

# Multi-Period Structural Model of a Mortgage Portfolio with Cointegrated Factors<sup>\*</sup>

Petr GAPKO, *corresponding author* (petr.gapko@seznam.cz)

Martin ŠMÍD

both authors: Econometric Department, Institute of Information Theory  
and Automation, Czech Academy of Sciences, Prague

## Abstract

*We propose a new dynamic two-factor model of a loan portfolio. Following the common approach, we quantify the credit risk associated with the portfolio by the probability of default and the loss given default, each of which is driven by a factor common for all debts in the portfolio, and a factor individual to each debt. In line with the empirical evidence, the individual factors are assumed to be AR(1) processes. The common factors, on the other hand, may be dependent on the external (macroeconomic) environment. We apply our model to the US nationwide mortgage portfolio, fitting the dynamics of the factors with a VECM model with several macroeconomic indicators as exogenous variables.*

## 1. Introduction

In 2009, when the financial crisis fully hit the US economy, losses from real estate loans in the US increased by a factor of ten compared with the relatively quiet period ending in 2007. The aim of our paper is to contribute to the discussion on the mechanics causing events of this type. One of possible ways of studying these mechanics is to identify the key risk factors driving losses and use an appropriate model of the factors to forecast future loss behavior.

We propose a structural factor model of the Merton-Vasicek type (for similar models, see Vasicek, 1987; Frye, 2000; Pykhtin, 2003; Jimenez and Mencia, 2009; or Witzany, 2011). In these structural models, defaults and losses given default of individual loans depend on the wealth of creditors and on the price of collaterals with which the loans are secured, respectively. In our approach, the wealth and the collateral price are each assumed to be driven by two factors, one individual to each debtor and the other common to all debtors. Generally, the common factors may be understood as a quantified influence of the external environment, e.g. the macroeconomic situation, and the individual factors might be interpreted as an individual's ability to maintain wealth, i.e. features specific to a collateral.

The idea of connecting structural credit risk models with macroeconomic models is quite common in the literature; see, for example, e.g. Pesaran *et al.* (2003) or Virolainen (2004). From the recent works, it is worth noting, for example, Hamerle *et al.* (2011), who showed, on a bond portfolio, that changes in the macroeconomic environment play a significant role by comparing a point-in-time multi-factor credit risk model with a through-the-cycle one, or Sommar and Shahnazarian (2009), who used the vector error correction model to estimate the dependency of the expected default frequency of a portfolio of nonfinancial listed companies

<sup>\*</sup> This work was supported by the Czech Science Foundation under Grant No. GA15-10331S.

on several macroeconomic factors, from which they found the most influential interest rate.

Compared to the aforementioned models, the original contribution of our approach is threefold. First, we take the dynamics of the debtors' wealth within the portfolio into account; in particular, we assume the individual factors to be an AR(1) process, which corresponds to the empirical evidence (see Hochguertel & Ohlsson, 2011). Second, we introduce a multi-generation approach, i.e. we let new debtors enter the portfolio in each period. Finally, we use cointegration analysis with exogenous macroeconomic variables to fit the joint dynamics of the common factors, which, to the best of our knowledge, has never been done before.

In the empirical part of our paper, we apply our model to the 30+ delinquency<sup>1</sup> and charge-off<sup>2</sup> rates of the US nationwide residential mortgage portfolio. As a result, we find that there exists a set of statistically significant macroeconomic influencers of the risk factors, namely GDP, CPI inflation and the FED base interest rate.

The paper is organized as follows: In Section 2, we introduce the model. In Section 3, we describe the empirical analysis, data and results. Section 4 concludes the paper.

## 2. The Model

As was already premised, we model the dynamics of a portfolio of loans. For simplicity, we assume that the financed amount is the same for all loans in the portfolio, a unit without loss of generality. New loans enter the portfolio at each time  $\tau$ ,  $N^\tau$ . The ratio of the newly coming loans ( $N^\tau$ ) to the overall size of the portfolio is assumed to be  $1/\min(m, \tau)$  for each  $\tau$ , where  $m$  is the duration of the loans, later assumed to be equal to 30. Similarly to Vasicek (1987), we assume that the disposable wealth  $A_t^{\tau,i}$  of the  $i$ -th debtor from the  $\tau$ -th generation (i.e. with the first repayment at time  $\tau$ ) follows a “trended” Geometrical Brownian Motion with a common factor as a “trend”, i.e.

