-0.730	*(0.490, 0.602)*	-0.159	*(-0.232, -0.003)*
β_2^d	0.575	*(0.333, 0.981)*	0.331	*(-0.571, -0.453)*	0.641	*(0.490, 0.670)*
γ^d	0.027	*(-0.005, 0.050)*	0.272	*(0.004, 0.104)*	-	-
av. $p^d(x10^-2)$	0.392	*(0.390, 0.393)*	0.395	*(0.393, 0.396)*	0.392	*(0.381, 0.401)*
m.e. of f_1^m	0.111		0.285		-0.062	
m.e. of f_2^m	0.225		0.130		0.250	
m.e. of f^d	0.110		0.106		-	
WAIC2	22.441		25.783		26.023	
Panel D: Loss given default						
μ_0	0.004	*(0.004, 0.004)*	0.005	*(0.004, 0.005)*	0.005	*(0.005, 05000)*
av. μ_1	0.314	*(0.214, 0.419)*	0.313	*(0.210, 0.503)*	0.314	*(0.233, 0.423)*
α_{μ_1}	0.264	*(0.015, 0.432)*	0.327	*(0.370, 0.517)*	0.265	*(0.223, 0.343)*
$\beta_{\mu_1 1}$	0.006	*(-0.001, 0.010)*	-0.030	*(-0.015, 0.046)*	0.007	*(0.001, 0.011)*
$\beta_{\mu_1 2}$	0.001	*(-0.010, 0.028)*	0.043	*(-0.028, 0.037)*	0.004	*(-0.004, 0.008)*
γ_{μ_1}	0.057	*(0.031, 0.081)*	0.031	*(0.050, 0.097)*	-	-
δ_{μ_1}	0.002	*(-0.025, 0.025)*	-	-	0.061	*(0.039, 0.101)*
$\sigma_0(x10^-2)$	0.050	*(0.050, 0.050)*	0.050	*(0.048, 0.051)*	0.050	*(0.050, 0.052)*
σ_1	0.078	*(0.078, 0.078)*	0.078	*(0.078, 0.078)*	0.078	*(0.078, 0.078)*
α^l	0.566	*(0.405, 0.701)*	0.961	*(1.241, 2.164)*	0.539	*(0.312, 1.303)*
β_1^l	0.081	*(-0.191, 0.501)*	-0.337	*(-0.164, 0.181)*	-0.064	*(-0.104, -0.017)*
β_2^l	-0.233	*(-0.413, 0.111)*	0.219	*(-0.347, 0.007)*	-0.114	*(-0.232, 0.030)*
γ^l	0.568	*(0.301, 0.705)*	0.389	*(0.359, 0.594)*	-	-
δ^l	0.035	*(0.010, 0.056)*	-	-	0.539	*(0.231, 0.786)*
av. p^l	0.681	*(0.643, 0.793)*	0.681	*(0.501, 0.794)*	0.681	*(0.400, 0.845)*
m.e. of f_1^m	1.756		-7.325		-1.385	
m.e. of f_2^m	-5.062		4.765		-2.474	
m.e. of f^d	12.334		8.450		-	
m.e. of f^l	0.754		-		12.804	
WAIC2	46813.45		43631.880		42460.710	

Table 11: This table presents the estimation results accompanying figure 14, displaying three model specifications with either three unique factors or a specification with no default or loss factor.

C Estimation

Priors

The prior for macro parameters is given below, with I_{N^m} an identity matrix of dimension N^m and $\boldsymbol{I}(\boldsymbol{A^m})$ the restriction on $\boldsymbol{A^m}$,

$$p(\boldsymbol{A^m}, \Sigma^{\boldsymbol{m}}) \propto iW(0.01\boldsymbol{I_{N^m}}, N^m)I(\boldsymbol{A^m}). \tag{11}$$

