TI 2015-050/III
Tinbergen Institute Discussion Paper

Cyclicality in Losses on Bank Loans

Bart Keijsers[1]
Bart Diris[1]
Erik Kole[1]

[1] Erasmus University Rotterdam; Tinbergen Institute, The Netherlands

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at http://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

Cyclicality in Losses on Bank Loans*

Bart Keijsers† Bart Diris
Erik Kole

Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam

Tinbergen Institute, Erasmus University Rotterdam

September 2017

Abstract

Based on unique data we show that macro variables, the default rate and loss given default of bank loans share common cyclical components. The innovation in our model is the distinction between loans with either severe or mild losses. The variation in the proportion of these two types drives the cyclic behavior of the loss given default, and constitutes the links with the default rate and macro variables. These links vary according to loan and borrower characteristics. During downturns, the proportion of defaults with severe losses increases, but the distribution of losses conditional on their being mild or severe does not change. Though loans are monitored more closely than bonds and are more senior, the cyclical variation in their losses resembles those for bonds, albeit around a lower average level. This variation leads to an increase in the capital reserves required for loan portfolios.

Keywords: Loss-given-default, default rates, credit risk, capital requirements, dynamic factor models

JEL classification: C32, C58, G21, G33

*The authors thank NIBC Bank, in particular Michel van Beest, for providing access to the Global Credit Data data and helpful comments. We thank Europlace Institute of Finance for financial support. We would like to thank Fabio Canova (the editor), three anonymous referees, Graham Elliot, Siem Jan Koopman, Allan Timmermann, Rossen Valkanov, and participants at the 3rd Annual IAAE Conference in Milan 2016, the Financial Econometrics and Empirical Asset Pricing Conference at the University of Lancaster 2016, the SoFiE Conference Aarhus 2015, the EFMA Breukelen 2015, the Financial Risks International Forum Paris 2015, the PECDC General Members Meeting The Hague 2014, the ESEM Toulouse 2014, the CEF Oslo 2014, the NESG Tilburg 2014, and seminar participants at ABN AMRO Bank and Erasmus University Rotterdam for comments and feedback. The opinions expressed in this article are the authors' own and do not reflect the view of NIBC Bank or Global Credit Data.

†Corresponding author. Address: Burg. Oudlaan 50, Room H08-11, P.O. Box 1738, 3000DR Rotterdam, The Netherlands, Tel. +31 10 408 86 59. E-mail addresses `keijsers@ese.eur.nl.`, `diris@ese.eur.nl`, `kole@ese.eur.nl`

1 Introduction

Recent advances in the risk management of bank loans, such as stress tests for the banking sector, highlight the importance of investigating their risk in relation to the macroeconomic environment. As stated in the Basel II Accord, risk measures should "reflect economic downturn conditions where necessary to capture the relevant risks" (BCBS, 2005). Though loan defaults occur more frequently during economic downturns, it is neither clear whether the resulting losses also show cyclical behavior, nor whether they are related to the business cycle.

In this paper, we analyze the cyclical variation in bank loan losses, their relation to the business cycle and differences across loan categories, and show that this information can improve the risk management of banks. We base our analysis on a large sample of 22,080 defaults from Global Credit Data[1], spanning the period 2003–2010. Their databases contain loans and defaults, and information on the recovery, the seniority of the loans, and the industry, country and size of the borrowers. To exploit this detailed information, we build a model that can accommodate both time and cross-sectional variation in the default rate and the loss given default (LGD), and link them to macroeconomic variables. We show how the model can be used for risk management.

Our research brings new insights for two reasons. First, the cyclicality of bank loan losses might be different from the more commonly studied bond losses, whose LGD and default rates are cyclical, related to the business cycle and positively correlated with each other.[2] Bank loans differ in several fundamental respects from bonds. Banks monitor their loans more closely than bond owners. Their loans are often more senior and backed by collateral. Further, they can postpone the sale of collateral until a favorable economic state, hoping to receive a higher price.[3] As a consequence, the default rate and LGD for bank loans can be lower, less cyclical and less interrelated. Besides, our research is based on the actual workout LGD, whereas research on bond losses mostly uses the expected or market-implied LGD shortly after default.

Second, research on bank loans default is scarce because data are not easily available and typically constitute samples that are either short or focus on a single country or

[1]In March 2015, the consortium changed its name to Global Credit Data from Pan European Credit Data Consortium.

[2]See the surveys by Allen and Saunders (2003) and Schuermann (2004). Pesaran et al. (2006), Duffie et al. (2007) and Azizpour et al. (2015) model the relation between the default rate and macro variables, whereas Frye (2000) and Creal et al. (2014) also include the LGD in their models.

[3]Acharya et al. (2007) shows the importance of the fire-sales effect for the LGD of bonds.

loan type, see, for example, Grunert and Weber (2009), Calabrese and Zenga (2010) and Hartmann-Wendels et al. (2014). We instead study a unique and rich data set that contains defaults for various countries and loan types over a period of eight years. Our model can reveal the influence of characteristics on time variation, in the form of different sensitivities to the same cycle or of adaption to different cycles, for example based on industry (see Shleifer and Vishny, 1992).

An important empirical difference between loan and bond LGD is their domain. Our data shows that bank loan LGD can exceed 100% or fall below 0%, whereas bond LGD always lies within this interval. If the LGD exceeds 100%, the bank loses more than the initial loan, for example because of principal advances (the bank lends an additional amount to the borrower for recovery). If the LGD falls below 0%, the bank recovers more than the initial loan, for example because of penalty fees, additional interest and recovered principal advances. Moreover, the LGD distribution is bimodal, with most loans being close to either a full recovery or a full loss. Schuermann (2004) shows this bimodality for bond LGD. Because of the differences in domain, we cannot use existing models for bond LGD in our analysis.

Our model links macroeconomic variables, the default rate, and the LGD via latent factors that follow autoregressive processes. While this set-up has been used before (see e.g. Pesaran et al., 2006; Koopman et al., 2012), the LGD component in our model is new. We model the LGD as a mixture of two normal distributions that differ in their means to capture the bimodality and LGD outside the $[0, 1]$ interval. So, losses can be mild or severe with a certain probability that depends on the factors.[4] The parameters that relate the LGD and the default rate to the latent factors can depend on loan characteristics.

We estimate our hierarchical model using a Bayesian Gibbs sampler. The main advantage of the Gibbs sampler is that it allows us to divide the complicated overall estimation problem into smaller subproblems (the different Gibbs steps) which are easier to deal with. The main complication in the estimation is that the probability of default and of severe losses depend in a nonlinear way on parameters and factors. We solve this issue using a new data augmentation technique proposed in Polson et al. (2013) which leads to easier to analyze Gaussian likelihoods conditional on the new Pólya-gamma latent variables.

Our results show the presence of a macro factor that captures the business cycle, and a

[4] Knaup and Wagner (2012) also distinguish severe and mild losses to derive a bank's credit risk indicator.

default-specific factor that captures variation in the credit cycle unrelated to the business cycle.[5] The LGD distribution varies over time via the probability of a loss being severe or mild. We do not find evidence that the average LGD for either severe or mild losses varies over time. In line with earlier research (see e.g., Frye, 2000; Schuermann, 2004), default rates and LGD of bank loans go up during economic downturns. However, the default-specific factor has an opposite effect, indicating that increases in the default rate unrelated to the business cycle are typically caused by borrowers that miss a payment but catch up later, and hence the LGD is low or zero.

Loan and borrower characteristics influence the cyclical variation in the default rate and LGD. The LGD of a collateralized loan is on average lower, but fluctuates more strongly over the business cycle as in Bruche and González-Aguado (2010). Loans to small and medium enterprises exhibit stronger fluctuations in both their default rates and LGDs compared to large corporates. While we find differences in the factor sensitivities of default rates and LGDs for different industries, we do not find evidence for industry-specific cycles.

We use our model to investigate the loss distribution of a fictional loan portfolio as in Miu and Ozdemir (2006) in a point-in-time setting. We calculate the expected loss and the economic capital (the difference the 99.9% quantile and the mean of the loss distribution). Both statistics show considerable variation over the business cycle. From peak to bottom of the cycle, the economic capital increases by a factor two. It is quite sensitive to the cyclical variation in the LGD. 22% of its increase over the cycle can be attributed to time-variation in the LGD. This result illustrates the importance of accounting for time-variation in the LGD that is related to the business cycle in risk management.

Our findings contribute to the literature on credit risk in two ways. First, we show that just as for bonds, the LGD of bank loans varies over the business cycle. The LGD for bank loans is generally much lower than for bonds, but can still double in times of distress. Whereas the average bond LGD varies from 25% to 80% as reported by Schuermann (2004), Altman et al. (2005), and Bruche and González-Aguado (2010), we find that loan LGD varies from 14% to 29% over time, though the periods that they study do not fully match with ours.[6] We also show how the cyclical behavior of the LGD is affected by characteristics. These results complements papers that only report how industry characteristics influence the (average) LGD (Schuermann, 2004; Acharya

[5] Koopman et al. (2012) refer to this default-specific factor as frailty factor.

[6] Schuermann (2004) also reports variation in LGD between 20% and 55% for traded bank loans, whereas our loans are not traded.

et al., 2007) or how the impact of seniority varies over the business cycle (Bruche and González-Aguado, 2010).

Second, our model exploits the panel structure of the LGD observations in a novel way. Though the time-varying mixture distribution for the LGD adds a layer of complexity to our model, it shows in detail how the LGD distribution changes. These insights would be lost if we would model the time-variation of the cross-sectionally averaged LGD or analyse the LGD distribution at each point in time. Our methods can be seen as an extension of Koopman and Lucas (2008), who only model the default rate, and Koopman et al. (2012), who add macro variables, by modeling the LGD as well. We also extend Calabrese (2014a), who only models the LGD, by linking the LGD to the default rate and macro variables. We deviate from Creal et al. (2014) and Bruche and González-Aguado (2010), who use the Beta distribution for the LGD, because the mixture of normals in our model can more easily accommodate observations outside $[0, 1]$. The default-specific factor also extends single Markov-switching business cycle of Bruche and González-Aguado (2010). Moreover, our model can flexibly include covariates and can easily be adopted in analyses of the risk of loan portfolios.

2 Data

We combine observations of macroeconomic variables, defaults of bank loans and their losses. Because we want to focus on the part for bank loans in our model, we make standard choices for the variables that represent the business cycle. In particular, we consider three macro variables that are also analyzed by Creal et al. (2014): the gross domestic product (GDP), industrial production (IP) and the unemployment rate (UR). The variables included are the growth rates with respect to the same quarter in the previous year. To match the mostly European loan data sets, we use macro variables of European countries. We provide an overview of the macro series in Appendix A in the online supplementary material.

Besides the macroeconomic component, our model contains components for the default rate and the LGD. We calculate the default rate as the number of defaulted loans divided by the number of loans at the start of a year. The LGD is the amount lost as a fraction of the exposure at default (EAD). We have observations of the workout LGD (also known as economic LGD), which is based on the actual cash flows after default. They are timed to coincide with the default date. By analyzing workout LGD, we further complement studies of the LGD of bonds, which mostly use the market prices of bonds soon after default (see e.g. Schuermann, 2004). Our sources for default and LGD observations are unique databases from Global Credit Data, to which we have access via NIBC, a Dutch bank. We first discuss the databases, and then turn to the data sets that we analyze.

2.1 Global Credit Data

Global Credit Data (GCD) is an international cooperation of banks to support statistical research for the advanced internal ratings-based approach (IRB) under the Basel accords.[7] The members pool information on loans and defaults to create two anonymous databases, the LGD database with information about the losses on resolved defaults, and the loan database with information to analyze the default rate. GCD has been founded in 2004 by 11 banks, and has grown to 53 members (April 2017). It focused originally on LGD, but later expanded its focus to the default rate.

Data quality is a crucial issue for GCD. It sets specific and detailed rules with regard to the default information that its members should submit. Default definitions are based on the Basel accords, and GCD uses its own standards to characterize further aspects, such as the size, industry and region of the borrower. Before default data is included in the databases, GCD conducts regular audits to check whether the data that a member submitted complies with its standards. A methodology committee regularly reviews these rules. To stimulate participation, data is available to members on the give-to-get principle. To obtain default data from a given year and loan category, members have to submit their own default data for that given year and category.

The LGD database contains the cash flows of all defaulted loans of the member banks. Because defaults are only included after the recovery process has ended, theses cash flows are final and realized. It also contains the default and resolution dates, the seniority of the loan, the presence of collateral, the size or type of the borrower as well as the industry and country to which the borrower belongs. GCD aims at a representative database with defaults dating from 1998, but some defaults go back as far as 1983. Members are obliged to submit defaults dating from 2001 onwards. Table B.I in Appendix B shows that a stable number of 40–45 banks contribute to the LGD database after 2001.

[7] See `https://www.globalcreditdata.org/` for general information and `https://www.globalcreditdata.org/index.php?page=members` for an up-to-date overview of the members.

The loan database contains information about borrowers and defaults, in particular their size or type, industry and country. Information about the seniority and the presence of collateral is not available in this database, as these characteristics are not seen as default drivers. GCD started with the construction of this database in 2009. It aims at a representative database from 2000 onwards. Table B.I shows that the number of banks contributing to this database is considerably less than to the LGD database. Because the number of defaults per bank is generally small, the need for pooling is larger for LGD than for loan data.

GCD provides a new version of the LGD database semi-annually, and of the loan database annually. Via NIBC we have access to the June-2014 version of the LGD database and the June-2013 version of the default database. While the parts of the databases that are available to NIBC vary per characteristic, they represent a large proportion of the total databases.

2.2 Sample selection

NIBC's LGD data set contains 92,797 loans with 46,628 counterparties We exclude non-representative observations based on Höcht and Zagst (2007) and NIBC's internal policy (see Appendix B.2 for details). Following NIBC's practice, we discount all cash flows by the two-year swap rate plus the spread from the loan. When the contractual spread is unavailable, we use the average spread of all defaulted loans. We transform the resulting workout LGD to a percentage of the EAD. We order the LGDs by quarter in line with the frequency of the macro variables.

The loan data set consists of in total 2.80 million loans of which 37,385 go into default, leading to an average default rate of 1.34%. The number of loans and defaults is specified per year. Because the number of contributing banks to this database is lower, the number of defaults is lower than in the LGD data set. Though we filter outliers from the LGD data set based on the size of the loan (measured by EAD), we cannot do so for the loan data set because the base value of the loan is not recorded.

Because the loan database starts in 2003, and the most recent LGD observations may be biased, we restrict our analysis to the period 2003–2010. The LGD is positively correlated with the workout period, i.e. the period needed to resolve the default. The most recent observations are few, have a short workout period by construction, and their LGDs are therefore typically small. After filtering the raw LGD data set, our sample contains

22,080 LGD observations of mostly European defaults, one of the most comprehensive datasets for bank loan LGD studied thus far. The largest data set that is reported by Grunert and Weber (2009) in their summary of empirical studies on bank loan recovery rates contains 5,782 observations over the period 1992–1995. More recently, Calabrese and Zenga (2010) and Calabrese (2014a,b) study a portfolio of 149,378 Italian bank loan defaults resolved in 1999. Hartmann-Wendels et al. (2014) consider 14,322 defaulted German lease contracts from 2001–2009. Though large, these studies focus on defaults from a single country or a single loan type whereas our dataset is more extensive.

2.3 Sample characteristics

The average LGD and the default rate both exhibit cyclical behavior (see Figure B.1 in Appendix B). Both increase during the financial crisis, though the peak of the LGD (of 28.4%) falls in 2008, whereas the peak of the default rate (of 2.2%) falls in 2009. By 2010, the average LGD is back at its pre-crisis level, but the default rate remains higher.

For our LGD sample we also investigate the cross-sectional distribution. When we pool all observations, the distribution shows bimodality, as is typical for LGD data (see e.g. Schuermann, 2004). Figure 1a shows that the LGD is either close to zero, or close to one when the complete value of the loan is lost. A substantial part (12.5%) of the LGD observations falls outside the $[0, 1]$ interval. These exceedances are related to principal advances, legal costs or penalty fees. A principal advance is an additional amount lent to aid the recovery of the defaulted borrower. If none of it is paid back, the losses exceed the EAD and the LGD exceeds one. If on the other hand the full debt is recovered, including penalty fees, legal costs and principal advances, the amount received during recovery exceeds the EAD and the LGD is negative. In line with Höcht and Zagst (2007) and Hartmann-Wendels et al. (2014), LGD observations below -0.5 or above 1.5 have been removed.

[Figure 1 about here.]

The bimodality in the distribution of the LGD is present in every quarter as shown by Figure 1b. In 2008, the financial crisis leads to higher peaks at zero and at one, indicating more defaults with either no loss or a full loss. In 2009, the peak around one is still present, but the peak at zero is substantially lower. The large proportion of full losses explains the large average LGD in those years. Our modeling framework exploits both the bimodality and the time variation of the LGD.

We report the effect of loan and borrower characteristics on the LGD statistics in Table I. The dip statistic of Hartigan and Hartigan (1985) indicates that all large subsamples are bimodal. Because some subsamples contain a small number of defaults, we limit our analysis to those subsets with at least 3,200 observations, which corresponds with 100 observations per quarter.

[Table I about here.]

Panel A shows that, as can be expected, loans of lower seniority have on average a higher LGD, and a increased probability of an LGD above 0.5. Most loans in our sample are senior, and 44% have some form of collateral. The non-parametric Kruskal-Wallis (KW) test rejects the hypothesis that the distributions of the subsets have the same location.

In Panel B we split the sample based on the size or type of the borrower. GCD distinguishes SMEs, large corporates and some more specific types of financing, for example for real estate, aircraft, or shipping. While these specific types are interesting, the number of observations is too small, and we concentrate on loans to SMEs and large corporates. The differences between those two loan categories are small, but the KW-test still indicates that they are significant.

Panel C categorizes the loans according to the industry of the borrower as indicated by GCD. A large part of the loans (67%) is concentrated in three industries, being industrials, financials or consumer staples. Industrials have the lowest average LGD and proportion of defaults with LGDs below 0.5, followed by consumer staples, and then financials. The KW-test indicates again significant differences in the locations of the distributions.