$$A_t^{\tau,i} = \exp\{Y_t + Z_t^{\tau,i}\}, \quad t \geq \tau \quad (1)$$

where

$Y$  (common factor) is a general stochastic process,

$Z^{\tau,i}$  (individual factor) is a stochastic process such that

$$Z_\tau^{\tau,i} \sim N(0, \sigma_1) \text{ for some } \sigma_1 > 0,$$

$$Z_t^{\tau,i} = \phi Z_{t-1}^{\tau,i} + U_t^{\tau,i}, U_t^{\tau,i} \sim N(0, \sigma), \quad t > \tau, \text{ for some constants } \phi \in R, \sigma > 0$$

(note that the distribution of the initial wealth differs among the generations, as it is dependent on the common factor).

Analogously to the wealth and similarly to Frontczak and Rostek (2015), we assume that the price of the collateral of the  $i$ -th loan from the  $\tau$ -th generation is

<sup>1</sup> Loans more than 30 days past due.

<sup>2</sup> Net charge-offs of loans from books.

initially equal to the (unit) size of the loan and later it follows dynamics similar to the wealth:

$$P_{\tau-1} = 1, \quad P_t^{\tau,i} = \exp\left\{(I_t - I_{\tau-1}) + E_t^{\tau,i}\right\}, \quad t \geq \tau \quad (2)$$

where

$I$  (the collateral common factor) is a general stochastic process, and

$E_t^{\tau,i}$  is a stochastic process fulfilling  $E_t^{\tau,i} = \psi E_{t-1}^{\tau,i} + V_t^{\tau,i}$ ,  $V_t^{\tau,i} \sim N(0, \rho)$ ,  $t \geq \tau$  and  $\psi \in R$  ( $E_{\tau-1}^{\tau,i} = 0$  by definition).

As is usual in structural factor models, we assume that

$$U_1^{1,1}, V_1^{1,1}, U_1^{1,2}, V_1^{1,2}, \dots, U_1^{1,N^1}, V_1^{1,N^1}, U_1^{2,1}, V_1^{2,1}, \dots, V_T^{1,N^1}, U_1^{2,1} \dots$$

are mutually independent and independent of  $Y, I$ .

The  $i$ -th loan from the  $\tau$ -th generation defaults at time  $t$  if the disposable wealth of the corresponding debtor does not suffice for repaying the mortgage, specifically if

$$A_t^{\tau,i} < B_t^{\tau,i} \quad (3)$$

where  $B_t^{\tau,i}$  denotes the liabilities of the debtor. Generally,  $B_t^{\tau,i}$  may include various liabilities; in the present paper, however, we take only accumulated instalments into account, i.e.  $B_t^{\tau,i} = (t - \tau + 1)b$ , where  $b$  the one-period installment (note that  $B_t^{\tau,i}$  is deterministic then). The corresponding percentage loss associated with the loan at  $t$  equals

$$G_t^{\tau,i} = \frac{Q_t^{\tau,i} \max(0, h_t - P_t^{\tau,i})}{h_t^\tau} \quad (4)$$

where  $Q_t^{\tau,i}$  is the indicator of inequality (3), i.e. the zero-one variable indicating the default, and  $h_t^\tau$  is the principal outstanding at  $t$ , computed in the standard way (see Šmíd, 2015).

The overall default rate and the charge-off rate on the portfolio are then defined as

$$Q_t = \frac{\sum_{1 \leq \tau \leq t} \sum_{1 \leq i \leq N^\tau} Q_t^{\tau,i}}{\sum_{1 \leq \tau \leq t} N_t^\tau}, \quad t > 1 \quad (5)$$

$$G_t = \frac{\sum_{1 \leq \tau \leq t} \sum_{1 \leq i \leq N^\tau} G_t^{\tau,i}}{\sum_{1 \leq \tau \leq t} N_t^\tau}, \quad t > 1 \quad (6)$$

respectively, where, for each  $t$ ,  $N_t^\tau$  is the number of loans from the  $\tau$ -th generation which did not default until  $t$ .

Once the initial sizes  $N^1, N^2, \dots$  grow to infinity, the influence of the individual factors on quantities  $Q_t$  and  $G_t$  is canceled out by the Law of Large Numbers, and both of the quantities become uniquely determined by the common factors; to be precise, there exists one-to-one mapping  $\Phi_t$  for each  $t$  fulfilling

$$(Q_1, G_1, Q_2, G_2, \dots, Q_t, G_t) = \Phi_t(Y_1, I_1, Y_2, I_2, \dots, Y_t, I_t) \quad (7)$$

(see Theorem 3 of Šmíd, 2015, where a rigorous description of the whole model may also be found).