The prior for the loading on the factors for the loan status and loss given default components also imposes restrictions,

$$p(\alpha^l, \alpha^d) \propto I(\alpha^l, \alpha^d). \tag{12}$$

The priors for the parameters of the loss given default component refers to both the variance and the mean of the normal mixture components. The priors for the location parameters impose restrictions to prevent label switching,

$$p(\mu_0, \mu_{1t}) \propto I(\mu_0, \mu_{1t}) \qquad \forall t. \tag{13}$$

The prior for the scale parameters are uninformative,

$$p(\sigma_0^2) \propto iG2(0.01, 0.01), \tag{14}$$

$$p(\sigma_1^2) \propto iG2(0.01, 0.01). \tag{15}$$

The prior for the persistence of the latent factors is

$$p(\phi_{jj}) \propto I(\phi_{jj}) \qquad \forall j. \tag{16}$$

The indicator functions $I(.)$ impose the relevant identification restrictions. For the macro and housing prior this is the negative loading of the first variable on either the macro or housing factor. For the loan status this implies a positive load of probability of default on the second default factor. For the probability of a severe loss this implies a positive load on the third loan factor. Other imposed restrictions are $\mu_0 \leq \mu_{1t}$ for all t and $-1 < \phi jj < 1$ for all j.

Likelihoods

The likelihood of the model consists of three components: macro, default and loss given default. The macro component is first discussed. $\boldsymbol{Y}^m$ is a T x N^m matrix, where T is the amount of observations and N^m the amount of macroeconomic time series. $\boldsymbol{X}^m$ is a T x (K^m+1) matrix with a constant and K^m macro factor(s) $\boldsymbol{f}_t^m$. The macro likelihood is given by,

$$p(\boldsymbol{Y^m}|\boldsymbol{A^m},\Sigma^{\boldsymbol{m}},\boldsymbol{f_t^m}) \propto |\Sigma^{\boldsymbol{m}}|^{-T/2} exp\left(-\frac{1}{2}tr((\Sigma^{\boldsymbol{m}-1})(\boldsymbol{Y^m}-\boldsymbol{X^m}\boldsymbol{A^m})'(\boldsymbol{Y^m}-\boldsymbol{X^m}\boldsymbol{A^m}))\right). \tag{17}$$

Furthermore, the likelihood of the default component is given as below. Where $\boldsymbol{y^d}$ contains all loan indicators y_{it}^d (default or no default), $\psi^{\boldsymbol{d}}$ a vector with elements $\psi_t^d = \alpha^d + \beta^{d'}\boldsymbol{f_t^m} + \gamma^{d'}\boldsymbol{f_t^d}$. Besides, D_t is defined as the number of defaulted loans in period t and L_t as the number of active loans in period t.

$$\begin{aligned} p(\boldsymbol{y^d}|\psi^{\boldsymbol{d}}) &= \prod_{i,t}^{N,T}(p_t^d)^{y_{it}^d}(1-p_t^d)^{1-y_{it}^d} \\ &= \prod_t^T \Lambda(\psi_t^d)^{D_t}(1-\Lambda(\psi_t^d))^{L_t-D_t} \\ &= \prod_t^T \frac{exp(\psi_t^d)^{D_t}}{(1+exp(\psi_t^d))^{L_t}}. \end{aligned} \tag{18}$$

Lastly, the LGD component models the observed losses, with $\boldsymbol{y^l}$ a vector with the observed LGDs and $\boldsymbol{s}$ the vector containing s_{it}, latent indicators corresponding with whether or not a loan is good or bad. The likelihood is given as

$$p(\boldsymbol{y^l}|\boldsymbol{s},\mu_0,\mu_{1t},\sigma_0^2,\sigma_1^2) = \prod_{i,t}^{N,T}\left(s_{it}f_{\mathcal{N}}(y_{it};\mu_{1t},\sigma_1^2)+(1-s_{it})f_{\mathcal{N}}(y_{it};\mu_0,\sigma_0^2)\right). \tag{19}$$