In Appendix B we compare our LGD data set to the LGD information of bonds in Moody's Ultimate Recovery Database (URD). The recovery of bank loans being different from bonds is an important motivation for our paper. Loans are typically more senior, more often have collateral, and lead to more closely monitoring (see Emery et al., 2004; Schuermann, 2004). We find that the LGDs on loans and bonds are bimodal, though bonds encounter more large LGDs. Our analysis will shed more light on the behavior of the LGD of bank loans, and the role that for example seniority plays.

3 Methods

3.1 Model specification

We propose a mixed measurement model in the style of Koopman et al. (2012) (whose notation we follow) and Creal et al. (2014), where the observations can follow different distributions and depend on latent factors. Our model contains a total of N variables at each point in time, though not all variables are always observed. We separate them in three sets being macro, loan and LGD variables, labeled m, l and d. We use y_{it}^{c} to denote the time t observation of variable i in category c, $c =$m, l, d. We use N^{c} to denote the size of a category. We collect all variables in the vector $\boldsymbol{y} = (\boldsymbol{y}^{\mathrm{m}\prime}, \boldsymbol{y}^{\mathrm{l}\prime}, \boldsymbol{y}^{\mathrm{d}\prime})'$.

The set of K latent factors $\boldsymbol{f}_t$ form the central part of our model, through which all observed processes are linked. We distinguish K^{m} macro factors $\boldsymbol{f}_t^{\mathrm{m}}$ that capture the business cycle, and affect all observed variables. Next to these macro factors, we introduce K^{l} loan factors $\boldsymbol{f}_t^{\mathrm{l}}$ that influence both the default and the LGD variables. The K^{d} LGD factors $\boldsymbol{f}_t^{\mathrm{d}}$ influence only the LGD variables. The factors $\boldsymbol{f}_t^{\mathrm{l}}$ and $\boldsymbol{f}_t^{\mathrm{d}}$ capture the dynamics of the credit cycle that are unrelated to the business cycle. Because of this general setup, we can investigate whether the LGD variables are related to the business cycle, the credit cycle, their own separate LGD cycle, or no cycle at all.

Following Koopman et al. (2012), we assume that $\boldsymbol{f}_t$ follows a VAR(1) process,

$$\boldsymbol{f}_{t+1} = \boldsymbol{\Phi} \boldsymbol{f}_t + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Omega}), \tag{1}$$

where the coefficient matrix $\boldsymbol{\Phi}$ is a diagonal matrix. The innovations are serially uncorrelated. These restrictions ensure that the loan and LGD factors are independent of the macro factors, in line with the literature on credit risk (see e.g. Duffie et al., 2009; Koopman et al., 2012). We impose that $\boldsymbol{f}_t$ is stationary, so $|\phi_{kk}| < 1$. The initial state vector $\boldsymbol{f}_1$ follows the unconditional distribution of the latent process, that is $\boldsymbol{f}_1 \sim \mathrm{N}(0, \boldsymbol{\Sigma}_f)$ with $\boldsymbol{\Sigma}_f$ solving $\boldsymbol{\Sigma}_f = \boldsymbol{\Phi}\boldsymbol{\Sigma}_f\boldsymbol{\Phi}' + \boldsymbol{\Omega}$. For identification, we impose that the unconditional variance equals the identity matrix $\boldsymbol{\Sigma}_f = \boldsymbol{I}$.

The first variable set contains the N^{m} macro variables, which depend linearly on the latent macro factors,

$$\boldsymbol{y}_t^{\mathrm{m}} = \boldsymbol{\alpha}^{\mathrm{m}} + \boldsymbol{B}^{\mathrm{m}} \boldsymbol{f}_t^{\mathrm{m}} + \boldsymbol{\nu}_t, \qquad \boldsymbol{\nu}_t \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma}^{\mathrm{m}}), \tag{2}$$

where $\boldsymbol{\alpha}^{\mathrm{m}}$ is a vector of size N^{m} containing the intercepts, and $\boldsymbol{B}^{\mathrm{m}}$ is a $N^{\mathrm{m}} \times K^{m}$ matrix with coefficients. The innovations in the macro variables follow a normal distribution with mean zero and variance matrix $\boldsymbol{\Sigma}^{\mathrm{m}}$ and are independent of the latent process. We standardize the macro variables to have zero mean and unit variance to ease the comparison of their relation with the latent factor (cf. Stock and Watson, 2002). For identification, we impose that $\boldsymbol{B}^{\mathrm{m}}$ is lower triangular with a sign restriction on the diagonal elements, and $K^{\mathrm{m}} < N^{\mathrm{m}}$.

The second variable set contains the status of the loans, y^{l}_{it}. Loan i at time t can either be performing ($y^{\mathrm{l}}_{it} = 0$) or in default ($y^{\mathrm{l}}_{it} = 1$). Conditional on $\boldsymbol{f}_t$, y^{l}_{it} follows a Bernoulli distribution with default probability p^{l}_{it},

$$y^{\mathrm{l}}_{it} | \boldsymbol{f}_t \sim \mathrm{Bernoulli}(p^{\mathrm{l}}_{it}) \tag{3}$$

$$p^{\mathrm{l}}_{it} = \Lambda(\alpha^{\mathrm{l}}_i + \boldsymbol{\beta}^{\mathrm{l}\prime}_i \boldsymbol{f}^{\mathrm{m}}_t + \boldsymbol{\gamma}^{\mathrm{l}\prime}_i \boldsymbol{f}^{\mathrm{l}}_t), \tag{4}$$

where, $\Lambda(z) = 1/(1 + \exp(-z))$ is the logistic function. The coefficients α^{l}_i, $\boldsymbol{\beta}^{\mathrm{l}}_i$ and $\boldsymbol{\gamma}^{\mathrm{l}}_i$ can depend on J^{l} loan-specific characteristics. Collecting them together with an intercept in the vector $\boldsymbol{x}^{\mathrm{l}}_i$, we obtain $\alpha^{\mathrm{l}}_i = \boldsymbol{\alpha}^{\mathrm{l}\prime}\boldsymbol{x}^{\mathrm{l}}_i$, $\boldsymbol{\beta}^{\mathrm{l}}_i = \boldsymbol{B}^{\mathrm{l}\prime}\boldsymbol{x}^{\mathrm{l}}_i$ and $\boldsymbol{\gamma}^{\mathrm{l}}_i = \boldsymbol{\Gamma}^{\mathrm{l}\prime}\boldsymbol{x}^{\mathrm{l}}_i$, where $\boldsymbol{\alpha}^{\mathrm{l}}$ is a vector and $\boldsymbol{B}^{\mathrm{l}}$ and $\boldsymbol{\Gamma}^{\mathrm{l}}$ are matrices. For identification, we impose that $\boldsymbol{\Gamma}^{\mathrm{l}}$ is lower triangular with sign-restricted diagonal elements. The number of loan factors should not exceed the number of characteristics, $K^{\mathrm{l}} \leq J^{\mathrm{l}}$.

We assume that conditional on $\boldsymbol{f}_t$, the default status of the loans are mutually independent. When no loan-specific characteristics are used, the default rate at time t is the same for all loans, $p^{\mathrm{l}}_{it} = p^{\mathrm{l}}_t = \Lambda(\alpha^{\mathrm{l}} + \boldsymbol{\beta}^{\mathrm{l}\prime}\boldsymbol{f}^{\mathrm{m}}_t + \gamma^{\mathrm{l}}\boldsymbol{f}^{\mathrm{l}}_t)$, and the number of defaulted loans follows a binomial distribution. When the characteristics are categorical and separate the loans into groups, for example based on industry, country or borrower type, the number of defaulted loans within a particular group also follows a binomial distribution.

Whereas the structure of our model thus far is similar to Koopman et al. (2012), we propose a novel part for the final set of variables, which are the LGDs of a defaulted loan, y^{d}_{it}. Based on the empirical distribution in Figure 1, we distinguish defaulted loans with a severe loss (close to a full loss) from those with a mild loss (close to a full recovery). We model the default type by a latent binary variable s_{it} that takes a value zero (one) to indicate a mild (severe) loss.[8] Conditional on $\boldsymbol{f}_t$, s_{it} follows a Bernoulli distribution with parameter p^{d}_{it}. Conditional on s_{it}, y^{d}_{it} follows a normal distribution. Mathematically, this part of the model can be written as

$$y^{\mathrm{d}}_{it} \sim \begin{cases} \mathrm{N}(\mu_{i0}, \sigma_i^2) & \text{if } s_{it} = 0 \\ \mathrm{N}(\mu_{i1}, \sigma_i^2) & \text{if } s_{it} = 1 \end{cases} \tag{5}$$

$$s_{it} | \boldsymbol{f}_t \sim \mathrm{Bernoulli}(p^{\mathrm{d}}_{it}) \tag{6}$$

$$p^{\mathrm{d}}_{it} = \Lambda(\alpha^{\mathrm{d}}_i + \boldsymbol{\beta}^{\mathrm{d}\prime}_i \boldsymbol{f}^{\mathrm{m}}_t + \boldsymbol{\gamma}^{\mathrm{d}\prime}_i \boldsymbol{f}^{\mathrm{l}}_t + \boldsymbol{\delta}^{\mathrm{d}\prime}_i \boldsymbol{f}^{\mathrm{d}}_t). \tag{7}$$

We assume that conditional on $\boldsymbol{f}_t$, the LGDs are independent.

Conditional on $\boldsymbol{f}_t$, y^{d}_{it} follows a mixture of two normal distributions[9] that differ in their means,

$$\mu_{is} = \boldsymbol{\mu}_s' \boldsymbol{x}^{\mathrm{d}}_i, \quad s = 0, 1 \tag{8}$$

with $\boldsymbol{\mu}_s$ a vector of size $J^{\mathrm{d}} + 1$. These means can again be a function of the J^{d} default characteristics that we collect together with an intercept in the vector $\boldsymbol{x}^{\mathrm{d}}_i$ of size $J^{\mathrm{d}} + 1$. To prevent label switching, we impose $\mu_{i0} < \mu_{i1}$ over the support of $\boldsymbol{x}^{\mathrm{d}}_i$. We also allow the variance to be a function of the loan characteristics,

$$\ln \sigma_i^2 = \boldsymbol{\lambda}' \boldsymbol{x}^{\mathrm{d}}_i, \tag{9}$$

with $\boldsymbol{\lambda}$ a vector of size $J + 1$. We do not allow the variances to depend on the default type s_{it}. Because of this restriction and $\mu_{i0} < \mu_{i1}$, the probability $\Pr[s_{it} = 0 | y^{\mathrm{d}}_{it}, \boldsymbol{f}_t]$ is a decreasing function of y^{d}_{it}. So, the probability of a default being labeled mild decreases for increasing LGD. Without this restriction, losses that become more and more extreme would be inferred with an increasing probability to be of the type with the largest variance.

The (conditional) probability of a severe loss, p^{d}_{it} depends via a logit transformation on the latent factors (cf. Equation (4)). The coefficients can depend on the default-specific characteristics, that is $\alpha^{\mathrm{d}\prime}_i = \boldsymbol{\alpha}^{\mathrm{d}} \boldsymbol{x}^{\mathrm{d}}_i$, $\boldsymbol{\beta}^{\mathrm{d}}_i = \boldsymbol{B}^{\mathrm{d}\prime} \boldsymbol{x}^{\mathrm{d}}_i$, $\boldsymbol{\gamma}^{\mathrm{d}}_i = \boldsymbol{\Gamma}^{\mathrm{d}\prime} \boldsymbol{x}^{\mathrm{d}}_i$ and $\boldsymbol{\delta}^{\mathrm{d}}_i = \boldsymbol{\Delta}^{\mathrm{d}\prime} \boldsymbol{x}^{\mathrm{d}}_i$, where $\boldsymbol{\alpha}^{\mathrm{d}}$ is a vector and $\boldsymbol{B}^{\mathrm{d}}$, $\boldsymbol{\Gamma}^{\mathrm{d}}$ and $\boldsymbol{\Delta}^{\mathrm{d}}$ are matrices. For identification, we impose that $\boldsymbol{\Delta}^{\mathrm{d}}$ is lower triangular with positive diagonal elements. The number of default factors should

[8] This model component of latent default types is similar to the latent distinction between good and bad loans or investment projects that is used to include asymmetric information in models of capital structure, see e.g. Flannery (1986).

[9] In section 4.4, we consider a mixture of Student's t distributions.

not exceed the number of characteristics, $K^{\mathrm{d}} \leq J^{\mathrm{d}}$.

The latent factors influence the LGD via the probability p_{it}^{d} with which the default is severe. We do not include an effect of the factors on the average LGD for a given type μ_{is}, because the location of the modes of the distributions in Figure 1b stay approximately constant. The variation in the relative heights of the peaks is driven via the mixture probability p_{it}^{d}.

3.2 Estimation

We use uninformative priors that are specified in Appendix C. We impose the identification restrictions explained in the previous section using the priors.

We use a Gibbs sampler to estimate the model in a Bayesian way. This means that we simulate from the conditional posterior distributions for each parameter to obtain draws from the full posterior distribution of all parameters. The main advantage of the Gibbs sampler is that it allows us to divide our complicated estimation problem into smaller subproblems (the different Gibbs steps), which makes the estimation feasible. The following explains how we estimate the model without loan and default characteristics. The estimation of the model with these characteristics is a straightforward extension as explained in Appendix C.

The main complication in the simulation is that the probabilities of default and of a severe loss depend in a nonlinear way on the parameters and the latent factors via the logistic function. We tackle this complication by using new results in Polson et al. (2013) and Windle et al. (2013), who show that one can easily sample the parameters and latent factors from their conditional posterior distributions once one adds auxiliary latent variables - denoted ω_t^{l} and ω_t^{d} here - to the model. The main idea is that we obtain an easier to analyze linear Gaussian (state space) model conditional on the auxiliary latent variables. This new result makes our analysis feasible.

For ease of notation, we define $\boldsymbol{A}^{\mathrm{m}}$ as the matrix that collects the intercepts and slopes in Equation (2), $\boldsymbol{\alpha}^{\mathrm{l}}$ as the vector which collects all parameters in Equation (4), and $\boldsymbol{\alpha}^{\mathrm{d}}$ as the vector which collects all parameters in Equation (7).

The Gibbs sampler consists of the following steps:

- Macro module

  Sample $\boldsymbol{A}^{\mathrm{m}}$ from the matricvariate normal distribution (re-draw until $\boldsymbol{A}_{2,1}^{\mathrm{m}} = \boldsymbol{B}_{1,1}^{\mathrm{m}} < 0$ for identification of the first factor) and $\boldsymbol{\Sigma}^{\mathrm{m}}$ from an inverse Wishart distribution.

- Loan status module

  Sample latent variable ω_t^{l} from a Pólya-gamma distribution and $\boldsymbol{\alpha}^{\mathrm{l}}$ from a multivariate normal distribution (re-draw until $\boldsymbol{\alpha}_3^{\mathrm{l}} = \gamma^{\mathrm{l}} > 0$ for identification of the second factor).

- Loss given default module

  Sample latent variable ω_t^{d} from a Pólya-gamma distribution for all t and $\boldsymbol{\alpha}^{\mathrm{d}}$ from a multivariate normal distribution (re-draw until $\boldsymbol{\alpha}_4^{\mathrm{d}} = \delta^{\mathrm{d}} > 0$ for identification of the third factor). Sample s_{it} from a Bernoulli distribution for all i, t, μ_0 and μ_1 from normal distributions (re-draw until $\mu_0 < \mu_1$) and σ^2 from an inverse gamma-2 distribution.

- Factor module

  Simulate the latent factors using the simulation smoother of Durbin and Koopman (2002). Sample ϕ_{jj} for factors $j = 1, .., K$ using a Metropolis-Hastings step that imposes that $|\phi_{jj}| < 1$.

We refer to Appendix C for the exact distributions and derivations. We retain 100,000 draws after a burn-in of 50,000 draws to obtain results. Increasing the number of simulations does not impact results.

3.3 Motivation of modeling choices

Our model is designed to get a detailed view of the variation in LGDs and default rates, both over time and in relation to loan characteristics, and also of the interplay between these two sources of variation. Though simpler analyses are available, they cannot satisfactorily answer our research questions, because they do not fully exploit the richness of our data set. We first highlight the appealing properties of our model, and then indicate how it deviates from the other advanced alternatives that have recently been proposed.

Arguably, our model has two layers of complexity. The first is the latent factor structure that drives the time-series dynamics and dependence of our variables. Alternatively, the probability of default and of a severe loss can be linked directly to the macro variables. An additional default-specific (frailty) factor would then be difficult to include. The literature on credit and default risk shows that a latent factor structure can accurately capture this issue.[10] We do not use a Markov-switching process as in Bruche and González-Aguado (2010) or Calabrese (2014a), because autoregressive processes more naturally link to the gradual changes in macro variables. Our model is more advanced than the models proposed by Frye (2000); Gordy (2003); Pykhtin (2003) as we can include more factors and explicitly model their behavior and effect on the probability of default and LGD.

The second layer stems from the panel structure of our LGD observations, which we model by a mixture distribution with time-varying weights. Easier solutions could consist of data reduction by modeling the time-variation of the cross-sectional average of the LGD, or a separate or two-step analysis of the LGD distribution at each point in time. Because we fully model the LGD distribution, we can incorporate the determinants of both the cross-sectional variation and the time-variation. We think that in particular the effect of the loan and default characteristics add interesting insights to our analysis. Besides, we circumvent a generated regressor problem as in Pagan (1984).

Our model component for the LGD differs from existing models. We do not use a standard Beta distribution for the LGD as in Creal et al. (2014) and Bruche and González-Aguado (2010) or a mixture of point masses at 0 and 1 and a Beta distribution as in Calabrese (2014b). These distributions only have support on the unit interval. Because over 10% of the LGD observations are outside the $[0, 1]$ interval (see Figure 1a), using these distributions would require a transformation of the data. The results for the discrete-continuous distribution are difficult to interpret, since the LGDs drawn from the Beta component can be arbitrarily close to 0 and 1.

