Univerzita Karlova v Praze
Fakulta sociálních věd
Institut ekonomických studií

Petr Gapko

Three Essays on Credit Risk Quantification

Disertační práce

Praha 2014

Charles University in Prague
Faculty of Social Sciences
Institute of Economic Studies

Petr Gapko

Three Essays on Credit Risk Quantification

Dissertation Thesis

Praha 2014

Author: PhDr. Petr Gapko

Advisor: RNDr. Martin Šmíd, PhD

Academic Year: 2013/2014

Hereby I declare that the thesis is my original work. All used resources and datasets used are listed in the list of references. The dissertation thesis was used solely for the purpose of acquiring a PhD degree in Economics at the Charles University.

PhDr. Petr Gapko

Acknowledgment

I would like to take this opportunity to thank my advisor, Martin Šmíd, for all his support, patience and valuable advices during my PhD studies. My thanks also belongs to my student colleagues, who participated on the research by adding valuable comments or co-authorship. Namely, these are Jan Voříšek and Jozef Baruník.

Additionally, the financial help from the following grants are greatly appreciated: No. P402/12/G097, 402/09/0965 and No. 402/09/H045 of the Czech Science Foundation and No. 46108 of Charles University (GAUK). Without this support, I would not be able to visit conferences, purchase datasets and acquire a lot of valuable knowledge and contacts in the academic sphere.

My last thanks is given to my family, especially my wife, who supported me with her patience and a spare time to continue my research.

Abstrakt

Tématem dizertační práce je modelování a odhadování úvěrového rizika. V práci se konkrétněji zaměřujeme na úvěrové riziko retailového, zejména pak hypotečního, portfolia dlužníků. Práce je rozdělena do tří oddělených vědeckých článků se společným tématem, čímž je vývoj metodologie měření úvěrových rizik od jednoduchých rozšíření v současné praxi používaných modelů až po vytvoření modelu, který je schopen pracovat s takovými detaily, jako je např. struktura durace portfolia hypotečních úvěrů. Všechny tři články používají stejnou podkladovou časovou řadu delikvencí a podílu vymáhaných úvěrů národního portfolia hypoték ve Spojených státech. Protože byl výzkum prováděn několik let, pracují novější části dizertační práce s dodatečnými pozorováními.

V prvním článku demonstrujeme, že současné regulatorní standardy pro kvantifikaci úvěrových rizik jsou založeny na předpokladech, které nutně nereflektují realitu. Zobecněním dobře známého Vašíčkova modelu, který stojí za Basel II, konstruujeme model pro odhadování úvěrových rizik. Náš model, podobně jako Vašíčkův, dekomponuje úvěrové riziko (které vyjadřujeme jako portfoliovou pravděpodobnost selhání) na dva rizikové faktory, z nichž jeden je společný pro všechny dlužníky v portfoliu a druhý individuální pro každého dlužníka. Náš model obsahuje dynamiku společného faktoru, který ovlivňuje aktiva dlužníků a u kterého, na rozdíl od Vašíčkova modelu, povolujeme nenormalitu. Popisujeme, jak se mohou odhadnout parametry našeho modelu a navíc dokládáme statistickou evidenci, že model založený na nenormálních rozděleních lépe odpovídá pozorovaným měrám delikvencí na hypotékách ve Spojených státech.

Druhý článek je pokračováním našeho výzkumu. V tomto článku představujeme vylepšený vícefaktorový model, který simultánně popisuje míru selhání a ztrátu v selhání. Naše metodologie je znovu založena na Vašíčkově modelu, který zobecňujeme ve třech směrech. Za prvé, přidáváme model ztráty v selhání (loss given default, LGD). Za druhé, do modelu vnášíme dynamiku a za třetí, pro všechny faktory povolujeme nenormální rozdělení. Jak pravděpodobnost selhání, tak i ztráta v selhání jsou řízeny společným a individuálním faktorem. Individuální faktory jsou vzájemně nezávislé, ale umožňujeme závislost společných faktorů jakéhokoliv druhu. Náš model testujeme na národním portfoliu hypotečních delikvencí v USA, závislost společných faktorů modelujeme pomocí VECM metodologie a naše výsledky porovnáváme se

současnými regulatorními modely z Basel II. Naše nálezy ukazují, že metodologie, která je schopna popsat závislost mezi rizikovými faktory, je schopna přesněji predikovat střední hodnotu a kvantil ztrát.

Nejnovější část našeho výzkumu je popsána ve třetím článku. Podobně jako ve druhé části předpokládáme, že dlužníci drží aktiva, která pokrývají splátky dluhu, a vlastní nemovitosti, které slouží jako kolaterál. Hodnota aktiv i ceny nemovitostí sledují obecný stochastický proces, který je řízen společným a individuálním faktorem. Popisujeme vztahy mezi společnými faktory a podílem selhání, resp. ztrátou v selhání, a zároveň navrhujeme ekonometrický proces odhadu modelu. Na rozdíl od předešlého výzkumu přidáváme vícegenerační aspekt a modelujeme aktiva jednotlivých generací odděleně. Ukazujeme, že přesnější odhad vývoje společných faktorů může vést, v porovnání s Basel II rámcem, v úsporám kapitálu drženého proti kvantilovým ztrátám.

Abstract

The dissertation thesis deals with modeling and estimating credit risk. In the thesis we particularly focus on the credit risk of retail, and more exactly mortgage, debtors. The thesis is organized into three separate papers with a common theme, which is a development of a credit risk measurement methodology from simpler enhancements of the current research to a model able to capture such details as e.g. the duration structure of the mortgage portfolio. All three papers use the same underlying dataset, a time series of the national US mortgage portfolio delinquency and foreclosure rates. As the research was done during several years, the latter parts of the thesis work with additional observations.

In the first paper, we demonstrate that the current regulatory standards for credit risk quantification are based on assumptions that do not necessarily match the reality. Generalizing the well-known Vasicek's model, standing behind the Basel II, we build a model of a credit risk of a loan portfolio. The model, similarly to the Vasicek's model, decomposes the credit risk (expressed as the portfolio probability of default) into two risk factors, one common for all borrowers in the portfolio, and one individual for each single borrower. Our model involves dynamics of the common factor, which influences the borrowers' assets, and which we allow, in contrary to the Vasicek's model, to be non-normal. We show how the parameters of our model may be estimated, and additionally, we provide a statistical evidence that the non-normal model is able to fit better the observed US mortgage delinquency rates than a normal one.

The second paper is a continuation of the research. In this paper, 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 Vasicek's 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 probability of default 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, the Basel II. Our findings show, that a methodology, which is able to describe the dependency between the risk factors, can predict the mean and the quantile losses more precisely.

The most recent development in our research is described in the third paper. Similarly to the second paper, we assume borrowers hold assets covering the instalments and own real estate which serves as collateral. Both the value of the assets and the price of the estate follow general stochastic processes driven by common and individual factors. We describe the correspondence between the common factors and the percentage of defaults, and the loss given default, respectively, and we suggest a procedure of econometric estimation in the model. On the contrary to the second paper, here we add a multigenerational aspect and we model the assets of different generations separately. We show that a more accurate estimation of common factors can lead to savings in capital needed to hold against a quantile loss, compared to the Basel II framework.

Contents

1. Introduction

The dissertation thesis was inspired by one particular problem, which in the last decade influenced banking regulation, financial markets and the trustworthiness of large financial institutions – modeling and estimating credit risk. In the thesis, we propose models of estimation of credit risk with a particular focus on the credit risk of retail, and more specifically, mortgage debtors.

We built our models on the textbook approach of the risk modeling of the portfolio of loans of Vasicek (Vasicek, 1987), who deduces the default rates of borrowers and thus the credit risk of the loan portfolio from the value of the borrowers' assets, which follow a geometric Brownian motion. Further, we followed the extensions of Vasicek's model of Frye (Frye, 2000), who assumes that the loss given default (LDG) is a second determinant of credit risk as well as, Pykhtin (Pykhtin, 2003), who suggests a model where LGD is driven by one systematic and two idiosyncratic underlying variables. Among other most influential models, we can include the CreditMetrics model, in which the default frequency is modeled by transition matrices and probabilities or the CreditRisk+ model (Wilde, 1997), which, in contrary to the CreditMetrics model, assumes a Poisson distribution for the default frequency.

Our research adopted the above mentioned assumptions, namely that credit risk is based on the fact that the credit losses are a function of PD and LGD which are further decomposed to underlying factors. Also, the similarity between our research and the described approaches might be found in the fact that PD and LGD are both driven by systematic and idiosyncratic factors, specific for both variables. The main contribution of our work lies in several improvements. Firstly, we bring dynamics to the systematic and idiosyncratic factors. Moreover, these factors, in contrary to the current research, are estimated from macroeconomic indicators and only the remaining variance is considered to be an element of uncertainty. Secondly, the evolution of the residuals from the estimated models is allowed to be non-normal. Finally, in the last of the three models we constructed, we switch from the single portfolio approach to a multi-generation approach, which enables us to also model the duration structure of the loan portfolio.

The thesis is organized into three separate papers with a common theme, which is the development of a credit risk measurement methodology from simpler enhancements of the current research to a model capable of capturing such details as e.g. the duration structure of the mortgage portfolio. All three papers use the same underlying dataset, a time series of the national US mortgage portfolio delinquency and foreclosure rates. As the research was done over several years, the latter parts of the thesis utilize a longer dataset with additional observations.

Our research was, from the very beginning, focused on relaxing tightening assumptions in current models; however, as the research proceeded and we discovered further and further ways of how to describe the development of the credit risk with a higher accuracy, the final model, even though based on the same basis as the most common and used credit risk methodology, Vasicek's model, is a standalone method of estimating credit risk with a significantly lower estimated variance than Vasicek's approach.

In the first paper, we relax several obviously unrealistic assumptions of Vasicek's model, standing behind the Basel II. Our model, similarly to Vasicek's, assumes that assets of debtors follow the geometric Brownian motion. If the value of assets of a borrower falls under a certain threshold, commonly interpreted as a value of the borrower's debt, the debtor defaults. Also, as in Vasicek's model, we decompose the credit risk (expressed as the portfolio probability of default) into two risk factors, one common for all borrowers in the portfolio, and one individual for each single borrower. The proportion of defaults in the portfolio is calculated as a limit if the portfolio is sufficiently large. Additionally, by the Law of large numbers the individual factor on a large portfolio cancels out and enters the final loss (or, more exactly default) distribution only by its own distribution, which is assumed to be standard normal.

In contrary to Vasicek's, our model involves dynamics of the common factor, which influences the borrower's assets. For this factor we proposed an AR process and constructed an empirical model (estimated by the maximum likelihood estimator), in which the factor depends on macroeconomic development. The model was estimated on an empirical dataset of US mortgage delinquency rates and macroeconomic indicators. Our analysis shows that the residuals (i.e. the remaining unexplained variance in the common factor) have heavier tails than the originally proposed normal distribution. Thus, we allow the residuals of the process of factors to be non-normal, with the Generalized Hyperbolic Distribution, which we show to have the best fit. In particular, we provide statistical evidence that the non-normal model is able to fit the observed US mortgage delinquency rates better than ones with normal, lognormal or beta distributed residuals. We point out how the assumption, that risk factors follow a normal distribution, can be dangerous, especially during volatile periods comparable to the crisis in 2007-2009. The methodology based on the normal distribution can underestimate the impact of changes in tail losses. However, on the other hand, in periods with low volatility, the model showed lower capital requirements. This is due to the fact that we estimate the future loss distribution from historical information, which is, in fact, neglected by Vasicek's model.

The first paper is a joint research with Martin Šmíd, a supervisor of the dissertation, and was published in the Journal of Economics in 2012.

The second paper describes another model of credit risk, which is an extension of the research from the first paper. In this paper, we introduce an improved multi-factor credit risk model, describing simultaneously the default rate and the loss given default. Our methodology is again based on Vasicek's model (and thus the assumption that assets of borrowers follow a geometric Brownian motion), which we generalize in three ways this time. Firstly, we add a model for loss given default (LGD), which is also an improvement compared to the first model. Secondly, we bring dynamics to the model, and thirdly, we allow non-normal distributions of risk factors. Both the probability of default 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. Thus the modeling of LGD is an analogous to PD modeling, with the assumption that, analogously to the assets of the borrowers in the case of PD, the real estate prices follow a geometric Brownian motion. Based on this, we build an analytically trackable function, which maps the relationship between the factor and the LGD. The factors in this model are allowed to have a general shape with any kind of statistical distribution.

We tested our model on a nationwide portfolio of US mortgage delinquencies; however, as to our knowledge, there was no comparable LGD time series publically available, therefore, we constructed a proxy for LGD, based on the proportion of foreclosed on defaulted mortgages. We modeled the interdependence of the two common factors by a VECM model, and compared our results with the current regulatory framework, which is described in the Basel II. The results demonstrated that there is a statistically significant relationship between the actual values of the factors and their past values, and moreover, these two time series are cointegrated. Similarly to the first model, the normality of the residuals from the VECM was rejected and thus we used the fitted generalized hyperbolic distribution to arrive at the quantiles of PDs and LGDs. The final results show that, in contrary to the first paper, the capital requirement is lower than in the case of Vasicek's model. This is again caused by a more accurate model (and inclusion of the LGD model) than in the first case. This is a clear implication for risk management and quantification of credit risk, because our model, compared to Vasicek's framework, brings capital savings.

The second paper is also a product of joint research with Martin Šmíd. It was published in the Czech Journal of Economics and Finance in 2012.

The most recent development in our research, our most advanced model of quantification of credit risk, is described and estimated in the last paper. Similarly to the first and the second models, we assume that borrowers hold assets, from which they repay the instalments, and own real estate, which serves as a collateral. This model fixes the most significant drawback of our previous models, the single-generation approach. Particularly, the third model still assumes that the value of the assets, as well as the price of the real estate, follow a geometric Brownian motion driven by common and individual factors but, in contrary to the preceding research, portfolios last for more than a single period. In particular, in each period (or more specifically, in each data point) new debtors enter the examined portfolio, while a part of the examined portfolio exits the model by one of two possible exit states, which are a full repayment of a loan and a default state. However, a price for an increased accuracy in the duration of individual generations in the model is the loss of the analytical trackability of functions mapping factors to PD, LGD, respectively which have to be calculated numerically by simulation.