Furthermore, the likelihood of the latent variable $\boldsymbol{s}$ is written as

$$\begin{aligned} p(s|\psi^{\boldsymbol{d}}) &= \prod_{i,t}^{N,T}(p_t^d)^{s_{it}}(1-p_t^d)^{1-s_{it}} \\ &= \prod_t^T \Lambda(\psi_t^d)^{N_t}(1-\Lambda(\psi_t^d))^{T_t-N_t} \\ &= \prod_t^T \frac{exp(\psi_t^d)^{N_t}}{(1+exp(\psi_t^d))^{T_t}}. \end{aligned} \tag{20}$$

Following Polson et al. (2013) and the description of Keijsers et al. (2017) the pseudo variables ω_t^d and ω_t^l are sampled to easen the simulation of the default and loss components

$$p(\omega_t^l|L_t,\psi_t^l) = PG(L_t,\psi_t^l), \tag{21}$$

$$p(\omega_t^d|L_t,\psi_t^d) = PG(T_t,\psi_t^d), \tag{22}$$

which implies that the following holds,

$$p(^l|\psi^l\prod_t^T p(\omega_t^l|L_t,\psi_t^l) \propto \prod_t^T exp\left(-\frac{\omega_t^d}{2}\left(\frac{\kappa_t^d}{\omega_t^d}-\psi_t^d\right)^2\right)exp\left(\frac{\omega_t^d}{2}\left(\frac{\kappa_t^d}{\omega_t^d}\right)^2\right)p(\omega_t^d), \tag{23}$$

this result is used to sample from a posterior that is Gaussian distributed, see Keijsers et al. (2017) for a detailed derivation. A linear model for $\boldsymbol{\mu_{1t}}$ is specified to capture the time-varying dynamics of the mean of the component corresponding with $s_{it} = 1$. In case a model specification assumes a fixed μ_{1t} over time, μ_{1t} is equal to $\mu_1 \ \forall \ t$. The likelihoods are given below. $\boldsymbol{X^{\mu_1}}$ is a matrix containing a constant and the latent factors used in the entire model specification and is in general the same as $\boldsymbol{X^l}$, in that it contains all latent factors. The likelihood for $\boldsymbol{\mu_{1t}}$ is as below,

$$p(\boldsymbol{\mu_{1t}}|\beta_{\mu_1}, \sigma^2_{\mu_1}) = \left(\frac{1}{2\pi\sigma^2_{\mu_1}}\right)^{\frac{n}{2}} exp\left(-\frac{1}{2\sigma^2_{\mu_1}}(\boldsymbol{\mu_1} - \boldsymbol{X}\beta_{\mu_1})'(\boldsymbol{\mu_1} - \boldsymbol{X}\beta_{\mu_1})\right) \tag{24}$$

where β_{μ_1} contains all relevant factor loadings of the linear model for μ_{1t}.

Posteriors

Macro component

$$p(\boldsymbol{A^m}, \Sigma^{\boldsymbol{m}}|...) \propto |\Sigma^{\boldsymbol{m}}|^{-(T+N^m+N^h+1)/2}$$
$$exp\left(-\frac{1}{2}tr((\Sigma^{\boldsymbol{m}})^{-1}[0.01\boldsymbol{I_{N^m}} + (\boldsymbol{Y^m} - \boldsymbol{X^m A^m})'(\boldsymbol{Y^m} - \boldsymbol{X^m A^m})])\right) I(\boldsymbol{A^m}), \quad (25)$$

standard results for multivariate regression models show that $\boldsymbol{A^m}$ and $\Sigma^{\boldsymbol{m}}$ can be drawn from the following two distributions,

$$p(\boldsymbol{A^m}|...) = MN((\boldsymbol{X^{m'} X^m})^{-1}(\boldsymbol{X^{m'} Y^m}), \Sigma^{\boldsymbol{m}} \otimes (\boldsymbol{X^{m'} X^m})^{-1}), \quad (26)$$

$$p(\Sigma^{\boldsymbol{m}}|...) = iW(0.01\boldsymbol{I_{N^m}} + (\boldsymbol{Y^m} - \boldsymbol{X^m A^m})'(\boldsymbol{Y^m} - \boldsymbol{X^m A^m}), T + N^m). \quad (27)$$

