JEL Classification: G32
Keywords: credit risk, mortgage, loan portfolio, dynamic model, cointegration

Modeling Credit Losses for Multiple Loan Portfolios*

Petr Gapko – Econometric Department, Institute of Information Theory and Automation, Czech Academy of Sciences, (petr.gapko@seznam.cz) *corresponding author*

Martin Šmíd - Econometric Department, Institute of Information Theory and Automation, Czech Academy of Sciences

Abstract

We propose a dynamic structural model of credit risk of multiple loan portfolios. In line with Merton, Vasicek and Pykhtin, we assume that a loan defaults if the assets of the debtor fall below his liabilities, and the subsequent loss is determined by the collateral value. For each loan, the assets, liabilities and the collateral value each depends on a common and an individual factor. By applying our model to two nationwide United States loan portfolios with real estate collateral, we demonstrate its considerable predicting power and show that, similarly to calculations under prudential regulation, it can be used within financial institutions to measure credit risk under various macroeconomic situations and different probability levels. This makes the model usable for quantification of loan loss allowances under IFRS9[1] or for stress tests of credit risk.

1. Introduction

At the end of the last decade, when the financial crisis fully hit the United States (US) economy, losses from real estate loans in the US increased ten times, compared with the period of economic growth ending in 2007.[2] A natural challenge is to study the causes of crises of this kind to avoid them in the future, namely to construct realistic and practically usable models of credit risk. No doubt that these models have to consider the interconnectedness of different sources of credit risk and the macroeconomic environment. The aim of this paper is to introduce such a model, describing default rates (usually used to predict the probability of default, PD) and losses given default (LGD) of multiple loan portfolios secured by collaterals. To the best of our knowledge, there is a very limited amount of literature on the specific combination of model features proposed in our paper (i.e. combining PD, LGD, macroeconomic factors, their interconnectedness and dynamics).

The model we propose is a structural factor one, based on the approaches of Merton (1974), the author of the first factor model, Vasicek O. A. (2015a; 2015b), who first described the distribution of default rate within the model, and Pykhtin (2003), who incorporated the collateral value. Although there are different possible ways of building factor models, see, for instance, Frye (2000), Jimenez & Mencia (2009),

*The support from the Czech Science Foundation, grant No. 15-10331S, is kindly acknowledged.

[1] IFRS 9 is an International Financial Reporting Standard (IFRS) addressing the accounting for financial instruments

[2] According to the delinquency and charge-off rates published by the United States Federal Reserve System (FED)

Frontczak & Rostek (2015) or Witzany (2011), we stick to the Merton-Vasicek-Pykhtin concept of explaining defaults and LGD at micro level. In particular, the default happens when resources fall below a certain threshold, and the subsequent loss derives itself from the collateral price. Specifically, we assume that, for each loan, the log-assets, log-liabilities and log-collateral price are each a sum of an individual and a common factor. As, in a large portfolio, the individual factors diversify out, the PDs and LGDs depend only on the common factors, the dynamic of which is described by a linear vector model.

Multiple homogenous loan portfolios may be simultaneously treated in our model. The homogeneity of a portfolio means that the (joint) distributions of the assets, liabilities, and the collateral prices of individual loans are identical. The distributions may, however, differ between portfolios, each depending on its own common factors and having its own variability (standard deviations of the individual factors).

Our model is easily manageable in the sense that all its parameters except the variance of the collateral individual factor are estimated jointly. Moreover, analytical formulas exist for distributions of PD and LGD forecasts and confidence intervals, which, among other things, enables to express unbiased point forecasts analytically for both PDs and LGDs.

The model is general in the sense that the common factors may follow an arbitrary Vector Autoregression (VAR).[3] In the empirical part, we use a Vector Error Correction Model (VECM) style estimation, because it's still general and capable of handling integrated variables, which the factors themselves appear to be.

Last but not least, our model has a rather surprising theoretical implication which is inter-portfolio diversification of idiosyncratic risk. To be specific, there is no difference whether all the individual factors are unrelated or whether individual factors of loans are correlated between portfolios; the latter could happen, for instance, when people have mortgages and credit cards from the same bank. To be more exact, the losses of the portfolios may be stochastically dependent because of the dependence of their common factors, but the joint distribution of the losses would be the same, whether the individual factors of the loans are correlated or not.

We demonstrate the applicability of our approach by applying it to the nationwide residential and commercial US real estate loan portfolios. Several US macroeconomic indicators serve as explanatory variables, namely GDP, commercial and residential house price indices, FED base interest rate and unemployment. In explaining PDs and LGDs of both portfolios, our model is statistically significant. Notably, it predicts PDs significantly better than the LGDs, the factors of which are, on the other hand, cointegrated with the macro variables. This suggests that both PDs and LGDs are linked to the macroeconomic environment, each, however, in a different way: while there is some inertia in default rate, the severity of losses is interconnected with the current state of economics.

Not surprisingly, the losses of both the examined portfolios are highly related. They display similar trends over time and the correlations between the increments of their underlying factors are positive. Moreover, they are interconnected beyond explanation by their history and the exogenous variables. To demonstrate this, we show

[3] In principle, any linear model with normal distribution of forecasts may be used, and the requirement of normality may be relaxed in exchange for loosing analytical formulas for PD and LGD forecasts.

that the full model is highly statistically significant in comparison with its version without mutual relations of the portfolios.

Finally, we show that the amount of hypothetical economic capital recommended by our model for both portfolios can be compared to that resulting from the Internal Rating Based (IRB) regulatory approach, which allows to use the model as an internal credit risk model for calculation of loan loss allowances, economic capital and stressed levels of credit risk under adverse macroeconomic conditions within the Internal Capital Adequacy Assessment Process (ICAAP).

There are many different works modelling interconnectedness of credit risk and macroeconomic environment. For instance, Hamerle et al. (2011) showed, studying a bond portfolio, that it is necessary to consider changes in the macroeconomic environment, and they compare their point-in-time multi-factor credit risk model with the usual through-the-cycle approach. Similarly, Sommar & Shahnazarian (2009) used the vector error correction model to estimate the dependency of expected default frequency of a portfolio of nonfinancial listed companies on several macroeconomic factors, from which they found the interest rate to be the most influential. The influence of the interest rate is in line with the findings of Virolainen (2004), who also model a dependency of the credit risk on the key macroeconomic variables. None of these models, however, simultaneously treats PDs and LGDs of more portfolios.

The model closest to ours is that of Pesaran et al. (2006), which also treats losses of multiple portfolios (of corporate loans in their case) in dependence on a macroeconomic Global Vector Autoregressive (GVAR) model. Similar to us, a loan defaults if the log-value of a debtor's assets falls below a certain threshold. In contrary to us, LGD is constant. Moreover, their model is applied to a large dataset of international portfolios, which are heterogeneous by geography, credit quality and legal framework, on top consisting of loans with different ratings.

Apart from the different macroeconomic sub-model and constant LGD, the model of Pesaran et al. (2006) is technically equivalent to ours, with the conditional mean of the debtor's return replacing our common factor and the difference of the return and its conditional mean replacing our individual factor. In both the models, the common factors stand on the left hand side of a linear (macroeconomic) model. Thus, if we assume constant factors underlying LGD, use a GVAR model for the factors and apply bucketing, i.e. assume separate portfolios for loans with different ratings (Gordy, 2003), we can emulate their model by ours.

