

# Dynamic Multi-Factor Credit Risk Model with Fat-Tailed Factors<sup>\*</sup>

Petr GAPKO—Institute of Economic Studies, Faculty of Social Sciences, Charles University  
in Prague; Institute of Information Theory and Automation, Academy of Sciences  
of the Czech Republic (petr.gapko@seznam.cz)

Martin ŠMÍD—Institute of Information Theory and Automation, Academy of Sciences  
of the Czech Republic (martinsmid.eu@gmail.com)

## Abstract

*We introduce an improved multi-factor credit risk model describing simultaneously the default rate and the loss given default. Our methodology is based on the KMV model, which we generalize in three ways. First, we add a model for loss given default (LGD), second, we bring dynamics to the model, and third, we allow non-normal distributions of risk factors. Both the defaults and the LGD are driven by a common factor and an individual factor; the individual factors are mutually independent, but we allow any form of dependence of the common factors. We test our model on a nationwide portfolio of US mortgage delinquencies, modeling the dependence of the common factor by a VECM model, and compare our results with the current regulatory framework, which is described in the Basel II Accord.*

## 1. Introduction

The recent financial crisis showed significant shortfalls in banks' credit risk management and measurement processes. In particular, investments in mortgage-backed securities appeared to be much riskier than banks originally anticipated. Consequently, the subprime mortgage crisis in the US caused lots of banks to crash and triggered a worldwide debate on financial market regulation.

Current credit risk measurement techniques are mostly based on evaluation of the value-at-risk of a creditor, i.e., the amount the creditor will lose with a certain probability as a result of delinquency of debtors. The distribution of the losses is usually assumed to depend on several risk indicators, usually linked to the riskiness of the debtor and the conditions of the loan. Most credit risk models are based on two indicators: the (conditional) probability of default (PD) and the loss given default (LGD),<sup>1</sup> both of which are supposed to depend on other underlying factors. In particular, the probability of default of an individual is dependent on his/her solvency, which is usually assumed to be driven by a factor common to all debtors (i.e., the macro-economic environment) and a factor reflecting the specifics of the individual (i.e., his/her ability to increase the value of his/her own assets). The loss given default, on the other hand, is dependent on the contractual conditions of the loan, mainly on

<sup>\*</sup> Support from the Czech Science Foundation under grants 402/09/H045 and 402/09/0965 and from Charles University under grant GAUK 46108 is gratefully acknowledged. The authors are also grateful to Jan Voříšek for valuable comments and the idea of using the VECM model for the description of the factors.

<sup>1</sup> PD and LGD are usually referred to as risk factors; however, in this paper we call them "indicators" in order to verbally distinguish between these main quantities and the factors that drive them.

the value of the collateral. Collateral value is typically assumed to be driven by one or two (the common and the individual) factors; the simplest models, however, take LGD as fixed.

The Basel II (Bank for International Settlements, 2006) “Internal Rating Based” (IRB) approach to credit risk measurement assumes that LGD is fixed, while PD is modeled by the famous KMV (Merton-Vasicek) model (Vasicek, 1987, 1991, 2002). In this model, the solvency of a debtor is supposed to be driven by two standard normal factors (the common and the individual one).<sup>2</sup>

In our paper, we question three of the most restrictive assumptions of the IRB approach: the normal distribution of all factors, the fixed LGD, and the static nature of the approach. In our model, the (two) factors driving PD may follow any distribution, LGD is random and driven by two factors, and, moreover, our model is multi-periodic with the underlying factors allowed to follow a stochastic process of an arbitrary type. We show how a suitable version of our model is able to explain the credit losses observed in reality. In our opinion, our results might be useful for credit risk management in banks, specifically to determine more precisely the capital that banks need to hold to protect themselves against unexpectedly large credit losses.

This paper is organized as follows. In the first part, we summarize the current state of knowledge in the field of credit risk modeling. In the second part, we describe our proposed methodology and extensions of the current regulatory framework. Then we test our approach using empirical data and compare our results with the Basel II IRB model. Finally, we conclude and provide ideas for further research.

## **2. Current Credit Risk Measurement Methodologies**

In this section, we describe more precisely the idea of value-at-risk models for credit risk, summarize the basic facts about the Basel II requirements for credit risk modeling, and suggest ways of overcoming their shortfalls.

### **2.1 Current Credit Risk Models**

In the past three decades, the methods used by banks to determine the riskiness of their loan portfolios have evolved from simple averaging of past losses to complex models that combine the estimated riskiness of individual loans. The most influential models include CreditMetrics (RiskMetrics Group, 1997), which uses transition matrices to determine the level of defaults in a portfolio, CreditRisk+ (Wilde, 1997), which assumes a Poisson distribution for the default frequency, and the KMV model (Vasicek, 1987, 1991, 2002), used by the Basel II IRB approach and generalized in this paper. A comprehensive comparison of these methodologies can be found in Crouhy et al. (2000) and in Gordy (2000).