Default component

As ω_t^d does not influence any distributions besides its own, we can simply draw it from a Polya-Gamma distribution as below,

$$p(\omega_t^d|...) = PG(L_t, \psi_t^d). \quad (28)$$

The terms involving $\boldsymbol{\alpha^d}$ are collected from the likelihood, prior and latent variable distributions, such that $\boldsymbol{\alpha^d}$ is proportional as follows,

$$p(\boldsymbol{\alpha^d}|...) \propto \prod_t^T exp\left(-\frac{\omega_t^d}{2}\left(\frac{\kappa_t^d}{\omega_t^d} - \psi_t^d\right)^2\right) exp\left(\frac{w_t^d}{2}\left(\frac{\kappa_t^d}{\omega_t^d}\right)^2\right) p(\omega_t^d) I(\boldsymbol{\alpha^d}). \quad (29)$$

Following the results in Polson et al. (2013) and Keijsers et al. (2017) we can simulate α^l and likewise α^d as follows,

$$\boldsymbol{\alpha^d} \sim N(\boldsymbol{m^l}, \boldsymbol{V}^l), \quad (30)$$

where

$$\boldsymbol{V^d} = (\boldsymbol{X^l} \Omega^{\boldsymbol{l}} \boldsymbol{X^l})^{-1}, \quad (31)$$

$$\boldsymbol{m^d} = (\boldsymbol{X^l} \Omega^{\boldsymbol{l}} \boldsymbol{X^l})^{-1}(\boldsymbol{X^l \kappa^l}), \quad (32)$$

where $\boldsymbol{X^d}$ is a matrix with a constant in the first column and the relevant factors in the remaining columns, $\boldsymbol{\kappa^d}$ contains the elements κ_t^d and $\Omega^{\boldsymbol{d}}$ a diagonal matrix with ω_t^d being the diagonal elements.

Loss given default component

Similar to ω_t^d, ω_t^l also does not influence any distributions besides its own, we thus also draw it from a Polya-Gamma distribution,

$$p(\omega_t^l|...) = PG(T_t, \psi_t^l), \quad (33)$$

and proceed drawing $\boldsymbol{\alpha^l}$ similarly as for the default component.

Collecting terms involving s_{it} in the prior and likelihood the following is obtained,

$$p(s_{it}|...) \propto (p_t^l)^{s_{it}}(1-p_t^l)^{(1-s_{it})} exp\left(s_{it}\left(-\frac{y_{it}^l-\mu_0}{2\sigma_0}\right)^2\right) exp\left((1-s_{it})\left(-\frac{y_{it}^l-\mu_{1t}}{2\sigma_1}\right)^2\right). \tag{34}$$

therefore s_{it} can be sampled from the Bernoulli distribution,

$$p(s_{it}=1|...) = \frac{p_t^l N(\mu_0,\sigma_0^2)}{(1-p_t^l)N(\mu_0,\sigma_1^2)+p_t^l N(\mu_{1t},\sigma_1^2)} \tag{35}$$

Collecting terms involving μ_0, the following is obtained,

$$p(\mu_0|...) \propto \prod_{i,t}^{N,T} \sigma_0^{-1} exp\left(-\frac{1}{2\sigma_0^2}(y_{it}^l-\mu_0 s_{it})^2\right), \tag{36}$$

therefore μ_0 can be sampled from a normal distribution with sample mean $\bar{y}_0$ and standard deviation $\frac{\sigma^2}{N_0}$, where N_0 is the amount of observations with a latent indicator of 0.

Standard results for a regression model show that β_{μ_1} can be drawn from a normal distribution,

$$p(\beta_{\mu_1}|...) = N(\hat{\beta}_{\mu_1},\sigma^2(X'X)^{-1}). \tag{37}$$

Collecting terms involving σ_0^2 from the prior and likelihood I obtain

$$p(\sigma_0^2|...) \propto \sigma_0^{-\frac{N_0T+0.01+2}{2}} exp\left(-\frac{1}{2\sigma_0^2}\left(0.01+\sum_{it}^{N_0T}(y_{it}^d-\mu_0(1-s_{it})-\mu_{1t}s_{it})^2\right)\right), \tag{38}$$

Which means σ_0^2 can be sampled from an inverse gamma-2 distribution,

$$p(\sigma_0^2|...) = iG2(0.01+\sum_{it}^{N_0T}(y_{it}^d-\mu_0(1-s_{it})-\mu_{1t}s_{it})^2, 0.01+N_0T), \tag{39}$$

and similarly for σ_1^2.