Besides explicit modelling of LGD and analytic formulas for PD and LGD distributions, the main advantage of our model in comparison with Pesaran et al. (2006) is that our model integrates its sub-model describing the factors with the part which transforms the factors into the losses more closely. While they estimate parameters of both the sub-models separately, we estimate them jointly.[4] Moreover, the number of parameters in our model is lower; there is no need to estimate the variance of the individual factor underlying PD, because it is combined with the parameters of the sub-model for factors. Contrary to them, we give analytic formulas for the predictions and confidence intervals of the losses. Last but not least, the dynamics of PDs is endogenous in our model with non-trivial dependence of the

[4] The only exception is the variance of the individual collateral factor, which, however, the other model does not treat at all.

corresponding factors on their past, while it is exogenous, estimated only by averaging of the past values, in Pesaran et al. (2006).

On the other hand, we cannot compete with the empirical results of Pesaran et al. (2006). While they adopt a well-established GVAR model for the international macroeconomic environment, we technically construct an ad hoc one for a single economy (for which, as we believe. that VECM is sufficient). While Pesaran et al. (2006) bring valid empirical results, we regard our empirical part rather as an illustration of our approach. Nevertheless, even these limited empirical results demonstrate a promising potential of our approach because, despite the simplicity of our macroeconomic sub-model, it gives quite meaningful results. It should be also noted our empirical study is rather preliminary and it serves mainly as a demonstration of our approach, which can, in principle, be used with any other macroeconomic VAR sub-model including a GVAR one.

The paper is organized as follows. In the next section, we provide a description of the model´s methodology. In Section 3, we describe the dataset used. Section 4 describes our empirical analysis and its results. Finally, Section 5 concludes. More technical parts of the text and some detailed empirical results are postponed into the Appendix.

2. The Model

Similar to Vasicek O. A. (2015a), we say that a loan *defaults* when

$$A < B \tag{1}$$

where A is the value the debtor's (hypothetical) assets and B is the value of his debts. Here,

$$A = \exp\{Y^A + Z^A\}, \qquad B = \exp\{Y^B + Z^B\}, \tag{2}$$

where Y^A, Y^B are factors, common to all the loans in the portfolio, and Z^A, Z^B are jointly normal individual factors.[5]

The *relative recovery* is, in line with Pykhtin (2003), computed as

$$R = \frac{\min(P, p)}{p} = \min(p^{-1}P, 1) \qquad P = \exp\{X + E\}, \tag{3}$$

where p is the outstanding principal of the loan, P is the price of the collateral, X is another common factor and E is a normally distributed individual factor, independent of (Z^A, Z^B).[6]

Now, consider an infinitely large portfolio of loans and define three important quantities: the *default rate*, usually called (conditional) probability of default (PD):

[5] In the credit risk literature, B is usually assumed to be deterministic. If this is the case, then we can put $Y^B = \log(B), Z^B = 0$ without affecting our future results.

[6] Dependence could be assumed for the price of losing analytical expression for LGD, see Pykhtin (2003).

$$Q = p\lim_{\#\text{ of loans}\to\infty} \frac{\#\text{ of defaults in the portolio}}{\#\text{ of loans in the portolio}} \tag{4}$$

the *loss given default*, abbreviated as LGD:

$$G = p\lim_{\#\text{ of loans}\to\infty} \frac{\text{total loss of the portfolio}}{\#\text{ of defaults}} \tag{5}$$

and the *charge-off rate* (relative loss):

$$L = p\lim_{\#\text{ of loans}\to\infty} \frac{\text{total loss of the portfolio}}{\#\text{ of loans}} \tag{6}$$

Here, $p\lim$ denotes limit in probability. Not surprisingly,

$$L = QG. \tag{7}$$

(see Kallenberg, 2002, Corollary 4.5.)
Assume further, that the portfolio is homogeneous in the sense that its loans have the same principal p and their individual factors are independent between the loans. Then it follows that

$$Q = \varphi(-Y), \qquad Y = \frac{Y^A - Y^B}{\rho}, \tag{8}$$

where φ is a standard normal c.d.f. and ρ is the standard deviation of $Z^A - Z^B$, and

$$\begin{aligned} G = h(I;\sigma), \qquad I = X - \log p, \qquad & h(\iota;\sigma) \\ = \varphi(-\frac{\iota}{\sigma}) - \exp\{\iota + \frac{1}{2}\sigma^2\}\varphi(-\frac{\iota}{\sigma} - \sigma) & \end{aligned} \tag{9}$$

where σ is the standard deviation of E (see Proposition 1, Appendix A.1. for the proof within a more general setting discussed below). Thanks to the strict monotonicity of φ and h (see Appendix of Gapko & Šmíd, 2012 for the latter), the correspondence between (Q, G) and (Y, I) is one-to-one.

Now consider n homogeneous portfolios evolving in time and assume that the j-th loan from the i-th portfolio defaults at time t in the case that $A_{i,j,t} < B_{i,j,t}$, where

$$A_{i,j,t} = \exp\{Y_{i,t}^A + Z_{i,j,t}^A\}, \qquad B_{i,j,t} = \exp\{Y_{i,t}^B + Z_{i,j,t}^B\}, \qquad 1 \le i \le n, \quad j \ge 1, \quad t \ge 1, \tag{10}$$

and that the corresponding relative recovery is $R_{i,j,t} = \min(p_{i,t}^{-1} P_{i,j,t}, 1)$ where $p_{i,t}$ is the outstanding principal and

$$P_{i,j,t} = \exp\{X_{i,t} + E_{i,j,t}\}, \qquad 1 \le i \le n, \quad j \ge 1, \quad t \ge 1. \tag{11}$$

Here, Y_i^A, Y_i^B, X_i are general stochastic processes and, for each i, vectors

$$\left(Z_{i,1,1}^{A}, Z_{i,1,1}^{B}, E_{i,1,1}\right), \left(Z_{i,1,2}^{A}, Z_{i,1,2}^{B}, E_{i,1,2}\right), \ldots \tag{12}$$

$$\left(Z_{i,2,1}^{A}, Z_{i,2,1}^{B}, E_{i,2,1}\right), \left(Z_{i,2,2}^{A}, Z_{i,2,2}^{B}, E_{i,2,2}\right), \ldots \tag{13}$$

are Gaussian i.i.d. with $(Z_{i,j,t}^{A}, Z_{i,t}^{B})$ independent of $E_{i,j,t}$. Note that we did not assume the independence of individual factors from different portfolios, so our model can handle the situations when different portfolios consist of the same subjects, having taken different types of loans.

Analogously to the static case, we have

$$Q_{i,t} = p\lim \frac{\#\text{ of defaults in } i-\text{th portfolio at } t}{\#\text{of loans in } i-\text{th portfolio at } t} = \varphi(-Y_{i,t}), \quad Y_{i,t} = \frac{Y_{i,t}^{A} - Y_{i,t}^{B}}{\rho_i}, \tag{14}$$

$$G_{i,t} = p\lim \frac{\text{total loss of } i-\text{th portfolio at } t}{\#\text{of defaults in } i-\text{th portfolio at } t} = h\left(I_{i,t}; \sigma_i\right), \quad I_{i,t} = X_{i,t} - \log p_{i,t}, \tag{15}$$

$$L_{i,t} = p\lim \frac{\text{total loss of } i-\text{th portfolio at } t}{\#\text{of loans in } i-\text{th portfolio at } t} = Q_{i,t} G_{i,t} = \varphi\left(-Y_{i,t}\right) h\left(I_{i,t}; \sigma_i\right), \quad (1) \tag{16}$$

where $\rho_i = \mathrm{stdev}(Z_{i,t})$, $Z_{i,t} = Z_{i,t}^{A} - Z_{i,t}^{B}$ and σ_i is the standard deviation of $E_{i,t}$ – for the proof, see Proposition 1, Appendix A.1.