### **2.2 The KMV Model**

The KMV (Vasicek) model assumes that the wealth of an individual follows geometrical Brownian motion and that the values of the assets of individuals are

<sup>2</sup> Basel II is a widely known and accepted set of principles for banking capital regulation. IRB is one of several credit risk quantification methods described and allowed in Basel II. The currently proposed Basel III—the supposed successor of Basel II—uses the same risk quantification model as Basel II.

correlated, which is equivalent to saying that the individual's wealth can be decomposed into a systemic and an idiosyncratic part (see (1) and (2)). While the systemic part might be interpreted as the macroeconomic environment, the individual factor may be viewed as an ability to change one's personal wealth over time (education, health conditions, etc...).<sup>3</sup>

In particular, the KMV model assumes that the logarithm of the assets of the  $i$ -th individual fulfills

$$\log A_{i,1} = \log A_{i,0} + \eta + \gamma X_i \quad (1)$$

Here,  $A_{i,0}$  is the individual's wealth at time zero,  $\eta$  and  $\gamma$  are constants, and  $X_i$  is a random variable fulfilling

$$X_i = Y + Z_i \quad (2)$$

where  $Y$  is the common factor and  $Z_1, Z_2, \dots$  are i.i.d. individual factors, independent of  $Y$ .

*Default* is defined the state where the value of an individual's assets decreases below a certain threshold  $B_i$ ; this threshold is usually interpreted as the sum of the individual's debts (including installments at least). The probability of default is then

$$PD_i = \mathbb{P}[A_{i,1} < B_i] = \mathbb{P}[X_i < c_i] \quad c_i = \frac{\log B_i - \log A_{i,0} - \eta}{\gamma} \quad (3)$$

The KMV model assumes that the factors  $Y$  and  $Z_i, i = 1, 2, 3, \dots, n$ , are centered normal with such variances that  $\text{corr}(X_i, X_j) = \rho$  for some prescribed  $\rho$  and each  $i \neq j$ .

After some calculations we obtain the *default rate* (DR)<sup>4</sup>, defined as

$$RD = \frac{\text{number of defaults}}{\text{number of loans}}$$

which approximately fulfills

$$\mathbb{P}[RD \leq x] = N \left( \frac{(\sqrt{1-\rho})N^{-1}(x) - N^{-1}(PD)}{\sqrt{\rho}} \right) \quad (4)$$

given a sufficiently large number of loans. Here,  $N$  denotes the standard normal cumulative distribution function and  $PD = PD_1$ <sup>5</sup> (for more details of the calculation see Vašíček, 1987). It follows that the distribution of  $DR$  is heavy-tailed, with the heaviness of the tail dependent on the correlation  $\rho$ .

<sup>3</sup> The systemic factor is exogenous to both the KMV and to our model. For interesting research into the relations of systemic factors among various financial and insurance sectors, see Billio et al. (2012).

<sup>4</sup> The quantity which we call DR is sometimes called the *empirical* or *observed PD*. We use a different name so as not to suggest that DR is an estimate of PD (it is clear from (4) that DR is neither unbiased nor consistent).

<sup>5</sup> Note that  $PD_i = PD$  for any  $i$  because the individual factors are equally distributed.

Finally, since LGD is fixed, we may take it as a unit without any loss of generality. Thus, in the KMV model the credit loss  $L$  of the portfolio equals  $DR$ .

### 2.3 Existing Models with Random LGD

The biggest shortfall of the original Vasicek model usually discussed in the literature (see, for example, Cipollini and Missaglia, 2008) is the absence or randomness of LGD. Several recent models assume a non-random LGD; however, as far as we know, none of these studies challenged the assumption of standard normal distribution of the risk factors. In this sub-section we describe several of the most popular models of this kind.

The simplest (and the most natural) enhancement of the Vasicek model for LGD is the one proposed in Frye (2000), which assumes that LGD is a second risk indicator driving credit losses. In this model, LGD is a function of collateral:

$$LGD_i = \max[0; 1 - \text{Collateral}_i]$$

while the collateral value is expressed as

$$\text{Collateral}_i = \mu_i(1 + \sigma_i C_i)$$

where  $C_i$  is the risk factor, which can be further expressed as a function of a systemic risk factor  $Y$  identical to that driving defaults and a specific risk factor  $E_i$ , i.e.,

$$C_i = \sqrt{q}Y + \sqrt{1-q}E_i \quad (5)$$

The loss distribution is taken from the Vasicek framework (i.e., fulfilling (1)) with

$$X_i = \sqrt{p}Y + \sqrt{1-p}Z_i \quad (6)$$

which implies that the correlation between defaults and LGD is determined by how factors  $X_i$  and  $C_i$  depend on factor  $Y$ .

An extension of the Frye model can be found in Pykhtin (2003), who supposes that the risk factor driving LGD depends on one systemic and two idiosyncratic factors:

$$\begin{aligned} C_i &= \sqrt{q}Y + \sqrt{1-q}E_i \\ E_i &= \sqrt{w}Z_i + \sqrt{1-w}E_i' \end{aligned} \quad (7)$$

where the systemic factor  $Y$  is common to both defaults and LGD. In this framework factor  $Z_i$  also influences the idiosyncratic factor driving defaults (factor  $E_i'$  is specific to LGD). The correlation between the two idiosyncratic factors is  $w$ . In practice, this approach is used by the Moody model (Meng et al., 2010).

Another extension of the KMV model can be found in Witzany (2011). In this model LGD is assumed to be driven by a specific factor different from the one driving defaults and by two systemic factors, one common to the defaults and the other specific to LGD.