Thanks to the one-to-one property, the values of the common factors may be uniquely retrieved from the values of the loss rates in the limit case. Consequently, having a time series of the loss rates at hand, its future values may be predicted by transforming it into the factors, predicting the factors and transforming the predictions back into the loss rates.

### 3. Data and Empirical Estimation

As previously mentioned, we applied our model to the US nationwide portfolio of mortgages.

#### 3.1 The Dataset

The dataset was provided by the United States Federal Reserve System. The time range covered was the period between 1991 and 2014 in quarterly granularity, so the number of observations was 96. The dataset consists of time series of two quantities, namely the mortgage delinquency rate, which is a proportion of loans more than 30 days past due (30+) on the total mortgage balance, and the mortgage charge-off rate, which is an annualized proportion of loans charged off net of recoveries on the average total balance.

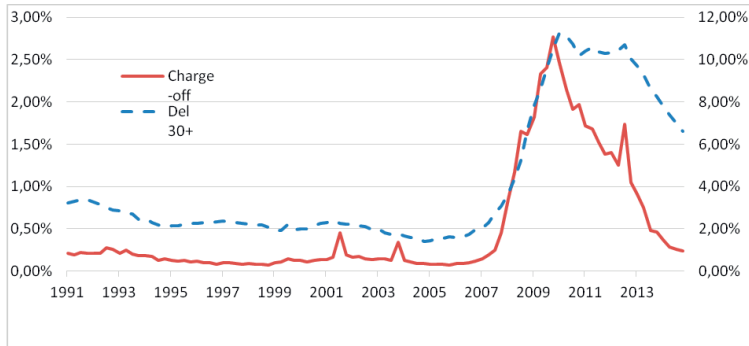
In practice, there exist accounts which after becoming delinquent are not charged off, but are “cured”, i.e. the amount past due is paid and subsequent installments continue to be paid on time. As cured account would bring additional complexity to the model, we omit the cure rate and assume that the cure amount is repaid wholly in the 30+ delinquency state.

*Table 1* and *Figure 1* summarize the descriptive statistics and show the development of the input data. The 30+ delinquency rate was used as a proxy for the default rate and the charge-off rate represented the real loss from the unpaid balance. From *Figure 1* it is obvious that there exists a visual suspicion that the two time series are strongly 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 clearly visible as both time series rocketed upward in the period between 2007 and 2010 to multiples of their previous values.

#### 3.2 Underlying Factors Extraction

We applied the inverse transformation (vii) to our input data with the following parameters:

**Figure 1 Development of the 30+ Delinquency Rate (right axis) and Charge-Off Rate (left axis)**



**Table 1 Descriptive Statistics of the Input Data**

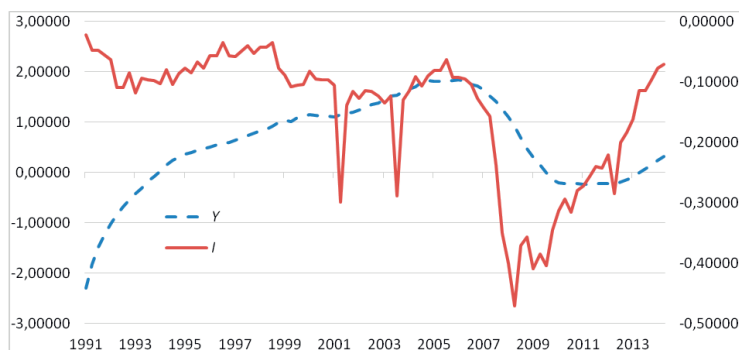
| Statistic          | 30+ delinquency rate | Charge-off rate |
|--------------------|----------------------|-----------------|
| Mean value         | 0.0417               | 0.0051          |
| Median             | 0.0231               | 0.0017          |
| Minimum            | 0.0139               | 0.0007          |
| Maximum            | 0.1127               | 0.0277          |
| Standard deviation | 0.0325               | 0.0069          |
| Variance           | 0.7790               | 1.3452          |
| Skewness           | 1.1517               | 1.7144          |
| Kurtosis           | -0.3822              | 1.6548          |
| 5% percentile      | 0.0158               | 0.0008          |
| 95% percentile     | 0.1060               | 0.0218          |

- $\sigma_1$ , the standard deviation of the initial wealth, was set to 0.5, which roughly corresponds to a long-term standard deviation of family income in the US.
- $\rho$ , the standard deviation of the loss given default, was set to 0.12 according to Gapko & Šmíd (2012b).
- $\varphi$ , the autocorrelation coefficient of the default rate individual factor, was set to 0.8, which corresponds to findings in Hochguertel & Ohlsson (2011).
- $\psi$ , the autocorrelation coefficient of the loss given default individual factor, was set to 0.1.<sup>3</sup>
- The mortgage interest rate was set to zero for simplicity.