Further, assume that the common factors follow a VAR model, i.e.

$$(Y_{1,t}^{A}, \ldots Y_{n,t}^{A}, Y_{1,t}^{B}, \ldots, Y_{n,t}^{B}, X_{1,t}, \ldots, X_{n,t})^{T} = \Gamma U_t + \mathcal{E}_t \qquad t = 1,2, \ldots \tag{17}$$

where Γ is a deterministic matrix, $\mathcal{E}_t$ is a Gaussian white noise and U_t is a matrix of regressors possibly including trend, constants, lagged values of Y_i^{A}, Y_i^{B}, X_i, their differences, and exogenous variables. Consequently, the dynamics of Qs and Gs is given by

$$Q_{i,t} = \varphi(-Y_{i,t}) = \varphi\left(-\left[\Gamma_Q U_t + \epsilon_t\right]\right), \tag{18}$$

$$G_{i,t} = h\left(I_{i,t}; \sigma\right) = h\left(\Gamma_G U_t^{p} + \eta_t; \sigma_i\right) \tag{19}$$

where $\Gamma^{Q} = \frac{1}{\rho} \star (\Gamma_1 - \Gamma_2)$, $\Gamma^{G} = (\Gamma_3, \mathcal{I}_n)$, $U_t^{p} = (U_t^{T}, \log p_{1,t}, \ldots \log p_{n,t})^{T}$ and (ϵ, η) is a Gaussian white noise. Here, $\rho = (\rho_1, \ldots, \rho_n)^{T}$, $\sigma = (\sigma_1, \ldots, \sigma_n)^{T}$, $\star$ is componentwise multiplication, $\mathcal{J}_n$ is $n \times n$ identity matrix, and Γ_j, $j = 1,2,3$, is the matrix consisting of the j-th third of rows of Γ.

As $(Q_{1,t}, \dots, Q_{n,t}, G_{1,t}, \dots, G_{n,t})$ are uniquely determined by the common factors and the parameters (ρ, σ, p_t), their distribution depends only on these values and the parameters of the factors' distribution; in particular, it does not depend on either $\mathrm{corr}(Z_{i,t}, Z_{j,t})$ or $\mathrm{corr}(E_{i,t}, E_{j,t})$, $1 \leq i < j \leq n$, i.e. mutual inter-portfolio correlations of individual factors. Thus, for example, if two portfolios consist of loans of the same debtors, the first of mortgages secured by houses, the second of leasing loans secured by cars, and with $\mathrm{corr}(Z_{1,t}, Z_{2,t}) = 1$ and $\mathrm{corr}(E_{1,t}, E_{2,t}) > 0$, then the distribution of $(Q_{1,t},, G_{1,t}, Q_{2,t},, G_{2,t})$ is the same as if the debtors from the portfolios were not related at all and their individual factors were independent.

Once $(Q_{i,t}, G_{i,t})_{i \leq n, t \leq \tau}$ and $(U_t^p)_{i \leq n, t \leq \tau}$ (i.e. the PD, LGD, the exogenous variables, and the outstanding principals in the individual portfolios) are observed, it is easy to estimate the dynamics of Q_i and G_i, provided that the parameters σ_i (the standard deviations of the collateral individual factors) are known. The estimation procedure consists of two steps. First, the adjusted factors Y and I are retrieved by inverse relations

$$Y_{i,t} = -\varphi^{-1}(Q_{i,t}), \qquad I_{i,t} = h^{-1}(G_{i,t}; \sigma_i) \tag{20}$$

and, second, the parameters Γ_Q , Γ_G and $\mathrm{var}(\epsilon, \eta)$ are estimated from the equations

$$Y_t = \Gamma_Q U_t + \epsilon_t, \qquad I_t = \Gamma_G U_t^p + \eta_t \tag{21}$$

by standard techniques. Note that the knowledge of ρ_i (the standard deviations of the wealth individual factors] is not necessary for the estimation, as they ``melt '' with the parameter $\Gamma_Q$. This, actually, is one of the main advantades of our model in comparison with Pesaran et al. (2006) who estimate the risk models separately from the econometrical ones, so they cannot ``spare '' the variability of the risk model parameters as we do. The knowledge of σ_i, on the other hand, is necessary for the procedure, as the transformation of the LGD into its corresponding factor is non-linear in σ_i; once it is unknown, however, the estimation procedure may be performed for various its values and the version of the model, exhibiting the best fit, may be chosen.

As for the outstanding principals $p_{i,t}$, their value usually known; if it is not the case, then an approximation, $\log p_{i,t} = \pi_i + \varpi_i t$, may be used. Given this approximation, the parameters π_i and ϖ_i become the trend coefficients of the VAR model.

Forecasting of Q and G is a relatively easy task because analytical formulas for conditional distributions of $Q_{i,T}$ and $G_{i,T}$ given the information up to $\tau < T$ may be derived. Namely, as

$$Y_{i,T} | \mathcal{U}_\tau \sim \mathcal{N}(\mu, v^2), \qquad \mathcal{U}_\tau = (U_t)_{t \leq \tau}, \tag{22}$$

for some ($\mathcal{U}_\tau$-measurable) μ and v in the VAR model, we have, according to Proposition 2, Appendix A.1,

$$\mathbb{P}[Q_{i,T} < \theta | \mathcal{U}_\tau] = \varphi\left(\frac{\varphi^{-1}(\theta) + \mu}{v}\right), \qquad \theta \in (0,1), \tag{23}$$

with the point forecast given by

$$\mathbb{E}(Q_{i,T} | \mathcal{U}_\tau) = \mathbb{P}[A_{i,T} < B_{i,T} | \mathcal{U}_\tau] = \varphi\left(\frac{-\mu}{\sqrt{v^2 + 1}}\right). \tag{24}$$

Similarly, as

$$I_{i,T} | \mathcal{U}_\tau \sim \mathcal{N}(\nu, w^2) \tag{25}$$

for some ν, w, we have

$$\mathbb{P}[G_{i,T} < \theta | \mathcal{U}_\tau] = \varphi\left(\frac{h^{-1}(\theta; \sigma_i) - \nu}{w}\right). \tag{26}$$

$$\mathbb{E}(G_{i,T} | \mathcal{U}_\tau) = h(\nu; \sqrt{\sigma_i^2 + w^2}) \tag{27}$$

the mean loss of a loan given (see Proposition 3 in Appendix A.2).

Moreover, as we may equivalently express,

$$\mathbb{P}[Q_T < \theta | \Omega_t] = \varphi\left(\frac{1}{\sqrt{\vartheta}}\left(\sqrt{1-\vartheta}\varphi^{-1}(\theta) - \varphi^{-1}(\mathbb{P}[A_{i,1,T} < B_{i,1,T} | \mathcal{U}_t])\right)\right), \quad \vartheta = \frac{v^2}{v^2 + 1}, \tag{28}$$

we see that our formula generalizes the well known Vasicek's formula for the distribution loss (Vasicek O. A., 2015b). To see it, note that, by (3), $\mu = -\sqrt{v^2 + 1}\varphi^{-1}(\mathbb{P}[A_{i,1,T} < B_{i,1,T} | \mathcal{U}_\tau])$. The second equality of the second line follows from the fact that $\log A_{i,j,T} - \log B_{i,j,T} = Y_{i,t}^A - Y_{i,t}^B + Z_{i,j,t}^A - Z_{i,j,t}^B = \rho_i Y_{i,t} + N(0, \rho_i)$.