Latent factor component

We obtain a state space model with the transition (state) equation that involves solely the latent factors and observation equation that uses the estimated parameters and observed time series. The transition equation is,

$$\boldsymbol{f_{t+1}} = \Phi \boldsymbol{f_t} + \boldsymbol{\eta_{t+1}}, \qquad \boldsymbol{\eta_{t+1}} \sim N(\boldsymbol{0},\Omega), \tag{40}$$

where $\Omega = \boldsymbol{I} - \Phi\Phi'$ due to the restriction on the unconditional covariance matrix. The observation equations are as follows,

$$\begin{aligned}
\boldsymbol{y_t^m} &= \boldsymbol{\alpha^m} + \boldsymbol{B^m}\boldsymbol{f_t^m} + \boldsymbol{\epsilon_t^m}, && \boldsymbol{\epsilon_t^m} \sim N(\boldsymbol{0},\Sigma^m),\\
\frac{\kappa_t^d}{\omega_t^d} &= \alpha^d + \boldsymbol{\beta^{d'}}\boldsymbol{f_t^m} + \boldsymbol{\gamma^{d'}}\boldsymbol{f_t^l} + \zeta_t^d, && \zeta_t^d \sim N(0,1/\omega_t^d),\\
\frac{\kappa_t^l}{\omega_t^l} &= \alpha^l + \boldsymbol{\beta^{l'}}\boldsymbol{f_t^m} + \boldsymbol{\gamma^{l'}}\boldsymbol{f_t^d} + \boldsymbol{\delta^{l'}}\boldsymbol{f_t^l} + \zeta_t^l, && \zeta_t^l \sim N(0,1/\omega_t^l).
\end{aligned} \tag{41}$$

The latent factor $\boldsymbol{f_t}$ of the transition equation is obtained by using a simulation smoother. The results of Durbin & Koopman (2002) are used, which is simpler and more computational efficient than the simulation smoother by De Jong & Shephard (1995). The implementation of Durbin and Koopman uses a standard Kalman filter and disturbance smoother which then allows for simulation smoothing.

The coefficient matrix of the transition equation, Φ, is obtained by collecting the terms involving ϕ_{jj} and implementing a Metropolis-Hastings step in the Gibbs sampler that draws from the following posterior,

$$p(\phi_{jj}|...) \propto (1-\phi_{jj}^2)^{-1/2} exp\left(-\frac{1}{2(1-\phi_{jj}^2}(\boldsymbol{f_{t+1}}-\phi_{jj}\boldsymbol{f_t})^2\right), \qquad \forall j, \qquad j \in m,d,l. \tag{42}$$

The Metropolis-Hastings step is necessary since ϕ_{jj} appears in both the mean and the variance, which does not allow for straightforward sampling from a distribution. The Metropolis-Hastings step uses as a proposal density a normal distribution with mean and variance equal to respectively the mode and inverse of the negative Hessian of the log of the probability density function of equation (42).

Gibbs sampling pseudocode

Algorithm 1: Mixed Measurement Dynamic Factor Model

Initialize parameters $\mathbf{x^{(0)}}$

for iteration i = 1, 2, ..., 100,000 **do**

Macro component

x_1^i ~Matricvariate Normal ($\mathbf{A^m}$)

x_2^i ~inverse Wishart ($\Sigma^{\mathbf{m}}$)

Default component

x_3^i ~Pólya-Gamma ($\omega^{\boldsymbol{d}}$)

x_4^i ~Normal ($\alpha^{\boldsymbol{d}}$)