4 Results

4.1 Models without loan and default characteristics

We start our analysis by investigating the general relation between the macro variables, the defaults and the LGDs. In the basic specification we investigate whether the defaults and LGDs exhibit cyclical behavior, and how many factors are needed to capture it. We do not include loan nor default characteristics in this analysis. Based on the evidence in Duffie et al. (2007); Koopman et al. (2012); Azizpour et al. (2015) that favor at least one

[10] See among others Frye (2000); Gordy (2003); Pykhtin (2003); Pesaran et al. (2006); Koopman and Lucas (2008); Koopman et al. (2012); Creal et al. (2014); Azizpour et al. (2015).

macro and one default-specific factor, we take the two-factor model with a macro and a loan factor as our starting point. The macro factor can affect all processes, and the loan factor can only influence the probabilities of default and of a severe LGD. We then investigate the added value of an additional default factor. We present our estimation results in Table II.

[Table II about here.]

Our results for the two-factor specification show a persistent macro factor, and a loan factor with much quicker mean reversion. The macro variables show clear exposures to the macro factor. A high value signals a recessionary state of the world, with low growth rates for GDP and industrial production, and a high unemployment rate. Because the macro variables have been normalized with unconditional variances equal to one, an increase of the factor by one leads to changes equal to one standard deviation times their loadings, so decreases of GDP and IP by $0.969 \cdot 2.57\% = 2.49\%$ and $0.802 \cdot 6.41\% = 5.14\%$, and a increase of the unemployment rate by $0.879 \cdot 0.90\% = 0.79\%$. The factor has a slightly stronger effect on GDP growth and the unemployment rate than on the growth of industrial production.

Unsurprisingly, the business cycle factor positively affects the probability of default. On average its marginal effect is 0.11%, which is economically large, compared to the average default probability of 0.31% per quarter. The sensitivity to the loan factor is smaller than to the macro factor, and consequently the average marginal effect is smaller (0.080%) as well, though still sizable.

Our main interest is the component for the LGD. Here we also see a clear effect of the business cycle. The posterior distribution of β^{d} has a mean of 0.328, but the spread is wide. So, during a recession when the latent factor is positive, the probability of defaulted loans with a severe loss increases. A mild LGD has a mean of 7.2%, whereas severe losses are on average much larger at 82.9%. The two loss types are clearly different, as indicated by the standard deviation of 13.1%. On average, the probability of a severe loss is 17.4%, and the average LGD equals 20.4%. The marginal effect of the macro factor on the severe loss probability is on average 4.5%, which translates to an increase of the average LGD by 3.4%. Though this effect is less strong than for the default rate, it is still quite substantial. The loan factor has a negative effect on the LGD, as indicated by the negative posterior mean and small standard deviation for γ^{d}. A positive shock to this factor leads to more defaults, but decreases the probability of a severe loss. These effects indicate that these defaults are related to firms that miss a loan payment (interest or repayment), but catch up afterwards.

To get a better understanding of the factors, we plot their evolution in Figure 2a. The macro factor starts negative, which indicates the benign economic environment of the first part of our sample period. After 2008, it shows a sharp increase, corresponding with the credit crisis. The loan factor shows more erratic behavior, and is less persistent than the macro factor. It is high around 2006 and low around 2008. So, given the sensitivities the factor has an upward effect on the default probabilities but downward on the LGDs around 2006. Around 2008, the effects are reversed, indicating fewer defaults related to temporary delays in payments.

[Figure 2 about here.]

The plot of the fit over time of the macro variables, the realized default rate, and the average LGD in Figure 3 shows which part of their variation is captured by the factors. The macro factor reasonably tracks the macro variables, in particular during the great recession. The model-implied and realized default rate series almost coincide. While the macro factor captures the long-term swings in the default rate, the loan factor captures the more short-lived fluctuations. The deviations between the average model-implied and realized LGD are also relatively small, though larger than for the default rate. The combination of the factors captures the low average LGD in 2006-2007, and the subsequent pronounced upswing.

[Figure 3 about here.]

The addition of a third factor that can only influence the LGD decouples it from the other variables. The default factor is persistent, and strongly affects the LGD. The posterior mean of δ^{d} is 0.512, much larger than the mean of β^{d} of 0.031, which captures the effect of the macro factor. Compared to the two-factor specification, the influence of the macro factor decreases by a factor 10, and the effect of the loan factor is halved. Figure 2b shows that the default factor seems to lead the macro factor. A longer sample period may shed more light on this issue. Both the plot and the Widely Applicable Information Criterion, version 2 (WAIC2, Watanabe, 2010), which corrects for the number of parameters, indicate that the fit of the three-factor model is better for the macro variables, in particular for the unemployment rate. For the default rate changes are negligible. Though the fit for the average LGD in Figure 3e looks slightly better for the three-factor structure, the WAIC2 actually deteriorates.[11] Because we are mostly interested in capturing the LGD part in relation to the other variables, we do not favor the three-factor specification.

We investigate the importance of time-variation in the probability of a severe loss by estimating a two-factor specification with the restriction $\beta^{\mathrm{d}} = \gamma^{\mathrm{d}} = 0$. The WAIC2 value for the unrestricted specification is lower, indicating that the improvement in the fit outweighs the additional two parameters.

We conclude that the combination of a macro and a loan factor accurately captures the dynamics in the macro variables, probability of default and LGD. The macro factor captures the business cycle with relatively long swings, whereas the loan factor captures more short-lived fluctuations. During recessions, both the probability of default and the LGD increase, so the macro factor leads to positive dependence between default rates and LGD. However, the loan factor negatively affects their dependence, because it captures the defaults related to mere delays in loan payments.

4.2 The effects of loan and default characteristics

Loan and default characteristics affect the probability of default and the LGD of a loan. Their influence can take the form of a fixed effect, may influence the sensitivity to the latent factors, or give rise to a completely new latent factor. For example, Shleifer and Vishny (1992) argue that credit cycles are industry specific.

We use the richness of our data set to investigate the effect of seniority, and the size and industry of the borrower. Because this information in the GCD databases is categorical, we include it in our model by dummy variables. We require a minimum of 3,200 observations (100 per quarter) for a group to include the corresponding dummy variable.

For each characteristic we estimate a two-factor model and extensions with additional loan and default factors. To save space we present the results for the two-factor models here. We show the full results in Appendix D.

[Table III about here.]

[Figure 4 about here.]

[11]Because of differences in the scales of the variables, the sum of the WAIC2 values cannot be used to evaluate the overall fit.

4.2.1 Effects of seniority

In our analysis of seniority, we distinguish defaults of senior secured and senior unsecured loans (see the number of observations in Table I). Because seniority is only available for the LGD observations, it only influences the LGD component of our model.

The left panels of Table III show considerable differences between the LGD of senior secured and senior unsecured loans. First, unsecured loans suffer on average a larger LGD. Both the average severe LGD, and the average probability of the loss being severe are higher for unsecured loans (85.8% vs. 76.6% and 19.1% vs. 15.1%). The average values for a mild LGD do not differ much. Together, these differences translate to an average LGD of 22.2% for an unsecured loan and 17.5% for a secured loan.

Second, secured loans are more sensitive to the business cycle. Its sensitivity is about 2 times as large, and consequently the marginal effect on the average LGD is 7.2% (5.9%) for the secured (unsecured) loans. The LGD of secured loans does not respond strongly to the loan factor. The posterior distribution of γ^{d} is wide and close to zero. The corresponding sensitivity of unsecured defaults is much larger and clearly negative. Apparently, delays in loan payments are concentrated in unsecured loans.

The behavior of the loan factor and the probability of default have also changed in comparison with the results without characteristics. The loan factor has become more persistent, and the default rate has lost is sensitivity to the macro factor. The posterior mean of β^{l} is negative, but its distribution is very wide. The plot in Figure D.1a shows that the loan factor resembles the macro factor much more than in Figure 2a, except for the crisis period where it lags the macro factor, and looks again more like the loan factor of the model without characteristics. Because we do not know the seniority of each loan, we cannot further investigate this issue.

The fit over time of the default rate (Figure 4a) is similar to the model without characteristics in Figure 3d. Figure 4a clearly shows the differences in the LGD behavior. The average LGD is generally higher for unsecured loans than for secured loans, except during the credit crisis. These results are in line with Bruche and González-Aguado (2010) and extend Hamerle et al. (2011). The sensitivity to the macro factor explains the pronounced increase around 2008 in the LGD of secured loans. Its effect on the LGD of unsecured loans is partly offset by the loan factor. During the credit crisis, the LGD of both loan types increases, but less for unsecured loans. The dependence of the value of collateral on the business cycle may explain part of this effect.

Our results for a three factor specification, which has an additional default factor that only affects the LGD of unsecured loans do no support a different cycle for unsecured loans. While the fit of the average unsecured LGD improves, the WAIC2 actually increases. We conclude that a two-factor structure, with a strong effect of the macro factor on the secured LGD, and of the loan factor on the unsecured LGD is the best model.

4.2.2 Effects of borrower size

In our analysis of the effect of borrower size, we distinguish SMEs and large corporates, both for the loans and the defaults in our sample. Comparing the middle panels of Table III to the two-factor results in Table II shows that borrower size mostly affects the probability of default. It is much larger when lending to SMEs (0.35%) than to large corporates (0.14%). The sensitivities to both the macro and the loan factor are also higher for SMEs. This result is in line with the higher riskiness, both systematic and idiosyncratic that is documented for the equities of small firms. The distribution of the LGD does not vary in relation to borrower type. The differences in the coefficients and the implied statistics in the middle of panel D are small compared to their posterior standard deviations.

The macro and loan factors in Figure D.2a largely resemble those in Figure 2a. The default rates implied by the two-factor structure track the realized rates closely for SMEs and a bit less for large corporates (Figure 4c). However, in particular in the first part of our sample period, there is room for improvement. Because of the large number of loans to SMEs (12,000 vs. 6,500 large corporates), the loan factor reflects the SME default rate more. The factors fit the dynamics of the average LGD for SME also better than for large corporates (Figure 4d).

In Appendix D, we report the three-factor specification that has an additional SME loan factor. The comparison by WAIC2 values favors this specification, though the improvement over the two-factor model is small. The sensitivities of the SME probability of default and LGD to the general loan factor become less at the expense of the sensitivities to the new SME loan factor. The additional factor has most influence at the beginning and end of our sample period, but also points to different behavior during the credit crisis. We interpret these results as some evidence for separate default-specific factors that depend on size.

4.2.3 Effects of industry

The industry in which the borrower is active substantially influences the probability of default and the LGD. Due to the number of observations, we can distinguish the industries Consumer Staples (CS), Industrials (IND) and Financials (FIN). The results for this analysis in the right panels of Table III indicate that these differences pertain to both the components for the default probabilities and for the LGD.

For the probability of default, both the average and the factor sensitivities vary over the industries. Loans to borrowers in FIN (IND) have the lowest (highest) average default probability of 0.16% (0.41%). Sensitivities to the business cycle is highest for FIN, followed by IND. The average default probability for CS is in the middle (0.27%), but least sensitive to the business cycle. The sensitivities to the loan factor are about equal.

The LGDs differ mainly in their sensitivity to the business cycle. It is highest for FIN, followed by CS and lowest for IND. The differences in marginal effects on the average LGD, which equal 7.0%, 4.3% and 3.2%, are substantial. The differences in the other parameters are less consequential, and the average LGD is about the same for the different industries.

The factor estimates are again not much different than in our baseline specification. The ability of the model to fit the default rates remains remarkably good, as deviations of the model-implied series from the realisations are small (Figure 4e). For the LGD series, the deviations are a bit larger.

Our analysis of a four-factor specification with two additional loan factors does not provide evidence in favor of separate credit cycles for each industry, as proposed by Shleifer and Vishny (1992). The three loan factors are difficult to distinguish from white noise, and the factor loadings do not indicate that each industry has its own factor. The four-factor specification improves the fit, though perhaps less than expected.

We conclude that industry characteristics have an important effect on default rates and LGD. Loans to industrials are generally most risky, as their probability of default is highest, but loans to financials vary most related to business cycle fluctuations. Loans to CS firms are more in the middle.

4.3 Implications for risk management

We illustrate the implications of our model for risk management by the calculation of the expected loss and the economic capital, which measures the risk of unexpected losses on a loan portfolio. We calculate it as the difference between 99.9% quantile and the mean of the loss distribution. We show how they change during our sample period for a portfolio that corresponds with our data set and based on the latent factors we have inferred. Though other papers[12] have already shown the importance of the positive dependence between default rates and LGD for economic capital in an unconditional setting, our analysis gives additional insights in the dynamics of the loss distribution.

At each point in time we consider a portfolio of 2,000 loans, each with an exposure at default of 1 euro, as in Miu and Ozdemir (2006). After every full iteration of the Gibbs sampler, we simulate the loss on the portfolios, conditional on the values drawn for the parameters and factors in that iteration. This yields the posterior loss distribution per time period, from which we get the quarterly expected loss and economic capital.

Figure 5 shows the cyclical variation in the expected loss distribution and the economic capital, both for our base two-factor model and the restricted version with constant LGD as presented in of Section 4.1. We clearly see that both the expected loss and the economic capital fluctuate stronger when the LGD is also time-varying. When the LGD cannot vary over time, the expected loss varies from 0.61 (0.03%) to 2.37 (0.12%), but when the LGD can vary, the maximum is at 2.98 (0.15%), a substantial increase. In good times, time variation in the LGD leads to a slightly lower expected loss, but in bad times to a pronounced increase. These effects carry over to the economic capital, which is clearly lower from 2005 to 2006, and rapidly increases during 2008. During 2008, economic capital based on time-varying LGD is 5.64 (0.28%) compared to 5.03 (0.25%) with constant LGD. While these number may seem small, it means that 0.60 of the cyclical increase of 5.64 - 2.91 = 2.73 comes from time-variation in the LGD, so 22%. The increase of 0.60 is similar to the results in Bruche and González-Aguado (2010, Table 3).

[Figure 5 about here.]

These results illustrate how our model can be used in a risk management setting. By using values for the latent factors at a specific point in time, or prespecified values, stress tests can be conducted. Based on the point-in-time estimates of portfolio losses, or by constructing an unconditional loss distribution, our model can be used to construct through-the-cyle economic capital (see Miu and Ozdemir, 2006).

[12] See Frye (2000); Pykhtin (2003); Gordy (2003); Düllmann and Gehde-Trapp (2004); Miu and Ozdemir (2006) among others.

4.4 Alternative specifications

Figure 6 shows the fit of our model for the LGD distributions at two points in time, Q4 of 2005 and Q2 of 2008. The probability of a severe loss is much higher in 2008Q2, and accordingly, we see an increase in the right mode and a decrease in the left mode. While this effect captures some of the changes in the empirical distributions, other more flexible distributions may lead to a better fit. We therefore investigate a replacement of the normal distribution in Equation (8) by the skewed Student's t of Azzalini and Capitanio (2003). We show the full results in Appendix E and discuss the main consequences here.

[Figure 6 about here.]

The dotted lines in Figure 6 show an improvement in the fit. The low degrees of freedom leads to more peakedness for the mild losses. Both distributions show considerable skewness, which helps to fit the observations in the middle. Though we observe an increase in the average probability of a severe loss, the overall factor structure, and the factor sensitivities change only marginally (see Figure E.1).

Though the mixture of skewed Student's t distributions offers a better cross-sectional fit, it has an important theoretical disadvantage. The mixture of normal distributions with the same variance has the property that $\Pr[s_{it} = 0|y_{it}^{\mathrm{d}}, \boldsymbol{f}_t]$ is a decreasing function of y_{it}^{d}. So, if the LGD grows larger, the probability with which it is inferred as mild decreases. The (skewed) Student's t distribution does not have this property because of its fat tails, even when the degrees of freedom of the mixture components are equal.[13] For this reason, we do not replace the normal distributions in our specifications. Because the consequences for the other model components are small, we conclude that our results are robust to this choice.

Our specification does not allow for lead-lag relations between the macro variables and the loan and default variables. While we leave a full investigation of models with lead-lag effects, as well as richer VAR dynamics for future research, we briefly investigate the potential of such extensions by simply leading and lagging the macro series. We find that using past values of the macro variables does not lead to a better fit of the default rates and LGD series. Using future values slightly increases the fit of the LGD series but not of the default rate. This is likely related to the workout period. Though interesting, a model that needs future information is of course less useful in practice.

[13] We illustrate this effect in Figure E.2.

5 Conclusion

The loss given default and the default rate on bank loans are both cyclical. We show that this variation stems from a macro factor capturing the business cycle and a default-specific factor that captures variations in the credit cycle on top of the business cycle. The time variation in the LGD is explained by changes in the probability of a severe versus a mild loss. While different from bonds, bank loans are also sensitive to the business cycle with higher default rates and LGD during downturns.

Our model describes the stylized facts of the LGD on bank loans well. It captures the bimodal shape of the empirical distribution and provides an interpretation of the components, by explicitly modeling the extremes of no and full loss. It is flexible enough to include the differences across loan characteristics. Further, the model has applications in risk management, such as the calculation of economic capital.

References

Acharya, V. V., Bharath, S. T., and Srinivasan, A. (2007). Does industry-wide distress affect defaulted firms? Evidence from creditor recoveries. *Journal of Financial Economics*, 85(3):787–821.

Allen, L. and Saunders, A. (2003). A survey of cyclical effects in credit risk measurement models. BIS Working Paper 126, Bank of International Settlements, Basel, Switzerland.

Altman, E., Brady, B., Resti, A., and Sironi, A. (2005). The Link between Default and Recovery Rates: Theory, Empirical Evidence, and Implications. *Journal of Business*, 78(6):2203–2228.

Azizpour, S., Giesecke, K., and Schwenkler, G. (2015). Exploring the Sources of Default Clustering. Technical report, Working Paper, Boston University Questrom School of Business.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):367–389.

Basel Committee on Banking Supervision (2005). Guidance on Paragraph 468 of the Framework Document. *Bank for International Settlements*.