Thanks to strict monotonicity of the functions transforming the factors to the rates, the confidence intervals for Q and G may be obtained by the same transformations by which we are getting the losses from the factors. In particular, once $[Y_{i,T}^L, Y_{i,T}^H]$, $[I_{i,T}^L, I_{i,T}^H]$ are confidence intervals for future values of $Y_{i,T}, I_{i,T}$, the intervals $[\varphi(-Y_{i,T}^H), \varphi(-Y_{i,T}^L)]$, $[h(I_{i,T}^H; \sigma_i), h(I_{i,T}^L; \sigma_i)]$, may serve as confidence sets for $Q_{i,t}, G_{i,t}$, respectively.

Contrary to $Q_{i,T}$ and $G_{i,T}$, the distribution of $L_{i,T}$ is not generally analytically tractable. In particular, an analytical formula for $\mathbb{P}[G_{i,T} < \theta | \mathcal{U}_\tau]$ exists only when $Y_{i,T}$ and $I_{i,T}$ are conditionally independent given $\mathcal{U}_\tau$, which is generally not true in a VAR model. However, the distribution of $L_{i,T}$ may be efficiently computed by a Monte Carlo simulation.

3. Data

Detailed loan portfolios data are usually a subject of the banking secret and thus strictly confidential, therefore extremely difficult to obtain. Due to this restriction we decided to apply our model to two US nationwide portfolios: the residential and commercial real estate loans (mortgages), for which loss data are publicly available. For both portfolios, we used delinquency rates, which are proportions of loans more than 30 days past due (30+) on the total balance, as proxies for Q and charge off rates, which are proportions of loans charged off (net of recoveries) on the average total balance, as proxies for L. Consequently, we computed G as the ratio of L and Q. The datasets are available at the United States Federal Reserve System. The time period covered ranges from 1991 to 2016 in a quarterly frequency.

Table 1 and Figure 1 summarize descriptive statistics and display the time series of the delinquency rates and charge-off rates of the residential and commercial portfolios. The Figure 1 clearly points at the fact that the time series are correlated. Also, the recent economic crisis, which started in the US in late 2007 and impacted the US mortgage and real estate markets excessively is visible, as all the time series rocketed up to multiples of their preceding values between 2007 and 2010.

Table 1 Descriptive Statistics of Input Data

Statistic	*30+ delinquency rate residential*	*Charge-off rate residential*	*30+ delinquency rate commercial*	*Charge-off rate commercial*
Mean value	0.041	0.005	0.038	0.009
Median	0.023	0.002	0.023	0.003
Minimum	0.013	0.001	0.009	0.0001
Maximum	0.110	0.027	0.121	0.036
Standard deviation	0.031	0.007	0.031	0.011
Variance	0.748	1.386	0.815	1.162
Skewness	1.115	1.840	1.131	1.185
Excess kurtosis	-0.332	2.134	0.091	-0.082
5% percentile	0.011	0.001	0.016	0.001
95% percentile	0.111	0.031	0.107	0.022

Having the Qs and Gs, we extracted the factors by (2) with the values of $\sigma = 0.135$ for the commercial loans and $\sigma = 0.056$ for mortgages; the values were computed from the series of the residential and commercial house price indices the way described in Appendix A.3. The resulting time series of the extracted common factors Y (PD) and I (LGD) for both commercial (Y^c, I^c) and residential (Y^r, I^r) mortgage portfolios are illustrated in Figure 2. As we obtained the factors from a monotone transformation of the loss rates, there is again a strong visual correlation, especially between $Y^r$and Y^c, and $I^r$and I^c. On the other hand, the correlation matrix calculated on first differences of the four factors, summarized in the Table 2, showed a strong correlation only between Y^r and Y^c.

Figure 1 Development of the 30+ Delinquency Rates (Q^r = Residential, Q^c = Commercial) and Charge-Off Rates (G^r = Residential, G^c= Commercial)

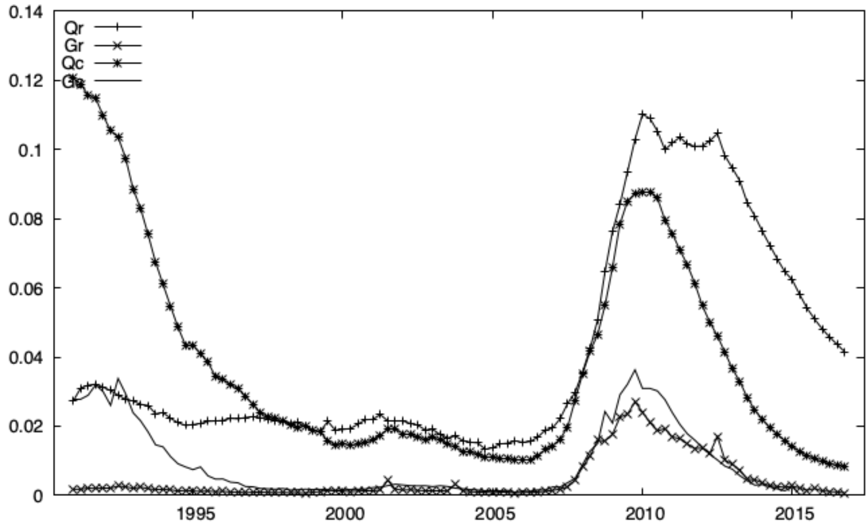


Figure 2 The Development of the Extracted Common Factors $Y^r$and I^r (Left Axis), and $Y^c$and I^c (Right Axis)

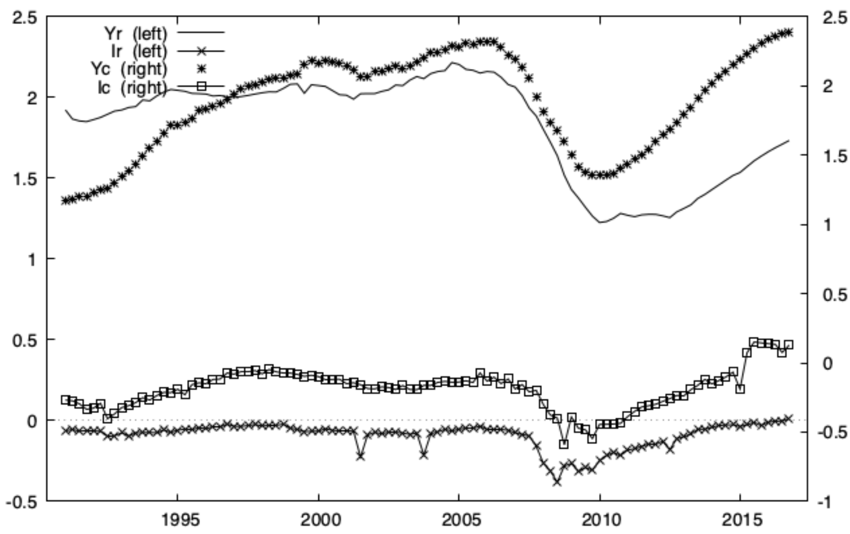


Table 2 The Correlation Matrix of First Differences Of I^r, I^c, Y^r, Y^c, Sample Size $n = 103$

	dI^c	dY^c	dI^r	dY^r
dI^c	1.0000	0.2680***	0.1182	0.2420**
dY^c		1.0000	0.2869***	0.7268***
dI^r			1.0000	0.1198
dY^r				1.0000

Our choice of the set of explanatory macroeconomic variables was inspired by Pesaran et al. (2006), who used the output (GDP), inflation, equity prices, foreign exchange rate, interest rate and real money balances as explanatory variables in their factor sub-model. Similarly to them, we chose the output (GDP), industrial production (IP), residential and commercial HPI indices as the representatives for the asset value (instead of the equity indices in Pesaran et al., 2006), FED base interest rate (FEDR), unemployment (U) and finally personal income (PI). We ignored the foreign exchange rate as our analysis is based purely on the US domestic indicators and the instabilities in the economy are represented by the remaining variables. We used the CPI to obtain real indicators from nominal. The commercial HPI index was subsequently dropped as it did not prove to have a significant effect in the final model. Except for FEDR, logarithms of all the variables were used in the actual estimation.