Loss given default component

x_5^i ~Pólya-Gamma ($\omega^{\boldsymbol{l}}$)

x_6^i ~Normal ($\alpha^{\boldsymbol{l}}$)

x_7^i ~Bernoulli ($\mathbf{s}$)

x_8^i ~Normal ($\boldsymbol{\mu_{1t}}$)

x_8^i ~Normal ($\boldsymbol{\mu_0}$)

x_8^i ~inverse gamma-2 ($\boldsymbol{\sigma_1}$)

x_8^i ~inverse gamma-2 ($\boldsymbol{\sigma_0}$)

Latent factor component

x_9^i ~ simulation-smoother ($\mathbf{f_t}$)

x_{10}^i ~MH (normal proposal) (Φ)

D Data

Data filtering

In order to exclude observations that do not represent the data without any imposed bias several filters are applied.

- *Unrepresentative states.* Observations belonging to the three sub-national administrative divisions Puerto Rico (PR), Guam (GU) and the Virgin Islands (VI) are filtered due to significant differences with the remaining states, resulting in the removal of 307,103 loan observations over time.

- *Workout period bias.* Observations imposing a workout period are filtered by removing observations after 2015. The workout period is generally referred to as the time between the default date and the foreclosure date. When a loan is in default the foreclosure process starts. In a judicial-foreclosure state a court is involved to sell the property during this process, for non-judicial states a local attorney is hired to auction the mortgage's property. As a result the length of the foreclosure process and related costs influences the loss given default considerably. To prevent a workout period bias, mortgage observations in 2015 are not considered in the analysis as loss observations in this year have a short workout and therefore impose a bias. Figure 19 shows the workout period distribution, with workout period observations mostly below 2 and on average 1.483 years. This further results in a removal of 28,103 default observations and 26,703 loss observations from the data set without PR, GU and VI.

- *Data sparseness.* The first years of the data contain few quarterly loss observations which provides difficulty when estimating parameters of the loss given default component, as a result 2000 and 2001 are not considered in the analysis of the default rate and loss given default, omitting 4,816 default observations and 1,342 loss given default observations.

- *Loss outlier influence.* To limit the influence of outliers, loss given default observations outside of the $[-0.5, 1.5]$ range are removed, this filters out 28,099 of the remaining 486,644 loss given default observations. Höcht & Zagst (2007), Hartmann-Wendels et al. (2014) and Keijsers et al. (2017) use a similar range for loss filtering. Given the small fraction of loss observations outside of the $[-0.5, 1.5]$ range it seems acceptable to filter in a similar way.

The resulting data considers the period 2002-2014 and has a total of 798,012 default observations and 458,599 loss given default observations. Table 12 shows the amount of observations and results of filtering.

Loan observations

Year	Active loans	*Filtered*	Defaults	Losses	$\not\subset [-0.5, 1.5]$
2000	1,170,506	3,276	422	71	0
2001	2,950,200	5,676	4,394	1,271	3
2002	4,277,629	7,685	10,733	4,471	25
2003	5,186,266	11,395	17,307	9,018	74
2004	5,234,521	13,244	19,025	11,912	142
2005	5,385,193	13,883	18.566	11,346	333
2006	5,659,185	15,658	20,546	11,630	498
2007	6,162,060	17,944	21,831	13,167	1,008
2008	6,770,716	20,701	50,993	19,723	1,500
2009	7,274,501	22,759	165,142	39,080	2,650
2010	7,230,763	24,255	166,605	82,593	4,749
2011	7,030,153	25,804	115,350	79,765	3,911
2012	6,852,386	28,462	88,668	77,350	4,691
2013	6,941,010	30,939	55,997	58,440	4,645
2014	7,268,666	32,373	34,781	40,104	3,873
2015	7,525,538	33,049	28,103	26,703	2,425
Total	92,919,293	307,103	830,931	486,644	30,527

Table 12: This table presents the loan observations per year. Loan observations are filtered if the property is located in the Virgin Islands, Guam or Puerto Rico. The table also displays the number of default and loss observations per year. In case a loss observation is smaller than -0.5 or larger than 1.5 it is considered an outlier and removed from the analysis.