In the empirical part, we describe the correspondence between the common factors and the percentage of defaults, and the loss given default, respectively, and we suggest a procedure of econometric estimation of the model. Similarly to the second model, we chose the VECM procedure to model the relationships between the two common factors, and also the external environment, represented by a set of macroeconomic variables. For this we used the same dataset as in preceding cases, however, enriched with recent observations. The VECM results showed that the two common factors are cointegrated, and moreover, also depended on two macroeconomic variables – GDP and unemployment rates. Using the enhanced framework, we extracted more information from the empirical datasets, which induced that, as opposed to the prior model, the normality of residuals was not rejected in the VECM model. The more accurate estimation of common factors led to lower variation of the quantile estimate, which, translated to the regulatory language, means savings in capital, which is needed to cover unexpected losses, as compared to the Basel II framework. The second implication of the model is that the mean value of the loss can be forecasted by means of forecasts of common factors, GDP and unemployment, which enables to calculate expected and unexpected losses under various macroeconomic scenarios. This feature can be used e.g. for stress testing.

The third paper is research conducted together with Martin Šmíd and Jan Vořížek. This part of the research was not published at the time of the submission of the dissertation, but had been submitted to the Journal of Credit Risk.

In the three papers we have shown that the current commonly used credit risk quantification methodology is a very gross estimation of the mean and quantile values of credit losses. The framework can be improved by relaxing several of its assumptions, which, on the other hand, brings mathematical and computational complications, particularly in the case of the multigenerational approach. In a nutshell, we have managed to bring dynamics into the evolution of credit losses in time, and we have described the mapping of risk factors into PD and LGD. Additionally, our approach is compatible with the econometric estimation of the factors model, if it can be estimated by MLE. Lastly, we have shown that a clear link exists between the credit risk and macroeconomic environment, and that this link can be incorporated into the quantification of credit losses. Even if the calculations are complex, usage of our model leads to a more exact evolution of underlying risk factors, which also leads to a lower variance in the loss distribution and therefore, a lower difference between the mean and the quantile losses. In particular, our enhanced credit risk measurement methodology can save a portion of capital.

The complexity of our approach has also introduced a space for further improvements. Among the main challenges, we can point out the appropriateness of the used data, especially representing the LGD. For some portfolios, more accurate LGD datasets, e.g. for traded bonds, may be found. Also using an internal dataset from a bank could lead to a better estimate of the LGD. Secondly, the computational time of estimation of the most recent model version is quite time consuming. Introducing several simplifications, or fine tuning the code, could lead to acceleration of the numerical calculation. Lastly, the model can be enhanced to calculate expected and unexpected losses for multiple portfolios by creating a module which would be capable of joining the intra- and inter-portfolio correlations. We believe that these sets of models can contribute to a better understanding of credit risk and might, therefore, be implemented in banking practices.

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*

Frye, J. (2000). Collateral Damage. *Risk.*

Pykhtin, M. V. (2003). Unexpected Recovery Risk. *Risk, 74-78.*

RiskMetrics Group, I. (1997). CreditMetrics Technical Document. *New York: JP Morgan & Co.*

Vasicek, O. A. (1987). Probability of Loss on Loan Portfolio. *KMV.*

Wilde, T. (1997). Encyclopedia of Quantitative Finance. *John Wiley & Sons, Ltd.*

2. Modeling a distribution of mortgage credit losses

2.1 Introduction

In our paper, we will focus on credit risk quantification methodology. Because banking is heavily regulated in developed countries, the minimum standards for credit risk quantification are often summarized in directives. The current recommended system of financial regulation was developed and is maintained by international supervisory institutions located in Europe (Basel Committee on Banking Supervision, CEBS – Committee of European Banking Supervisors) and its standards are formalized in the Second Basel Accord ("Basel II," Bank for International Settlements, 2006) and is implemented into European law by the Capital Requirements Directive (CRD) (European Commission, 2006).

For credit risk, Basel II allows only two possible quantification methods – a "Standardized Approach" (STA) and an "Internal Rating Based Approach" (IRB) (for more details on these two methods see Bank for International Settlements, 2006). The main difference between STA and IRB is that under IRB banks are required to use internal measures for both the quality of the deal (measured by the counterparty's "probability of default – PD") and the quality of the deal's collateral (measured by the deal's "loss given default – LGD"). The counterparty's probability of default is the chance that the counterparty will default (or, in other words, fail to pay back its liabilities) in the upcoming 12 months. A common definition of default is that the debtor is more than 90 days delayed in its payments (90+ days past due). LGD is an estimate of how much of an already defaulted amount a bank would lose. LGD takes into account expected recoveries from the default, i.e., the amount that the creditor expects to be able collect back from the debtor after the debtor defaults. These recoveries are mainly realized from collateral sales and bankruptcy proceedings.

PD and LGD are two major and common measures of deal quality and basic parameters for credit risk measurement. PD is usually obtained by one of the following methods: from a scoring model, from a Merton-based distance-to-default model (e.g. Moody's KMV, mainly used for commercial loans; Merton, 1973 and 1974) or as a long-term stable average of past 90+ delinquencies.[1] The model, presented later in the paper, provides a connection between the scoring models and those based on past delinquencies. LGD can be understood as a function of collateral value.

Once PDs and LGDs are obtained, we are able to calculate the "expected loss." The expected loss is the first moment, the mean, of a loss distribution, i.e., a mean measure of the credit risk. It is a sufficiently exact measure of credit risk at the long-term horizon. However, in the short term (e.g., the one-year horizon), it is insufficient to protect against expected losses only. The problem is that losses on a portfolio follow a certain probability distribution in time. Thus, to protect itself against credit losses, a bank not only has to cover the expected loss (mean), but also should look into the right tail and decide which quantile (probability level) loss should be covered by holding a sufficient amount of capital.

Banks usually cover a quantile that is suggested by a rating agency, but with the condition that they have to observe the regulatory level of probability of 99.9% at minimum. The regulatory level may seem a bit excessive, as it can be interpreted as meaning that banks should cover a loss which occurs once in a thousand years. The fact is that such a far tail in the loss distribution was chosen because of an absence of data. The quantile loss is usually calculated by a Value-at-Risk type model (Saunders & Allen, 2002; Andersson et al., 2001). The IRB approach is a type of Value-at-Risk model and approximates the loss distribution with a mixture of two standardized normal distributions. The IRB model assumes that credit losses are caused by two risk factors: first is a credit quality of the debtor and the second is a common risk factor for all debtors, often interpreted as macroeconomic environment. For both factors, the IRB model assumes the standard normal distribution in time.

In this paper, we will introduce a new approach to quantifying credit risk which can be classed with the Value-at-Risk models. Our approach is different from the IRB method in the assumption of the loss distribution. In the general version of our model, we assume that risk factors can be distributed not only standard normal but can follow a more general distribution in time, the

[1] Delinquency is often defined as a delay in installment payments, e.g., 90+ delinquencies can be interpreted as a delay in payments of more than 90 days.

distribution of the common factor possibly depending on its history (allowing us to model a dynamics of the factor which appeared to be necessary especially during periods like the present financial crisis). In the simpler version, we keep the IRB assumption that the individual risk factor (credit quality of a debtor) follows a standard normal distribution. In its general form, the new approach can be used to measure the credit risk of many types of banking products, i.e., consumer loans, mortgages, overdraft facilities, commercial loans with a lot of variance in collateral, exposures to sovereign counterparties and governments, etc. To test our model, we will demonstrate its goodness-of-fit on a nationwide mortgage portfolio. Moreover, we will compare our results with the IRB approach, prove that the assumption of normal distribution of the common factor can be outperformed, and comment on what difficulties can arise when an inappropriate assumption of normality is made.

The paper is organized as follows. After the introduction we will describe the usual credit risk quantification methods and Basel II-embedded requirements in detail. Then we will derive a new method of measuring credit risk, based on the class of generalized hyperbolic distributions and Value-at-Risk methodology. In the last part, we will focus on the data description and verification of the ability of the class of generalized hyperbolic distributions to capture credit risk more accurately than the Basel II IRB approach. Moreover, we will compare the class of distributions we use with several distributions that are, alongside the IRB's standard normal distribution, commonly used for credit risk quantification. At the end we summarize our findings and offer recommendations for further research.

At the time of the dissertation defense, the Basel III enhanced banking regulation was adopted in Europe. However, as there were no significant changes in the Basel III regarding the calculation of the credit risk, our approach discussed in this paper remains still valid and up-to-date.

2.2 Credit risk measurement methodology

The Basel II document is organized into three separate pillars. The first pillar requires banks to quantify credit risk, operational risk, and market risk by a method approved by a supervisor.[2] For credit risk there are two possible quantification methods: the "Standardized Approach" (STA) and the "Internal Rating Based Approach" (IRB). Both methods are based on quantification of risk-weighted assets for each individual exposure. The STA method uses measures defined by the supervisor, i.e., each deal is assigned a risk-weight based on its characteristics. Risk-weighted assets are obtained by multiplying the assigned risk-weight by the amount that is exposed to default. The IRB approach is more advanced than STA. It is based on a Vasicek-Merton credit risk model (Vasicek, 1987) and its risk-weighted assets calculation is more complicated than the STA case. First of all, PD and LGD are used to define the riskiness of each deal. These measures are then used to calculate risk-weighted assets based on the assumption of normal distribution of asset value. In both cases, the largest loss that could occur at the 99.9% level of probability[3] is calculated as 8% of the risk-weighted assets (for more details on risk-weighted assets calculations see (Bank for International Settlement, 2006)). The loss itself is defined as the amount that is really lost when a default occurs. Default is a delay in payments of more than 90 days (90+ delinquencies).

2.2.1 Expected and unexpected loss for an individual exposure

Expected and unexpected losses are the two basic measures of credit risk. The expected loss is the mean loss in the loss distribution, whereas the unexpected loss is the difference between the expected loss and a chosen quantile loss. In this part we will focus on expected and unexpected loss quantification for a single exposure, e.g., one particular loan. Calculation of both expected and unexpected losses requires PD and LGD. As there is no PD or LGD feature in the STA

---

[2] A supervisor is an institution supervising a certain country's financial market, for the Czech Republic the supervisor is the Czech National Bank.

[3] The 99.9% level of probability is defined by the Basel II document and is assumed to be a far-enough tail for calculating losses that do not occur with a high probability. Note that a 99.9% loss at the one-year horizon means that the loss occurs once in 1,000 years on average. Because the human race lacks such a long dataset, 99.9% was chosen based on rating agencies' assessments.

method, and because supervisory institutions are interested in unexpected losses only, under STA it is impossible to calculate the expected loss, and even the unexpected loss calculation is highly simplified and based on benchmarks only. On the other hand, the advantage of this method is its simplicity. The IRB approach uses PDs and LGDs and thus is more accurate than the STA but relatively difficult to maintain. A bank using the IRB method has to develop its own scoring and rating models to estimate PDs and LGDs. These parameters are then used to define each separate exposure.[4] The average loss that could occur in the following 12 months is calculated as follows:

$$\mathrm{EL} = \mathrm{E(PD)} \cdot \mathrm{E(LGD)} \cdot \mathrm{EAD}, \qquad (2.1)$$

where EAD is the exposure-at-default[5] and EL is the abbreviation for "Expected Loss." The mean value of the expected loss is based on the mean value of the counterparty PD, the mean value of the deal LGD and the EAD. The EAD is usually also a variable as it is a function of a "Credit Conversion Factor" (CCF)[6]. However, for mortgage portfolios, CCF is prescribed by the regulator. For our calculations we assume that if a default is observed, it happens on a 100% drawn credit line. Thus we don't treat EAD as a variable but a constant. EL is the average loss that would occur each year and thus is something that banks incorporate into their loan-pricing models. It necessarily has to be covered by ordinary banking fees and/or interest payments. However, EL is the "mean loss" and thus is unable to capture any volatility in losses. To protect themselves against loss volatility, banks should hold capital to cover the maximum loss that could occur at the regulatory probability level at minimum. To capture the variability in credit losses over time and to calculate the needed quantile of the loss distribution, we need a second moment of the loss distribution, the standard deviation and the shape of the loss distribution at minimum.

On the deal level, the standard deviation calculation can be derived from the properties of default. Default is a binary variable – it either occurs (with a probability equal to *PD*) or does not occur (with a probability equal to *(1-PD)*). If the LGD is positive, the loss occurs with the same

[4] Exposure is the usual expression for the balance on a separate account that is currently exposed to default. We will adopt this expression and use it in the rest of our paper.
[5] Exposure-at-default is a Basel II expression for the amount that is (at the moment of the calculation) exposed to default.
[6] CCF is a measure of what amount of the loan (or a credit line) amount is in average withdrawn in the case of a default. It is measured in % of the overall financed amount and is important mainly for off-balance sheet items (e.g. credit lines, credit commitments, undrawn part of the loan, etc…).

probability as the default, but is usually lower than the defaulted amount (due to the fact that the bank sells its collateral and partly collects the defaulted amount – this is, in fact, the LGD) and thus follows a binomial distribution[7]. We can calculate the standard deviation of a loss by substituting into the formula for the binomial distribution's standard deviation. Finally, to protect itself at a given probability level, a bank has to hold a stock of capital equal to the unexpected loss: the difference between a certain quantile (equal to the chosen probability level) and the mean of the loss distribution.

2.2.2 Expected and unexpected loss for a portfolio

On the portfolio level (constructed from a certain number of individual deals), the expected loss calculation can be performed in the same way as for an individual deal. We either sum the expected losses for the deals included in the portfolio or calculate a portfolio-weighted average PD and LGD, where the weights are the EADs of the individual deals. The portfolio EAD is then calculated as the sum of the EADs for the deals included. Therefore, we can use formula (2.1) to calculate the portfolio expected loss.