4. Results

First, we tested all factors and exogenous variables for unit roots. Using the Augmented Dickey-Fuller (ADF) test, we could not reject the unit root except for FEDR. Thus, we treated all exogenous variables (IP, PI, U, HPI^r, GDP) and the four factors (I^r, I^c, Y^r, Y^c) as integrated with order one. The detailed results of the ADF tests can be found in the Appendix in Table A.4.1.

In line with common practice, we continued with tests for cointegration between the integrated variables. As the first step, we tested the cointegration among all exogenous variables IP, PI, U, HPI^r and GDP. This test produced mixed results, as the Engle-Granger test did not confirm the cointegration while the Johansen test suggested cointegration rank one. In the next step, we tested the cointegration between the pairs of factors, namely I^r, I^c and Y^r, Y^c. The cointegration was confirmed between I^r, I^c , but not between Y^r, Y^c. Then, we examined the cointegration among all the four factors, where the Johansen test suggested the cointegration rank two. Next, we tested the cointegration between individual factors and the set of exogenous variables. The cointegration was found between I^r and exogenous variables, between I^c and exogenous variables, but not between Y^r, Y^c and exogenous variables. The resulting relations may be found in Table A.5.1. Finally, we performed the test for all variables altogether. The Johansen test suggested cointegration rank 3 (by the Lmax test) or higher (by the trace test). Consequently, we decided to continue with the cointegration rank 3. See the Appendix A.5 for more details of all cointegration tests.

Consequently, we constructed the final model. First, we estimated a nine-equation VECM model for $I^c, I^r, Y^c, Y^r, IP, PI, U, HPI^r, GDP$ with rank 3 and 2 lags and with $FEDR$ as exogenous variable. The cointegration matrix β was restricted to reflect the relations tested by the EG tests, see Table A.3.1 for the results of re-estimation. Finally, we removed insignificant variables and re-estimated the four equations with factors on the right hand side.[7] The results are summarized in Table 3. EC_n represents the error correction term of the n-th cointegration equation.

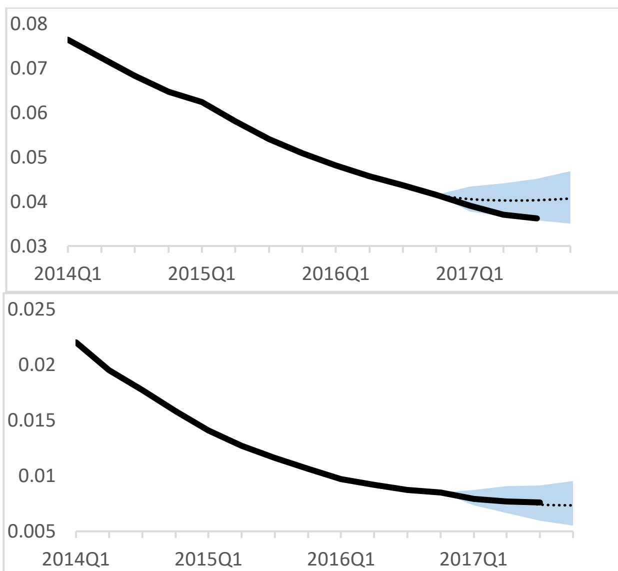
[7] Thus, our final model is not a VECM but rather a restricted VAR model with exogeneous variables including the three error correction terms from the original VECM.

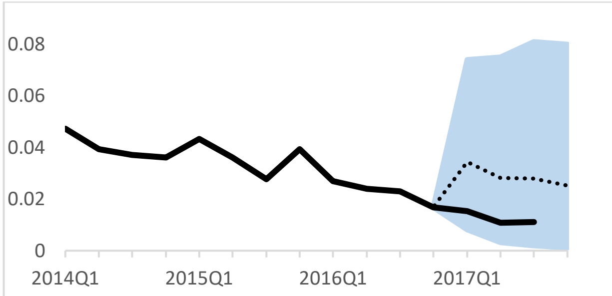
Table 3 Results of the VECM Estimation (Significance: * - 90%, ** - 95%, * - 99%)**

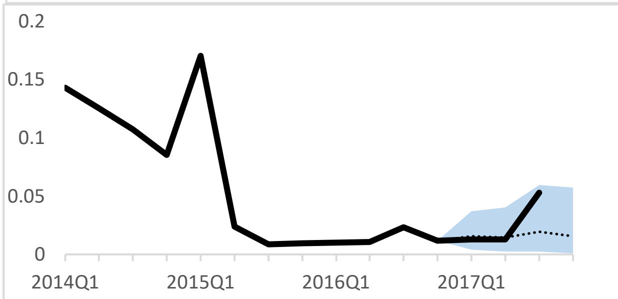
Variable	*ΔIc*	*ΔIr*	*ΔYc*	*ΔYr*
Constant	-1.944 ***	1.117 **	-0.719 **	-2.437 ***
ΔIc (lag1)	-0.369 ***	-	-	0.061 *
ΔIr (lag1)	-	-	-	-0.138 ***
ΔYc (lag1)	0.658 ***	0.779 ***	0.348 ***	-
ΔHPI (lag1)	-1.355 ***	-	0.719 ***	-
ΔU (lag1)	-	-	-0.184 ***	-0.222 ***
FEDR (lag1)	-0.006 ***	0.009 ***	-0.005 ***	-0.009 ***
ΔIP (lag1)	-	-	-0.365 **	-
EC1	-0.103 ***	0.073 ***	0.072 ***	-0.039 ***
EC2	-	-0.657 ***	0.087 *	0.199 ***
EC3	0.335 ***	0.229 ***	-0.087 **	0.161 ***
Adjusted R-square	31 %	36 %	82 %	73 %

The predicting power of the model is strong at Y^c, Y^r, but quite weak at I^r, I^c. This is evident from the R-square values and also visible in Figure 3, which shows the fan charts of the mean predictions of Q^r, Q^c, G^r and G^c for four periods, compared with the actual values of the respective delinquency rates and LGDs. Our results, similarly to Virolainen (2004) or Pesaran et al. (2006) show that credit risk is significantly dependent on interest rates. Moreover, in all the mentioned works, the coefficient sign of the interest rate was negative. This means that periods with low interest rates tend to overlap with those with high default rates.

Figure 3 The Comparison of the Predictions (Dotted Line) of Q^r, Q^c, G^r and G^c (From Top) with the Actuals (Solid Line) and the 95% Prediction Confidence Intervals (Light Area)







A clear distinction between the default rate and LGD factors emerges from our results. The LGD factors are cointegrated with the macroeconomic environment, but more imprecisely predicted, while the default rate factors are predicted well, but not significantly cointegrated with the macroeconomic environment. This suggests that the LGDs are determined by the macroeconomic development whereas the default rates exhibit stronger inertia. In other words, the default rates are dependent more on the pas while the LGDs depend on the present.