The resulting time series of the extracted common factors  $Y$  (default rate) and  $I$  (loss given default) are depicted in *Figure 2*.

<sup>3</sup> According to Guren (2014), the annual AR coefficient of house prices in the US ranges between 0.4 and 0.7. We expect that most of the autocorrelation is caused by common factors, such as cyclical development on the real estate market. Thus, we set the AR coefficient of the LGD individual factor to 0.1.

**Figure 2 Development of the Extracted Common Factors  $Y$  (default rate, left axis) and  $I$  (loss given default, right axis)**



### 3.3 Estimation of the Factors' Dynamics

In the next step, we studied the interconnectedness of the two common factors,  $Y$  and  $I$ . According to the nature of the credit risk, the default rate and the charge-off rate<sup>4</sup> should be interconnected, i.e. during times of economic expansions, when the financial income of households and real estate prices increase, the expected decrease of the default rate should be accompanied by the expected decrease of the charge-off rate, and *vice versa* for the periods of economic downturns. The hypothesis of interconnectedness is supported by the fact that the mutual correlation between the factors is 57%.

To examine the mutual relationship between the factors  $Y$  and  $I$  in more detail, we used cointegration analysis. Existence of the unit root in both time series, which is one of the necessary conditions for cointegration, wasn't rejected. We further ran the Johansen cointegration test, which confirmed that cointegration of rank one is present.<sup>5</sup>

As the next step, we estimated the dynamics of  $Y$  and  $I$  and the dependence on the external environment using the VECM model, with  $Y$  and  $I$  being endogenous variables and a set of macroeconomic indicators representing exogenous variables. We considered GDP, the house price index (HPI), consumer prices (inflation), the FED base interest rate, industrial production and personal income as representatives of the external environment. The dataset of macroeconomic variables was obtained from the US Bureau of Economic Analysis. The estimation results are summarized in *Table 2*.

The predictive power of the estimate, measured by  $R$ -square, is high in the case of  $Y$  (over 90%), but quite low in the case of  $I$  (just below 23%). The results show how strong the mutual relationship of  $Y$  and  $I$  is and they also prove the dependence of both  $Y$  and  $I$  on the external environment. The mean squared errors of the equations for  $Y$  and  $I$  are 0.0009 and 0.0019, respectively. The discovered relationship between the factors and the macroeconomic environment is largely in line with our expectations and is easily interpretable—credit risk grows with a rising interest rate

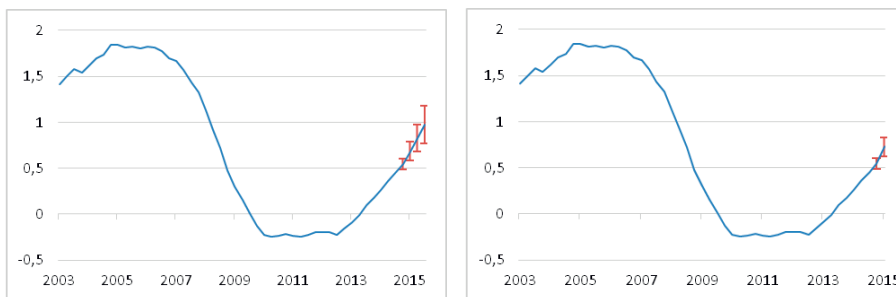
<sup>4</sup> LGD can be easily obtained by dividing the charge-off rate by the default rate.

<sup>5</sup> The details of the Johansen cointegration test are provided in the *Appendix*.