However, the calculation of the unexpected loss on the portfolio level is not so straightforward. Generally, the unexpected loss of a portfolio on a certain probability level can be calculated as a decrease of the loan portfolio value on the same percentile. However, deals are correlated among each other. We have a complicated correlation structure that is usually unknown and thus we do not even know how the individual deals in our portfolio interact. There are two ways of constructing an unexpected loss calculation model. If the correlation structure among the individual deals is known, we can multiply the vector of the unexpected losses by the correlation matrix to get a portfolio unexpected loss. This approach is often referred to as a "bottom-up" one.

Often, the correlation matrix of the individual deals is not known and thus a different approach has to be chosen to determine the unexpected loss of the loan portfolio. The second approach is widely known as a "top-down" approach and the main idea is to estimate the loss distribution

[7] Please note that the LGD variable can in some cases turn to positive values. This is for example a situation when a loan's collateral covers the loan value and a bank collects some additional cash on penalty fees and interest.

based on historical data or assume a distribution structure and determine the standard deviation or directly the difference between the chosen quantile and the mean value.[8]

2.3 Our approach

2.3.1 The distribution of Loan Portfolio Value

The usual approach to modelling the loan portfolio value is based on the famous paper by Vasicek (2002) assuming that the value $A_{i,1}$or the i-th's borrower's assets at the time one can be represented as

$$\log A_{i,1} = \log A_{i,0} + \eta + \gamma \mathrm{X}_i \qquad (2.2)$$

where $A_{i,0}$ is the borrower's wealth at the time zero, η and γ are constants and X_i is a (unit normal) random variable, which may be further decomposed as

$$X_i = Y + Z_i$$

where Y is a factor, common for all the borrowers, and Z_i is a private factor, specific for the borrower (see Vasicek (2002) for details).

2.3.2 The generalization

We generalize the model in two ways: we assume a dynamics of the common factor Y and we allow non-normal distributions of both the common and the private factors. Similarly to the original model, we assume that

$$\log A_{i,t} = \log A_{i,t-1} + Y_t + U_{i,t} \qquad (2.3)$$

[8] Remember that the loss mean value equals the expected loss of a deal

where $A_{i,t}$ is the wealth of the i-th borrower at the time $t \in \mathbb{N}$, $U_{i,t}$ is a random variable specific to the borrower and Y_t is the common factor following a general (adapted) stochastic process with deterministic initial value Y_0. Further, for simplicity, we assume that the duration of the debt is exactly one period and that the initial wealth fulfils

$$\log A_{i,t-1} = \Sigma_{j=1}^{t-1} Y_j + V_{i,t}$$

for all $i \leq n$ where $V_{i,t}$ is a centered variable specific to the borower - such an assumption makes sense, for instance, if Y_t stands for log-returns of a stock index which corresponds to the situation when the borrower owns a portfolio with the same composition as the index plus some additional assets.

Further, we suppose that all $(U_{i,t}, V_{i,t})_{i \leq n, t \in \mathbb{N}}$ are mutually independent and idependent of $(Y_t)_{t \in \mathbb{N}}$, and that all $Z_{i,t} = U_{i,t} + V_{i,t}$, $i \leq n$, $t \in \mathbb{N}$, are identically distributied with $\mathbb{E} Z_{1,1} = 0$, $\operatorname{var}(Z_{1,1}) = \sigma$, $\sigma > 0$, having a strictly increasing continuous cummulative distribution function Ψ (here, n is the number of borrowers). Note that we do not require increments of Y_t to be centered (which may be regarded a compensation for the term η present in (1) but missing in (2)).

2.3.3 Percentage loss in the generalized model

Denote $\bar{Y}_t = (Y_\tau)_{\tau \leq t}$ the history of the common factor up to the time t. Analogously to the original model, the conditional probability of the bankruptcy of the i-th borrower at the time t given $\bar{Y}_t$ equals to

$$\mathbb{P}(A_{i,t} < B_{i,t} | \bar{Y}_t) = \mathbb{P}(Z_{i,t} < log B_{i,t} - \Sigma_{j=1}^{t} Y_j | \bar{Y}_t) = \Psi\big(\log B_{i,t} - \Sigma_{j=1}^{t} Y_j\big),$$

where $B_{i,t}$ are the borrower's debts (installments) - we assume the debts to be the same for all the borrowers and all the times, i.e. $\log B_{i,t} = b, t \in \mathbb{N}, i \leq n$, for some b.

Ten primary topic of our interest is the percentage loss L_t of the entire portfolio of the loans at the time t. After taking the same steps as Vasicek (1991) (with conditional non-normal c.d.f.'s instead of the unconditional normal ones), we get, for a very large portfolio, that

$$L_t \doteq \Psi\left(b - \Sigma_{j=1}^{t} Y_j\right), \quad t \in \mathbb{N},$$

furter implying that

$$Y_t \doteq \Psi^{-1}(L_{t-1}) - \Psi^{-1}(L_t) \qquad (2.4)$$

and

$$L_t \doteq \Psi(\Psi^{-1}(L_{t-1}) - Y_t) \qquad (2.5)$$

the latter formula determining roughly the dynamics of the process of the losses, the former one allowing us to do statistical inference of the common factor based on the time series of the percentage losses.

To see that the Merton-Vasicek model is a special version of the generalized model, see the Appendix.

In our version of the model we assume $Z_{i,t}$ to be normally distributed and the common factor to be an ARCH process

$$Y_t = \sqrt{Y_{t-1}^2 + c}\ \varsigma_t,$$

where $\varsigma_1, \varsigma_2, \ldots$ are i.i.d. (possibly non-normal) variables and c is a constant.

Since the equation (2.3) may be rescaled by the inverse standard deviation of Z without loss of generallity, we may assume that Ψ is the standard normal distribution function.

As it was already mentioned, we assume the distribution of ς_1 to be generalized hyperbolic and we use the ML estimation to get its parameters - see the Appendix for details. In addition of the estimation of the parameters, we compare our choice of the distribution to several other classes of distributions.

2.3.4 The class of generalized hyperbolic distributions

Our model is based on the class of generalized hyperbolic distributions first introduced in Barndorff-Nielsen et al. (1985). The advantage of this class of distributions is that it is general enough to describe fat-tailed data. It has been shown (Eberlein, 2001, 2002, 2004) that the class of generalized hyperbolic distributions is better able to capture the variability in financial data than the normal distribution, which is used by the IRB approach. Generalized hyperbolic distributions have been used in an asset (and option) pricing formula (Rejman et al., 1997; Eberlein, 2001; Chorro et al., 2008), for the Value-at-Risk calculation of market risk (Eberlein, 2002; Eberlein, 1995; Hu & Kercheval, 2008) and in a Merton-based distance-to-default model to estimate PDs in the banking portfolio of commercial customers (e.g., Oezkan, 2002). We will show that the class of generalized hyperbolic distributions can be used for the approximation of a loss distribution for the retail banking portfolio with a focus on the mortgage book.

The class of generalized hyperbolic distributions is a special, quite young class of distributions. It is defined by the following Lebesque density:

$$gh(x;\lambda,\alpha,\beta,\delta,\mu) = a(\lambda,\alpha,\beta,\delta)(\delta^2 + (x\text{-}\mu)^2)^{\frac{\lambda\text{-}0,5}{2}} \times K_{\lambda\text{-}0,5}(\alpha\sqrt{((\delta^2 + (x\text{-}\mu)^2)})\exp(\beta(x\text{-}\mu)) \quad (6)$$

where

$$a(\lambda,\alpha,\beta,\delta) = \frac{(\alpha^2\text{-}\beta^2)^{0,5\lambda}}{\sqrt{2\pi}\cdot\alpha^{(\lambda\text{-}0,5)}\delta^{\lambda}K_{\lambda}(\delta\sqrt{\alpha^2\text{-}\beta^2})}$$

and K_λ is a Bessel function of the third kind (or a modified Bessel function – for more details on Bessel functions see Abramowitz, 1968). The GH distribution class is a mean-variance mixture of the normal and generalized inverse Gaussian (GIG) distributions. Both the normal and GIG distributions are thus subclasses of generalized hyperbolic distributions. μ and δ are scale and location parameters, respectively. Parameter β is the skewness parameter, and the transformed parameter $\bar{\alpha} = \alpha\delta$ determines the kurtosis. The last parameter λ is a determination of the distribution subclass. There are several alternative parameterizations in the literature using transformed parameters to obtain scale- and location-invariant parameters. This is a useful feature that will help us with the economic capital allocation to individual exposures. For the moment-generating function and for more details on the class of generalized hyperbolic distributions, see the Appendix.

Because the class of generalized hyperbolic distributions has historically been used for different purposes in economics as well as in physics, one can find several alternative parameterizations in the literature. In order to avoid any confusion, we list the most common parameterizations. These are:

$$\zeta = \delta\sqrt{\alpha^2 - \beta^2}, \ \ \rho = \frac{\beta}{\alpha}$$

$$\xi = (1 + \zeta)^{-0,5}, \quad \chi = \xi\rho$$

$$\overline{\alpha} = \alpha\delta, \ \ \overline{\beta} = \beta\delta$$

The main reason for using alternative parameterizations is to obtain a location- and scale-invariant shape of the moment-generating function (see the Appendix).

2.4 Data and results

2.4.1 Data description

To verify whether our model based on the class of generalized hyperbolic distributions is able to better describe the behavior of mortgage losses, we will use data for the US mortgage market. The dataset consists of quarterly observations of 90+ delinquency rates on mortgage loans collected by the US Department of Housing and Urban Development and the Mortgage Bankers Association.[9] This data series is the best substitute for losses that banks faced from their mortgage portfolios, relaxing the LGD variability (i.e. assuming that LGD = 100%). The dataset begins with the first quarter of 1979 and ends with the third quarter of 2009. The development of the US mortgage 90+ delinquency rate is illustrated in *Figure 2.1* and its descriptive statistics in

[9] 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.).

Table 2.1. We observe an unprecedentedly huge increase in the 90+ delinquency rate beginning with the second quarter of 2007.

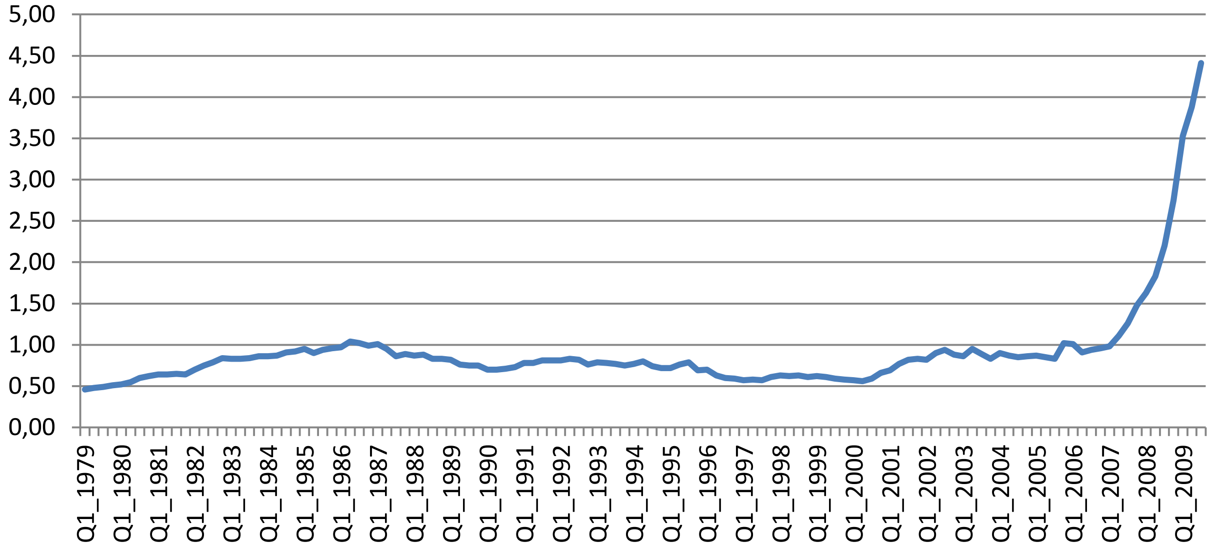

Figure 2.1: Development of US 90+ delinquency rate

Time Series Statistic	Value (90+ delinquency)
Mean	0,9417
Median	0,8100
Minimum	0,5300
Maximum	4,4100
Standard Deviation	0,6112
Skewness	4,0317
Kurtosis	17,0240
5th percentile	0,5600
95th percentile	2,1260

Table 2.1: Descriptive statistics of US 90+ delinquency rate

Starting our analysis, we have computed the values of the common factor "Y" using the formula (4). Quite interestingly, its evolution is indeed similar to the one of US stock market – see *Figure 2.2*, displaying the common factor (left axis), adjusted for inflation, against the S&P 500 stock index. A simple correlation analysis indicates that the common factor is lagged behind the index by two quarters (the value of the Pearson correlation coefficient is about 30%).