In order to test whether it is worth to treat the two portfolios simultaneously instead of their separate modelling, we made a likelihood ratio (LR) test of the final model with respect to its version without mutual dependence. In particular, EC_2 was removed from the equation for ΔI^c, further, $EC_1, \Delta Y^c$ were removed from the equation for ΔI^r, and EC_1, ΔI^c were removed from the equation for ΔY^r. Even though the "mutual" term EC_3 was not removed for simplicity, the high value LR=59.52 of the statistics suggests for the interconnectedness of the two portfolios beyond the explanation by the macro variables.

Finally, we computed the amount of hypothetical regulatory capital for both portfolios and compared it to the amount prescribed by the IRB approach. According to usual practice, the amount was computed as the 99.9% quantile of the yearly loss L. In the case of our model, we computed the value by Monte Carlo (MC) simulation (see the end of Section 2) while the IRB regulatory capital was computed in a standard way, by Vasicek's formula for loss distribution with fixed correlation of 15 % between log-assets; as the (fixed) LGD, is latest observed value was used. The results of the two models is summarized in Table 4. In both cases, our model recommends lower economic capital than the standard approach. This, together with the fact that our model appears to be quite realistic, suggests that the IRB approach may be too conservative, and using our model may lead to savings by not requiring excessively high economic capital.

Table 4 Comparison of the Predictions of IRB Vs. our Model – 12 Month Loss on the 99.9% Probability Level

Segment/Model	*IRB*	*Our*
Retail	0.47%	0.40%
Commercial	0.12%	0.07%

The correlations of the underlying factors, computed by the second formula of (4), came out 0.25% for the commercial loans and 0.12% for the residential ones. If we used these correlations instead of the IRB prescribed 15%, then the retail IRB charge would be 0.09% and the commercial one 0.01%. Lower values of the correlations might be expected because the common factors are explained by the VECM model, hence their (conditional) variances are lower than the unconditional ones.[8] On the other hand, the resulting values of the IRB charges, being substantially lower than these given in Table 4, might rise doubts about the internal consistency of our approach. This discrepancy, however, may be explained by the facts that, contrary to the IRB formula, we take the uncertainty of the LGDs into account and that this uncertainty is much greater than that of the PDs in our model. These results may suggest that, in banking practice, the factor correlations are being overestimated in order to substitute for the stochasticity of LGD. The detailed procedure replicating the empirical part of our paper may be found at https://github.com/utia-econometrics/GS2017.

5. Conclusion

We constructed a multi-period multi-portfolio dynamic model of credit losses and applied it to two US national loan portfolios. Our model is unique; except for Pesaran et al. (2006), no comparable model has been published to our knowledge. In comparison with Pesaran et al. (2006), our model has several advantages: it's compact in the sense that most of its parameters are estimated jointly, it requires less parameters while being comparably general, and it models the LGD explicitly. Compared to our approach, the advantage of Pesaran et al. (2006) is a concrete well estimated macroeconomic model; however, our approach is flexible enough to possibly adopt their macroeconomic sub-model while preserving our advantages.

We applied our model to real-life data, namely to two nationwide US loan portfolios. Although we proceeded purely technically when calibrating our model, we achieved relatively high accuracy of forecasts, especially in case of the PDs.

Finally, we made a thought experiment where we constructed hypothetical economic capital charges to two nationwide portfolios. For both portfolios, the economic capital recommended by our model, which was calculated on regulatory probability level, was lower compared to the IRB regulatory approach. This suggests that applying our model might lead to economic capital savings. Moreover, as our model results can be compared to those of IRB and as our model enables to calculate credit risk under various settings of macroeconomic conditions on various probability

[8] The correlations, computed from the sample (unconditional) variances of the factors, are 8.7%, 11.0%, respectively.

levels, it can be used in internal processes within financial institutions, such as calculation of loan loss allowances under IFRS9 and/or stress testing of credit risk.

Finally, our model is a straightforward generalization of the widely used Vasicek-Merton approach; in particular, both models share the same set of assumptions (e.g. the evolution of the borrowers' assets). Therefore, and for the reasons discussed above, our model could be a suitable candidate for a fully-fledged economic capital model in a financial institution.

APPENDIX

A.0 Auxiliary Results

Lemma 1. Let $\mathcal{R}$ and $\mathcal{S}$ be measurable spaces and let $S \in \mathcal{S}$, $R_1 \in \mathcal{R}, R_2 \in \mathcal{R}^2, \ldots$ be random elements such that that S is independent of $(R_1, R_2, \ldots)$. Let $f_n: \mathcal{S} \times \mathcal{R}^n \to \mathbb{R}$, $n \in \mathbb{N}$, and $g: \mathcal{S} \to \mathbb{R}$ be measurable functions. Let $\operatorname{plim}_n f_n(s, R_n) = g(s)$ for any $s \in \mathcal{S}$. Then $\operatorname{plim}_n f_n(S, R_n) = g(S)$.

Proof. By (Kallenberg, 2002), p.63, random variables $X_1, X_2, \ldots$ converge to a random variable X in probability iff $\mathbb{E}(|X_n - X| \wedge 1) \to 0$. Thus, denoting $\phi_n(s) = \mathbb{E}(|f_n(s, R_n) - g(s)| \wedge 1)$, we have $\phi_n(s) \to 0$ for any $s \in \mathcal{S}$. To prove the Lemma, note that, for any n,

$$\mathbb{E}(|f_n(R_n, S) - g(S)| \wedge 1) = \mathbb{E}(\mathbb{E}(|f_n(R_n, S) - g(S)| \wedge 1|S)) = \mathbb{E}(\phi_n(S)) \qquad (29)$$

by the Law of Iterated Expectation, 6.8.14. of (Hoffmann-Jorgensson, 1994), respectively, and that

$$\mathbb{E}(\phi_n(S)) = \int \phi_n(s) d\mathbb{P}_S(s) \to 0 \qquad (30)$$

by the Dominated Convergence Theorem (with the integrable upper bound equal to one). Q,E,D,

Lemma 2. (i) Denote $k(e, \iota) = 1 - \min(\exp\{\iota + e\}, 1)$. For any deterministic ι and normal E,

$$\mathbb{E}k(E, \iota) = h\big(\iota; \operatorname{stdev}(E)\big). \qquad (31)$$

(ii) If $I = I_1 + I_2$, where I_1 is Ω-measurable for some sigma field Ω, and where $I_2|\Omega \sim \mathcal{N}(0, s^2)$ for some $s > 0$, then, for any $\sigma > 0$,

$$\mathbb{E}(h(I; \sigma)|\Omega) = h\left(I_1; \sqrt{s^2 + \sigma^2}\right). \qquad (32)$$

Proof. (i) For any ι,

$$\begin{aligned} \mathbb{E}k(E, \iota) &= 1 - \int \min(\exp\{e + \iota\}, 1) d\mathbb{P}_E(e) \\ &= 1 - \int_{-\infty}^{-\iota} \exp\{e + \iota\} d\mathbb{P}_E(e) + \int_{-\iota}^{\infty} d\,\mathbb{P}_E(e) \\ &= \mathbb{P}[E \leq -\iota] - \exp\{\iota\} \int_{-\infty}^{-\iota} \exp\{e\} d\mathbb{P}_E(e) = h(\iota; \sigma) \end{aligned} \qquad (33)$$

where the last equality follows by a straightforward calculation (for details, see e.g. Gapko & Šmíd, 2012 or Pykhtin, 2003).

(ii) By Kallenberg (2002), Proposition 6.8., I_2 is independent of Ω, which implies that

$$(I_1, I_2) \text{ is conditionally independent of } \Omega \text{ given } I_1. \qquad (34)$$

(see (Kallenberg, 2002), Propositions 6.6 and 6.8).