Bruche, M. and González-Aguado, C. (2010). Recovery rates, default probabilities, and the credit cycle. *Journal of Banking & Finance*, 34(4):754–764.

Calabrese, R. (2014a). Downturn Loss Given Default: Mixture distribution estimation. *European Journal of Operational Research*, 237(1):271–277.

Calabrese, R. (2014b). Predicting bank loan recovery rates with a mixed continuous-discrete model. *Applied Stochastic Models in Business and Industry*, 30(2):99–114.

Calabrese, R. and Zenga, M. (2010). Bank loan recovery rates: Measuring and nonparametric density estimation. *Journal of Banking & Finance*, 34(5):903–911.

Creal, D., Schwaab, B., Koopman, S. J., and 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.

Duffie, D., Horel, G., Saita, L., and Eckner, A. (2009). Frailty correlated default. *Journal of Finance*, 64(5):2089–2123.

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

Düllmann, K. and Gehde-Trapp, M. (2004). Systematic risk in recovery rates: an empirical analysis of US corporate credit exposures. Discussion paper 02/2004, Deutsche Bundesbank, Frankfurt.

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

Emery, K., Cantor, R., Ou, S., Solomon, R., and Stumpp, P. (2004). Recent bank loan research: Implications for Moody's bank loan rating practices. Technical report, Moody's Investor Service.

Flannery, M. J. (1986). Asymmetric information and risky debt maturity choice. *The Journal of Finance*, 41(1):19–37.

Frye, J. (2000). Depressing Recoveries. *Risk*, pages 108–111.

Gordy, M. B. (2003). A risk-factor model foundation for ratings-based bank capital rules. *Journal of Financial Intermediation*, 12(3):199–232.

Grunert, J. and Weber, M. (2009). Recovery rates of commercial lending: Empirical evidence for German companies. *Journal of Banking & Finance*, 33(3):505–513.

Hamerle, A., Knapp, M., and Wildenauer, N. (2011). Modelling loss given default: a "point in time"-approach. In Engelmann, B. and Rauhmeier, R., editors, *The Basel II Risk Parameters*, chapter 8, pages 137–150. Springer.

Hartigan, J. A. and Hartigan, P. M. (1985). The Dip Test of Unimodality. *The Annals of Statistics*, 13(1):70–84.

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

Höcht, S. and Zagst, R. (2007). Loan Recovery Determinants - A Pan-European Study. *Working Paper.*

Knaup, M. and Wagner, W. (2012). A Market-Based Measure of Credit Portfolio Quality and Bank's Performance During the Subprime Crisis. *Management Science*, 58(8):1423–1437.

Koopman, S. J. and Lucas, A. (2008). A non-Gaussian panel time series model for estimating and decomposing default risk. *Journal of Business & Economic Statistics*, 26(4):510–525.

Koopman, S. J., Lucas, A., and 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.

Miu, P. and Ozdemir, B. (2006). Basel Requirement of Downturn LGD: Modeling and Estimating PD & LGD Correlations. *Journal of Credit Risk*, 2(2):43–68.

Pagan, A. (1984). Econometric Issues in the Analysis of Regressions with Generated Regressors. *International Economic Review*, pages 221–247.

Pesaran, M. H., Schuermann, T., Treutler, B.-J., and Weiner, S. M. (2006). Macroeconomic Dynamics and Credit Risk: A Global Perspective. *Journal of Money, Credit and Banking*, 38(5):1211–1261.

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

Pykhtin, M. (2003). Recovery rates: Unexpected recovery risk. *RISK*, 16(8):74–79.

Schuermann, T. (2004). What Do We Know About Loss Given Default? In Shimko, D., editor, *Credit Risk Models and Management*, chapter 9. Risk Books, London, 2nd edition.

Shleifer, A. and Vishny, R. W. (1992). Liquidation Values and Debt Capacity: A Market Equilibrium Approach. *The Journal of Finance*, 47(4):1343–1366.

Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

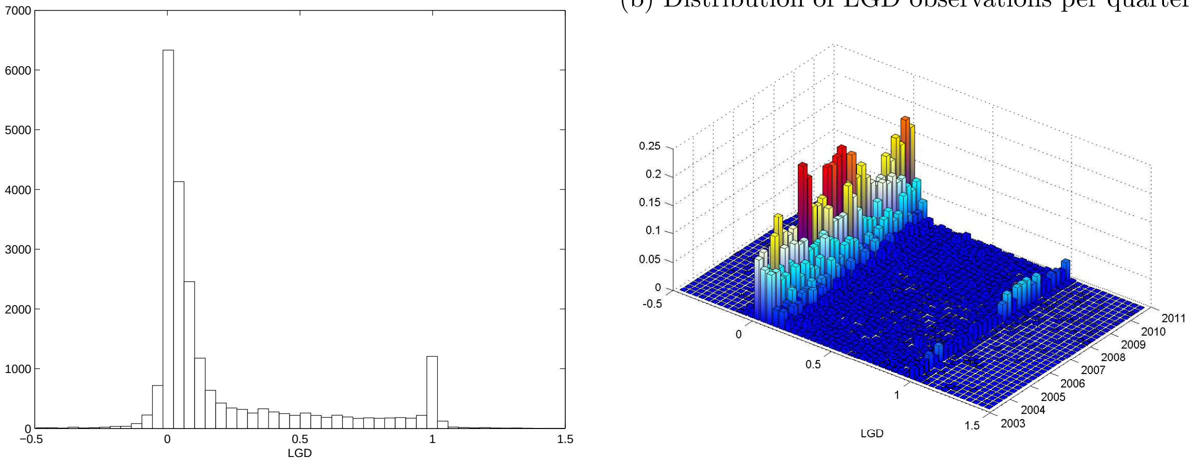
Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(Dec):3571–3594.

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

Figure 1: Empirical distribution of LGD

(a) Histogram of pooled LGD observations

(b) Distribution of LGD observations per quarter



Panel a shows the histogram of the pooled set of LGD observations. Panel b shows the empirical distribution of the LGD observations per quarter. We use all LGD observations over the period 2003–2010 after applying the filters in Appendix B.2.

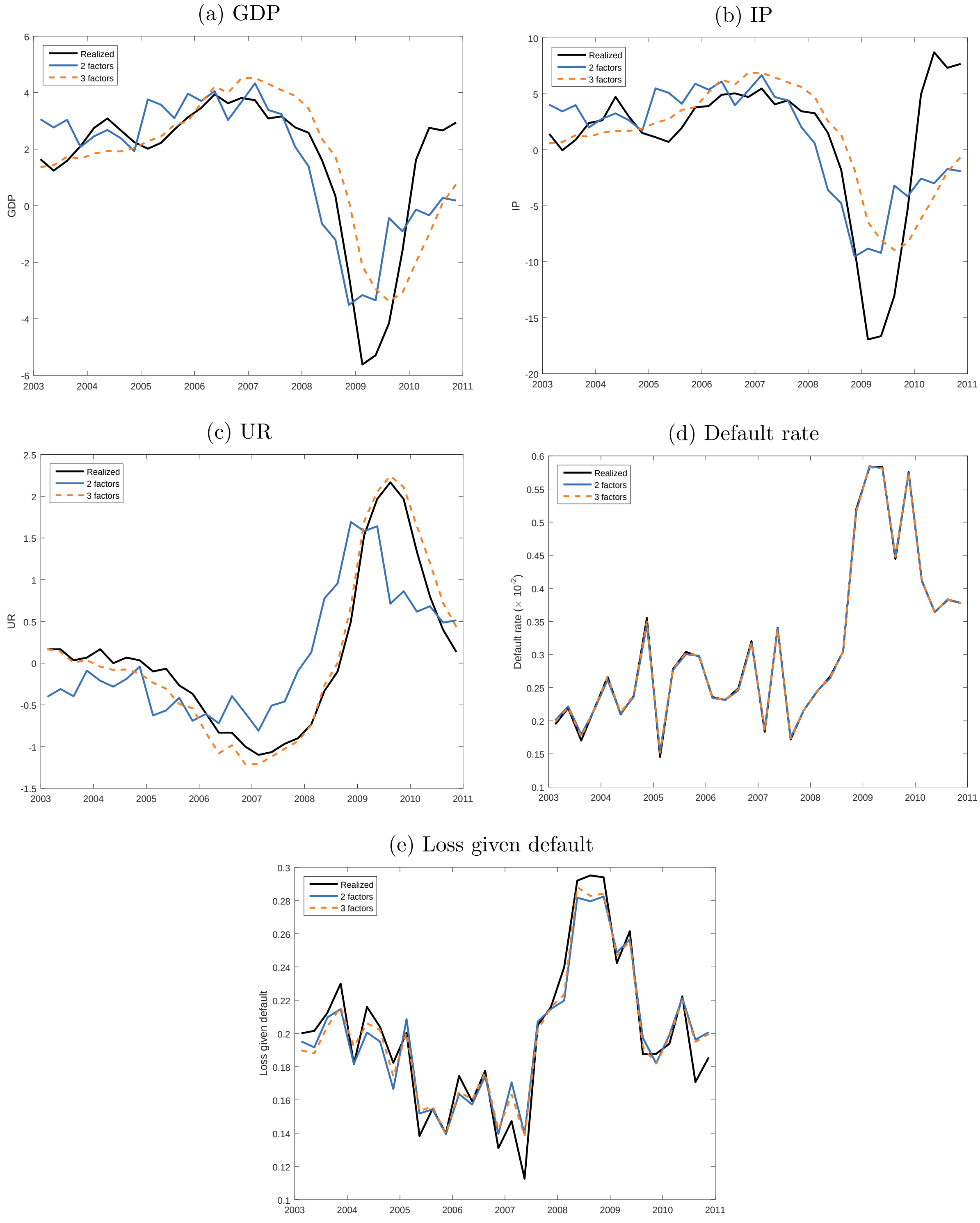
Figure 2: Latent factors

(a) Two factors

(b) Three factors

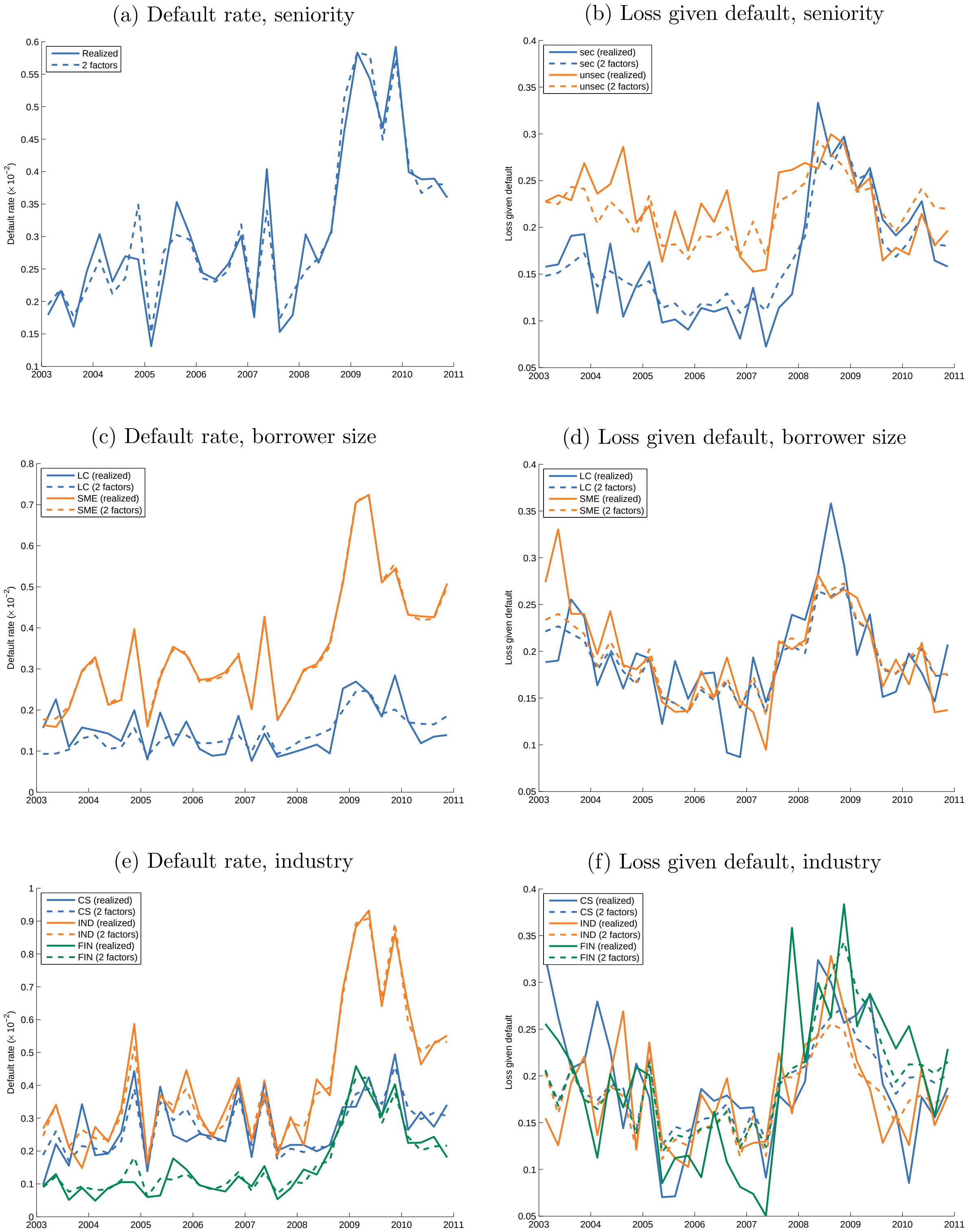
The figures show the smoothed latent factors for the model defined in Section 3.1. We report results for specifications with a macro and a loan factor (a), and an additional default factor (b). The specifications do not include loan or default characteristics.

Figure 3: Time series fit



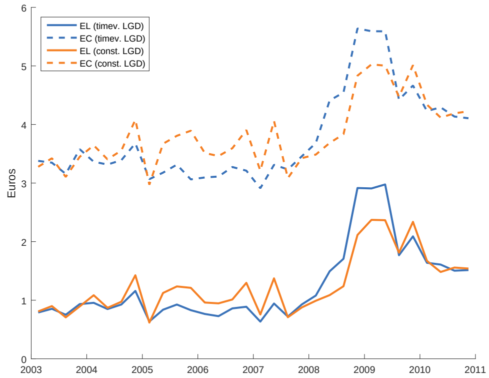
The panels show the time series fit of the model without cross-sectional differences for the growth rate of GDP (a), the growth rate of industrial production (b), the year-on-year change in the unemployment rate (c), the default rate (d) and the cross-sectional average of the LGD (e).

Figure 4: Time series fit, loan characteristics



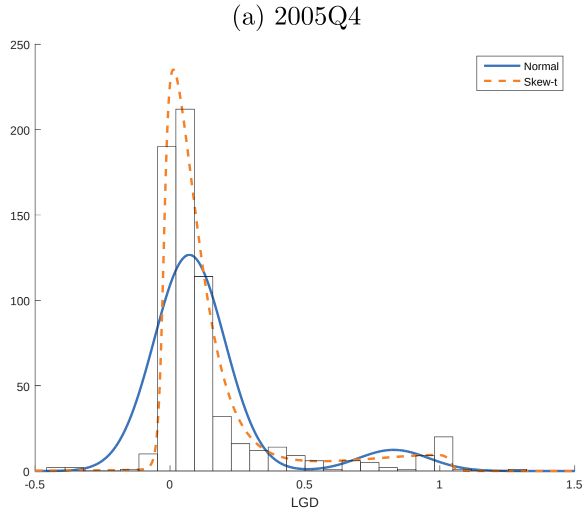
The panels show the time series fit of the two-factor model with different loan and default characteristics for the default rate and the cross-sectional average of the LGD.

Figure 5: Portfolio simulation results



The figure presents the portfolio loss (EL) and economic capital (EC) for a portfolio of loans, each quarter consisting of 2,000 loans and each loan with an exposure of 1 euro. The EL and EC are based on the two-factor model without cross-sectional variation, with a time-varying LGD (solid line) and a constant LGD (dashed lines).

Figure 6: Mixture fit



Panel a shows the cross-sectional fit on the LGDs of the mixture of normals and mixture of skew-t for the fourth quarter of 2005, and panel b shows the fit for the second quarter of 2008. The fitted distributions are based on the posterior means of the parameters of the two-factor model without cross-sectional variation.

Table I: LGD statistics per loan and borrower characteristics

Characteristic	Defaults	Average	Fraction LGD > 0.5	HDS p-value
Total	22,080	0.204	0.170	0.000
Panel A: Seniority				
Senior secured*	9,723	0.175	0.138	0.000
Senior unsecured*	12,011	0.222	0.191	0.000
Subordinated secured	110	0.289	0.255	0.002
Subordinated	236	0.427	0.419	0.000
Panel B: Borrower size or type				
SME*	12,028	0.193	0.164	0.000
Large Corporate*	6,496	0.199	0.159	0.000
Real Estate Finance	2,068	0.326	0.284	0.000
Aircraft Finance	556	0.088	0.045	0.000
Shipping Finance	331	0.077	0.054	0.100
Project Finance	302	0.177	0.132	0.002
Banks	276	0.286	0.286	0.000
Public Services	23	0.246	0.174	0.234
Panel C: Industries				
Industrials*	6,944	0.178	0.150	0.000
Financials*	4,629	0.217	0.178	0.000
Consumer Staples*	3,232	0.186	0.162	0.000
Unknown	2,817	0.309	0.279	0.000
Information Technology	1,384	0.188	0.155	0.000
Consumer Discretionary	1,089	0.196	0.128	0.034
Other	606	0.147	0.102	0.000
Telecommunication Services	410	0.203	0.183	0.304
Utilities	391	0.145	0.079	0.280
Health Care	366	0.123	0.082	0.086
Materials	212	0.147	0.127	0.534

This table presents the number of defaults, the average LGD, the fraction of defaults with an LGD larger than 0.5 and the p-value of Hartigan and Hartigan's (1985) dip statistic (HDS) using 500 bootstraps, to test the null hypothesis of a unimodal distribution versus the alternative of a multimodal distribution, for different subsets of the LGD data set. Subsets with more than 3,200 observations (indicated by a *) are selected for analysis with our model in Section 4.2.