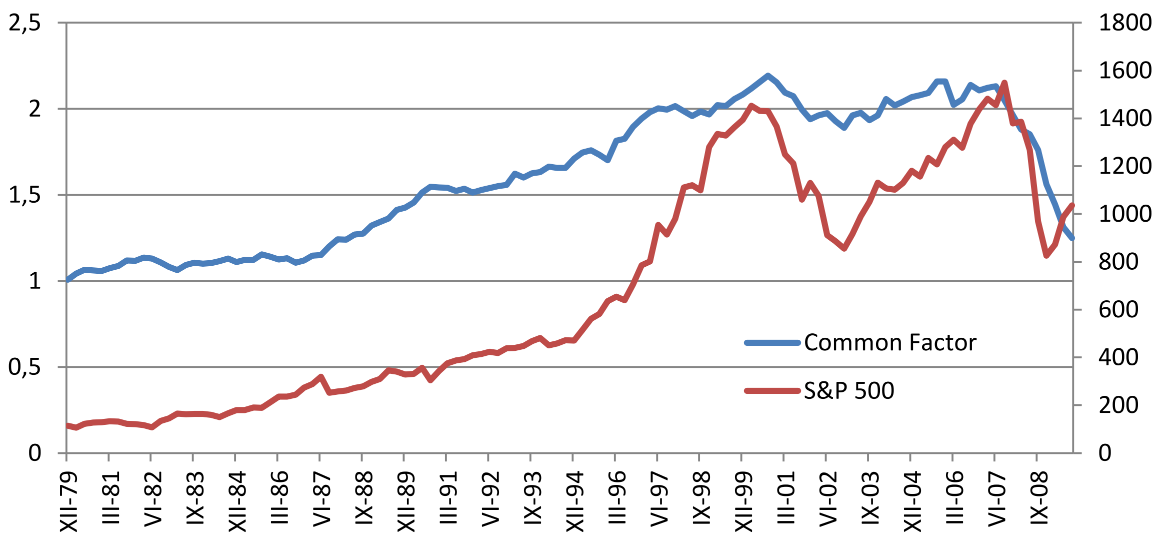

Figure 2.2: Comparison of the development of the common factor and lagged S&P 500 returns

A more exact estimation of the potential relationship, the autoregressive estimation, performed on log-changes (Y dependent on S&P), showed that there is a significant dependence of a change in the common factor on the change of the S&P 500 stock index, lagged by one quarter. The detailed results of the autoregressive estimation can be found in the *Table 2.2*. The regression R^2 was 23%.

Variable	Coefficient	Standard Error	P-value
Intercept	0.00024	0.00399	0.9520
S&P 5000	0.05631	0.02851	0.0507

Table 2.2Results of the autoregressive estimation of dependence of Y on S&P 500

2.4.2 Results

We considered several distributions for describing the distribution of ς_1 (hence of $(L_t)_{t\geq 1}$), namely loglogistic, logistic, lognormal, Pearson, inverse Gaussian, normal, lognormal, gamma, extreme value, beta and the class of generalized hyperbolic distributions. In the set of distributions compared, we were particularly interested in the goodness-of-fit of the class of generalized hyperbolic distributions and their comparison to other distributions. For more information on the MLE estimation we have performed, see the Appendix.

The second step is to test the hypothesis that the empirical dataset comes from the tested distribution. We used the chi-square goodness-of-fit test in the form:

$$\chi^2 = \sum_{i=1}^{t} (O_i - E_i)^2 / E_i, \qquad (2.6)$$

where O_i is the observed frequency in the i-th bin, E_i is the frequency implied by the tested distribution, and k is the number of bins. It is well known that the test statistic asymptotically follows the chi-square distribution with $(k - c)$ degrees of freedom, where c is the number of estimated parameters. In general, only the generalized hyperbolic distribution from all considered distributions was not rejected to describe the dataset based on the chi-square statistic (on a 99% level).

Figure 2.3 shows graphically the difference between the estimated generalized hyperbolic and normal distributions. From *Figure 2.3* we can see that the GHD is able to describe better both the skewness and the kurtosis of the dataset.

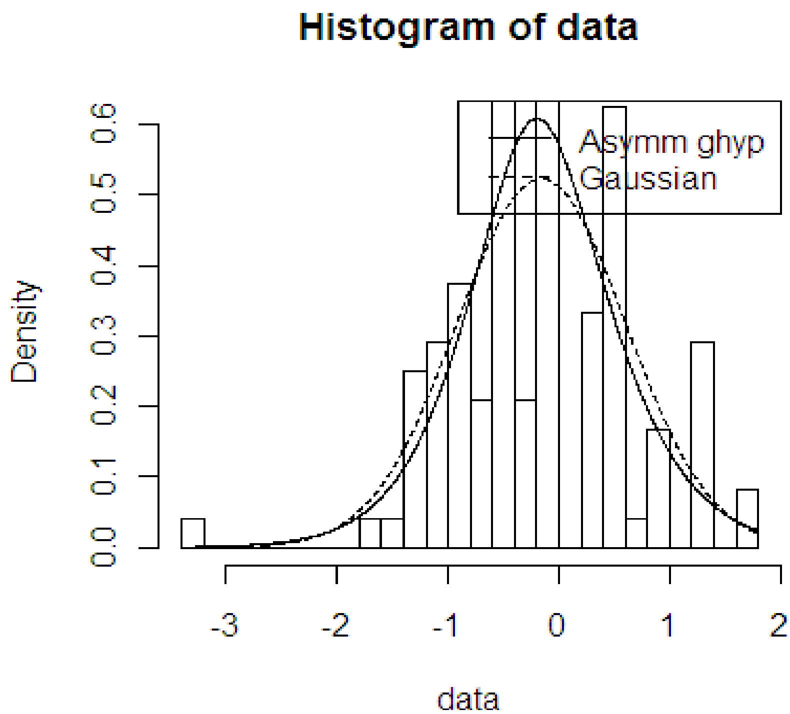

Figure 2.3: Compared histograms: GHD vs. Normal vs. dataset

The chi-square statistic show that the class of generalized hyperbolic distributions is the only one suitable to describe the behavior of delinquencies, even if we considered the dynamics of the common factor when using them. This fact can have a large impact on the economic capital requirement, as the class of generalized hyperbolic distributions is heavy-tailed and thus would imply a need for a larger stock of capital to cover a certain percentile delinquency. We will now demonstrate the difference between the economic capital requirements calculated under the assumption that mortgage losses follow a generalized hyperbolic distribution and under the Basel II IRB method (assuming standard normal distributions for both risk factors and a 15% correlation between the factors[10]). Note that we assume that all loans last only one period of time, therefore all loans enter the calculation as entrants at the beginning of the period and exit the calculation either by defaulting or a full repayment at the end of the period. Even though this is a significant limitation to our approach, it keeps our model simple and might be partially justified by the fact that some mortgages might be repaid at the time of interest rate re-fixation.

[10] The correlation 15% is a benchmark set for the mortgage exposures in the Basel II framework.

2.4.3 Economic capital at the one-year horizon: implications for the crisis

The IRB formula, defined in Pillar 1 of the Basel II Accord, assumes that losses follow a distribution that is a mix of two standard normal distributions describing the development of risk factors and their correlation. The mixed distribution is heavy-tailed and the factor determining how heavy the tails are is the correlation between the two risk factors. However, because the common factor is considered to be standard normally distributed, the final loss distribution's tails could be not heavy enough. If a heavy-tailed distribution will be considered for the common factor, the final loss distribution would probably have much heavier tails. Because the regulatory capital requirement is calculated at the 99.9% probability level, this disadvantage may lead to serious mistakes in the assessment of capital needs. To show the difference between the regulatory capital requirement (calculated by the IRB method) and the economic capital requirement calculated by our model, we will perform the economic capital requirement calculations at the 99.9% probability level as well.

When constructing loss forecasts, we repeatedly used (2.5) to get

$$L_{t+4} \doteq \psi(\psi^{-1}(L_t) - \sum_{1 \le i \le 4} Y_{t+1})$$

If we wanted to describe the distribution of the forecasted value we would face complicated integral expressions. We therefore decided to use simulations to obtain yearly figures. We were particularly interested in the following: the capital requirement based on average loss and the capital requirement based on last experienced loss. The average loss is calculated as a mean value from the original dataset of 90+ delinquencies and serves as a "through-the-cycle" PD estimate. This value is important for the regulatory-based model (Basel II) as a "through-the-cycle" PD should be used there. The last experienced loss is, on the second hand, important for our model with GHD distribution due to the dynamical nature of the model. The next Table summarizes our findings. To illustrate how our dynamic model would predict if the normal distribution of the common factor was used, we added this version of the dynamic model as well.

Model	Basel II IRB (through-the-cycle PD)	Our dynamic model with normal distribution	Our dynamic model with GHD
Distribution used for the individual factor	Standard Normal	Standard Normal	Standard Normal
Distribution used for the common factor	Standard Normal	Normal	Generalized Hyperbolic
99.9% loss	10.2851%	9.5302%	12.5040%

Table 2.3: Comparison of Basel II, Dynamic Normal and Dynamic GHD models tail losses

The first column in the *Table 2.3* relates to the IRB Basel II model, i.e. a model with a standard normal distribution describing the behavior of both risk factors and the correlation between these factors set at 15%. The PD used in the IRB formula (see Vasicek, 2002 for details) was obtained from the original dataset as an average default rate through the whole time period. The second column contains results from the dynamic model where a standard normal distribution of the individual risk factor is supplemented by the normal distribution, which describes the common factor and its parameters were estimated in the same way as those of GHD. The last column is related to our dynamic model where the GHD is assumed for the common factor. The results in the *Table 2.3* show that the dynamic model, based on the last experience loss, predicts higher quantile losses in the case of GHD and slightly lower in the case of Normal distribution, compared to the IRB formula. Thus, heavy tails of the GHD distribution evoke higher quantile losses than the current regulatory IRB formula, which at the end lead to a higher capital requirement.

2.5 Conclusion

We have introduced a new model for quantification of credit losses. The model is a generalization of the current framework developed by Vasicek and our main contribution lies in two main attributes: first, our model brings dynamics into the original framework and second, our model is generalized in that sense that any statistical distribution can be used to describe the behavior of risk factors.

To illustrate that our model is able to better describe past risk factor behavior and thus better predicts future need of capital, we compared the performance of several distributions common in credit risk quantification. In this sense, we were particularly interested in the performance of the class of Generalized Hyperbolic distributions, which is often used to describe heavy-tail financial data. For this purpose, we used a quarterly dataset of mortgage delinquency rates from the US financial market. Our suggested class of Generalized Hyperbolic distributions showed much better performance, measured by the Wasserstein and Anderson-Darling metrics, than other "classic" distributions like normal, logistic or gamma.

In the next section, we have compared our dynamic model with the current risk measurement system required by the regulation. The current banking regulation, summarized and formalized in the Second Basel Accord (Basel II, translated to Credit Requirements Directive or CRD in the EU), uses the standard normal distribution as an underlying distribution that drives risk factors for credit risk assessment. In the loss distribution, the mean value (expected loss) should be covered by banking fees and interest and the difference between the mean value and the 99.9^{th} quantile (unexpected loss) should be covered by the stock of capital. We were particularly interested in the difference between our dynamic model and the current IRB regulatory model, which is used to calculate the required stock of capital in every advanced bank subject to the Basel II regulation.

Our results show that the mix of standard normal distributions used in the Basel II regulatory framework was, at the 99.9% level of probability, underestimating the potential unexpected loss on the one-year horizon. Therefore, introducing the dynamics with a heavy-tailed distribution describing the common factor may lead to a better capturing of tail losses.

We have proved that using the normal distribution of risk factors development to quantify credit risk is an assumption that could be easily outperformed by choosing a different, alternative distribution, such as the class of generalized hyperbolic distributions. However, there are still several questions that need to be answered before the class of generalized hyperbolic distributions can be used for credit risk assessment. First question points at the use of the 99.9^{th} quantile. As this was chosen by the Basel II framework based on benchmarks from rating agencies, it is not sure, whether particularly this quantile should be required in our dynamic generalized model. Second, more empirical studies have to be performed to prove the goodness-of-fit of the class of generalized hyperbolic distributions. Third, the assumption that all loans last only one period is limiting. The final suggestion is to add an LGD feature to the calculation to obtain a general credit risk model.

Appendix

The moment-generating function for the class of generalized hyperbolic distributions is of the form:

$$M(u) = e^{u\mu}\left(\frac{\alpha^2-\beta^2}{\alpha^2-(\beta+u)^2}\right)^{\lambda/2}\frac{K_\lambda(\delta\sqrt{\alpha^2-(\beta+u)^2})}{K_\lambda(\alpha^2-\beta^2)}, \quad (2.1a)$$

where *u* denotes the moment. For the first moment, the formula simplifies to (see e.g. Eberlein, 2001 for details):

$$M'(0) = E(x) = \mu + \frac{\beta\delta}{\sqrt{\alpha^2-\beta^2}}\frac{K_{\lambda+1}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}{K_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)}, \quad (2.2a)$$

The second moment is calculated in a (technically) more difficult way:

$$M''(0) = Var(x) = \delta^2\left(\frac{K_{\lambda+1}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}{\left(\delta\sqrt{\alpha^2-\beta^2}\right)K_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)}\right) +$$

$$+\frac{(\beta\delta)^2}{\alpha^2-\beta^2}\left(\frac{K_{\lambda+2}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}{K_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)} - \left(\frac{K_{\lambda+1}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}{K_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)}\right)^2\right) \quad (2.3a)$$

By substituting from equations (2.2a) and (2.3a) into equation (2.1a) we obtain much simpler expression for the first and second moments of the class of generalized hyperbolic distributions. The following equations express the first and the second moment of the class of generalized hyperbolic distributions in their scale- and location-invariant shape:

$$M(1) = E(x) = \mu + \frac{\beta\delta}{\sqrt{\alpha^2-\beta^2}}\frac{K_{\lambda+1}(\zeta)}{K_\lambda(\zeta)},$$

$$M(2) = Var(x) = \delta^2\left(\left(\frac{K_{\lambda+1}(\zeta)}{\zeta K_\lambda(\zeta)}\right) + \frac{\beta^2}{\alpha^2-\beta^2}\left(\frac{K_{\lambda+2}(\zeta)}{K_\lambda(\zeta)} - \left(\frac{K_{\lambda+1}(\zeta)}{K_\lambda(\zeta)}\right)^2\right)\right)$$

On MLE estimation of the parameters

To estimate the parameters of the model, i.e. the constant c and the vector of the parameters Θ of (the distribution of) ς_1, we apply the (quasi) ML estimate to the sample $Y_2, Y_3, \ldots$ computed from (4), using the fact, that the conditional density of Y_t given $\bar{Y}_{t-1}$ is

$$f(y; c.\Theta) = \rho_t(c)\varphi(\rho_t y; \Theta) \quad \rho_t(c) = [Y_{t-1}^2 + c]^{-1/2}$$

where $\varphi(z; \Theta)$ is the p.d.f. of the generalized hyperbolic distribution with parameters Θ. The (quasi) log-likelihood function is then

$$L(c.\Theta) = \sum_{i=2}^{T} \log(\rho_t(c)) + \sum_{i=2}^{T} \log(\varphi(\rho_t(c)Y_i; \Theta))$$

Therefore, we may find its maximum in two steps: maximize $K(c) = \max_\Theta L(c.\Theta)$ where the right hand side is determined using the standard ML procedure for g.h. distributions.