### 3. Our Approach

In our proposed model, we, similarly to Frye (2000) and Pykhtin (2003), assume a random LGD. However, we look at defaults and LGD separately first and then offer ideas about how these two can be linked through dynamic dependence of their underlying factors. While the sub-model for defaults is a generalization of Vasicek's approach, the LGD sub-model is a new one, making few assumptions but naturally explaining LGD as a function of the price of collateral. As to the evolution of the factors, we allow maximum generality; in fact, we only show how to "plug in" any model of the factors into our approach.

#### 3.1 Model for Defaults

Analogously to Vasicek, we assume that

$$\log A_{i,t} = \log A_{i,t-1} + \Delta Y_t + U_{i,t}, \quad i \leq n \quad (8)$$

where  $n$  is the number of borrowers,  $A_{i,t}$  is the wealth of the  $i$ -th borrower at time  $t \in \mathbb{N}$ ,  $U_{i,t}$  is a random variable specific to the  $i$ -th borrower, and  $\Delta Y_t = Y_t - Y_{t-1}$  is the first difference of the common factor  $Y_t$  following a general (adapted) stochastic process. Such a setting makes sense, for instance, if  $Y_t$  stands for (the logarithm of) a stock index; then, our model corresponds to the situation where a borrower owns a portfolio with the same composition as the index plus some additional assets.

For simplicity, we assume that the duration of the debt is exactly one period and that the initial wealth in each period equals

$$\log A_{i,t-1} = Y_{t-1} + V_{i,t}, \quad i \leq n \quad (9)$$

where  $V_{i,t}$  is a random variable specific to the  $i$ -th borrower. Further, we assume all  $(U_{i,t})_{i \leq n, t \in \mathbb{N}}$  to be mutually independent and independent of  $(\Delta Y_t)_{t \in \mathbb{N}}$ , and all  $Z_{i,t}, Z_{i,t} = U_{i,t} + V_{i,t}$ ,  $i \leq n$ ,  $t \in \mathbb{N}$  to be identically distributed with  $\mathbb{E}Z_{1,1} = 0$ ,  $\text{var}(Z_{1,1}) = \sigma$ ,  $\sigma > 0$ ,  $Z_{1,1}$ , having a strictly increasing continuous cumulative distribution function  $\Psi$ . Since the equation for wealth may be scaled, we can assume that  $\sigma = 1$ . Note that we do not require the increments of  $Y_t$  to be centered.

Even though the assumption of one-period duration of debts may seem very restrictive, in fact it is not; even if the total duration of a mortgage is measured in decades, the periods between the re-fixing of interest rates, at the end of which the mortgage may be repaid, are much shorter (sometimes as little as one year).<sup>6</sup>

It follows from our independence assumptions that the (conditional) probability of default of the  $i$ -th borrower at time  $t$  given  $\bar{Y}_t := (\Delta Y_1, \dots, \Delta Y_{t-1})$  equals

$$\mathbb{P}[A_{i,t} < B_{i,t} | \bar{Y}_t] = \mathbb{P}[Z_{i,t} < \log B_{i,t} - Y_t | \bar{Y}_t] = \Psi(\log B_{i,t} - Y_t) \quad (10)$$

where  $B_{i,t}$  are the debts of the  $i$ -th borrower at time  $t$ .

<sup>6</sup> A multi-period version of our model may also be formulated (see Šmíd and Gapko, 2010). However, this is tractable only by means of Monte Carlo simulation.

Our primary topic of interest is the default rate (DR), defined as:

$$R_t = \frac{\text{number of defaults at } t}{n}$$

If we assume the debts to be the same for all borrowers and at all times, i.e.,  $\log B_{i,t} = b, t \in \mathbb{N}, i \leq n$ , for some  $b$ , and if we approximate  $R_t \doteq \lim_n \frac{\text{number of defaults at } t}{n}$ , we may apply the Law of Large Numbers to the conditional probabilities described in (10) (we may do this since  $A_{1,t}, A_{2,t} \dots$  are conditionally independent given  $\bar{Y}_t$ ) to obtain (for a very large portfolio):

$$R_t \doteq \mathbb{P}[A_{i,t} < b \mid \bar{Y}_t] = \Psi(b - Y_t), \quad t \in \mathbb{N}$$

further implying that

$$\Delta Y_t \doteq \Psi^{-1}(R_{t-1}) - \Psi^{-1}(R_t)$$

and

$$R_t \doteq \Psi(\Psi^{-1}(R_{t-1}) - \Delta Y_t) \quad (11)$$

The latter formula roughly determines the dynamics of the process of losses, while the former one allows us to statistically infer the common factor based on the time series of the rates of default.

Furthermore, we shall assume that factor  $Z$  is normal, i.e.,  $\Psi$  is the cumulative distribution function (CDF) of the standard normal distribution.

### 3.2 Model for LGD

Our model for LGD is analogous to our version of the default model. However, contrary to the Frye and Pykhtin models, we assume a separate common factor driving LGD. This choice is quite natural, as the systemic conditions driving defaults are different from those driving LGD: while defaults depend on many different variables (e.g. average wage, unemployment rate, and real estate prices), losses given default depend mainly on real estate prices. Note that we do not assume independence of the factors driving defaults and LGD; as we show below, we allow for any form of stochastic dependence on each other as well as on the past values of both factors.