**Table 2 VECM Estimation Results**

| Variable          | Factor     |            |
|-------------------|------------|------------|
|                   | Y          | I          |
| Constant          | 0.132 ***  | -          |
| $Y_{t-1}$         | 0.319 ***  | 0.083 **   |
| $I_{t-1}$         | -0.253 *** | -0.231 **  |
| GDP YoY           | 0.756 ***  | 0.937 ***  |
| HPI               | -          | -          |
| CPI inflation YoY | -          | -1.462 *** |
| FED int. rate     | -0.014 *** | -          |
| Ind. production   | -          | -          |
| Personal income   | -          | -          |
| Correction term   | -0.029 *** | 0.009 **   |

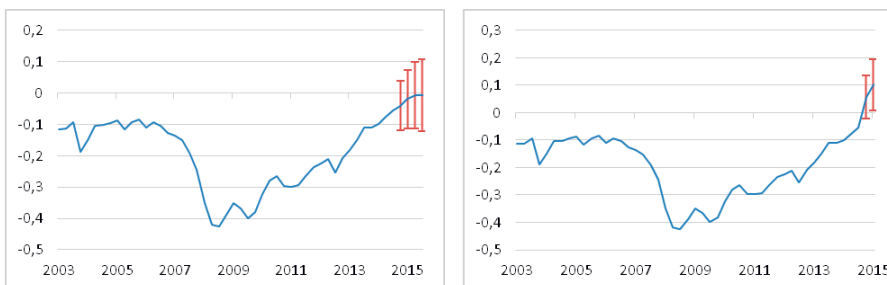
**Figure 3 Prediction of Y Given an Unchanged Macroeconomic Environment (left) and Real Macroeconomic Development in Q1 and Q2 2015 (right)**



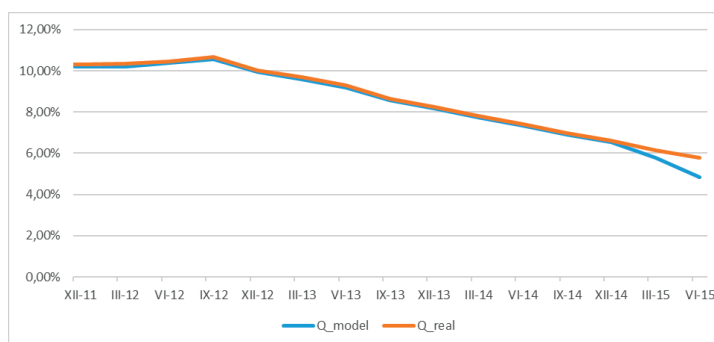
and growing inflation and decreases with increasing GDP. Note that there is a negative relationship between the factors and credit risk, i.e. the higher the value of a factor, the lower the credit risk. On the other hand, the negative sign at the coefficient of  $I_{t-1}$  is counterintuitive, as it suggests negative autocorrelation of the  $I$  time series and negative correlation between  $Y$  and  $I$ . The reason for this might be the nature of the charge-offs: as the charge-off is a purely accounting operation, there exists higher proportional variance in the charge-off rate compared to the delinquency rate, which is smoother. Thus, the historical variation in the charge-off rate most probably introduced a negative autocorrelation in the  $I$  time series, as well as the negative correlation between  $Y$  and  $I$ .

From our estimation, we constructed two sets of predictions of  $Y$  and  $I$ . In the first set, the prediction horizon was four quarters and the exogenous variables were set so that the external environment was unchanged. The second prediction horizon was two quarters and it was based on the values of the macroeconomic indicators (GDP, inflation and interest rate) from Q1 and Q2 2015. The former prediction allows us to see the development of  $Y$  and  $I$  ceteris paribus, whereas the latter can be compared to the real development and thus is able to backtest the model. The predictions of  $Y$  and  $I$  are shown in *Figures 3* and *4*.

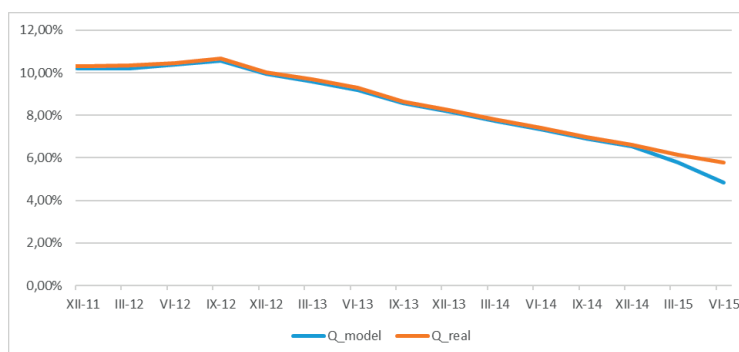
**Figure 4 Prediction of I Given an Unchanged Macroeconomic Environment (left) and Real Macroeconomic Development in Q1 and Q2 2015 (right)**