The Merton-Vasicek model as a special case of our generalized framework

In the present section, we show how our generalized model relates to the original one. Let us start with the computation of the loss's distribution, given that the probability of default

$$p_t = \mathbb{P}(A_{i,t} < B_{i,t} | \bar{Y}_{t-1})$$

is known (e.g. estimated by a credit scoring): In this case then

$$F(\theta|\bar{Y}_{t-1}) = 1 - \Phi_t(\chi_t^{-1}(p_t) - \Psi^{-1}(\theta)).$$

where χ_t is the conditional c.d.f. of the variable $\xi_t := Y_t + Z_{1,t}$ and Φ_t is the conditional distribution function of Y_t .

To see it, note that

$$p_t = \mathbb{P}(\xi_t < b - \Sigma_{j=1}^{t} Y_j | \bar{Y}_{t-1}) = \chi(b - \Sigma_{j=1}^{t-1} Y_j)$$

and that

$$\mathbb{P}(L_t < \theta|\bar{Y}_{t-1}) = \mathbb{P}(\Psi\big(b - \Sigma_{j=1}^{t-1} Y_j\big) < \theta|\bar{Y}_{t-1})$$

$$= \mathbb{P}(\Psi(\chi^{-1}(p_t) - Y_t) < \theta) = \mathbb{P}(Y_t > \chi^{-1}(p_t) - \Psi^{-1}(\theta))$$

$$= 1 - \Phi_t(\chi^{-1}(p_t) - \Psi^{-1}(\theta)).$$

Now, turn our attention to the correlations of the risk factors of different loans: Denoting $X_{i,t} := Y_t + Z_{i,t}$, we get

$$\mathrm{cov}(X_{i,t}, X_{j,t} \big| \bar{Y}_{t-1}) = \mathrm{var}(X_{i,t}, \big| \bar{Y}_{t-1}) = \mathrm{var}(Y_t \big| \bar{Y}_{t-1})$$

and, consequently,

$$corr(X_{i,t}, X_{j,t} \big| \bar{Y}_{t-1}) = \frac{\mathrm{var}(Y_t \big| \bar{Y}_{t-1})}{\mathrm{var}(Y_t \big| \bar{Y}_{t-1}) + \mathrm{var}(Z_t)}$$

In particular, if we assume Y_1, Y_2, ... to be i.i.d. and

$$Y_1 \colon N(0, \rho), \qquad Z_{1,1} \colon N(0, 1 - \rho)$$

for some ρ, then clearly $\xi_t \colon N(0,1)$ implying

$$\mathbb{P}(L_t < \theta|\bar{Y}_{t-1}) = 1 - N\left(\frac{N^{-1}(p_t) - \sqrt{1-\rho}N^{-1}(\theta)}{\sqrt{\rho}}\right).$$

$$= N\left(\frac{\sqrt{1-\rho}N^{-1}(\theta) - N^{-1}(p_t)}{\sqrt{\rho}}\right)$$

and

$$\mathrm{corr}(\mathrm{X}_{i,t}, X_{j,t}|\bar{Y}_{t-1}) = \rho$$

i.e. the formuls of Vasicek (2002).

References

Abramowitz, S. (1968). *Handbook of Mathematical Functions.* New York: Dover publishing.

Andersson, F., Mausser, H., Rosen, D., & Uryasev, S. (2001). Credit Risk Optimization with Conditional Value-at-Risk Criterion. *Mathematical Programming*, 273–271.

Anderson, T. W., & Darling, D. A. (1952). Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. *Annals of Mathematical Statistics 23* , 193-212.

Bank for International Settlements (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. E., Blæsild, P., & Jensen, J. L. (1985). The Fascination of Sand. *A Celebration of Statistics*, 57–87.

Chorro, C., Guegan, D., & Ielpo, F. (2008). *Option Pricing under GARCH Models with Generalized Hyperbolic Innovations (II): Data and Results.* Paris: Sorbonne University.

Eberlein, E. (2001). Application of Generalized Hyperbolic Lévy Motions to Finance. Chapter in Barndorff-Nielsen et al.: *Lévy Processes: Theory and Applications*, pp. 319–337.

Eberlein, E. (2002). The Generalized Hyperbolic Model: Financial Derivatives and Risk Measures. *Proceedings from the Mathematical Finance-Bachelier Congress 2000*, 245–267.

Eberlein, E. (2004). Generalized Hyperbolic and Inverse Gaussian Distributions: Limiting Cases and Approximation of Processes. *Seminar on Stochastic Analysis, Random Fields and Applications IV, Progress in Probability 58*, 221–264.

Eberlein, E., & Keller, U. (1995). Hyperbolic Distributions in Finance. *Bernoulli, 1*, 281–299.

European Commission (2006). *DIRECTIVE 2006/49/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 June 2006.* European Commission.

Hu, W., & Kercheval, A. (2008). *Risk Management with Generalized Hyperbolic Distributions.* Florida State University.

JP Morgan. (n.d.). *CreditMetrics*.

Merton, R. C. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *Journal of Finance 29*, Chapter 12.

Merton, R. C. (1973). Theory of Rational Option Pricing. *Bell Journal of Economics and Management Science, 4*, Chapter 8.

Moody's. (n.d.). *Moody's KMV*. Retrieved from www.moodyskmv.com

Oezkan, F. (2002). *Lévy Processes in Credit Risk and Market Models.* University of Freiburg.

Rejman, A., Weron, A., & Weron, R. (1997). Option Pricing Proposals under the Generalized Hyperbolic Model. *Stochastic Models, Volume 13, Issue 4*, 867–885.

Saunders, A., & Allen, L. (2002). *Credit Risk Measurement: New Approaches to Value at Risk and Other Paradigms.* John Wiley and Sons.

Vasicek, O. A. (1987). *Probability of Loss on Loan Portfolio.* KMV.

Vasicek, O. A. (1991). *Limiting loan loss probability distribution,* KMV

Vasicek, O. A. (2002). *The distribution of loan portfolio value, RISK 15, 160-162*

3. Dynamic Multi-Factor Credit Risk Model with Fat-Tailed Factors

3.1 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)[11], 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 macroeconomic 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 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

[11] 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.

(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).[12]

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.

3.2 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.

[12] 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.

3.2.1 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).

3.2.2 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 correlated, which is equivalent to saying that the individual's wealth can be decomposed into a systematic and an idiosyncratic part (see (3.1) and (3.2)). While the systematic 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…).[13]

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

$$\log A_{i,1} = \log A_{i,0} + \eta + \gamma X_i. \qquad (3.1)$$

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

$$X_i = Y + Z_i, \qquad (3.2)$$

[13] The systematic factor is exogenous to both the KMV and to our model. For interesting research into the relations of systematic factors among various financial and insurance sectors, see Billio et al. (2012).

where Y is the common factor and $Z_1, Z_2, \ldots$ 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

$$PD_i = \mathrm{P}\left[A_{i,1} < B_i\right] = \mathrm{P}[X_i < c_i] \qquad c_i = \frac{\log B_i - \log A_{i,0} - \eta}{\gamma}. \tag{3.3}$$

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

After some calculations we obtain the *rate of default* (RD)[14], defined as

$$RD = \frac{number\ of\ defaults}{number\ of\ loans}, \tag{3.4}$$

which approximately fulfills

$$\mathrm{P}\left[RD \le x\right] = \mathrm{N}\left(\frac{\left(\sqrt{1-\rho}\right)\mathrm{N}^{-1}(x) - \mathrm{N}^{-1}(PD)}{\sqrt{\rho}}\right), \tag{3.5}$$

given a sufficiently large number of loans. Here, N denotes the standard normal cumulative distribution function and $PD = PD_1$ [15] (for more details of the calculation see Vašíček, 1987). It follows that the distribution of *RD* is heavy-tailed, with the heaviness of the tail dependent on the correlation ρ.

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 R.

[14] The quantity which we call RD is sometimes called the *empirical* or *observed PD*. We use a different name so as not to suggest that RD is an estimate of PD (it is clear from (3.5) that RD is neither unbiased nor consistent).

[15] Note that $PD_i = PD$ for any i because the individual factors are equally distributed.

3.2.3 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 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:

$$LGD_i = \max[0; 1 - Collateral_i]$$

while the collateral value is expressed as

$$Collateral_i = \mu_i (1 + \sigma_i C_i),$$

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

$$C_i = \sqrt{q} Y + \sqrt{1-q} E_i . \qquad (3.6)$$

The loss distribution is taken from the Vasicek framework (i.e., fulfilling (3.1)) with

$$X_i = \sqrt{p} Y + \sqrt{1-p} Z_i , \qquad (3.7)$$

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 systematic and two idiosyncratic factors, starting from the same point as Frye:

$$C_i = \sqrt{q}Y + \sqrt{1-q}E_i \quad , \qquad (3.8)$$

$$E_i = \sqrt{w}Z_i + \sqrt{1-w}E_i' \ , \qquad (3.9)$$

where the systematic 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 systematic factors, one common to the defaults and the other specific to LGD.

3.3 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.

3.3.1 Model for Defaults

Analogously to Vasicek, we assume that

$$\log A_{i,t} = \log A_{i,t-1} + \Delta Y_t + U_{i,t}, \quad i \le n, \qquad (3.10)$$

where n is the number of borrowers, $A_{i,t}$ is the wealth of the i-th borrower at time $t \in \mathrm{N}$, $U_{i,t}$ 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[16] and that the initial wealth in each period equals

$$\log A_{i,t-1} = Y_{t-1} + V_{i,t}, \quad i \le n, \qquad (3.11)$$

where $V_{i,t}$ is a random variable specific to the i-th borrower. Further, we assume all $(U_{i,t},)_{i\le n, t\in \mathrm{N}}$ to be mutually independent and independent of $(\Delta Y_t)_{t\in\mathrm{N}}$, and all $Z_{i,t}, Z_{i,t} = U_{i,t} + V_{i,t}$, $i \le n,\ t \in \mathrm{N}$ to be identically distributed with $\mathrm{E}Z_{1,1} = 0$, $\mathrm{var}(Z_{1,1}) = \sigma$, $\sigma > 0$, $Z_{1,1}$, having a strictly increasing continuous cumulative distribution function Ψ. 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).[17]

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

[16] This is a very restrictive assumption, which is a point of our further research; however, the assumption is the tax paid for the model's simplicity

[17] 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.

$$\mathrm{P}[A_{i,t} < B_{i,t} \mid \bar{Y}_t] = \mathrm{P}[Z_{i,t} < logB_{i,t} - Y_t \mid \bar{Y}_t] = \Psi\left(\log B_{i,t} - Y_t\right), \qquad (3.12)$$

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

Our primary topic of interest is the rate of default (RD), which we define in our framework as $R_t = \frac{number\ of\ defaults\ at\ t}{n}$. As n is the number of borrowers, the definition is equivalent to that in the Section 3.2.2. If we assume the debts to be the same for all borrowers and at all times, i.e., $\log B_{i,t} = b, t \in \mathrm{N}$, $i \leq n$, for some b, and if we approximate $R_t \doteq \lim_n \frac{number\ of\ defaults\ at\ t}{n}$, we may apply the Law of Large Numbers to the conditional probabilities described in (3.12) (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):

$$R_t \doteq \mathbb{P}\left(A_{i,t} < b \middle| \bar{Y}_t\right) = \Psi(b - Y_t), \quad t \in \mathbb{N}, \qquad (3.13)$$

further implying that

$$\Delta Y_t \doteq \Psi^{-1}(L_{t-1}) - \Psi^{-1}(L_t) \qquad (3.14)$$

and

$$R_t \doteq \Psi(\Psi^{-1}(R_{t-1}) - \Delta Y_t). \qquad (3.15)$$

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., Ψ is the cumulative distribution function (CDF) of the standard normal distribution.

3.3.2 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 systematic 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

$$\log P_{i,t} = \log a_i + I_t + E_{i,t}, \qquad (3.16)$$

(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 $(I_t, Y_t)_{t>0}$ and all the individual factors described in subsection 3.3.1 (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

$$G_i = \frac{\min(P_{i,t}; C_i)}{C_i}. \qquad (3.17)$$

Furthermore, let us say that $C_i = C$, $a_i = a, i \le N$ 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 ΔI). Then

$$G_i = \min\left(e^{I_t + E_{i,t}}; 1\right) = \exp\left\{\min\left(I_t + E_{i,t}; 0\right)\right\}. \qquad (3.18)$$

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

$$\tilde{G}_t = \lim_N \frac{1}{N} \sum_{i=1}^{N} G_i = \mathrm{E}(G_1 \mid I_t)\,. \qquad (3.19)$$

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

$$\tilde{G} = \mathrm{E}\left(e^I e^{\min(E_1;-I)} \mid I\right) = e^I \mathrm{E}\left(e^{\min(E_1;-I)} \mid I\right) = e^I \left[\int_{-\infty}^{-I} e^x \mathrm{d}F(x) + e^{-I}\left(1 - F(-I)\right)\right]$$

$$= e^I \int_{-\infty}^{-I} e^x \mathrm{d}F(x) + 1 - F(-I) \,, \qquad (3.20)$$

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

$$D_t = 1 - \tilde{G}_t = h(I_t)\,, \qquad (3.21)$$

where

$$h(\iota) = F(-\iota) - e^{\iota} \int_{-\infty}^{-\iota} e^x \mathrm{d}F(x)\,, \qquad (3.22)$$

or, after integrating by parts,

$$h(\iota) = e^{\iota} \int_{-\infty}^{-\iota} F(x) e^x \mathrm{d}x\,. \qquad (3.23)$$

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

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

$$h(\iota) = h_\sigma(\iota) = \Phi\left(-\frac{\iota}{\sigma}\right) - \exp\left(\iota + \frac{1}{2}\sigma^2\right)\Phi\left(-\frac{\iota}{\sigma} - \sigma\right), \qquad (3.24)$$

$$h_{\sigma}^{'}(\iota) = \exp\left(\iota + \frac{1}{2}\sigma^2\right)\left[\frac{1}{\sigma}\varphi\left(-\frac{\iota}{\sigma} - \sigma\right) - \Phi\left(-\frac{\iota}{\sigma} - \sigma\right)\right] - \frac{1}{\sigma}\varphi\left(-\frac{\iota}{\sigma}\right), \qquad (3.25)$$

where φ is the standard normal probability density function and where the derivative of h is with respect to ι. For the calculation of (3.24), see the *Appendix*.