Coming to the definitions, we assume that the property price of the  $i$ -th defaulted debtor is

$$\log P_{i,t} = \log a_i + I_t + E_{i,t} \quad (12)$$

(or, equivalently,  $P_{i,t} = a_i \exp\{I_t\} \exp\{E_{i,t}\}$ ), where  $I_t$  is an (unobservable) common factor underlying LGD following a general adapted process,  $E_{i,t}$  is a centered individual factor independent of  $(I_t, Y_t)_{t \geq 0}$  and all the individual factors described in subsection 3.1 (i.e.,  $U_i$ ,  $V_i$ , and  $Z_i$ ), and  $a_i$  is a constant reflecting the ratio of the  $i$ -th debtor's property price to the common factor.

Let  $C_i$  be the size of the  $i$ -th debt, including the cost of recovery. Then the recovered percentage of the  $i$ -th debt at time  $t$  is

$$G_i = \frac{\min(P_{i,t}, C_i)}{C_i}$$

Furthermore, let us say that  $C_i = C$ ,  $a_i = a, i \leq N$  and let  $E_{1,t}, E_{2,t}, \dots$  be i.i.d. Given all this, we may assume without any loss of generality that  $C = 1, a = 1$  (the constants may now be incorporated into  $\Delta I$ ). Then

$$G_i = \min(e^{I_t + E_{i,t}}; 1) = \exp\{\min(I_t + E_{i,t}, 0)\}$$

If there is a large number of defaulted debtors, then the average of  $G_i$  is, by the Law of Large Numbers,

$$\tilde{G}_t = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N G_i = \mathbb{E}(G_1 | I_t)$$

Evaluating the right-hand side (and omitting the time index), we get

$$\begin{aligned} \tilde{G} &= \mathbb{E}\left(e^I e^{\min(E_1, -I)} | I\right) = e^I \mathbb{E}\left(e^{\min(E_1, -I)} | I\right) = e^I \left[ \int_{-\infty}^{-I} e^x dF(x) + e^{-I} (1 - F(-I)) \right] \\ &= e^I \int_{-\infty}^{-I} e^x dF(x) + 1 - F(-I) \end{aligned} \quad (13)$$

where  $F$  is the cumulative distribution function (CDF) of  $E_1$ . Consequently, the LGD equals

$$D_t = 1 - \tilde{G}_t = h(I_t) \quad (14)$$

where

$$h(t) = F(-t) - e^t \int_{-\infty}^{-t} e^x dF(x) \quad (15)$$

or, after integrating by parts,

$$h(t) = e^t \int_{-\infty}^{-t} F(x) e^x dx \quad (16)$$

As shown in the *Appendix*,  $h$  is strictly decreasing, hence its inverse exists.

Assume further that  $E_1$  is normal with variance  $\sigma^2$ . Then  $F(x) = \Phi(x/\sigma)$ , where  $\Phi$  is the standard normal CDF and

$$h(t) = h_\sigma(t) = \Phi\left(-\frac{t}{\sigma}\right) - \exp\left(t + \frac{1}{2}\sigma^2\right) \Phi\left(-\frac{t}{\sigma} - \sigma\right) \quad (17)$$

$$h'_\sigma(t) = \exp\left(t + \frac{1}{2}\sigma^2\right) \left[ \frac{1}{\sigma} \varphi\left(-\frac{t}{\sigma} - \sigma\right) - \Phi\left(-\frac{t}{\sigma} - \sigma\right) \right] - \frac{1}{\sigma} \varphi\left(-\frac{t}{\sigma}\right)$$

where  $\varphi$  is the standard normal probability density function and where the derivative of  $h$  is with respect to  $t$ . For the calculation of (17), see the *Appendix*.

### 3.3 Econometrics of the Model

As already said, we place no special requirements on the (vector) process  $(Y_t, I_t)$ . We will only assume that the process may be transformed into independent residuals in the sense that there exist mappings  $Q_1, Q_2, \dots$  such that

$$Q_t(\omega_t; \lambda) = \varepsilon_t, \quad \omega_t = (Y_1, I_1, Y_2, I_2, \dots, Y_t, I_t)$$

for each  $t$ , where  $\lambda$  is a (vector) parameter and  $\varepsilon_1, \varepsilon_2, \dots$  is a sequence of i.i.d. two-dimensional random variables whose density  $\eta$  possibly depends on a (vector) parameter  $\mu$ . Given this assumption and some invertibility and differentiability conditions (which would be better tested in concrete cases), the conditional density of  $(Y_t, I_t)$  given  $\omega_{t-1}$  is, by the formula for transformed density,

$$v_t(y, t; \lambda, \mu) = \eta(Q_t(\omega_{t-1}, y, t; \lambda); \mu) |D_t(y, t)|$$

where  $D_t(y, t)$  is the Jacobian determinant of  $Q_t$ , restricted to the last two variables.

Suppose now that we have a sequence of historical RDs and LGDs  $R_1, D_1, R_2, D_2, \dots, R_T, D_T$  at our disposal and we want to estimate all parameters of our model, i.e.,  $\lambda, \mu$ , and  $\sigma$ . A straightforward way to do this is by maximum likelihood estimation, with the likelihood function taking the form of

$$L(R_1, D_1, \dots, R_T, D_T; \lambda, \mu, \sigma) = \tag{18}$$

$$= \sum_{t=1}^T \left[ \log(\eta(Q_t(\omega_t))) + \log(|D_t(Y_t, I_t)|) - \log(\Psi'(\Psi^{-1}(R_t))) - \log(-h'_\sigma(h_\sigma^{-1}(D_t))) \right]$$

(recall that  $Y_t = \Psi^{-1}(R_{t-1}) - \Psi^{-1}(R_t)$ ,  $I_t = h_\sigma^{-1}(D_t)$ ). Note that the third term in the square brackets may be omitted during the maximization because it does not depend on any parameter.