Let $E \sim \mathcal{N}(0, \sigma^2)$ be independent of (I_2, Ω). By the Chain Rule (Kallenberg, 2002, 3.8), then

$$E, I_2, \Omega, \text{are mutually independent.} \tag{35}$$

Further, put

$$\eta_1(\iota_1, \iota_2) = h(\iota_1 + \iota_2; \sigma). \tag{36}$$

As $h(I_1 + I_2; \sigma)$ is $\sigma(I_1, I_2)$ measurable,

$$\mathbb{E}(h(I_1 + I_2; \sigma)|I_2, I_2) = \eta(I_1. I_2). \quad (7)$$

Further, we have

$$\eta_1(\iota_1, \iota_2) \overset{(i)}{=} \mathbb{E}(k(E, \iota_1 + \iota_2)) \overset{h.j}{=} \mathbb{E}(k(E, I_1 + I_2)|I_1 = \iota_1, I_2 = \iota_2) \tag{37}$$

where "h.j." stands for 6.8.14 of Hoffmann-Jorgensson (1994). Finally, abbreviating the Law of Iterated Expectation as "l.i.e",

$$\begin{gathered}
\mathbb{E}(h(I_1 + I_2; \sigma)|\Omega) \overset{(5)}{=} \mathbb{E}(h(I_1 + I_2; \sigma)|I_1) \overset{l.i.d}{=} \mathbb{E}(\mathbb{E}(h(I_1 + I_2; \sigma)|I_1, I_2)|I_1) \\
\overset{(7)}{=} \mathbb{E}(\eta(I_1, I_2)|I_1) = \mathbb{E}(\mathbb{E}(k(E, I_1 + I_2)|I_1, I_2)|I_1)) \\
= \mathbb{E}(\mathbb{E}(k(I_2 + E, I_1)|I_1, I_2)|I_1) \overset{l.i.d}{=} \mathbb{E}(k(I_2 + E, I_1)|I_1) \overset{h.j}{=} \mu(I_1) \\
\mu(\iota_1) = \mathbb{E}(k(I_2 + E, \iota_1)) \overset{(i)}{=} h(\iota_1; \sqrt{s^2 + \sigma^2}).
\end{gathered} \tag{38}$$

Q.E.D.

A.1 Transformation of Factors

Proposition 1.

(i) For any $1 \leq i \leq n$, $t \geq 1$, $Q_{i,t} = \varphi(-Y_{i,t})$.

(ii) For any $1 \leq i \leq n$, $t \geq 1$, $G_{i,t} = h(I_{i,t}; \sigma)$.

Proof. Fix i and t.

(i) For any j, denote

$$D_{i,j,t} = \mathbf{1}[A_{i,j,t} < B_{i.j.t}] \tag{39}$$

the indicator of the default of the j-th loan from the i-th portfolio at t, and note that

$$D_{i,j,t} = d(Z_{i,j,t}, Y_{i.t}), \qquad d(z, y) = \mathbf{1}\left[\frac{z}{\rho_i} < -y\right]. \tag{40}$$

Using this, we can write

$$Q_{i,t} = \operatorname{plim}_{n\to\infty} f_n(Y_{i,j}, Z_{i,1,t}, \dots, Z_{i,n,t}) \tag{41}$$

where

$$f_n(y, z_1, \dots, z_n) = \frac{\sum_{j=1}^{n} d\left(z_j, y\right)}{n}. \tag{42}$$

As, by the Law of Large Numbers,

$$f_n(y, Z_{i,1,t}, \dots, Z_{i,n,t}) \to \mathbb{E}d\left(Z_{i,1,t}, y\right) = \mathbb{P}\left[\frac{Z_{i,1,t}}{\rho_i} < -y\right] = \varphi(-y), \tag{43}$$

the assertion follows from Lemma 1.
(ii) For any j, denote

$$K_{i,j,t} = D_{i,j,t}(1 - R_{i,j,t}), \qquad R_{i,j,t} = \min(p_{i,t}^{-1} P_{i,j,t}, 1), \tag{44}$$

the relative loss from the j-th loan of the i-th portfolio at t. Alternatively,

$$\begin{aligned} K_{i,j,t} &= D_{i,j,t}(1 - \min(\exp\{X_{i,t} + E_{i,j,t} - \log p_{i,t}, 1)) \\ &= d\left(Z_{i,j,t}, Y_{i,t}\right) k\left(E_{i,j,t}, I_{i,j}\right), \end{aligned} \tag{45}$$

where k is defined at Lemma 2 (i). Using this, we have

$$L_{i,t} = \mathrm{plim}_{n\to\infty} g_n((I_{i,t}, Y_{i,t}), (E_{i,1,t}, Z_{i,1,t}), \dots, (E_{i,n,t}, Z_{i,n,t})) \tag{46}$$

where

$$g_n(\iota, y, (e_1, z_1), \dots, (e_n, z_n)) = \frac{\sum_{j=1}^{n} d\,(z_j, y) k(e_j, \iota)}{n}. \tag{47}$$

By the Law of Large Numbers and the independence of the factors, we have, for any ι, y,

$$\begin{aligned} \mathrm{plim}_n g_n((E_{i,1,t}, Z_{i,1,t}), \dots, (E_{i,n,t}, Z_{i,n,t})) &= \mathbb{E}(d(Z_{i,1,t}, y) k(E_{i,1,t}, \iota)) \\ &= \mathbb{E}d(Z_{i,1,t}, y)\mathbb{E}k(E_{i,1,t}, \iota). \end{aligned} \tag{48}$$

Further, by Lemma 2 (i), and due to the fact that $\mathbb{E}d(Z_{i,1,t}, y) = \varphi(-y)$ (see the proof of (i)), we have, by Lemma 1, that $L_{i,t} = \varphi(-Y_{i,t}) h(I_{i,t}; \sigma_i)$. The assertion now follows from (i) and the fact that $G_{i,t} = \frac{L_{i,t}}{Q_{i,t}}$ by (see Kallenberg, 2002, Corollary 4.5). Q.E.D.

A.2 Distribution of Forecasts

Proposition 2. Let Ω be a sigma field and let $Q = \varphi(-Y)$ where $Y|\Omega \sim \mathcal{N}(\mu, v^2)$ for some Ω-measurable μ, v. Then
(i) $\mathbb{P}[Q \le \theta|\Omega] = \varphi\left(\frac{\varphi^{-1}(\theta)+\mu}{v}\right)$,
(ii) $\mathbb{E}[Q|\Omega] = \varphi\left(\frac{-\mu}{\sqrt{v^2+1}}\right) = \mathbb{P}[Z < -Y]$ for any $Z \sim \mathcal{N}(0,1)$ independent of Y.
Proof. (i)

$$\mathbb{P}[Q \leq \theta|\Omega] = \int \mathbf{1}\{\varphi(y) \leq \theta\}\mathrm{d}\mathbb{P}_{-Y|\Omega}(y) = \int \mathbf{1}\{\varphi(y) \leq \theta\}\mathrm{d}\varphi\left(\frac{y+\mu}{v}\right)$$
$$= \int \mathbf{1}\{z \leq \theta\}\mathrm{d}\varphi\left(\frac{\varphi^{-1}(z)+\mu}{v}\right) = \varphi\left(\frac{\varphi^{-1}(\theta)+\mu}{v}\right). \tag{48}$$

(ii)

$$\mathbb{E}(Q|\Omega) = \int z\mathrm{d}\varphi\left(\frac{\varphi^{-1}(z)+\mu}{v}\right) = \int \varphi(x)\mathrm{d}\psi(x), \tag{49}$$

where $\psi(x) = \varphi\left(\frac{x+\mu}{v}\right)$ is the c.d.f. of $\mathcal{N}(-\mu, v^2)$. The r.h.s. is, however, nothing else but the formula for the probability that a difference of independent $\mathcal{N}(0,1)$ and $\mathcal{N}(-\mu, v^2)$ is less than zero, i.e.

$$\mathbb{E}(Q|\Omega) = \mathbb{P}[\mathcal{N}(0,1) + \mathcal{N}(\mu, v^2) \leq 0] = \mathbb{P}[\mathcal{N}(\mu, 1+v^2) \leq 0]$$
$$= \varphi\left(\frac{-\mu}{\sqrt{v^2+1}}\right). \tag{50}$$

Q.E.D.