3.3.3 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 $Q_1, Q_2, \ldots$ such that

$$Q_t(\omega_t; \lambda) = \varepsilon_t, \quad \omega_t = (Y_1, I_1, Y_2, I_2, \ldots, Y_t, I_t), \qquad (3.26)$$

for each *t*, where λ is a (vector) parameter and $\varepsilon_1, \varepsilon_2, \ldots$ is a sequence of i.i.d. two-dimensional random variables whose density η possibly depends on a (vector) parameter μ. 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 ω_{t-1} is, by the formula for transformed density,

$$\upsilon_t(y, \iota; \lambda, \mu) = \eta\left(Q_t(\omega_{t-1}, y, \iota; \lambda); \mu\right) | D_t(y, \iota) |, \qquad (3.27)$$

where $D_t(y, \iota)$ 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 $R_1, D_1, R_2, D_2 \ldots, R_T D_T$ at our disposal and we want to estimate all parameters of our model, i.e., λ, μ, and σ. A straightforward way to do this is by maximum likelihood estimation, with the likelihood function taking the form of

$$L\left(R_1, D_1, \ldots, R_T, D_T; \lambda, \mu, \sigma\right) =$$

$$= \sum_{t=1}^{T}\left[\log\left(\eta\left(Q_t\left(\omega_t\right)\right)\right) + \log\left(\left|D_t\left(Y_t, I_t\right)\right|\right) - \log\left(\Psi'(\Psi^{-1}\left(R_t\right))\right) - \log\left(-h_\sigma^{'}\left(h_\sigma^{-1}\left(D_t\right)\right)\right)\right] \quad (3.28)$$

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

3.4 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.

3.4.1 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.[18] All data start with the first quarter of 1979 and end with the third quarter of 2009. Thus, the difficult period of the subprime mortgage crisis and the subsequent real recession is included.

3.4.2 Estimation

To estimate our model, we proceeded as follows. First, we extracted factor *Y* from the values of R_t. Second, we computed factor *I* from the values of *D* by employing *h* specified in (3.24); since the function *h*, which maps D_t to *I*, depends also on parameter σ, we estimated the model for a sufficient number of values of σ. Third, we found a suitable model for the dynamics of the pair

[18] 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.)

(Y, I). Finally, we estimated the model of the series (*Y*, *I*) for each σ and chose the version with the highest likelihood.

3.4.2.1 Extraction of Y

As a proxy for the rate of default (denoted by R_t), we used the series of 90+ delinquency rates[19] depicted in *Figure 3.1*. 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.

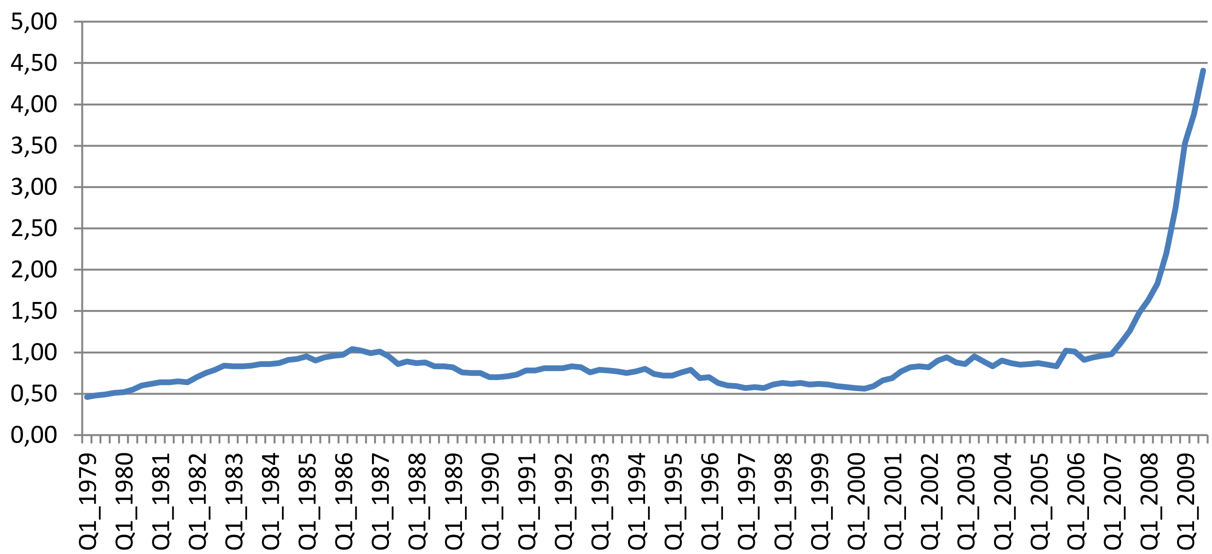

Figure 3.1: The US 90+ delinquency rates – the proxy for RD (R_t)

The values of the common factor "*Y*" were computed by means of (3.15). 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 3.2*). 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 one to two quarters and that both datasets are significantly correlated (the value of the Pearson

[19] The 90+ delinquency rate is the proportion of all receivables 90 or more days past due in a given quarter.

correlation coefficient is about 30%, which is significant at 5%). Additionally, the autoregressive analysis in (Gapko & Šmíd) showed a strong dependence of Y on the S&P 500 lagged by one quarter.

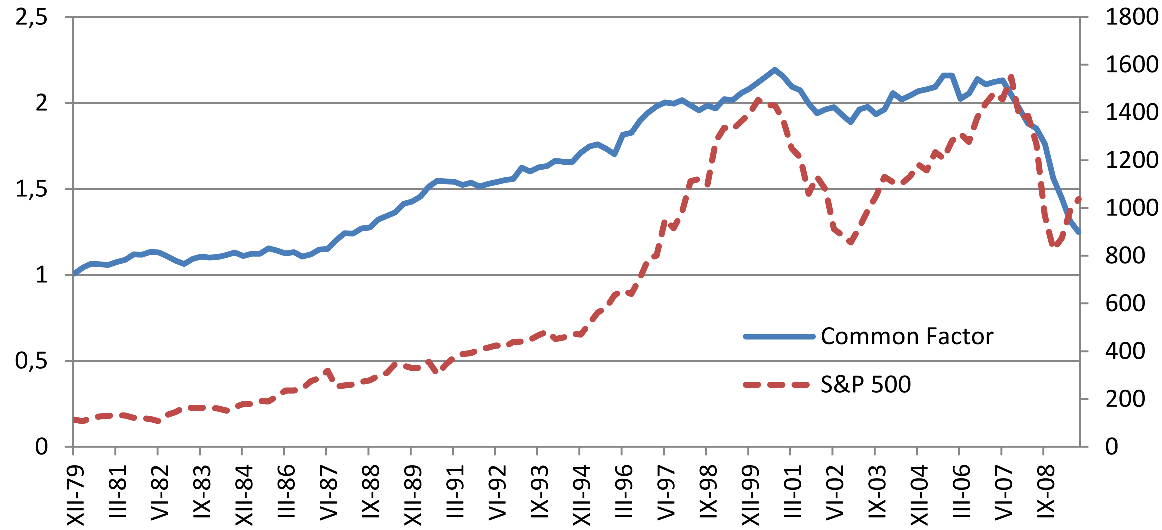

Figure 3.2: Comparison of the common factor *Y* and the lagged S&P 500 index (values of the common factor on the left-hand scale; values of the S&P 500 on the right-hand scale)

3.4.2.2 Extraction of I

As a proxy for the LGD (denoted by D_t in our paper), the proportion of started foreclosures[20] 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. In other words, the proxy represents all possible factors except changes in the collateral (residential real estate) price movements. We are aware that this is a simplification, however, the provided dataset is the (to our knowledge) best available approximation of LGD for the overall US mortgage market. The resulting series of D_t is plotted in *Figure 3.3*.

[20] 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.

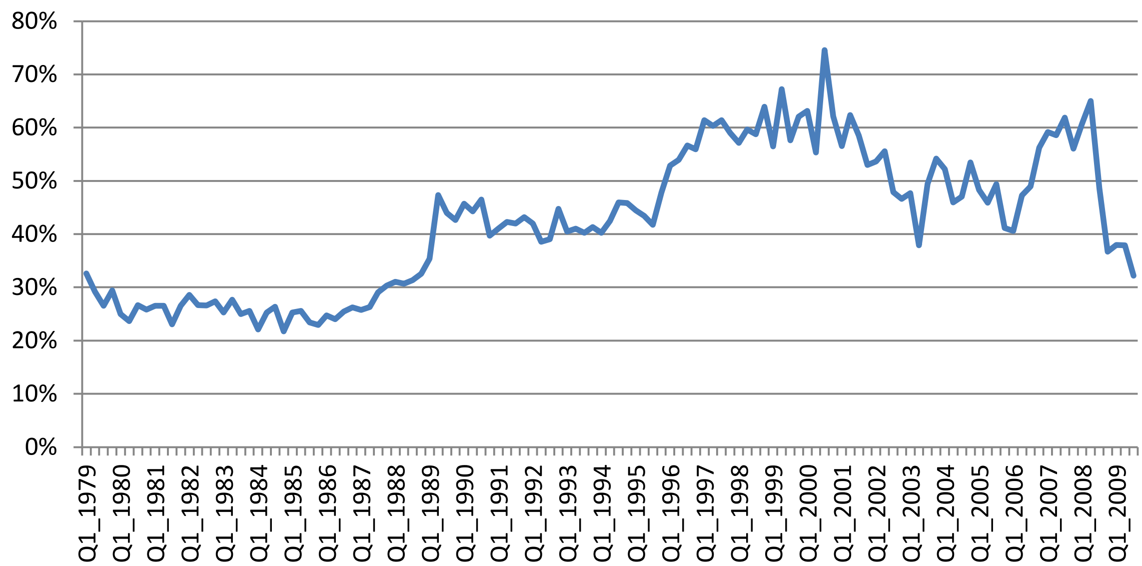

Figure 3.3: Foreclosures/90+ delinquencies – the proxy for LGD (D_t)

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.

3.4.2.3 Selection of the Model for (Y, I)

The two time series used to estimate the joint model of PD and LGD behave in a different way, which is illustrated in the *Table 3.1*, where the descriptive statistics of both R_t and D_t are summarized. Thus we analyzed the datasets separately and then estimate the mutual relationship. After a preliminary analysis of the series of Y we found clear ARCH behavior of the factor, hence we decided to analyze the transformed version of the factor

$$y_t = \frac{\Delta Y_t}{|\Delta Y_{t-1}|}$$

instead of its original values.

Time Series Statistic	Value (90+ delinquency)	Value (foreclosures)
Mean	1,1329	0,4343
Median	0,8300	0,3400
Minimum	0,4600	0,1300
Maximum	5,0200	1,4700
Standard Deviation	0,9869	0,3054
Skewness	2,5856	1,8730
Kurtosis	5,5353	2,5721
5th percentile	0,5570	0,1500
95th percentile	3,7190	1,2090

Table 3.1: Descriptive statistics of R_t and D_t

The stationarity of both time series was rejected as the Augmented Dicky-Fuller's test didn't reject the unit root hypothesis. Therefore, we suspected that the factors Y and I can be potentially nonstationary as well, which was confirmed by the Augmented Dicky Fuller's test.

For a sufficiently dense set of the values of σ, 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.,

$$\Delta y_{t-1} = \alpha_1 + \beta_1 \Delta y_{t-1} + \gamma_1 \Delta I_{t-1} + \delta_1 e_{t-1} + \varepsilon_{1,t} \qquad (3.29)$$

$$dI_t = \alpha_2 + \beta_2 \Delta y_{t-1} + \gamma_2 \Delta I_{t-1} + \delta_2 e_{t-1} + \varepsilon_{2,t}, \qquad (3.30)$$

where Δy and ΔI are the first differences of y and I and e is an error correction term. For each of the examined values of σ, we computed the maximum likelihood function of the VECM model by means of (3.28) and chose σ = 12% as the estimate of σ since this value gave the greatest likelihood. We found it very interesting that the estimated σ intuitively corresponds to the standard deviation of real estate prices (Quigley, 1999).