## 4. Empirical Results

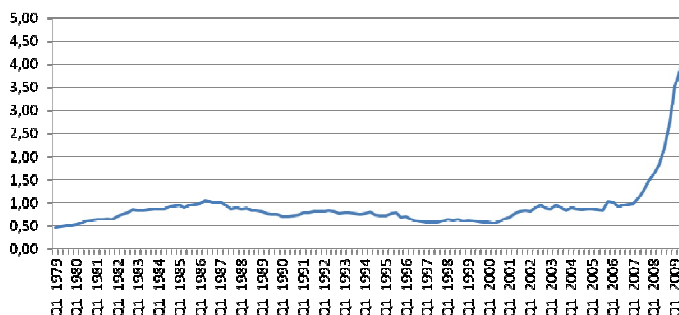
We empirically tested our proposed methodology on a nationwide retail mortgage portfolio and compared the results with the Basel II IRB framework. In this section, we provide a detailed description of the datasets we used, the estimation process, and the results.

### 4.1 Description of the Data

The dataset for our empirical work consists of quarterly delinquency rates on mortgage loans from the whole US economy and was provided by the US Department of Housing and Urban Development and the Mortgage Bankers Association.<sup>7</sup> All data start with the first quarter of 1979 and end with the third



**Figure 1 The US 90+ Delinquency Rates—the Proxy for DR**



quarter of 2009. Thus, the difficult period of the subprime mortgage crisis and the subsequent real recession is included.

## 4.2 Estimation

To estimate our model, we proceeded as follows. First, we extracted factor  $Y$  from the values of  $R$ . Second, we computed factor  $I$  from the values of  $D$  by employing  $h$  specified in (17); since the function  $h$ , which maps  $D$  to  $I$ , depends also on parameter  $\sigma$ , we did the transformation for a sufficient number of values of  $\sigma$ . Third, we found a suitable model for the dynamics of the pair  $(Y, I)$ . Finally, we estimated the model of the series  $(Y, I)$  for each  $\sigma$  and chose the version with the highest likelihood.

### 4.2.1 Extraction of $Y$

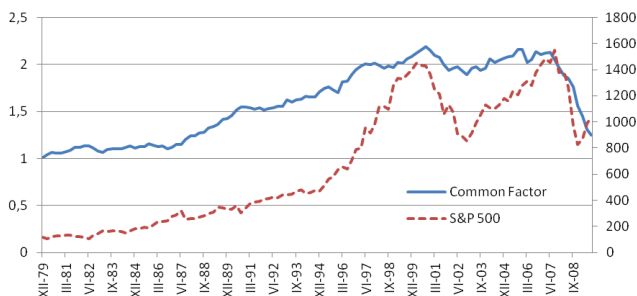
As a proxy for the default rate (denoted by  $R$ ), we used the series of 90+ delinquency rates<sup>8</sup> depicted in *Figure 1*. We can see that the number started growing significantly at the end of 2007. During the estimation process, we used two types of delinquency rates: quarterly delinquencies and their yearly averages. The average delinquencies were used for the computation of the Basel II IRB capital requirement because the IRB method requires a long-term average probability of default as an input. The quarterly delinquency rates, on the other hand, served as the input data for our model.

The values of the common factor “ $Y$ ” were computed by means of (11). To verify our conjecture that the common factor may coincide with a stock index, we compared graphically the values of the common factor with the S&P 500 stock index (see *Figure 2*). It can be seen that the evolution of the common factor exhibits similarities to the stock index. A simple linear correlation analysis indicates that the common factor is lagged behind the stock index by two quarters and that both datasets are significantly correlated (the value of the Pearson correlation coefficient is about 30%, which is significant at 5%).

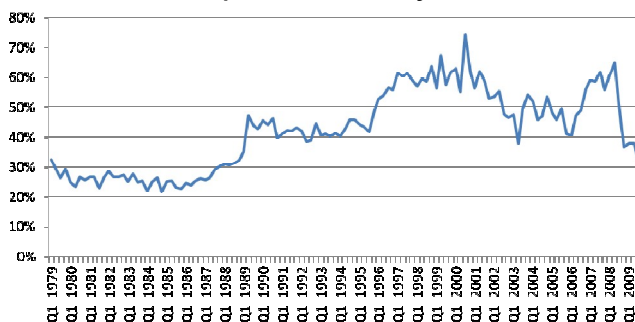
<sup>7</sup> The Mortgage Bankers Association is the largest US society representing the US real estate market, with over 2,400 members (banks, mortgage brokers, mortgage companies, life insurance companies, etc.)

<sup>8</sup> The 90+ delinquency rate is the proportion of all receivables 90 or more days past due in a given quarter.

**Figure 2 Comparison of the Common Factor and the Lagged S&P 500 Index**  
(values of the common factor on the left-hand scale;  
values of the S&P 500 on the right-hand scale)



**Figure 3 Foreclosures/90+ Delinquencies—the Proxy for LGD**



#### 4.2.2 Extraction of $I$

As a proxy for the LGD (denoted by  $D$  in our paper), the proportion of started foreclosures<sup>9</sup> in the 90+ delinquency rates was used. Unfortunately, the proxy cannot be exact, because it does not include income collected from the sale of debtors' property; however, it at least gives us an idea of how large the losses would be in the case of no real estate collateral. The resulting series of  $D$  is plotted in *Figure 3*.