Proposition 3. Let Ω be a sigma field and let $G = h(I;\sigma)$ where for $I|\Omega \sim \mathcal{N}(\nu, w^2)$ for some Ω-measurable μ, w. Then
(i) $\mathbb{P}[G < \theta|\Omega] = \varphi\left(\frac{h^{-1}(\theta;\sigma)-\nu}{w}\right)$
(ii) $\mathbb{E}(G|\Omega) = h(\nu, \sqrt{\sigma^2 + w^2})$
Proof. (i)

$$\mathbb{P}[G < \theta|\Omega_t] = \int \mathbf{1}\{h(\iota;\sigma) < \theta\}d\mathbb{P}_{I|\Omega}(\iota) = \int \mathbf{1}\{h(\iota;\sigma) < \theta\}d\varphi\left(\frac{\iota-\nu}{w}\right)$$
$$= \int \mathbf{1}\{x < \theta\}d\varphi\left(\frac{h^{-1}(x;\sigma)-\nu}{w}\right) = \varphi\left(\frac{h^{-1}(\theta;\sigma)-\nu}{w}\right) \tag{51}$$

(ii) Follows from Lemma 2 (ii) with $I_1 = \nu$ and $I_2 \sim \mathcal{N}(0, w^2)$. Q.E.D

A.3 Determination of σ

Assume that, at time t, the portfolio contains multiple "generations" of loans namely the loans originated at $t-1, t-2, \dots, t-k$ (the loans older than k are no longer present in the portfolio). Assume further that the inflow of fresh loans into the portfolio is constant in time. Finally, assume that all the collaterals securing loans from the generation which started at s have been bought for the same price $\exp\{H_s\}$ and that the price of each of them at t is $\exp\{H_t + (S_t - S_s)\}$ where S is a normal random walk, specific to the loan, with variance θ^2,

Denote G_t the age of a loan randomly chosen at t. Clearly, after k periods, the ratio of the generations within the portfolio is: $(1-q): \dots : (1-q)^{k-1}$, which uniquely determines $\pi_i = \mathbb{P}[G_t = i]$.

Let P_t be the price of a randomly chosen collateral. By the Law of Iterated Variance, we then get

$$\tilde{\sigma}^2 = var(\log P_t|H) = var(\mathbb{E}(\log P_t|G_t, H)|H) + \mathbb{E}(var(\log P_t|G_t, H)|H) = var(\mathbb{E}(S_t - S_G|G,H)|H) + \mathbb{E}(var(S_t - S_G|G,H)|H) = \theta^2\, \mathbb{E}G_t = \theta^2 \sum_{i=1}^{k} i\pi_i, \quad (52)$$

Even though the $\mathcal{L}(\log P_t|H)$ is a mixture of normal distributions rather than a normal distribution, it is thin tailed so it will not make a big harm to approximate it by $N(H_t, \tilde{\sigma}^2)$.

A.4 Results of the ADF Tests and the Engle-Granger Cointegration Tests for Individual Factors

Table A.4.1 Results of the ADF Unit Root Tests

Variable	*P-value of the ADF unit root test*
Ic	0.87
Ir	0.52
Yc	0.19
Yr	0.11
U (unemployment)	0.97
PI (personal income)	0.79
IP (industrial production)	0.83
GDP	0.93
HPI	0.97
FEDR (FED interest rate)	0.005

A.5 Cointegration Relations

Table A.5.1 Cointegration Relations. EG – Cointegrating Regression from the Engle-Granger Test (Trends Omitted), J – Estimate by Johansen Procedure with Cointegration Matrix Restricted so as to Reflect Structure of the EG Regressions (Trends Omitted)

Explained variable	***Ic***		***Ir***			***Yc***	
Regressor	**EG**	**J**	**EG**	**J**		**EG**	**J**
Yr						1.26461***	0.95105***
U	-0.36938***	-0.55046	0.0223668	-0.1299		-0.072549	-0.4824***
PI	-1.81483***	-2.0682	-1.09478***	-0.39289		-1.21729**	-4.1344***
IP	2.11787***	2.3236	2.17336***	0.30487		1.75668***	0.65825
GDP	-2.79685***	-1.3035	-2.899***	-0.73101		-0.350578	2.8717***
HPIr	0.299283***		1.2083***	0.39988***	0.40402***	-0.6515***	-0.31533*
ADF p-value	0.013			0.012		0.229	

REFERENCES

Frontczak R, Rostek S (2015): Modeling Loss Given Default with Stochastic Collateral. *Economic Modelling*, 44(2015):162-170.

Frye J (2000): Collateral Damage. *Risk,* 13(4):91-94.

Gapko P, Šmíd M (2012): Dynamic Multi-Factor Credit Risk Model with Fat-Tailed Factors. *Czech Journal of Economics and Finance,* 62(2):125-140.

Gordy M (2003): A Risk-Factor Model Foundation for Ratings-Based Bank Capital Rules. *Journal of financial intermediation,* 12(3):199–232.

Hamerle A, Dartsch A, Jobst R, Plank K (2011): Integrating Macroeconomic Risk Factors Into Credit Portfolio Models. *The Journal of Risk Model Validation,* 5(2):3-24.

Jimenez G, Mencia J (2009): Modelling the Distribution of Credit Losses with Observable and Latent Factors. *Journal of Empirical Finance,* 16(2):235-253.

Kallenberg O (2002): *Foundations of modern probability.* New York City: Springer-Verlag New York.

Merton RC (1974): On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *The Journal of Finance,* 29(2):449-470.

Pesaran MH, Schuermann T, Treutler BJ, Weiner SM (2006, August): Macroeconomic Dynamics and Credit Risk: A Global Perspective. *Journal of Money, Credit and Banking,* 38(5):1211-1261.

Pykhtin MV (2003): Unexpected Recovery Risk. *Risk,* 16(8):74-78.

Sommar PA, Shahnazarian H (2009, September): Interdependencies between Expected Default Frequency and the Macro Economy. *International Journal of Central Banking,* 5(3):83-110.

Vasicek OA (2015a): Probability of Loss on Loan Portfolio. In *Finance, Economics and Mathematics* (pp. 143-146). Hoboken, NJ, USA: John Wiley & Sons.

Vasicek OA (2015b): Limiting Loan Loss Probability Distribution. In *Finance, Economics and Mathematics* (pp. 147-148). Hoboken, NJ, USA: John Wiley & Sons.

Virolainen K (2004): Macro Stress Testing with a Macroeconomic Credit Risk Model for Finland. *Bank of Finland Discussion Paper No. 18/2004*.

Witzany J (2011): A Two-Factor Model for PD and LGD Correlation. In *Bulletin of the Czech Econometric Society,* 18(28):1-19.