Table II: Parameter estimates, model without loan and default characteristics

	Macro and loan factor		Macro, loan and default factor		Macro and loan factor, constant LGD	
			Panel A: Factor			
ϕ_{11}	0.856	(0.078)	0.917	(0.050)	0.939	(0.045)
ϕ_{22}	0.322	(0.207)	0.271	(0.220)	0.290	(0.215)
ϕ_{33}			0.778	(0.130)		
			Panel B: Macro variables			
α_{GDP}	−0.008	(0.317)	−0.198	(0.391)	0.580	(0.668)
β_{GDP}	−0.969	(0.531)	−0.904	(0.562)	−1.102	(0.824)
α_{IP}	−0.007	(0.284)	−0.156	(0.338)	0.475	(0.572)
β_{IP}	−0.802	(0.465)	−0.725	(0.483)	−0.889	(0.714)
α_{UR}	0.008	(0.298)	0.253	(0.467)	−0.707	(0.801)
β_{UR}	0.879	(0.491)	1.125	(0.673)	1.376	(0.971)
WAIC2	110.3		−2.4		−9.3	
			Panel C: Loan status			
α^{l}	−5.850	(0.097)	−5.759	(0.122)	−6.057	(0.197)
β^{l}	0.371	(0.201)	0.318	(0.200)	0.360	(0.271)
γ^{l}	0.258	(0.058)	0.298	(0.065)	0.292	(0.070)
av. p^{l} ($\times 10^{-2}$)	0.309	(0.002)	0.310	(0.002)	0.310	(0.002)
m.e. of f^{m}	0.114	(0.062)	0.098	(0.092)	0.111	(0.084)
m.e. of f^{l}	0.080	(0.018)	0.062	(0.020)	0.090	(0.022)
WAIC2	323.2		319.8		320.5	
			Panel D: Loss given default			
μ_0	0.072	(0.001)	0.072	(0.001)	0.072	(0.001)
μ_1	0.829	(0.002)	0.829	(0.002)	0.829	(0.003)
σ	0.131	(0.001)	0.131	(0.001)	0.131	(0.001)
α^{d}	−1.643	(0.159)	−1.697	(0.302)	−1.560	(0.018)
β^{d}	0.328	(0.202)	0.031	(0.211)		
γ^{d}	−0.293	(0.075)	−0.153	(0.072)		
δ^{d}			0.512	(0.339)		
av. p^{d}	0.174	(0.003)	0.174	(0.003)	0.174	(0.003)
m.e. of f^{m}	0.045	(0.028)	0.004	(0.029)		
m.e. of f^{l}	−0.040	(0.010)	−0.021	(0.010)		
m.e. of f^{d}			0.070	(0.046)		
av. LGD	0.204	(0.002)	0.204	(0.002)	0.204	(0.002)
m.e. of f^{m}	0.034	(0.021)	0.003	(0.022)		
m.e. of f^{l}	−0.030	(0.008)	−0.016	(0.007)		
m.e. of f^{d}			0.053	(0.035)		
WAIC2	−25,978.3		−25,969.3		−25,958.4	

This table presents the posterior mean and standard deviation (in parentheses) of the parameters of the model in Section 3.1. We report results for specifications with a macro and a loan factor, an additional default factor. and macro and loan factor that do not influence the LGD component. The specifications do not include loan or default characteristics. Panel A presents the elements of $\boldsymbol{\Phi}$ of the factor component. Panel B presents the macroeconomic component with the intercepts α and factor sensitivities β for the gross domestic product (GDP), industrial production (IP) and unemployment rate (UR). Panel C presents the loan status component. The probability of default has fixed effect α^{l} and factor sensitivities β^{l} for the macro factor, and γ^{l} for the loan factor. Panel D presents the LGD component. The LGD type can be either mild or severe. Conditional on the type, the LGD follows a normal distribution with mean μ_0 or μ_1, and volatility σ. The probability of severe loss has fixed effect α^{d} and factor sensitivities β^{d} for the macro factor, γ^{d} for the loan factor, and δ^{d} for the default factor. We also report time-series averages of the probability of default p^{l}, the probability of a severe loss p^{d} and the average LGD, the marginal effects that the factors have on these variables, and the Widely Applicable Information Criterion, version 2 (WAIC2) for each component. The number of observations, N, given by the sum of the macroeconomic, default rate, LGD observations is 22,208.

Table III: Parameter estimates, model with loan and default characteristics

	Seniority				Borrower size				Industry					
					Panel A: Factor									
ϕ_{11}	0.813	(0.120)			0.845	(0.098)					0.910	(0.060)		
ϕ_{22}	0.775	(0.138)			0.510	(0.217)					0.221	(0.269)		
					Panel B: Macro variables									
α_{GDP}	0.241	(0.631)			-0.433	(0.628)					0.526	(0.625)		
β_{GDP}	-0.935	(0.694)			-1.225	(0.885)					−1.299	(0.880)		
α_{IP}	0.187	(0.518)			-0.362	(0.543)					0.430	(0.532)		
β_{IP}	-0.725	(0.583)			-1.025	(0.756)					−1.062	(0.753)		
α_{UR}	-0.207	(0.556)			0.385	(0.572)					−0.490	(0.589)		
β_{UR}	0.792	(0.612)			1.087	(0.814)					1.206	(0.832)		
					Panel C: Loan status									
					Large corp.		SME		Cons. staples		Industrials		Financials	
α^{l}	-5.665	(0.233)			−6.441	(0.174)	−5.533	(0.237)	−6.118	(0.173)	−5.929	(0.302)	−6.990	(0.377)
β^{l}	-0.179	(0.199)			0.404	(0.273)	0.559	(0.383)	0.203	(0.293)	0.602	(0.499)	0.787	(0.600)
γ^{l}	0.628	(0.354)			0.231	(0.101)	0.366	(0.143)	0.362	(0.103)	0.364	(0.106)	0.342	(0.108)
av. p^{l} ($\times 0.01$)	0.309	(0.002)			0.143	(0.003)	0.347	(0.002)	0.273	(0.005)	0.414	(0.005)	0.162	(0.004)
m.e. of f^{m}	-0.055	(0.061)			0.058	(0.039)	0.193	(0.132)	0.055	(0.080)	0.248	(0.205)	0.128	(0.097)
m.e. of f^{l}	0.193	(0.109)			0.033	(0.014)	0.126	(0.049)	0.099	(0.028)	0.150	(0.044)	0.055	(0.018)
					Panel D: Loss given default									
	Sen. Secured		Sen. Unsecured		Large corp.		SME		Cons. staples		Industrials		Financials	
μ_0	0.070	(0.002)	0.072	(0.001)	0.075	(0.002)	0.062	(0.001)	0.056	(0.002)	0.056	(0.002)	0.085	(0.003)
μ_1	0.766	(0.004)	0.858	(0.003)	0.848	(0.005)	0.849	(0.003)	0.851	(0.006)	0.837	(0.004)	0.796	(0.006)
σ	0.129	(0.001)	0.130	(0.001)	0.126	(0.001)	0.124	(0.001)	0.121	(0.002)	0.119	(0.001)	0.145	(0.002)
α^{d}	−2.174	(0.580)	−1.665	(0.250)	−1.707	(0.291)	−1.564	(0.269)	−1.735	(0.265)	−1.786	(0.280)	−2.025	(0.465)
β^{d}	0.900	(0.619)	0.500	(0.320)	0.293	(0.302)	0.243	(0.272)	0.401	(0.314)	0.327	(0.335)	0.793	(0.549)
γ^{d}	0.041	(0.055)	−0.216	(0.133)	−0.404	(0.173)	−0.406	(0.167)	−0.327	(0.123)	−0.472	(0.147)	−0.492	(0.171)
av. p^{d}	0.151	(0.004)	0.191	(0.004)	0.160	(0.005)	0.166	(0.003)	0.163	(0.007)	0.156	(0.004)	0.185	(0.006)
m.e. of f^{m}	0.103	(0.070)	0.075	(0.048)	0.036	(0.037)	0.033	(0.037)	0.055	(0.043)	0.041	(0.042)	0.099	(0.068)
m.e. of f^{l}	0.005	(0.006)	−0.033	(0.020)	−0.050	(0.022)	−0.056	(0.023)	−0.045	(0.017)	−0.060	(0.019)	−0.061	(0.022)
av. LGD	0.175	(0.003)	0.222	(0.003)	0.199	(0.004)	0.193	(0.003)	0.186	(0.006)	0.178	(0.004)	0.217	(0.004)
m.e. of f^{m}	0.072	(0.049)	0.059	(0.038)	0.028	(0.029)	0.026	(0.030)	0.043	(0.034)	0.032	(0.033)	0.070	(0.049)
m.e. of f^{l}	0.003	(0.004)	−0.026	(0.016)	−0.039	(0.017)	−0.044	(0.018)	−0.035	(0.014)	−0.047	(0.015)	−0.044	(0.015)

This table presents the posterior mean and standard deviation (in parentheses) of the parameters of the model in Section 3.1 for different loan and default characteristics. All model specifications have a macro and a loan factor. The horizontal panels correspond with the panels in Table II. We distinguish defaults of senior secured and senior unsecured loans (left panels, labeled “Seniority”); loans and defaults of large corporates and SMEs (middle panels, labeled ”borrower size”); loans and defaults of borrowers in the sectors Consumer Staples, Industrials, and Financials (right panels, labeled “Industry”). The characteristics are included as dummy variables. Number of observations are 21,862, 18,684 and 14,997.

Supplementary Material to "Cyclicality in Losses on Bank Loans"*

Bart Keijsers† Bart Diris
Erik Kole
Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam
Tinbergen Institute, Erasmus University Rotterdam

September 2017

A Macroeconomic variables

[Table A.I about here.]

[Table A.II about here.]

[Figure A.1 about here.]

*The opinions expressed in this article are the authors' own and do not reflect the view of NIBC Bank or Global Credit Data.

†Corresponding author. Address: Burg. Oudlaan 50, Room H08-11, P.O. Box 1738, 3000DR Rotterdam, The Netherlands, Tel. +31 10 408 86 59. E-mail addresses `keijsers@ese.eur.nl.`, `diris@ese.eur.nl`, `kole@ese.eur.nl`

Table A.I: Macroeconomic variables description

Abbreviation	Subject	Measure	Country	Transformation
GDP	B1_GE: Gross domestic product - expenditure approach	GYSA: Growth rate compared to the same quarter of previous year, seasonally adjusted	OECD - Europe	-
IP	Industrial production, s.a.	Growth on the same period of the previous year	OECD - Europe	-
UR	Harmonised unemployment rate (monthly), Total, All persons	Level, rate or quantity series, s.a.	European Union (28 countries)	Difference with same quarter of previous year

The table presents the macroeconomic variables as defined in the OECD database (see http://stats.oecd.org/) and possible transformations.

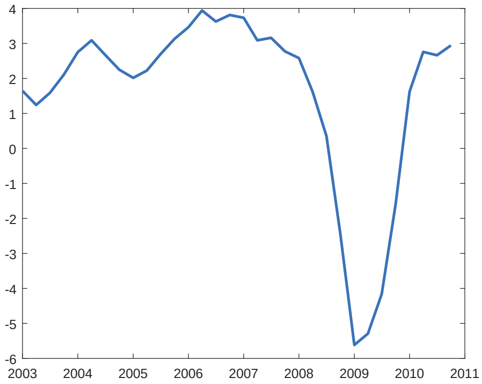
Table A.II: Macroeconomic variables statistics

	GDP	IP	UR
Mean	1.575	0.968	0.070
Median	2.621	2.806	0.017
Maximum	3.940	8.715	2.167
Minimum	-5.617	-16.945	-1.100
Standard deviation	2.567	6.412	0.903
Skewness	-1.755	-1.692	0.877
Kurtosis	4.981	5.086	3.084
AR(1)	0.894	0.861	0.938
AR(2)	0.646	0.535	0.786
AR(3)	0.343	0.135	0.581
AR(4)	0.070	-0.212	0.354

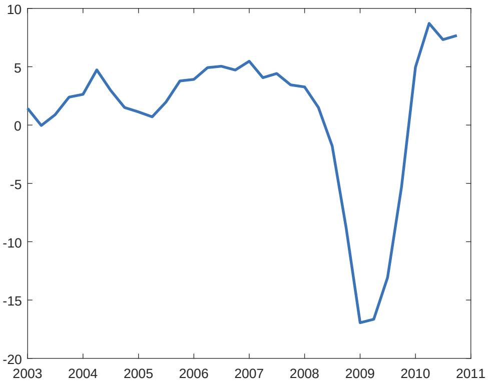
The table presents descriptive statistics for the macroeconomic variables gross domestic product, industrial production, and unemployment rate, in differences with the same quarter of the previous year, as defined in table A.I. AR(x) is the x-th order autocorrelation.

Figure A.1: Macroeconomic variables time series

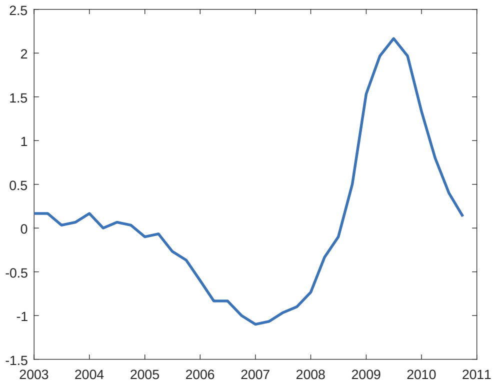
(a) GDP



(b) IP



(c) UR



The figures present the time series of the macroeconomic variables gross domestic product (a), industrial production (b) and unemployment rate (c), in differences with the same quarter of the previous year, as defined in table A.I.

B Loan and default data

B.1 GCD databases

[Table B.I about here.]

[Figure B.1 about here.]

B.2 Data Filter

Following Höcht and Zagst (2007), who also use data from the Global Credit Data Consortium, and NIBC's internal policy, we apply the following filters to the LGD database.

- *$EAD \geq$ €100,000.* The paper focuses on loans where there has been an actual (possible) loss, so EAD should be at least larger than 0. Furthermore, there are some extreme LGD values in the database for small EAD. To account for this noise, loans with EAD smaller than € 100,000 are excluded.

- $-10\% < \big((CF + CO) - (EAD - EAR)\big)/(EAD + PA) < 10\%$*, where CF cash flows, CO charge-offs and PA principal advances.* The cash flows that make up the LGD should be plausible, because they are the major building blocks of the LGD. A way of checking this is by looking at under-/overpayments. The difference between the EAD and the exposure at resolution (EAR), where resolution is the moment where the default is resolved, should be close to the sum of the cash flows and charge-offs. The cash flow is the money coming in and the charge-off is the acknowledgement of a loss in the balance sheet, because the exposure is expected not to be repaid. Both reduce the exposure and should explain the difference between EAD and EAR. There might be an under- or overpayment, resulting in a difference. To exclude implausible cash flows, these loans are excluded when they are more than or equal to 10% of the EAD and principal advances (PA). The 10% is a choice of the Global Credit Data Consortium.

- $-0.5 \leq LGD \leq 1.5$. Although theoretically, LGD is expected between 0 and 1, it is possible to have an LGD outside this range, e.g. due to principal advances or a profit on the sale of assets. Abnormally high or low values are excluded. They are implausible and influence LGD statistics too much.

- *No government guarantees.* The database contains loans with special guarantees from the government. Most of the loans are subordinated, but due to the guarantee, the average of the subordinated LGD is lower than expected. Because the loans are very different from others with the same seniority and to prevent underestimation of the subordinated LGD, these loans are excluded from the dataset.

Some consortium members also filter for high principle advances ratios, which is the sum of the principal advances divided by the EAD. Even though high ratios are plausible, they are considered to influence the data too much and therefore exclude loans with ratios larger than 100%. NIBC does include these loans, because they are supposed to contain valuable information and the influence of outliers is mitigated because they cap their LGD to 1.5. The data shows that the principal advances ratio does not exceed 100%, so applying the filter does not affect the data and is therefore not considered.

B.3 Comparison with Moody's Ultimate Resolved Database

Because we do not have direct access to Moody's URD, we use its discussion in Altman and Kalotay (2014) and Bastos (2014) to construct a comparison. Moody's URD contains information about some 5,200 resolved defaults of bonds (around 60%) and bank loans (around 40%), and about 1,000 borrowers. Figure B.2a shows that the LGD distribution of bonds is still bimodal, but has more probability mass at large losses. The LGD distributions of bank loans of both data sets are quite similar, even though URD focuses on "US non-financial corporations holding over $50 million in debt at the time of default" (Bastos, 2014). Consequently, the average LGD of bonds is much higher (55.1%) than of loans (19.5% based on URD and 20.1% in our data set). Figure B.2b shows that the average LGD of bonds exceeds that of loans. The behavior of both series is similar before the credit crisis, but differs after it. It also shows that our LGD data set is much larger than the URD. These results confirm that the LGD of bank loans differ substantially from bonds.

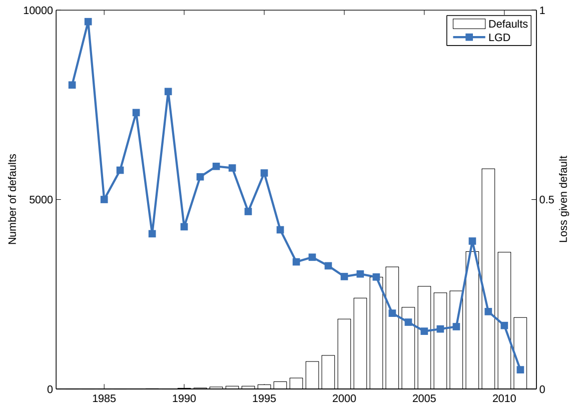
[Figure B.2 about here.]