**Figure 5 Comparison of the Model Prediction and Real Development of the 30+ Delinquency Rate (Q)**



**Figure 6 Comparison of the Model Prediction and Real Development of the Charge-Off Rate (G)**



Finally, we translated the predicted values of  $Y$  and  $I$  back into the 30+ delinquency rate ( $Q$ ) and charge-off rate ( $G$ ). Figures 5 and 6 show the input dataset  $Q$  and  $G$ , respectively, since the end of 2011, where the comparison of the model-predicted delinquency rate and charge-off rate in Q1 and Q2 2015 to the actual development of delinquency and charge-off rates is added. The results show that our model correctly estimates the direction of the out-of-sample development; however,



it slightly overreacts to changes in the macroeconomic environment and suggests bigger movements in  $Q$  and  $G$  than those observed.

#### 4. Conclusion

We constructed a multi-period multigenerational dynamic model of credit losses, the dynamics of which are dependent on the external environment. The model introduces several important original enhancements, namely the internal dynamics of the individual factors within the loan portfolio, a multi-generation approach and, finally, inclusion of macroeconomic variables in the estimation process by means of a VECM model.

We applied the model to the US nationwide portfolio of mortgage loans. The empirical analysis confirmed the well-known fact that there exists a clear and estimable relationship between the credit risk and the macroeconomic environment. We found out that the most significant macroeconomic variables influencing credit risk factors are GDP, inflation and the interest rate. Additionally, we proved the inter-connectedness between defaults and recorded losses (or, in other words, loss given default).

Finally we demonstrated the predictive power of our model. A simple backtest showed that the model gives reasonable results. On the other hand, the model was found to overreact to changes in macroeconomic variables.

Apart from the prediction, other applications of our model suggest themselves. Thanks to the inclusion of macroeconomic variables, the model may also be used for credit risk stress testing. Obviously, the model can also be used as a model of economic capital within a financial institution.

## APPENDIX

**Table A1 Results of the Johansen Cointegration Test**

| Rank | Eigenvalue | Trace test<br>( $p$ -value) | Lmax test<br>( $p$ -value) |
|------|------------|-----------------------------|----------------------------|
| 0    | 0.5146     | 70.732<br>(0.000)           | 67.946<br>(0.000)          |
| 1    | 0.0292     | 2.786<br>(0.095)            | 2.786<br>(0,095)           |

## REFERENCES

- Frontczak R, Rostek S (2015): Modeling loss given default with stochastic collateral. *Economic Modelling*, Volume 44, January 2015:162–170.
- Frye J (2000): Collateral Damage. *Risk*, Volume 13, Number 4.
- Gapko P, Šmíd M (2012a): Dynamic Multi-Factor Credit Risk Model with Fat-Tailed Factors. *Finance a úvěr-Czech Journal of Economics and Finance*, 62(2):125–140.
- Gapko P, Šmíd M (2012b): Modeling a Distribution of Mortgage Credit Losses. *Ekonomický časopis*, Vol. 60, no. 10 (2012), p. 1005–1023.
- Guren AM (2014): The Causes and Consequences of House Price Momentum. *Harvard University Working Paper*.
- Hamerle A, Dartsch A, Jobst R, Plank K (2011): Integrating macroeconomic risk factors into credit portfolio models. *Journal of Risk Model Validation*, Volume 5/ Number 2, :3–24.
- Hochguertel S, Ohlsson H (2011): Wealth Mobility and Dynamics over Entire Individual Working Life Cycles. *ECB Working Paper Series*, no. 1301.
- Jimenez G, Mencia J (2009): Modelling the distribution of credit losses with observable and latent factors. *Journal of Empirical Finance*, Volume 16, Issue 2, 235–253.
- Merton RC (1974): On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *Journal of Finance*, 29, Chapter 12.
- Pesaran MH, Schuermann T, Treutler B-J, Weiner SM (2003): Macroeconomic Dynamics and Credit Risk: A Global Perspective. *CEifo Working Paper*, no. 995.
- Pykhtin MV (2003): Unexpected Recovery Risk. *Risk*, August:74–78.
- Šmíd M (2015): *Model of risk and losses of a multigeneration mortgage portfolio*. 10th International Scientific Conference Financial management of firms and financial institutions. Available at: <http://ssrn.com>.
- Sommar PA, Shahnazarian H (2009): Interdependencies between Expected Default Frequency and the Macro Economy. *International Journal of Central Banking*, September:83–110.
- Vasicek OA (1987): *Probability of Loss on Loan Portfolio*. KMV.
- 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. *Bulletin of the Czech Econometric Society*, Volume 18, Issue 28.