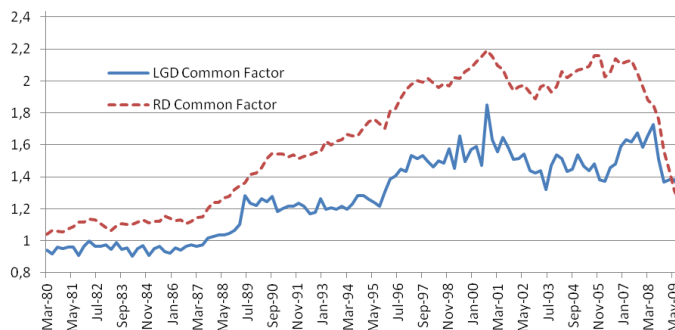
It is very interesting that, in the several recent periods, when the 90+ delinquency rate increased significantly, the ratio of seriously delinquent (defaulted) accounts which fell into the foreclosure process decreased. This can be intuitively explained by state aid under which the Fed bought a non-negligible amount of bad loans, especially from the mortgage market.

#### 4.2.3 Selection of the Model for $(Y, I)$

After a preliminary analysis of the series of  $Y_t$  we found clear ARCH behavior of the factor, hence we decided to analyze the transformed version of the factor instead of its original values  $y_t = \frac{\Delta Y_t}{|\Delta Y_{t-1}|}$ .

<sup>9</sup> Foreclosure is a process whereby a creditor ceases all attempts to force a debtor to repay a seriously delinquent debt. The loan is treated as a loss and a late collection process begins. The creditor collects the debtor's property and tries to sell it on the real estate market.

**Figure 4 Graphical Comparison of DR and LGD Common Factors**



Further, for a sufficiently dense set of the values of  $\sigma$ , we extracted  $I$  by means of the inversion of  $h$  and fitted the (vector) time series  $(y, I)$  using a vector error correction model (VECM) with one lag, i.e.,

$$\Delta y_{t-1} = \alpha_1 + \beta_1 \Delta y_{t-1} + \gamma_1 \Delta I_{t-1} + \delta_1 e_{t-1} + \varepsilon_{1,t}$$

$$\Delta I_t = \alpha_2 + \beta_2 \Delta y_{t-1} + \gamma_2 \Delta I_{t-1} + \delta_2 e_{t-1} + \varepsilon_{2,t}$$

where  $\Delta y$  and  $\Delta I$  are the first differences of  $y$  and  $I$  and  $e$  is an error correction term. For each of the examined values of  $\sigma$ , we computed the maximum likelihood function of the VECM model by means of (18) and chose  $\sigma = 12\%$  as the estimate of  $\sigma$  since this value gave the greatest likelihood. We found it very interesting that the estimated  $\sigma$  intuitively corresponds to the standard deviation of real estate prices (Quigley, 1999).

Figure 4 compares the two common factors. It is obvious at first glance that these two show some similarities.

The resulting VECM model with  $Y$  as the dependent variable in the first equation and  $I$  in the second one is summarized in Table 1 (in accordance with the definition of the model, cointegration rank 1 was assumed).

To determine whether our choice of cointegration rank was correct, we performed both the Engel-Granger and the Johansen cointegration tests. Both tests confirmed cointegration of rank one. Moreover, we see that the (transformed) factor  $Y$  depends on the past value of both factors, while factor  $I$  does not show dependence on the past (except the one caused by the cointegration). The  $R^2$  of the whole model is around 30%.

Since normality of the residuals from the VECM model was rejected (with  $p$ -value lower than 0.01), we additionally fitted the residuals using the generalized hyperbolic distribution. This distribution was first described in Barndorff-Nielsen (1977), and it has been shown that it is able to describe financial time series more realistically than, for example, the standard normal distribution (Eberlein and Keller, 1995). The choice of distribution is based on Gapko and Šmíd (2010), where the authors found that the class of generalized hyperbolic distributions best fits the increments of the  $Y$  factor.

**Table 1 Estimated Coefficients of the VECM Model**

1 <sup>st</sup> equation (y dependent)	Coefficient Value (SE)	p-value
Constant	0.552233 (0.101865)	3.44E-07
$\Delta y_{t-1}$	-0.169582 (0.0914292)	0.0663
$\Delta I_{t-1}$	0.111233 (0.0286587)	0.0002
Error Correction Term	-0.534066 (0.0982860)	3.26E-07
2 <sup>nd</sup> equation (I dependent)	Coefficient Value (SE)	p-value
Constant	-0.299802 (0.321560)	0.3532
$\Delta y_{t-1}$	-0.106660 (0.288617)	0.7124
$\Delta I_{t-1}$	-0.362746 (0.0904674)	0.0001
Error Correction Term	0.293693 (0.310262)	0.3459