Table B.I: Number of banks contributing to the databases

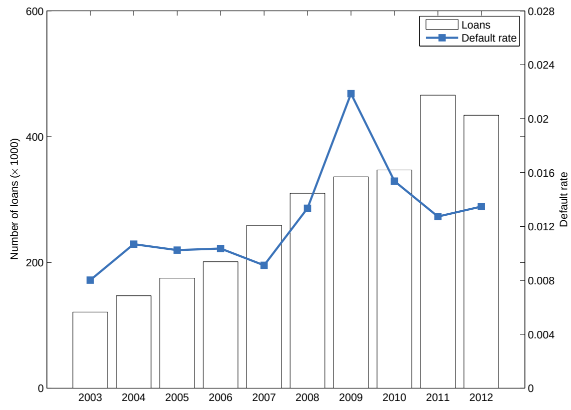
Year	LGD	Loan
2000	33	NA
2001	38	NA
2002	41	NA
2003	43	7
2004	39	9
2005	41	10
2006	45	10
2007	47	11
2008	46	14
2009	46	16
2010	43	17
2011	40	17
2012	42	16
2013	39	NA
2014	37	NA

This table shows how many banks contribute loans and defaults to the LGD and the loan databases for a given year. The versions of the databases correspond with June 2014 (LGD) and June 2013 (Loan).

Figure B.1: Evolution of LGD and default rate observations



(a) Number of defaults and average LGD over time

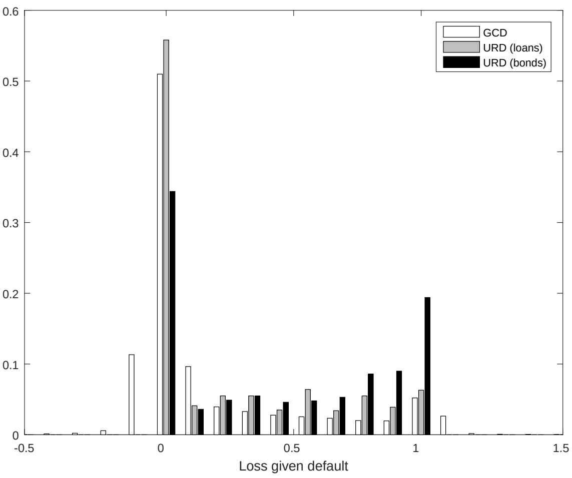


(b) Default rate

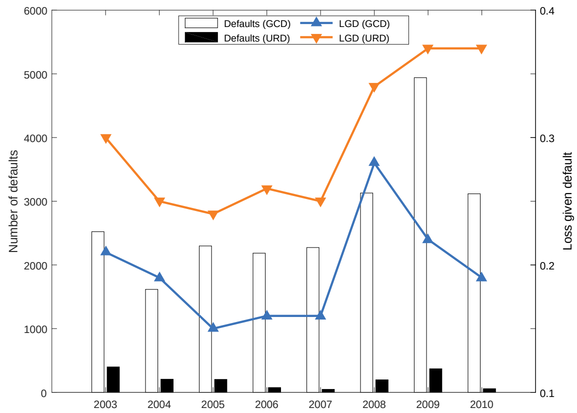
Panel a presents the average LGD and the number of observations per year for the period 1983–2011 from the Global Credit Data default database. Panel b presents the number of loans and the observed default rate per year for the period 2003–2012 from the Global Credit Data loan database.

Figure B.2: Comparison of LGD in our data set with Moody's URD

(a) Histogram of pooled LGD observations



(b) Number of defaults and average LGD per year



Panel a shows the pooled distribution of the LGD, and panel b shows the number of defaults and the average LGD per year, both for our GCD data set and Moody's Ultimate Recovery Database. We use all LGD observations in the GCD data set over the period 2003–2010 after applying the filters in appendix B.2. The URD distributions in panel a are based on Altman and Kalotay (2014), and the evolution in panel b on Bastos (2014).

C Bayesian estimation procedure

This section provides a description of the Bayesian estimation of the model in section 3. The likelihood, priors and posteriors are derived for the model without loan and default characteristics. We explain in appendix C.4 how to extend this to include loan and default characteristics, and how to replace the LGD component by a mixture of skew-t distributions.

C.1 Likelihood and latent variables

The likelihood consists of a macro, loan status and LGD component. First, we consider the macro component of the likelihood. Define $\boldsymbol{Y}^{\mathrm{m}}$ as the $T \times N^{\mathrm{m}}$ matrix with observations on $\boldsymbol{y}_t^{\mathrm{m}}$ and define $\boldsymbol{X}^{\mathrm{m}}$ as the $T \times (K^{\mathrm{m}}+1)$ matrix with a constant and observations on the macro factors $\boldsymbol{f}_t^{\mathrm{m}}$. We can write the likelihood as

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

Second, we derive the loan status component of the likelihood. Define $\boldsymbol{y}^{\mathrm{l}}$ as the vector with all loan indicators y_{it}^{l}, the vector $\boldsymbol{\psi}^{\mathrm{l}}$ with elements $\psi_t^{\mathrm{l}} = \alpha^{\mathrm{l}} + \beta^{\mathrm{l}\prime}\boldsymbol{f}_t^{\mathrm{m}} + \gamma^{\mathrm{l}\prime}\boldsymbol{f}_t^{\mathrm{l}}$, D_t as the number of defaulted loans in period t and L_t as the number of total loans in period t. We can write the likelihood as

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

Third, we analyze the LGD component given the severe loss indicator s_{it}. Define $\boldsymbol{y}^{\mathrm{d}}$ as the vector with all realized LGDs and define $\boldsymbol{s}$ as the vector which contains all s_{it} for all i, t. We can write the likelihood as

$$p(\boldsymbol{y}^{\mathrm{d}}|\boldsymbol{s}, \mu_0, \mu_1, \sigma^2) \propto \prod_{i,t}^{N,T} \sigma^{-1} \exp\left(-\frac{1}{2\sigma^2}(y_{it}^{\mathrm{d}} - \mu_0(1-s_{it}) - \mu_1 s_{it})^2\right).$$

We use three types of latent variables in our model next to the factors. First, we use the severe loss indicators s_{it} for all i, t. Define the vector $\boldsymbol{\psi}^{\mathrm{d}}$ with elements $\psi_t^{\mathrm{d}} = \alpha^{\mathrm{d}} + \beta^{\mathrm{d}\prime}\boldsymbol{f}_t^{\mathrm{m}} +$

$\gamma^{\mathrm{d}\prime}\boldsymbol{f}_t^{\mathrm{l}}+\delta^{\mathrm{d}\prime}\boldsymbol{f}_t^{\mathrm{d}}$, T_t as the total number of LGD observations in period t and N_t as the total number of severe losses in period t.

$$
\begin{aligned}
p(\boldsymbol{s}|\boldsymbol{\psi}^{\mathrm{d}}) &= \prod_{it}^{NT}(p_t^{\mathrm{d}})^{s_{it}}(1-p_t^{\mathrm{d}})^{(1-s_{it})} \\
&= \prod_{t}^{T}\Lambda(\psi_t^{\mathrm{d}})^{N_t}\big(1-\Lambda(\psi_t^{\mathrm{d}})\big)^{(T_t-N_t)} \\
&= \prod_{t}^{T}\frac{\exp(\psi_t^{\mathrm{d}})^{N_t}}{\big(1+\exp(\psi_t^{\mathrm{d}})\big)^{T_t}}.
\end{aligned}
$$

Second, we follow Polson et al. (2013) and use the auxiliary latent variables ω_t^{l} and ω_t^{d} to make the sampling of respectively the loan status and LGD components easier

$$
\begin{aligned}
p(\omega_t^{\mathrm{l}}|L_t,\psi_t^{\mathrm{l}}) &= \mathrm{PG}(L_t,\psi_t^{\mathrm{l}}), \\
p(\omega_t^{\mathrm{d}}|T_t,\psi_t^{\mathrm{d}}) &= \mathrm{PG}(T_t,\psi_t^{\mathrm{d}}),
\end{aligned}
$$

where the definition of the Pólya-gamma distribution is given in equation (1) of Polson et al. (2013).

Finally, we derive some useful results that help us with deriving the posterior distribution. Windle et al. (2013) show that the Pólya-gamma distribution has the special form

$$
p(\omega_t^{\mathrm{l}}|L_t,\psi_t^{\mathrm{l}}) = \cosh^{L_t}(\psi_t^{\mathrm{l}}/2)\exp(-\omega_t^{\mathrm{l}}(\psi_t^{\mathrm{l}})^2/2)p(\omega_t),
$$

and that the following holds

$$
\cosh^{L_t}(\psi_t^{\mathrm{l}}/2)/\big(1+\exp(\psi_t^{\mathrm{l}})\big)^{L_t} \propto \exp(-\psi_t^{\mathrm{l}}L_t/2).
$$

This implies the following result that we use in the next sections

$$
\begin{aligned}
p(\boldsymbol{y}^{\mathrm{l}}|\boldsymbol{\psi}^{\mathrm{l}})\prod_{t}^{T}p(\omega_t^{\mathrm{l}}|L_t,\psi_t^{\mathrm{l}}) &= \prod_{t}^{T}\frac{\exp(\psi_t^{\mathrm{l}})^{D_t}}{\left(1+\exp(\psi_t^{\mathrm{l}})\right)^{L_t}}\cosh^{L_t}(\psi_t^{\mathrm{l}}/2)\exp(-\omega_t^{\mathrm{l}}(\psi_t^{\mathrm{l}})^2/2)p(\omega_t^{\mathrm{l}}) \\
&\propto \prod_{t}^{T}\exp\left(D_t\psi_t^{\mathrm{l}}-\psi_t^{\mathrm{l}}L_t/2-\omega_t^{\mathrm{l}}(\psi_t^{\mathrm{l}})^2/2\right)p(\omega_t^{\mathrm{l}}) \\
&\propto \prod_{t}^{T}\exp\left(\kappa_t^{\mathrm{l}}\psi_t^{\mathrm{l}}-\omega_t^{\mathrm{l}}(\psi_t^{\mathrm{l}})^2/2\right)p(\omega_t^{\mathrm{l}}) \\
&\propto \prod_{t}^{T}\exp\left(-\frac{\omega_t^{\mathrm{l}}}{2}\left(\frac{\kappa_t^{\mathrm{l}}}{\omega_t^{\mathrm{l}}}-\psi_t^{l}\right)^2\right)\exp\left(\frac{\omega_t^{\mathrm{l}}}{2}\left(\frac{\kappa_t^{\mathrm{l}}}{\omega_t^{\mathrm{l}}}\right)^2\right)p(\omega_t^{\mathrm{l}}),
\end{aligned}
$$

where $\kappa_t^{\mathrm{l}} = D_t - L_t/2$.

Similarly, the following holds

$$
p(\boldsymbol{y}^{\mathrm{d}}|\boldsymbol{\psi}^{\mathrm{d}})\prod_{t}^{T}p(\omega_t^{\mathrm{d}}|T_t,\psi_t^{\mathrm{d}}) \propto \prod_{t}^{T}\exp\left(-\frac{\omega_t^{\mathrm{d}}}{2}\left(\frac{\kappa_t^{\mathrm{d}}}{\omega_t^{\mathrm{d}}}-\psi_t^{d}\right)^2\right)\exp\left(\frac{\omega_t^{\mathrm{d}}}{2}\left(\frac{\kappa_t^{\mathrm{d}}}{\omega_t^{\mathrm{d}}}\right)^2\right)p(\omega_t^{\mathrm{d}}),
$$

where $\kappa_t^{\mathrm{d}} = N_t - T_t/2$.

Refer to page 8-9 of Windle et al. (2013) for more details.

C.2 Prior

First, we consider the macro parameters

$$
p(\boldsymbol{A}^{\mathrm{m}},\boldsymbol{\Sigma}^{\mathrm{m}}) \propto \mathrm{iW}(0.01\boldsymbol{I}_{N^{\mathrm{m}}},N^{\mathrm{m}})I(\boldsymbol{A}^{\mathrm{m}}),
$$

where $\boldsymbol{I}_{N^{\mathrm{m}}}$ is an identity matrix of dimension N^{m}. Second, we consider the loadings on the factors for the loan status and LGD components

$$
p(\boldsymbol{\alpha}^{\mathrm{l}},\boldsymbol{\alpha}^{\mathrm{d}}) \propto I(\boldsymbol{\alpha}^{\mathrm{l}},\boldsymbol{\alpha}^{\mathrm{d}}).
$$

Third, we impose priors for the parameters of the LGD component

$$
p(\mu_0,\mu_1,\sigma^2) \propto \mathrm{iG2}(0.01,0.01)I(\mu_0,\mu_1).
$$

Finally, we use a prior for the persistence of the factor

$$p(\phi_{jj}) \propto I(\phi_{jj}) \text{ for all } j.$$

The indicator functions $I(\cdot)$ impose the identification restrictions mentioned in the main text. To be precise, they impose that the first macro variable loads negatively on the macro factor, that the probability of default loads positively on the second default factor and that the probability of a severe loss loads positively on the third loan factor. The functions also impose that $\mu_0 < \mu_1$ and $-1 < \phi_{jj} < 1$ for all j.

C.3 Posterior

C.3.1 Macro component

We collect the terms involving $\boldsymbol{\Sigma}^{\mathrm{m}}$ and $\boldsymbol{A}^{\mathrm{m}}$ from the likelihood and prior and get

$$\begin{aligned} p(\boldsymbol{A}^{\mathrm{m}}, \boldsymbol{\Sigma}^{\mathrm{m}}|\ldots) \propto\; & |\boldsymbol{\Sigma}^{\mathrm{m}}|^{-(T+N^{\mathrm{m}}+N^{\mathrm{m}}+1)/2} \\ & \times \exp\left(-\frac{1}{2}\operatorname{tr}\left((\boldsymbol{\Sigma}^{\mathrm{m}})^{-1}\left[0.01\boldsymbol{I}_{N^{\mathrm{m}}} + (\boldsymbol{Y}^{\mathrm{m}} - \boldsymbol{X}^{\mathrm{m}}\boldsymbol{A}^{\mathrm{m}})'(\boldsymbol{Y}^{\mathrm{m}} - \boldsymbol{X}^{\mathrm{m}}\boldsymbol{A}^{\mathrm{m}})\right]\right)\right) I(\boldsymbol{A}^{\mathrm{m}}) \end{aligned}$$

Using standard results for multivariate regression models we see that $\boldsymbol{A}^{\mathrm{m}}$ can be drawn from a matricvariate normal distribution and $\boldsymbol{\Sigma}^{\mathrm{m}}$ from an inverse Wishart distribution

$$\begin{aligned} p(\boldsymbol{A}^{\mathrm{m}}|\ldots) &= \mathrm{MN}\left((\boldsymbol{X}^{\mathrm{m}\prime}\boldsymbol{X}^{\mathrm{m}})^{-1}(\boldsymbol{X}^{\mathrm{m}\prime}\boldsymbol{Y}^{\mathrm{m}}), \boldsymbol{\Sigma}^{\mathrm{m}} \otimes (\boldsymbol{X}^{\mathrm{m}\prime}\boldsymbol{X}^{\mathrm{m}})^{-1}\right), \\ p(\boldsymbol{\Sigma}^{\mathrm{m}}|\ldots) &= \mathrm{iW}\left(0.01\boldsymbol{I}_{N^{\mathrm{m}}} + (\boldsymbol{Y}^{\mathrm{m}} - \boldsymbol{X}^{\mathrm{m}}\boldsymbol{A}^{\mathrm{m}})'(\boldsymbol{Y}^{\mathrm{m}} - \boldsymbol{X}^{\mathrm{m}}\boldsymbol{A}^{\mathrm{m}}), T + N^{\mathrm{m}}\right), \end{aligned}$$

where we redraw until $\boldsymbol{A}^{\mathrm{m}}$ satisfies the identification restrictions.

C.3.2 Loan status component

Since ω_t^{l} only occurs in its own distribution (and not in the priors or likelihoods), we sample ω_t^{l} from its Pólya-gamma distribution

$$p(\omega_t^{\mathrm{l}}|\ldots) = \mathrm{PG}(L_t, \psi_t^{\mathrm{l}}).$$

We collect the terms involving $\boldsymbol{\alpha}^l$ from the likelihood, prior and latent variable distributions

$$p(\boldsymbol{\alpha}^{\mathrm{l}}|\ldots) \quad\propto\quad \prod_{t}^{T} \exp\left(-\frac{\omega_t^{\mathrm{l}}}{2}\left(\frac{\kappa_t^{\mathrm{l}}}{\omega_t^{\mathrm{l}}}-\psi_t^{l}\right)^2\right)\exp\left(\frac{\omega_t^{\mathrm{l}}}{2}\left(\frac{\kappa_t^{\mathrm{l}}}{\omega_t^{\mathrm{l}}}\right)^2\right)p(\omega_t^{\mathrm{l}})I(\boldsymbol{\alpha}^{\mathrm{l}}),$$

where ψ_t^{l} is a function of $\boldsymbol{\alpha}^{\mathrm{l}}$.

We see that we can interpret the single term in the product as the likelihood of a pseudo data point $\frac{\kappa_t^{\mathrm{l}}}{\omega_t^{\mathrm{l}}}$ drawn from a normal distribution with mean ψ_t^{l} and variance $1/\omega_t^{\mathrm{l}}$ as in Windle et al. (2013). Following Polson et al. (2013) and using the standard results for a linear regression with heteroscedasticity, we simulate $\boldsymbol{\alpha}^{\mathrm{l}}$ from a normal distribution

$$\boldsymbol{\alpha}^{\mathrm{l}} \sim \mathrm{N}(\boldsymbol{m}^{\mathrm{l}}, \boldsymbol{V}^{\mathrm{l}}),$$

where

$$\begin{aligned}\boldsymbol{V}^{\mathrm{l}} &= (\boldsymbol{X}^{\mathrm{l}\prime}\boldsymbol{\Omega}^{\mathrm{l}}\boldsymbol{X}^{\mathrm{l}})^{-1},\\ \boldsymbol{m}^{\mathrm{l}} &= (\boldsymbol{X}^{\mathrm{l}\prime}\boldsymbol{\Omega}^{\mathrm{l}}\boldsymbol{X}^{\mathrm{l}})^{-1}(\boldsymbol{X}^{\mathrm{l}\prime}\boldsymbol{\kappa}^{\mathrm{l}}),\end{aligned}$$

where $\boldsymbol{X}^{\mathrm{l}}$ is a matrix that contains the constant and the relevant factors, where $\boldsymbol{\kappa}^{\mathrm{l}}$ is a vector that collects the elements κ_t^{l} and where $\boldsymbol{\Omega}^{\mathrm{l}}$ is a diagonal matrix with ω_t^{l} as diagonal elements. We redraw $\boldsymbol{\alpha}^{\mathrm{l}}$ until it fulfills the identification restrictions.