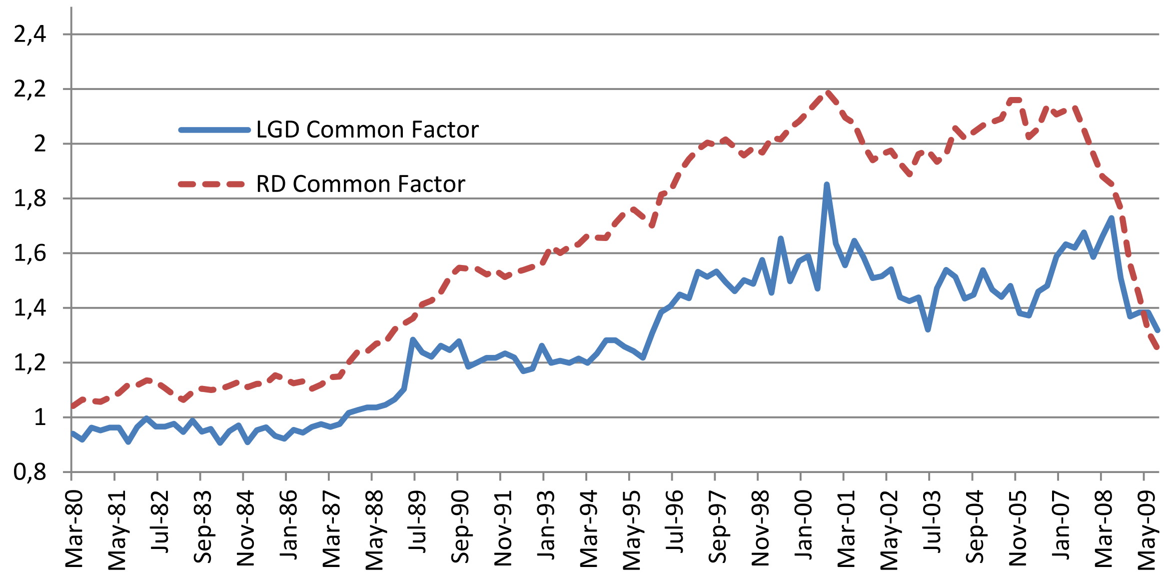

Figure 3.4: Graphical comparison of *Y* and *I* common factors

Figure 3.4 compares the two common factors. It seems obvious that these two show some similarities. To confirm whether a cointegration relationship exists between Δy and ΔI, we performed the Engel-Granger cointegration test. The results confirmed that both datasets are nonstationary; however, the unit root test of the cointegrating regression residuals showed that we can't reject the null hypothesis of unit root test. Despite the test not confirming fully our hypothesis of cointegration, we decided to estimate the VECM.

The resulting VECM model with *Y* as the dependent variable in the first equation and *I* in the second one is summarized in *Table 3.2* (in accordance with the definition of the model, cointegration rank 1 was assumed).

1st 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
2nd equation (*I* dependent)	**Coefficient Value (SE)**	**p-value**
Constant	-0.299802 (0.321560)	0.3532
Δ y (t-1)	-0.106660 (0.288617)	0.7124
Δ I (t-1)	-0.362746 (0.0904674)	0.0001
Error Correction Term	0.293693 (0.310262)	0.3459

Table 3.2: Estimated coefficients of the VECM model

From the *Table 3.2* 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). Also, it is worth mentioning that the dependence of *Y* on *I* is much stronger than the dependence of *I* on *Y*. The R^2 of the whole model is around 30%. Thus we found a cointegration between *Y* and *I*, which, on the other hand, is weaker than we expected (but still strong enough to show a time series inter-dependency).

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.

Before the end of this section, let us describe the derivation of the ML function (3.28) in detail. First, note that

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

where $\mathrm{S_t}$ and $\mathrm{M_t}$ are matrices possibly containing past values of both (transformed) factors *Y* and *I*. Since, in (3.28), the term $log\left|D_t(Y_t, I_t)\right| = -log\left(\left|\Delta Y_{t-1}\right|\right)$ does not depend on any parameter, it can be excluded from the maximization, so the ML estimate can be obtained by maximizing

$$\mathrm{L}\left(R_1, D_1, \ldots, R_T, D_T; \lambda, \mu, \sigma\right) = \sum_{\mathrm{t}=1}^{\mathrm{T}} \left[\log\left(\eta_1\left(\varepsilon_t^1\right)\right) + \log\left(\eta_2\left(\varepsilon_{\mathrm{t}}^2\right)\right) - \log\left(-h_{\sigma}^{'}\left(h_{\sigma}^{-1}\left(D_t\right)\right)\right) \right]$$

$$= \dot{\mathrm{L}}\left(Z_t\left(\varepsilon_t\right)_{t=1\ldots T}\right) - \sum_{\mathrm{t}=1}^{\mathrm{T}} \log\left(h_{\sigma}^{'}\left(h_{\sigma}^{-1}\left(D_t\right)\right)\right), \qquad (3.31)$$

where $\dot{\mathrm{L}}$ Is the likelihood function of the VECM model, $(\varepsilon_1^i, \varepsilon_2^i, \ldots)$ are the residuals from the *i*-th equation of the VECM model, and η_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).

3.4.3 Predictions

Having the model, we computed the quantiles of both RD and LGD on the 99.9$^{\mathrm{th}}$ percentile probability level, i.e., on the level used in the Basel II framework.

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

$$R_{t+4} \quad R_t) - \sum_{1 \le i \le 4} \Delta Y_{t+i}), \tag{3.32}$$

which can be easily achieved by using (3.15) four times consecutively. Technically, this was achieved by simulating Y four time periods to the future and deducting the sum of the predictions from the quantile at R_t.

3.4.3.1 Quantile of RD

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 RD 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[21]). The result is summarized in *Table 3.3*.

Model	**Basel II IRB (through-the-cycle PD)**	**Our dynamic model with GHD**
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 3.3: Comparison of Basel II and Dynamic GHD models tail RD

[21] The 15% correlation is the benchmark set for mortgage exposures in the Basel II framework.

The results show that our model predicts a lower value of the quantile of RD 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.

3.4.3.2 Quantile of LGD

Similarly to RD, we computed the quantiles of LGD (by means of simulations again). The resulting 99.9^{th} LGD quantile calculated by our model, 40.6%, is slightly below the regulatory 45% benchmark. The other computed quantiles are summarized in *Table 3.4.*

99^{th} quantile LGD	99.9^{th} quantile LGD	99.99^{th} quantile LGD
29.8%	40.6%	50%

Table 3.4: Selected LGD quantiles in our model

3.5 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 RD 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 RD 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 RD 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 RD 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 3.3. First we specify its derivative:

$$h'(\iota) = e^{\iota}\left(\int_{-\infty}^{-\iota} F(x)e^{x}\mathrm{dx} - F(-\iota)e^{-\iota}\right)$$

$$= e^{\iota}\left(\int_{-\infty}^{-\iota} F(x)e^{x}\mathrm{dx} - F(-\iota)\int_{-\infty}^{-\iota} e^{x}\mathrm{dx}\right)$$

$$= e^{\iota}\int_{-\infty}^{-\iota}[F(x) - F(\iota)]e^{x}\mathrm{dx} < 0.$$

Second, we evaluate the function given that further E_1 is normal with variance σ^2:

$$h_\sigma(\iota) = \Phi\left(-\frac{\iota}{\sigma}\right) - \exp(\iota)\int_{-\infty}^{-\iota}\frac{1}{\sqrt{2\pi\sigma}}\exp\left(-\frac{x^2}{\sigma^2}\right)\exp(x)\,\mathrm{dx}$$

$$= \Phi\left(-\frac{\iota}{\sigma}\right) - \exp(\iota)\frac{1}{\sqrt{2\pi\sigma}}\int_{-\infty}^{-\iota}\exp\left(-\frac{x^2}{\sigma^2} + x\right)\mathrm{dx}$$

$$= \Phi\left(-\frac{\iota}{\sigma}\right) - \exp(\iota)\frac{1}{\sqrt{2\pi\sigma}}\int_{-\infty}^{-\iota}\exp\left(-\frac{1}{2\sigma^2}(x^2 - 2x\sigma^2 + \sigma^4) + \frac{1}{2}\sigma^2\right)\mathrm{dx}$$

$$= \Phi\left(-\frac{\iota}{\sigma}\right) - \exp(\iota)\frac{1}{\sqrt{2\pi\sigma}}\int_{-\infty}^{-\iota}\exp\left(-\frac{1}{2\sigma^2}(x - \sigma^2)^2 + \frac{1}{2}\sigma^2\right)\mathrm{dx}$$

$$= \Phi\left(-\frac{\iota}{\sigma}\right) - \exp\left(\iota + \frac{1}{2}\sigma^2\right)\int_{-\infty}^{-\iota}\frac{1}{\sqrt{2\pi\sigma}}\exp\left(-\frac{(x-\sigma^2)^2}{2\sigma^2}\right)\mathrm{dx}$$

$$= \Phi\left(-\frac{\iota}{\sigma}\right) - \exp\left(\iota + \frac{1}{2}\sigma^2\right)\mathbb{P}[N(\sigma^2, \sigma^2) < -\iota]$$

$$= \Phi\left(-\frac{\iota}{\sigma}\right) - \exp\left(\iota + \frac{1}{2}\sigma^2\right)\Phi\left(-\frac{\iota}{\sigma} - \sigma\right)$$

(recall that Φ 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. *IES Working Papers*, WP 32/2000.

Gordy, M. B. (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.moodysanalytics.com/~/media/Insight/Quantitative-Research/Portfolio-Modeling/10-05-02-Implications-of-PD-LGD-Correlation-in-a-Portfolio-Setting.ashx

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

Quigley, J. M. (1999). Real Estate Prices and Economic Cycles. *International Real Estate Review*, 2(1): 1-20.

RiskMetrics Group, I. (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 th EURO Working Group on Financial Modelling* (pp. 153-162). České Budějovice: VŠB - Technical University of Ostrava.

Vasicek, O. A. (1987). *Probability of Loss on Loan Portfolio.* KMV. Retrieved from: http://www.moodysanalytics.com/~/media/Insight/Quantitative-Research/Portfolio-Modeling/87-12-02-Probability-of-Loss-on-Loan-Portfolio.ashx

Vasicek, O. A. (1991). Limiting Loan Loss Probability Distribution. *KMV*. Retrieved from: http://www.moodysanalytics.com/~/media/Insight/Quantitative-Research/Portfolio-Modeling/91-08-09-Limiting-Loan-Loss-Probability-Distribution.ashx

Vasicek, O. A. (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): .

4. Dynamic Model of Losses of a Creditor with a Large Mortgage Portfolio

4.1 Introduction

One of the sources of the recent financial crisis was the collapse of the mortgage business. Even if there are ongoing disputes about the causes of the collapse, wrong risk management seems to be one of them. Hence, realistic models of the lending institutions' risk are of great importance.

The textbook approach to the risk control of the loans' portfolio, which is also a part of the IRB standard (Bank for International Settlement, 2006), is that of Vasicek (Vasicek, The Distribution of Loan Portfolio Value, 2002) who deduces the rates of defaults of the borrowers, and consequently the losses of the banks, from the value of the borrowers' assets following a geometric Brownian motion.

In particular, the Vasicek's model assumes that the logarithm of the assets of the i-th individual fulfills

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

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

$$X_i = Y + Z_i,$$

where Y is the common factor having a centered normal distribution and $Z_1, Z_2, \dots$ are i.i.d. centered normal individual factors, independent of Y (Vasicek, Probability of Loss on Loan Portfolio, 1987).

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

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

After some calculations (cf. (Vasicek, Probability of Loss on Loan Portfolio, 1987)) we obtain the default rate (DR), defined as

$$DR = \frac{number\ of\ defaults}{number\ of\ loans},$$

approximately fulfilling

$$P[DR \le x] \doteq N\left(\frac{\left(\sqrt{1-\rho}\right) \cdot N^{-1}(x) - N^{-1}(PD_1)}{\sqrt{\rho}}\right)$$

given a sufficiently large number of loans. Here, N denotes the standard normal cumulative distribution function and

$$\rho = \operatorname{corr}\left(X_i, X_j\right) = \frac{\operatorname{var}(Y)}{\operatorname{var}(Y) + \operatorname{var}(Z_1)}.$$

It follows that the distribution of DR is "heavy-tailed,"[22] with the "heaviness" of the tail dependent on the correlation ρ.

We generalize the Vasicek's model in three ways:

1. We add dynamics to the model (note that the Vasicek's model is only one-period one).

2. We allow more general distribution of the assets. In a nutshell, the main advantage of our model is that asset increments can be described by any continuous distribution, which potentially enables us to use a distribution that is able to fit a particular dataset better than the normal one.

[22] This means that it cannot be successfully approximated by a light-tailed variable.

3. We add a sub-model of the losses given default which allows us to calculate the overall percentage loss of the bank.

Similarly as in the Vasicek's paper, in our model, there is a one-to-one correspondence between the common factors and the default rate (DR), and the loss given default (LGD), which allows for econometric estimation of the bivariate series of DR's and LGD's. Thus, these factors can have a general distribution of any kind.

To our knowledge, no dynamic generalization of the Vasicek's model incorporating the losses given default has been published yet. However, our approach to the dynamics and/or common modelling of DRs and LGDs is not the only one:

- There are more ways to get the relevant information from the past history of the system, e.g. credit scoring from which the distribution of the DR may be obtained in a standard way (Vasicek, The Distribution of Loan Portfolio Value, 2002) where the distribution of the losses is a function of the probability of default) or observing the credit derivatives (d'Ecclesia, 2008). Another approach to the dynamics could be to track the situation of individual clients (Gupton, Finger, & Bhatia, 1997) or to use affine processes (Duffie, 2005). The usefulness of our approach, however, could lie in the fact that it is applicable "from outside" in the sense that it does not require a bank's internal information.

- Numerous approaches to the joint modeling of DR and the LGD have been published (see e.g. (Witzany J. , 2010), (Yang & Tkachenko, 2012), (Frye, 2000) or (Pykhtin, 2003) and the references therein.) The novelty of our approach, however, is the fact that the form of the dependence of the LGD on the common factor driving the LGD, is not chosen ad-hoc, but it arises naturally from the matter of fact. In particular, it links the LGD to the price of the property serving as a collateral. (Gapko & Šmíd, 2012)

- In its general form, our approach does not assume particular dynamics of the common factors econometric model of which can thus be “plugged” into the model. In contrary to (Gapko & Šmíd, 2012) - a simpler version of our model - multiple generations of debtors are tracked in the presented paper.

Our results show that applying our multi-generational model to a specific dataset leads to a much lower variance in the forecasted credit losses than in the case of the single-generation model. Mainly thanks to the fact that our econometric model uses macroeconomic variables to explain common factors, which is supported by several recent articles, eg (Carling, Jacobson, Lindé, & Roszbach, 2007). It is able to explain changes in risk factors more accurately than a simple model based purely on extraction of common factors from the series of DRs and LGDs. The higher accuracy of the loss forecast then naturally leads to more realistic determination of a quantile loss. In our particular case, the 99.9th quantile loss is lower than in the Vasicek's model.