Before the end of this section, let us describe the derivation of the ML function (18) in detail. First, note that

$$Q_t = \begin{bmatrix} \frac{\Delta Y_t}{|\Delta Y_{t-1}|} - \alpha M_t & 0 \\ 0 & \Delta I_t - \beta S_t \end{bmatrix}$$

where  $S_t$  and  $M_t$  are matrices possibly containing past values of both (transformed) factors  $Y$  and  $I$ . Since, in (18), the term  $\log |D_t(Y_t, I_t)| = -\log(|\Delta Y_{t-1}|)$  does not depend on any parameter, it can be excluded from the maximization, so the ML estimate can be obtained by maximizing

$$\begin{aligned} L(R_1, D_1, \dots, R_T, D_T; \lambda, \mu, \sigma) &= \sum_{t=1}^T \left[ \log(\eta_1(\varepsilon_t^1)) + \log(\eta_2(\varepsilon_t^2)) - \log(-h'_\sigma(h_\sigma^{-1}(D_t))) \right] \\ &= \dot{L}(Q_t(\omega_t)_{t=1..T}) - \sum_{t=1}^T \log(h'_\sigma(h_\sigma^{-1}(D_t))) \end{aligned} \quad (19)$$

where  $\dot{L}$  is the likelihood function of the VECM model,  $(\varepsilon_1^i, \varepsilon_2^i, \dots)$  are the residuals from the  $i$ -th equation of the VECM model, and  $\eta_i$  is the density of the residuals (keep in mind, however, that the residuals depend on the parameters of the VECM model).

*Remark:* To be rigorous, we did not proceed exactly according to Section 3 because we did not maximize the parameters of the VECM model and of the residuals “at once”. However, since both estimations are already implemented (in *R* language), it seems reasonable to use the existing methods—to estimate the VECM first and then to fit the residuals. However, we pay a price for this simplification: our estimate becomes a quasi maximum likelihood one instead of a maximum likelihood one (because least squares estimation is an ML one only given normality of the residuals).

### 4.3 Predictions

Having the model, we computed the quantiles of both DR and LGD on the 99.9th percentile probability level, i.e., on the level used in the Basel II framework.

**Table 2 Comparison of Basel II and Dynamic GHD Models Tail DR**

Model	Basel II IRB	Our model
Distribution used for the individual factor	Standard Normal	Standard Normal
Distribution used for the common factor	Standard Normal	Generalized Hyperbolic
99.9% loss	10.3%	7.2%

**Table 3 Selected LGD Quantiles in Our Model**

99th quantile LGD	99.9th quantile LGD	99.99th quantile LGD
29.8%	40.6%	50%

During the estimation, we had to solve a technical problem. The common practice is to measure credit risk over a one-year horizon, while our dataset is based on quarterly observations. In order to get one-year predictions exactly, we would need to calculate convolutions of the (generalized hyperbolic) residuals, which would lead to complicated integral expressions. Therefore, we decided instead to use simulations for four consecutive quarters, using the formula

$$R_{t+4} \doteq \Psi \left( \Psi^{-1}(R_t) - \sum_{1 \leq i \leq 4} \Delta Y_{t+i} \right) \quad (20)$$

which can be easily achieved by using (11) four times consecutively.

#### 4.3.1 Quantile of DR

As was said in Section 3, we assumed that the distribution of the individual factor driving defaults, denoted by  $Z$ , is standard normal.

We compared the quantiles of DR calculated by our proposed methodology and those obtained by the Basel II IRB method (assuming standard normal distributions for both risk factors and a 15% correlation between the factors<sup>10</sup>). The result is summarized in *Table 2*.

The results show that our model predicts a lower value of the quantile of DR than the IRB formula, which may seem surprising in light of the fact that we rejected normality of the residuals in favor of a fat-tailed distribution. However, if we keep in mind that we use information from the past to estimate the distribution of the factor (which the static model does not), we are able to “predict” the factor more exactly. This decreases the uncertainty in the model and thus explains the lower value of the quantile.

#### 4.3.2 Quantile of LGD

Similarly to DR, we computed the quantiles of LGD (by means of simulations again). The resulting 99.9th LGD quantile calculated by our model, 40.6%, is slightly below the regulatory 45% benchmark. The other computed quantiles are summarized in *Table 3*.

<sup>10</sup> The 15% correlation is the benchmark set for mortgage exposures in the Basel II framework.

## 5. Conclusion

We proposed a new model for quantifying credit risk, widely generalizing the IRB approach implemented in the Basel II regulatory framework. In particular, we extended the original model framework so that both DR and LGD are considered, each being driven by one common and one individual factor. In our proposed methodology, nearly any dynamic stochastic model may be used to describe the dynamics of the (common) factors.

We applied our model to real data, specifically to the time series of serious credit delinquencies in the nationwide US mortgage market. We used a VECM model with generalized hyperbolic residuals as the model for the common factors. Based on the model, we evaluated the quantiles for both DR and LGD, finding that our results are comparable with the levels prescribed by Basel II. In particular, our results show that the Basel II framework gives both higher DR and higher LGD than our model. This is because our model, employing dynamics, gives more precise forecasts of both factors. In the Basel II methodology with static models, information from the past is not exploited. Consequently, our results show that the current regulatory framework may overestimate credit losses, which may result in higher capital requirements and thus higher customer interest rates on loans.

The proposed methodology could be used as part of internal capital adequacy measurement in banks or other financial institutions. However, there are still some unresolved questions and suggestions for future research, including more detailed analysis of the relationship between DR and LGD and an empirical analysis of the model on a single bank's portfolio.