C.3.3 LGD component

Using a similar reasoning as above, we sample ω_t^{d} from the Pólya-gamma distribution

$$p(\omega_t^{\mathrm{d}}|\ldots) = \mathrm{PG}(T_t, \psi_t^{\mathrm{d}})$$

and sample $\boldsymbol{\alpha}^{\mathrm{d}}$ from a normal distribution

$$\boldsymbol{\alpha}^{\mathrm{d}} \sim \mathrm{N}(\boldsymbol{m}^{\mathrm{d}}, \boldsymbol{V}^{\mathrm{d}}),$$

where

$$\begin{aligned}\boldsymbol{V}^{\mathrm{d}} &= (\boldsymbol{X}^{\mathrm{d}\prime}\boldsymbol{\Omega}^{\mathrm{d}}\boldsymbol{X}^{\mathrm{d}})^{-1},\\ \boldsymbol{m}^{\mathrm{d}} &= (\boldsymbol{X}^{\mathrm{d}\prime}\boldsymbol{\Omega}^{\mathrm{d}}\boldsymbol{X}^{\mathrm{d}})^{-1}(\boldsymbol{X}^{\mathrm{d}\prime}\boldsymbol{\kappa}^{\mathrm{d}}),\end{aligned}$$

where $\boldsymbol{X}^{\mathrm{d}}$ is a matrix that contains the constant and the relevant factors, where $\boldsymbol{\kappa}^{\mathrm{d}}$ is a vector that collects the elements κ_t^{d} for all t and where $\boldsymbol{\Omega}^{\mathrm{d}}$ is a diagonal matrix with ω_t^{d} as diagonal elements. We redraw $\boldsymbol{\alpha}^{\mathrm{d}}$ until it fulfills the identification restrictions.

We collect the terms involving s_{it} in the prior and likelihood and obtain

$$p(s_{it}|\ldots) \propto (p_t^{\mathrm{d}})^{s_{it}}(1-p_t^{\mathrm{d}})^{(1-s_{it})} \exp\left(-\frac{1}{2\sigma^2}(y_{it}^{\mathrm{d}} - \mu_0(1-s_{it}) - \mu_1 s_{it})^2\right).$$

Hence we can sample s_{it} from the following Bernoulli distribution

$$p(s_{it} = 1|\ldots) = \frac{p_t^{\mathrm{d}}\mathrm{N}(\mu_1, \sigma^2)}{(1-p_t^{\mathrm{d}})\mathrm{N}(\mu_0, \sigma^2) + p_t^{\mathrm{d}}\mathrm{N}(\mu_1, \sigma^2)}.$$

We collect the terms involving μ_1 from the posterior and get

$$p(\mu_1|\ldots) \propto \prod_{i,t}^{N,T} \sigma^{-1} \exp\left(-\frac{1}{2\sigma^2}(y_{it}^{\mathrm{d}} - \mu_1 s_{it})^2\right).$$

Hence, we can sample μ_1 (and μ_0 as well) from normal distributions

$$\begin{aligned} p(\mu_1|\ldots) &= \mathrm{N}\left(\bar{y}_1, \frac{\sigma^2}{N_1}\right), \\ p(\mu_0|\ldots) &= \mathrm{N}\left(\bar{y}_0, \frac{\sigma^2}{N_0}\right), \end{aligned}$$

where $\bar{y}_1$ is the sample mean for the N_1 observations with a latent indicator of 1 and where $\bar{y}_0$ is the sample mean for the N_0 observations with a latent indicator of 0. We redraw μ_0 and μ_1 from their posterior distributions until they fulfill the identification restrictions.

We collect the terms involving σ^2 from the prior and likelihood and get

$$p(\sigma^2|\ldots) \propto \sigma^{-\frac{NT+0.01+2}{2}} \exp\left(-\frac{1}{2\sigma^2}\left(0.01 + \sum_{it}^{NT}(y_{it}^{\mathrm{d}} - \mu_0(1-s_{it}) - \mu_1 s_{it})^2\right)\right),$$

which means that we can sample σ^2 from an inverse gamma-2 distribution

$$p(\sigma^2|\ldots) = \mathrm{iG2}\left(0.01 + \sum_{it}^{NT}(y_{it}^{\mathrm{d}} - \mu_0(1-s_{it}) - \mu_1 s_{it})^2, 0.01 + NT\right).$$

C.3.4 Factor component

If we collect the terms involving the factors from the prior, likelihood and latent variable distributions, we see that we have a linear Gaussian state space model as in Windle et al. (2013).

Our model has the following transition equation

$$\boldsymbol{f}_{t+1} = \boldsymbol{\Phi} \boldsymbol{f}_t + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega}),$$

where $\boldsymbol{\Omega} = \boldsymbol{I} - \boldsymbol{\Phi}\boldsymbol{\Phi}'$ because of the restriction on the unconditional covariance matrix.

We obtain the following observation equations

$$\begin{aligned}
\boldsymbol{y}_t^{\mathrm{m}} &= \boldsymbol{\alpha}^{\mathrm{m}} + \boldsymbol{B}^{\mathrm{m}} \boldsymbol{f}_t^{\mathrm{m}} + \boldsymbol{\varepsilon}_t, & \boldsymbol{\varepsilon}_t &\sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}^{\mathrm{m}}), \\
\frac{\kappa_t^{\mathrm{l}}}{\omega_t^{\mathrm{l}}} &= \alpha_i^{\mathrm{l}} + \boldsymbol{\beta}_i^{\mathrm{l}\prime} \boldsymbol{f}_t^{\mathrm{m}} + \boldsymbol{\gamma}_i^{\mathrm{l}\prime} \boldsymbol{f}_t^{\mathrm{l}} + \zeta_t^{\mathrm{l}}, & \zeta_t^{\mathrm{l}} &\sim \mathrm{N}(0, 1/\omega_t^{\mathrm{l}}), \\
\frac{\kappa_t^{\mathrm{d}}}{\omega_t^{\mathrm{d}}} &= \alpha_i^{\mathrm{d}} + \boldsymbol{\beta}_i^{\mathrm{d}\prime} \boldsymbol{f}_t^{\mathrm{m}} + \boldsymbol{\gamma}_i^{\mathrm{d}\prime} \boldsymbol{f}_t^{\mathrm{l}} + \boldsymbol{\delta}_i^{\mathrm{d}\prime} \boldsymbol{f}_t^{\mathrm{d}} + \zeta_t^{\mathrm{d}}, & \zeta_t^{\mathrm{d}} &\sim \mathrm{N}(0, 1/\omega_t^{\mathrm{d}}),
\end{aligned}$$

where the second and third observations are pseudo data points as explained in the derivations of $\boldsymbol{\alpha}^{\mathrm{l}}$ and $\boldsymbol{\alpha}^{\mathrm{d}}$. We sample the latent factor $\boldsymbol{f}_t$ using the simulation smoother of Durbin and Koopman (2002).

We collect the terms involving ϕ_{jj} from the equation for the latent factor and obtain

$$p(\phi_{jj}|\ldots) \propto (1-\phi_{jj}^2)^{-1/2} \exp\left(-\frac{1}{2(1-\phi_{jj}^2)} (f_{j,t+1} - \phi_{jj} f_{jt})^2 \right), \text{ for all } j.$$

Since ϕ_{jj} occurs in both the mean and variance, we cannot derive the posterior analytically. We use a Metropolis-Hastings step instead where we use a normal distribution as proposal density and where we calculate the acceptance probability in the usual way for the independence Metropolis-Hastings sampler. To calculate the mean and variance of the proposal density we maximize the log PDF of the above expression and use the mode as mean and the inverse of the negative Hessian as the covariance matrix. We redraw from the proposal density until the proposed draw satisfies the identification restriction.

C.4 Extensions

We consider a couple of alternative models in the main paper and appendix.

First, we consider models with loan and default characteristics that affect the probability of default and the LGD of a loan. It is straightforward to extend our method to allow for these differences. The only difference is that we need to draw all parameters and latent variables except for $\boldsymbol{f}_t$, ϕ_{jj}, $\boldsymbol{A}^{\mathrm{m}}$ and $\boldsymbol{\Sigma}^{\mathrm{m}}$ per group of characteristics.

Second, we consider a model with a mixture of two skewed student-t distributions instead of a mixture of normals. The only difference is that we draw the parameters in the LGD component (except for ω_t^{d} and $\boldsymbol{\alpha}^{\mathrm{d}}$) based on the derivations in Frühwirth-Schnatter and Pyne (2010). Please refer to the online appendix of Frühwirth-Schnatter and Pyne (2010) for more details on the conditional posterior distributions.

D Results for models with loan and default characteristics

[Figure D.1 about here.]

[Figure D.2 about here.]

[Figure D.3 about here.]

[Table D.I about here.]

[Table D.II about here.]

[Table D.III about here.]

[Table D.III (continued) about here.]

Table D.I: Parameter estimates, model with seniority

	Macro and loan factor		Macro and two loan factors	
Panel A: Factor				
ϕ_{11}	0.813	(0.120)	0.803	(0.115)
ϕ_{22}	0.775	(0.138)	0.720	(0.231)
ϕ_{33}			0.477	(0.309)
Panel B: Macro variables				
α_{GDP}	0.241	(0.631)	-0.035	(0.418)
β_{GDP}	-0.935	(0.694)	-0.843	(0.487)
α_{IP}	0.187	(0.518)	-0.027	(0.346)
β_{IP}	-0.725	(0.583)	-0.633	(0.411)
α_{UR}	-0.207	(0.556)	0.031	(0.388)
β_{UR}	0.792	(0.612)	0.756	(0.456)
WAIC2	129.4		123.3	
Panel C: Loan status				
α^{l}	-5.665	(0.233)	-5.780	(0.175)
β^{l}	-0.179	(0.199)	-0.174	(0.320)
γ_1^{l}	0.628	(0.354)	0.574	(0.292)
av. p^{l} ($\times 0.01$)	0.309	(0.002)	0.309	(0.002)
m.e. of f^{m}	-0.055	(0.061)	-0.054	(0.099)
m.e. of f_1^{l}	0.193	(0.109)	0.177	(0.090)
WAIC2	321.3		320.0	

Panel D: Loss given default

	Macro and loan factor: Sen. secured		Sen. unsecured		Macro and two loan factors: Sen. secured		Sen. unsecured	
μ_0	0.070	(0.002)	0.072	(0.001)	0.070	(0.002)	0.072	(0.001)
μ_1	0.766	(0.004)	0.858	(0.003)	0.766	(0.004)	0.858	(0.003)
σ	0.129	(0.001)	0.130	(0.001)	0.129	(0.001)	0.130	(0.001)
α^{d}	-2.174	(0.580)	-1.665	(0.250)	-1.930	(0.393)	-1.542	(0.315)
β^{d}	0.900	(0.619)	0.500	(0.320)	0.849	(0.450)	0.385	(0.247)
γ_1^{d}	0.041	(0.055)	-0.216	(0.133)	-0.037	(0.050)	0.146	(0.126)
γ_2^{d}							-0.257	(0.240)
av. p^{d}	0.151	(0.004)	0.191	(0.004)	0.151	(0.004)	0.191	(0.004)
m.e. of f^{m}	0.103	(0.070)	0.075	(0.048)	0.097	(0.051)	0.058	(0.037)
m.e. of f_1^{l}	0.005	(0.006)	-0.033	(0.020)	-0.004	(0.006)	0.022	(0.019)
m.e. of f_2^{l}							-0.039	(0.036)
av. LGD	0.175	(0.003)	0.222	(0.003)	0.175	(0.003)	0.222	(0.003)
m.e. of f^{m}	0.072	(0.049)	0.059	(0.038)	0.067	(0.036)	0.046	(0.029)
m.e. of f_1^{l}	0.003	(0.004)	-0.026	(0.016)	-0.003	(0.004)	0.017	(0.015)
m.e. of f_2^{l}							-0.030	(0.028)
WAIC2	$-11{,}645.0$		$-14{,}438.7$		$-11{,}648.0$		$-14{,}434.2$	

This table presents the posterior mean and standard deviation (in parentheses) of the parameters of the model in section 3.1 for different loan and default characteristics. We report results for specifications with a macro and a loan factor, and an additional loan factors. The horizontal panels correspond with the panels in table II. We distinguish defaults of senior secured and senior unsecured loans. The characteristics are included as dummy variables.

Table D.II: Parameter estimates, model with borrower size

	Macro and loan factor				Macro and two loan factors			
	Panel A: Factor							
ϕ_{11}	0.845	(0.098)			0.861	(0.090)		
ϕ_{22}	0.510	(0.217)			0.245	(0.262)		
ϕ_{33}					0.618	(0.190)		
	Panel B: Macro variables							
α_{GDP}	-0.433	(0.628)			-0.627	(0.492)		
β_{GDP}	-1.225	(0.885)			-1.265	(0.886)		
α_{IP}	-0.362	(0.543)			-0.535	(0.440)		
β_{IP}	-1.025	(0.756)			-1.083	(0.790)		
α_{UR}	0.385	(0.572)			0.568	(0.455)		
β_{UR}	1.087	(0.814)			1.149	(0.782)		
WAIC2	114.9				109.7			
	Panel C: Loan status							
	Large corp.		SME		Large corp.		SME	
α^{l}	−6.441	(0.174)	−5.533	(0.237)	−6.394	(0.127)	−5.474	(0.132)
β^{l}	0.404	(0.273)	0.559	(0.383)	0.335	(0.256)	0.426	(0.323)
γ_1^{d}	0.231	(0.101)	0.366	(0.143)	0.303	(0.088)	0.168	(0.080)
γ_2^{d}							0.392	(0.160)
av. p^{l} ($\times 0.01$)	0.143	(0.003)	0.347	(0.002)	0.147	(0.004)	0.346	(0.002)
m.e. of f^{m}	0.058	(0.039)	0.193	(0.132)	0.049	(0.038)	0.147	(0.111)
m.e. of f_1^{l}	0.033	(0.014)	0.126	(0.049)	0.045	(0.013)	0.058	(0.028)
m.e. of f_2^{l}							0.135	(0.055)
WAIC2	349.9		333.4		252.1		319.2	
	Panel D: Loss given default							
	Large corp.		SME		Large corp.		SME	
μ_0	0.075	(0.002)	0.062	(0.001)	0.075	(0.002)	0.062	(0.001)
μ_1	0.848	(0.005)	0.849	(0.003)	0.849	(0.005)	0.849	(0.003)
σ	0.126	(0.001)	0.124	(0.001)	0.126	(0.001)	0.124	(0.001)
α^{d}	−1.707	(0.291)	−1.564	(0.269)	−1.630	(0.268)	−1.517	(0.270)
β^{d}	0.293	(0.302)	0.243	(0.272)	0.333	(0.389)	0.348	(0.295)
γ_1^{d}	−0.404	(0.173)	−0.406	(0.167)	−0.490	(0.143)	−0.186	(0.110)
γ_2^{d}							−0.370	(0.159)
av. p^{d}	0.160	(0.005)	0.166	(0.003)	0.160	(0.005)	0.166	(0.003)
m.e. of f^{m}	0.036	(0.037)	0.033	(0.037)	0.042	(0.049)	0.048	(0.041)
m.e. of f_1^{l}	−0.050	(0.022)	−0.056	(0.023)	−0.062	(0.018)	−0.026	(0.015)
m.e. of f_2^{l}							−0.051	(0.022)
av. LGD	0.199	(0.004)	0.193	(0.003)	0.199	(0.004)	0.193	(0.003)
m.e. of f^{m}	0.028	(0.029)	0.026	(0.030)	0.033	(0.038)	0.038	(0.032)
m.e. of f_1^{l}	−0.039	(0.017)	−0.044	(0.018)	−0.048	(0.014)	−0.020	(0.012)
m.e. of f_2^{l}							−0.040	(0.017)
WAIC2	−8,207.3		−15,655.8		−8,194.0		−15,661.0	

This table presents the posterior mean and standard deviation (in parentheses) of the parameters of the model in section 3.1 for different loan and default characteristics. We report results for specifications with a macro and a loan factor, and an additional loan factors. The horizontal panels correspond with the panels in table II. We distinguish loans and defaults of large corporates and SMEs. The characteristics are included as dummy variables.

Table D.III: Parameter estimates, model with industries

	Macro and loan factor						Macro and three loan factors					
Panel A: Factor												
ϕ_{11}			0.910	(0.060)					0.882	(0.071)		
ϕ_{22}			0.221	(0.269)					−0.052	(0.300)		
ϕ_{33}									0.260	(0.272)		
ϕ_{44}									−0.274	(0.328)		
Panel B: Macro variables												
α_{GDP}			0.526	(0.625)					−0.412	(0.331)		
β_{GDP}			−1.299	(0.880)					−1.136	(0.675)		
α_{IP}			0.430	(0.532)					−0.336	(0.299)		
β_{IP}			−1.062	(0.753)					−0.928	(0.579)		
α_{UR}			−0.490	(0.589)					0.376	(0.316)		
β_{UR}			1.206	(0.832)					1.037	(0.632)		
WAIC2			100.0						97.7			
Panel C: Loan status												
	Cons. staples		Industrials		Financials		Cons. staples		Industrials		Financials	
α^{l}	−6.118	(0.173)	−5.929	(0.302)	−6.990	(0.377)	−5.884	(0.075)	−5.388	(0.140)	−6.328	(0.184)
β^{l}	0.203	(0.293)	0.602	(0.499)	0.787	(0.600)	0.202	(0.180)	0.545	(0.361)	0.717	(0.457)
γ_1^{d}	0.362	(0.103)	0.364	(0.106)	0.342	(0.108)	0.405	(0.087)	0.297	(0.089)	0.272	(0.109)
γ_2^{d}									0.233	(0.075)	0.216	(0.099)
γ_3^{d}											0.173	(0.048)
av. p^{l} ($\times 0.01$)	0.273	(0.005)	0.414	(0.005)	0.162	(0.004)	0.274	(0.005)	0.415	(0.005)	0.162	(0.004)
m.e. of f^{m}	0.055	(0.080)	0.248	(0.205)	0.128	(0.097)	0.055	(0.049)	0.225	(0.149)	0.116	(0.074)
m.e. of f_1^{l}	0.099	(0.028)	0.150	(0.044)	0.055	(0.018)	0.110	(0.024)	0.123	(0.037)	0.044	(0.018)
m.e. of f_2^{l}									0.096	(0.031)	0.035	(0.016)
m.e. of f_3^{l}											0.028	(0.008)
WAIC2	350.7		321.9		281.9		268.3		293.2		241.9	

This table presents the posterior mean and standard deviation (in parentheses) of the parameters of the model in section 3.1 for different loan and default characteristics. We report results for specifications with a macro and a loan factor, and an additional two loan factors. The horizontal panels correspond with the panels in table II. We distinguish loans and defaults of borrowers in the sectors Consumer Staples, Industrials, and Financials. The characteristics are included as dummy variables. The number of observations is 14,997.

Table D.III: Parameter estimates, model with industries – ***Continued***

	Macro and loan factor						Macro and three loan factors					
	Panel D: Loss given default											
	Cons. staples		Industrials		Financials		Cons. staples		Industrials		Financials	
μ_0	0.056	(0.002)	0.056	(0.002)	0.085	(0.003)	0.056	(0.002)	0.056	(0.002)	0.085	(0.003)
μ_1	0.851	(0.006)	0.837	(0.004)	0.796	(0.006)	0.851	(0.006)	0.837	(0.004)	0.796	(0.006)
σ	0.121	(0.002)	0.119	(0.001)	0.145	(0.002)	0.121	(0.002)	0.119	(0.001)	0.145	(0.002)
α^{d}	−1.735	(0.265)	−1.786	(0.280)	−2.025	(0.465)	−1.516	(0.141)	−1.690	(0.200)	−1.578	(0.255)
β^{d}	0.401	(0.314)	0.327	(0.335)	0.793	(0.549)	0.351	(0.250)	0.256	(0.296)	0.664	(0.443)
γ_1^{d}	−0.327	(0.123)	−0.472	(0.147)	−0.492	(0.171)	−0.293	(0.103)	−0.238	(0.137)	−0.350	(0.167)
γ_2^{d}									−0.490	(0.149)	−0.215	(0.158)
γ_3^{d}											−0.313	(0.099)
av. p^{d}	0.163	(0.007)	0.156	(0.004)	0.185	(0.006)	0.163	(0.007)	0.157	(0.004)	0.185	(0.006)
m.e. of f^{m}	0.055	(0.043)	0.041	(0.042)	0.099	(0.068)	0.048	(0.034)	0.033	(0.038)	0.084	(0.056)
m.e. of f_1^{l}	−0.045	(0.017)	−0.060	(0.019)	−0.061	(0.022)	−0.040	(0.014)	−0.030	(0.017)	−0.044	(0.021)
m.e. of f_2^{l}									−0.063	(0.019)	−0.027	(0.020)
m.e. of f_3^{l}											−0.040	(0.013)
av. LGD	0.186	(0.006)	0.178	(0.004)	0.217	(0.004)	0.186	(0.006)	0.178	(0.004)	0.217	(0.004)
m.e. of f^{m}	0.043	(0.034)	0.032	(0.033)	0.070	(0.049)	0.038	(0.027)	0.025	(0.029)	0.060	(0.040)
m.e. of f_1^{l}	−0.035	(0.014)	−0.047	(0.015)	−0.044	(0.015)	−0.032	(0.011)	−0.024	(0.014)	−0.032	(0.015)
m.e. of f_2^{l}									−0.049	(0.015)	−0.019	(0.014)
m.e. of f_3^{l}											−0.028	(0.009)
WAIC2	−4,361.6		−9,569.7		−4,385.9		−4,361.0		−9,574.1		−4,382.3	

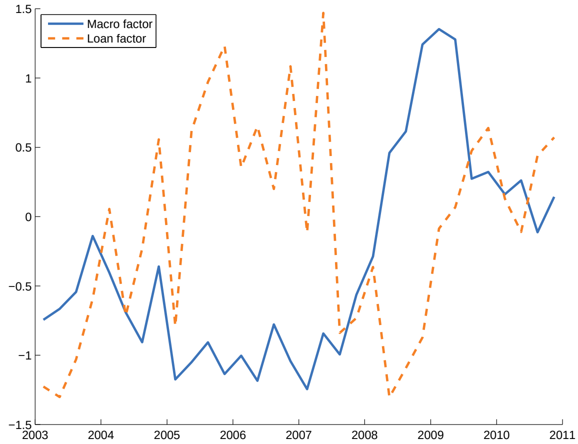
See table note on previous page.

Figure D.1: Latent factors, seniority

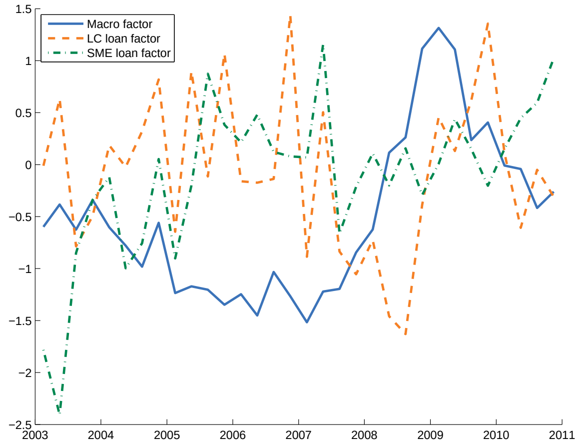
The figures show the smoothed latent factors for the model defined in section 3.1. We report results for specifications with a macro and a loan factor (a), and an additional loan factor (b). We distinguish defaults of senior secured and senior unsecured loans. The characteristics are included as dummy variables.

Figure D.2: Latent factors, borrower size

(a) Two factors



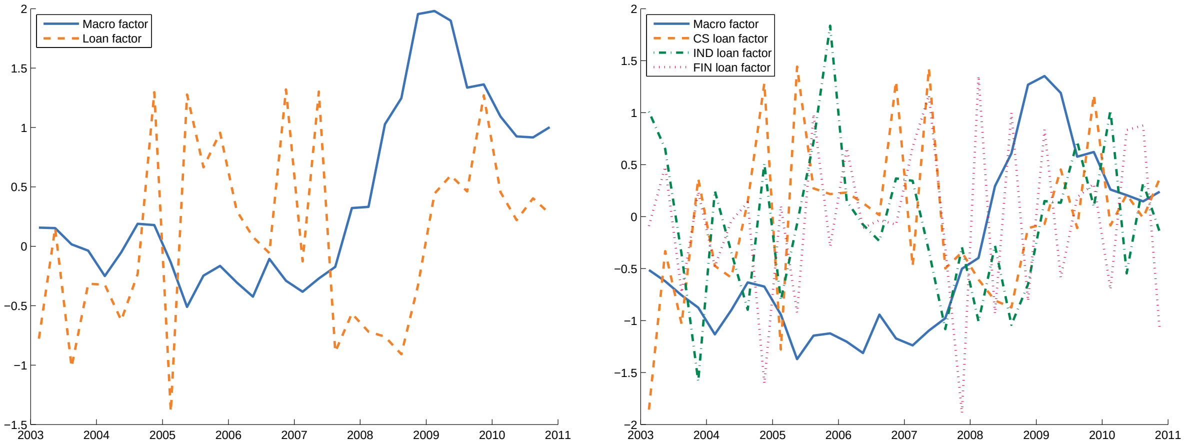
(b) Three factors



The figures show the smoothed latent factors for the model defined in section 3.1. We report results for specifications with a macro and a loan factor (a), and an additional loan factor (b). We distinguish loans and defaults of large corporates and SMEs. The characteristics are included as dummy variables.

Figure D.3: Latent factors, industry

(a) Two factors (b) Four factors



The figures show the smoothed latent factors for the model defined in section 3.1. We report results for specifications with a macro and a loan factor (a), and an additional two loan factors (b). We distinguish loans and defaults of borrowers in the sectors Consumer Staples, Industrials, and Financials. The characteristics are included as dummy variables.

E Results for model with mixture of skew-t

In section 4.4, we discuss replacing the LGD component's mixture of normals with a mixture of skewed Student's t, or skew-t, distributions.

Following Azzalini and Capitanio (2003), if a random variable X is skew-t distributed, $X \sim St(\xi, \omega, \alpha, \nu)$, then

$$f(x; \xi, \omega, \alpha, \nu) = \frac{2}{\omega} t_\nu(q_x) T_{\nu+1}\left(\alpha q_x \sqrt{\frac{\nu+1}{q_x^2+\nu}}\right), \tag{E.1}$$

with location parameter ξ, scale parameter ω, shape parameter α, and degrees of freedom ν, where $q_x = (x-\xi)/\omega$, and t_ν and T_ν the PDF and CDF of a standard Student's t distribution with ν degrees of freedom.

The expected value of X is not ξ, but is shifted to the left or right, depending on the skewness. It is given by

$$\mathrm{E}[X] = \xi + \omega\delta\sqrt{\frac{\nu}{\pi}}\frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}, \tag{E.2}$$

where $\delta = \alpha/\sqrt{1+\alpha^2}$. The normal distribution is nested, by setting $\alpha = 0$ and $\nu \to \infty$.

For the Bayesian estimation procedure, we follow the reparametrization by Frühwirth-Schnatter and Pyne (2010), $g(x; \boldsymbol{\theta}) = f(x; \xi, \omega, \alpha, \nu)$, with parameter vector $\boldsymbol{\theta} = (\xi, \sigma, \psi, \nu)$, where $\sigma^2 = \omega^2(1-\delta^2)$ and $\psi = \omega\delta$. The estimates for the model with the mixture of skew-t distributions are presented in table E.I. Plugging in the posterior mean of $\boldsymbol{\theta}$, we find that the mean of the mild (severe) loss is 0.090 (0.605).

The PDF of a mixture is given by the sum of the PDFs of the mixture components, weighted by the mixture probability, or mathematically, $h(x; p_t^{\mathrm{d}}, \boldsymbol{\theta}_0, \boldsymbol{\theta}_1) = (1-p_t^{\mathrm{d}})g(x; \boldsymbol{\theta}_0) + p_t^{\mathrm{d}} g(x; \boldsymbol{\theta}_1)$. From this, we can calculate the posterior mixture probabilities. The likelihood that, conditional on the latent factors $\boldsymbol{f}_t$, observation y_{it}^{d} is a severe loss is

$$\Pr[s_{it} = 1 | y_{it}^{\mathrm{d}}, \boldsymbol{f}_t] = \frac{p_t^{\mathrm{d}} g(x; \boldsymbol{\theta}_1)}{(1-p_t^{\mathrm{d}})g(x; \boldsymbol{\theta}_0) + p_t^{\mathrm{d}} g(x; \boldsymbol{\theta}_1)}, \tag{E.3}$$

and $\Pr[s_{it} = 0 | y_{it}^{\mathrm{d}}, \boldsymbol{f}_t] = 1 - \Pr[s_{it} = 1 | y_{it}^{\mathrm{d}}, \boldsymbol{f}_t]$. The mixture fit in figure 6 and the posterior mixture probabilities in figure E.2 are based on the posterior means of $\boldsymbol{\theta}$ and of the implied probability of a severe loss p_t^{d}.

[Table E.I about here.]

[Figure E.1 about here.]

[Figure E.2 about here.]

Table E.I: Parameter estimates, mixture of skew-t

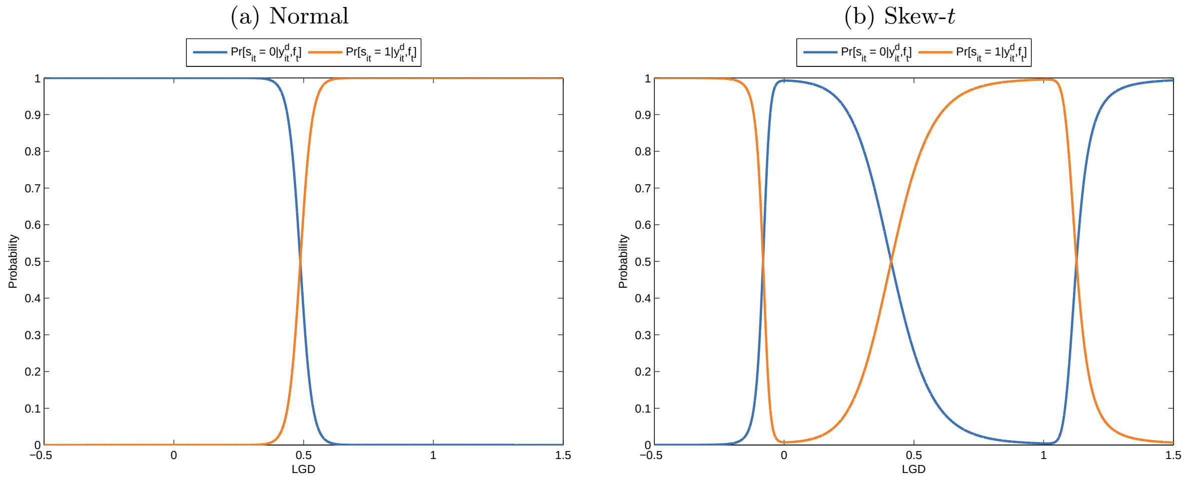
	Mixture of normals		Mixture of skew-t	
		Panel A: Factor		
ϕ_{11}	0.856	(0.078)	0.878	(0.075)
ϕ_{22}	0.322	(0.207)	0.230	(0.235)
		Panel B: Macro variables		
α_{GDP}	−0.008	(0.317)	−0.342	(0.355)
β_{GDP}	−0.969	(0.531)	−1.033	(0.611)
α_{IP}	−0.007	(0.284)	−0.273	(0.313)
β_{IP}	−0.802	(0.465)	−0.824	(0.517)
α_{UR}	0.008	(0.298)	0.341	(0.355)
β_{UR}	0.879	(0.491)	1.032	(0.617)
		Panel C: Loan status		
α^{l}	−5.850	(0.097)	−5.700	(0.118)
β^{l}	0.371	(0.201)	0.428	(0.260)
γ^{l}	0.258	(0.058)	0.254	(0.059)
av. p^{l} ($\times 10^{-2}$)	0.309	(0.002)	0.310	(0.002)
m.e. of f^{m}	0.114	(0.062)	0.132	(0.080)
m.e. of f^{l}	0.080	(0.018)	0.078	(0.018)
		Panel D: Loss given default		
ξ_0	0.072	(0.001)	−0.024	(0.001)
ξ_1	0.829	(0.002)	1.041	(0.002)
σ	0.131	(0.001)	0.017	(0.001)
ψ_0			0.119	(0.003)
ψ_1			−0.399	(0.008)
ν_0			4.891	(0.172)
ν_1			3.065	(0.143)
α^{d}	−1.643	(0.159)	−1.242	(0.187)
β^{d}	0.328	(0.202)	0.399	(0.264)
γ^{d}	−0.293	(0.075)	−0.308	(0.104)
av. p^{d}	0.174	(0.003)	0.220	(0.004)
m.e. of f^{m}	0.045	(0.028)	0.064	(0.043)
m.e. of f^{l}	−0.040	(0.010)	−0.050	(0.017)
av. LGD	0.204	(0.002)		
m.e. of f^{m}	0.034	(0.021)		
m.e. of f^{l}	−0.030	(0.008)		

This table presents the posterior mean and standard deviation (in parentheses) of the parameters of the model in section 3.1 with different distributions for the LGD component. All model specifications have a macro and a loan factor. The specifications do not include loan or default characteristics. The horizontal panels correspond with the panels in table II. The LGD type can be either mild or severe. In the left column, conditional on the type, the LGD follows a normal distribution with mean ξ_0 or ξ_1 (corresponding to μ_0 and μ_1 in table II), and volatility σ. In the right column, conditional on the type, the LGD follows a skew-t distribution with location parameter ξ_0 or ξ_1, scale parameter σ, shape parameter ψ_0 or ψ_1 and degrees of freedom ν_0 or ν_1.

Figure E.1: Latent factors, skew-t

The figure shows the latent factors of the two-factor model without cross-sectional variation, with the LGD distributed as a mixture of normals or mixture of skew-t.

Figure E.2: Smoothed posterior state probabilities



The figures present the smoothed state probabilities $\Pr[s_{it} = 0|y^{\mathrm{d}}_{it}, \boldsymbol{f}_t]$ (blue line) and $\Pr[s_{it} = 1|y^{\mathrm{d}}_{it}, \boldsymbol{f}_t]$ (orange line) for the two-factor model , with a mixture of normals (a) and with a mixture of skew-t distributions (b) for the LGD. The mixture probabilities are for the first quarter of 2003 and based on the posterior mean estimates of the parameters of the two-factor model without cross-sectional variation.

References

Altman, E. I. and Kalotay, E. A. (2014). Ultimate recovery mixtures. *Journal of Banking & Finance*, 40:116–129.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):367–389.

Bastos, J. A. (2014). Ensemble predictions of recovery rates. *Journal of Financial Services Research*, 46(2):177–193.

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

Frühwirth-Schnatter, S. and Pyne, S. (2010). Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions. *Biostatistics*, 11(2):317–336.

Höcht, S. and Zagst, R. (2007). Loan Recovery Determinants - A Pan-European Study. *Working Paper*.

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

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