## APPENDIX

In the Appendix, we provide mathematical details concerning the function  $h$  defined in Section 3.1. First we specify its derivative:

$$\begin{aligned}
 h'(t) &= e^t \left( \int_{-\infty}^{-t} F(x) e^x dx - F(-t) e^{-t} \right) \\
 &= e^t \left( \int_{-\infty}^{-t} F(x) e^x dx - F(-t) \int_{-\infty}^{-t} e^x dx \right) \\
 &= e^t \int_{-\infty}^{-t} [F(x) - F(t)] e^x dx < 0
 \end{aligned}$$

Second, we evaluate the function given that  $E_1$  is normal with variance  $\sigma^2$ :

$$\begin{aligned}
 h_\sigma(t) &= \Phi\left(-\frac{t}{\sigma}\right) - \exp(t) \int_{-\infty}^{-t} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{\sigma^2}\right) \exp(x) dx \\
 &= \Phi\left(-\frac{t}{\sigma}\right) - \exp(t) \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{-t} \exp\left(-\frac{x^2}{\sigma^2} + x\right) dx \\
 &= \Phi\left(-\frac{t}{\sigma}\right) - \exp(t) \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{-t} \exp\left(-\frac{1}{2\sigma^2}(x^2 - 2x\sigma^2 + \sigma^4) + \frac{1}{2}\sigma^2\right) dx \\
 &= \Phi\left(-\frac{t}{\sigma}\right) - \exp(t) \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{-t} \exp\left(-\frac{1}{2\sigma^2}(x - \sigma^2)^2 + \frac{1}{2}\sigma^2\right) dx \\
 &= \Phi\left(-\frac{t}{\sigma}\right) - \exp\left(t + \frac{1}{2}\sigma^2\right) \int_{-\infty}^{-t} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \sigma^2)^2}{2\sigma^2}\right) dx \\
 &= \Phi\left(-\frac{t}{\sigma}\right) - \exp\left(t + \frac{1}{2}\sigma^2\right) \mathbb{P}\left[N(\sigma^2, \sigma^2) < -t\right] \\
 &= \Phi\left(-\frac{t}{\sigma}\right) - \exp\left(t + \frac{1}{2}\sigma^2\right) \Phi\left(-\frac{t}{\sigma} - \sigma\right)
 \end{aligned}$$

(recall that  $\Phi$  is the standard normal CDF).

## REFERENCES

- Bank for International Settlement (2006): *Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework*. Retrieved from <http://www.bis.org/publ/bcbs128.htm>
- Barndorff-Nielsen O (1977): Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society London*, A353:401–419.
- Billio M et al. (2012): Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics* (in press).
- Cipollini A, Missaglia G (2008): Measuring bank capital requirements through dynamic factor analysis. *RECent Centre for Economic Research Working Paper Series*, WP 10.
- Crouhy M et al. (2000): A comparative analysis of current credit risk models. *Journal of Banking & Finance*, 24(1-2):59–117.
- Eberlein E, Keller U (1995): Hyperbolic distributions in finance. *Bernoulli*, 1(3):281–299.
- Frye J (2000): Collateral Damage. *Risk*, 13(April):91–94.
- Gapko P, Šmíd M (2010): Modeling a Distribution of Mortgage Credit Losses. (Charles University, Faculty of Social Sciences) *IES Working Papers*, WP 32/2000.
- Gordy MB (2000): A comparative anatomy of credit risk models. *Journal of Banking & Finance*, 24(1/2):119–149.
- Meng Q et al. (2010): *Implications of PD-LGD Correlation in a Portfolio Setting*. Moody's Analytics. Retrieved from: <http://www.moodyanalytics.com/~media/Insight/Quantitative-Research/Portfolio-Modeling/10-05-02-Implications-of-PD-LGD-Correlation-in-a-Portfolio-Setting.ashx>
- Pykhtin MV (2003): Unexpected Recovery Risk. *Risk*, 16(8):74–78.
- Quigley JM (1999): Real Estate Prices and Economic Cycles. *International Real Estate Review*, 2(1):1–20.
- RiskMetrics Group, Inc. (1997): *CreditMetrics Technical Document*. New York, JP Morgan & Co. Retrieved from: <http://www.ma.hw.ac.uk/~mcneil/F79CR/CMTD1.pdf>
- Šmíd M, Gapko P (2010): Dynamic model of losses of creditor with a large mortgage portfolio. *Proceedings of the 47th Meeting of the EURO Working Group on Financial Modelling* (pp. 153–162). České Budějovice, VŠB-Technical University of Ostrava.
- Vasicek OA (1987): *Probability of Loss on Loan Portfolio*. KMV. Retrieved from: <http://www.moodyanalytics.com/~media/Insight/Quantitative-Research/Portfolio-Modeling/87-12-02-Probability-of-Loss-on-Loan-Portfolio.ashx>
- Vasicek OA (1991): *Limiting Loan Loss Probability Distribution*. KMV. Retrieved from: <http://www.moodyanalytics.com/~media/Insight/Quantitative-Research/Portfolio-Modeling/91-08-09-Limiting-Loan-Loss-Probability-Distribution.ashx>
- Vasicek OA (2002): The Distribution of Loan Portfolio Value. *Risk*, 15(12):160–162.
- Wilde T (1997): *Encyclopedia of Quantitative Finance*. John Wiley & Sons, Ltd., New York.
- Witzany J (2011): A Two-Factor Model for PD and LGD Correlation. *Bulletin of the Czech Econometric Society*, 18(28).