The paper is organized as follows: after the general definitions (Section 4.2,) where the models of DRs and LGDs are constructed the procedure of econometric estimation of the model is proposed (Section 4.3.) Section 4.4 describes the empirical estimation and finally in Section 4.5, the paper is concluded.

4.2 The Model

In the present section, we introduce our model and discuss its estimation. Proofs and some technical details may be found in the Appendix.

4.2.1 Definition

Let there be (countably) infinitely many potential borrowers. At the time $S_i \in \mathbb{N}_0$, the i-th borrower takes out a mortgage of amount C^i, with help of which, he buys a real property with price $P^i_{S^i} = dC^i$ for some nonrandom $d > 0$. The mortgage is repaid by instalments amounting to $bC^i, b > 0$, at each of the times $S^i + 1, S^i + 2, \dots, S^i + r$, where $r \in \mathbb{N}$ - the duration of the mortgage - is the same for all the borrowers for simplicity.

The assets of the *i*-th borrower evolve according to stochastic process A_t^i such that, between the times the installments are paid, A follows a Geometrical Brownian Motion with stochastic trend, i.e.

$$A_{t-}^i = A_{t-1}^i \exp\{\Delta Y_t + \Delta Z_t^i\}, \quad t \in \mathbb{N}, \quad t > S^i,$$

where Y_t is a common factor (e.g. a log stock index) and Z_t^i, $\mathbb{E}\Delta Z_t^i = 0$, is a normally distributed individual factor for each $i < n$ with the same variance for each i (Δ stands for a one-period difference).

The instalments are paid by means of selling the necessary amount of the assets, i.e.

$$A_t^i = A_{t-}^i - bC^i, \quad t \in \mathbb{N}, \quad t > S^i.$$

If $A_t^i < 0$ then we say that the borrower defaults at t.

The price P_t^i of the real property serving as a collateral of the mortgage of the *i*-th debtor fulfils

$$P_t^i = \exp\{\Delta I_t + \Delta E_t^i\} P_{t-1}^i, \quad t > S^i,$$

(recall that $P_{S^i}^i = dC^i$), where I_t is another common factor (e.g. the logarithm of a real estate price index) and $\Delta E_t^i = \mathcal{N}(0, \sigma^2)$ is an individual factor.[23]

The exposure at default H_t^i (i.e. the remaining debt) of the *i*-th borrower at time *t* fulfils

$$H_t^i = p(t - S^i)C^i, \quad t > S^i$$

for some decreasing function fulfilling $p(1) = 1, p(\tau) = 0$ if $\tau \leq 0$ or $\tau > r$ (the shape of p may depend on the way of interest calculation and the accounting rules of the bank).

Finally, let

$$\pi_1, \pi_2, \ldots,$$

[23] It would not be difficult to have ΔZ_1^1 and ΔE_1^1 non-normal for the price of loosing closed form formula for function h (see further).

be the ratios of "newcomers" to the size of the overall portfolio at the times 1, 2, ….

Assume that the increments of the individual factors

$$\Delta X_1^1, \Delta E_1^1, \Delta X_1^2, \Delta E_1^2, \dots$$

$$\Delta X_2^1, \Delta E_2^1, \Delta X_2^2, \Delta E_2^2, \dots$$

$$\dots$$

are mutually independent and independent of $(Y_t, I_t, \pi_t)_{t\in\mathbb{N}}$ and that, for any *i*, the initial wealth and the size of each mortgage depend, out of all the remaining random variables, only on ω_{S^i}, where

$$\omega_t = (Y_1, I_1, \pi_1, Y_2, I_2, \pi_2, \dots, Y_t, I_t, \pi_t)$$

is the history of the common factors and the percentages of the newcomers up to the start of the mortgage (see (C) in Appendix [sec:Appendix] for details).

Until the end of the Section 4.2, fix $t \in \mathbb{N}$ and assume that the potential borrowers are numbered so that only those who are active since $t-1$ to t (i.e. those with $t-r \le S^i \le t-1$) and who did not default until $t-1$ are numbered.

4.2.2 Default rate

Introduce a zero-one variable Q_t^i indicating whether the *i*-th borrower defaults at *t*:

$$Q_t^i = \mathbf{1}[A_t^i < 0] = \mathbf{1}[A_{t-}^i < bC^i] = \mathbf{1}[a_{t-}^i < b] = \mathbf{1}[\log a_{t-}^i + \Delta Y_t + \Delta Z_t^i < \log b], \qquad (4.1)$$

where

$$a_t^i = \frac{A_t^i}{C^i}$$

is the value of assets per unit of the mortgage. The first topic of our interest will be the percentage of defaults (i.e., the percentage of the debtors who defaulted at *t*):

$$Q_t \coloneqq \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} Q_t^i.$$

It is clear from (4.1) that we may assume, without loss of generality, that $\log b = 0$ (if not than we may subtract $\log b$ from the increments of the common factor). Moreover, we may assume that the variance of ΔZ_t is unit (if not then we could divide $\log a_{t-1}^i$ and ΔY_t^i by its standard deviation).

Thanks to Lemma 8 (see Appendix A.1), we may, similarly to (Vasicek, 2002), apply the Law of Large Numbers to the conditional distribution of Q^i given ω_t to get

$$Q_t = \mathbb{E}(Q_t^1|\omega_t) = \mathbb{P}(Q_t^1 = 1|\omega_t)$$

and compute it, using the Complete Probability Theorem, by formula

$$\mathbb{P}(Q_t^1 = 1|\omega_t) = \sum_{s=t-r}^{t-1} \mathbb{P}(S^1 = s|\omega_t)\mathbb{P}(Q_t^1 = 1|S^1 = s, \omega_t)$$

From the definitions, and thanks to $A(t)$ (see Appendix A.1),

$$\mathbb{P}(Q_t^1 = 1|S^1, \omega_t) = \mathbb{P}\big(\log a_{t-1}^i + \Delta Y_t + \Delta Z_t^1 < 0\big|S^1, \omega_t\big) = \Psi_t^s(-\Delta Y_t|S^1, \omega_{t-1})$$

where $\Psi_t^s(\cdot\,|s,\omega)$ is the c.d.f. of $\log a_{t-1}^i + \Delta Z_t^i$ given $\omega_{t-1} = \omega$, $S^1 = s$, and because $\mathbb{P}(S^1 = s|\omega_t) = \mathbb{P}(S^1 = s|\omega_{t-1})$ by Lemma 7, we are getting:

Proposition 4.1

$$Q_t = \sum_{s=t-r}^{t-1} q_{t-1,s}(\omega_{t-1})\Psi_t^s(-\Delta Y_t|s, \omega_{t-1}), \qquad (4.2)$$

where

$$q_{t-1,s}(\omega_{t-1}) = \mathbb{P}(S^1 = s|\omega_{t-1})$$

♣

Note, that, by Lemma 6 (see Appendix A.1), $\Psi_t^s(\cdot\,|S^1, \omega_{t-1})$ is a strictly increasing c.d.f. of a convolution of two distributions, namely that of $\log a_{t-1}^1$ and the standard normal one. Note also that $q_{t-1,s}$ is in fact the percentage of debts, started at s, and present in the portfolio between times $t-1$ and t.

Corollary 4.2

For each ω_{t-1}, there exists a one to one mapping between Y_t and Q_t given by (4.2). In particular,

$$\Delta Y_t = -\Psi_t^{-1}(Q_t|\omega_{t-1}),$$

$$\Psi_t(y,\omega) = \sum_{s=t-r}^{t-1} q_{t-1,s}(\omega)\Psi_t^s(y|s,\omega). \qquad (4.3)$$

♣

4.2.3 Loss given default

Since the amount which the bank will recover in case of the default of the *i*-th debtor at time t is

$$G_t^i = \min(P_t^i, H_t^i)$$
$$= C^i \min\left(d \cdot \exp\left\{ \sum_{j=S^i+1}^{t} \left[\Delta I_j + \Delta E_j^i\right] \right\}, p\left(t - S^i\right) \right)$$
$$= C^i \exp\left\{ \min\left(d + \sum_{j=S^i+1}^{t} \left[\Delta I_j + \Delta E_j^i\right] \right), \log\left(p\left(t - S^i\right)\right) \right\}$$

we get that the percentage loss given default L_t, i.e. the ratio of the actual losses and the total exposure at default, is

$$L_t = \lim_{n\to\infty} \frac{\sum_{i=1}^n Q_t^i (H^i - G^i)}{\sum_{i=1}^n Q_t^i H^i} = 1 - \lim_{n\to\infty} \frac{\sum_{i=1}^n Q_t^i G^i}{\sum_{i=1}^n Q_t^i H^i}$$

Proposition 4.3

$$L_t = 1 - \frac{\sum_{s=t-r}^{t-1} \nu_{t,s,\omega_{t-1}} h_{t-s}\left(\Delta I_{s,t}\right)}{\sum_{s=t-r}^{t-1} \nu_{t,s,\omega_{t-1}}} \qquad \Delta I_{s,t} = I_t - I_s \qquad (4.4)$$

where

$$\nu_{t,s,\omega} = p(t-s)c_{s,\omega}\Psi_t^s(-\Delta Y_t|s,\omega)q_{t-1,s}(\omega), \qquad c_{s,\omega} = \mathbb{E}(C^1|S^1 = s,\ \omega_{t-1} = \omega)$$

and

$$h_\tau(\iota) = d\exp\left\{\frac{1}{2}\tau\sigma^2 + \iota\right\}\varphi\left(\frac{\omega_\tau - \iota}{\sqrt{\tau}\sigma} - \sqrt{\tau}\sigma\right) + p(\tau)\left[1 - \varphi\left(\frac{\omega_\tau - \iota}{\sqrt{\tau}\sigma}\right)\right]$$

$$\omega_\tau = \log\big(p(\tau)\big) - \log d,$$

and where φ is the standard normal distribution function. The function h_τ is strictly increasing.

Proof. See appendix A.2

♣

Corollary 4.4

For given ω_{t-1} there is one-to-one mapping between L_t and I_t, given by (4.4). *In particular,*

$$I_t = \Upsilon^{-1}_{t,\omega_{t-1}}(1 - L_t), \qquad (4.5)$$

where

$$\Upsilon_{t,\omega}(\iota) = \frac{1}{\sum_s \upsilon_{t,s,\omega}} \sum_{s=t-r}^{t-1} \nu_{t,s,\omega_t} h_{t-s}\big(\Delta I_{t-1,s} + \iota\big)$$

♣

4.2.4 Next period

Now, let us proceed to the portfolio at the next period: After renumbering (excluding the defaulted borrowers and adding the newcomers) we get.

Proposition 4.5

$$q_{t,s}(\omega_t) = \begin{cases} \pi_t & if\ s = t \\ (1-\pi_t)u_{t,s}(\omega_{t-1}) & if\ t-r < s < t \\ 0 & otherwise \end{cases}$$

where

$$u_{t,s}(\omega) = \mathbb{P}(S^1 = s | Q^1 = 0, \omega_{t-1} = \omega)$$
$$= q_{t-1,s}(\omega) \frac{1 - \Psi_t^s(-\Delta Y_t | s, \omega)}{1 - \Psi_t(-\Delta Y_t|\omega) - \big(1 - \Psi_t^s(-\Delta Y_t | t-r, \omega)\big) q_{t,t-r}(\omega)}$$

and

$$\mathbb{P}[\log a_t^1 \le z | S^1 = s, \omega_{t-1} = \omega] = \begin{cases} \vartheta_{t,\omega_{t-1}}(z) & if\ S^1 = t \\ \dfrac{\Psi_t^s(z - \Delta Y_t | s, \omega) - \Psi_t^s(-\Delta Y_t | s, \omega)}{1 - \Psi_t^s(-\Delta Y_t | s, \omega)} & otherwise \end{cases}$$

for each $z \ge 0$ *where* $\vartheta_{t,\omega}(z) = \mathbb{P}\big[log\ U^i \le z |\ \omega_{t-1} = \omega\big], U^i = A_{S_1}$

Proof. See appendix A.3

♣

4.2.5 Econometrics of the Model

Say we have the sample

$$\pi_1, Q_1, L_1, \pi_2, Q_2, L_2, \dots, \pi_T, Q_T, L_T \qquad (4.6)$$

at our disposal and want to infer (some of) the parameters of our model, whose complete list is

$$\mathbb{P}\big((X, Y, \pi) \in \cdot\big), c(\cdot), r, d, p(\cdot), \vartheta.(\cdot), \sigma \qquad (4.7)$$

Clearly, some further simplification of such a rich parameter space has to be done. For simplicity and computability, we decided to postulate values of all the parameters except of $\mathbb{P}\big((X, Y, \pi) \in \cdot\big)$ in the empirical part of our paper so that we are able (recursively) to evaluate the transforming function Ψ_t and Υ_t independently on unknown parameters and the econometrics of the model reduces to the one of the factors Y and I. In other words, the values of all parameters except of $\mathbb{P}\big((X, Y, \pi) \in \cdot\big)$ were chosen based on empirical observations or expert judgment.

4.2.6 Numerics of the Model

Generally, Ψ_t is a convolution of truncated (normal) distributions (the defaults are due to the truncations). We chose the Monte Carlo simulation as the easiest way of the functions evaluation which was done in the Mathematica software.

Since the formula for Ψ_t is recursive and involves $\Psi_{t-1}, \dots, \Psi_{t-r}$, which are unknown at the time t, we acted as if the borrowing began at $t = 1$, i.e. we took $q_{1,1} = 1$ and $q_{1,s} = 0$ for all $s < 1$.

4.3 Empirical estimation

In this part, we describe the estimation procedure of the previously introduced model. The final result of the estimation procedure is a loss distribution and, in particular, a mean predicted loss and a predicted loss quantile on a one-quarter horizon.

The estimation process can be divided into three separate parts: the extraction of both common factors from a historical dataset, a prediction of these factors based on an econometric model and finally, the calculation of future mean and quantile losses given the future values of the factors.

4.3.1 Data description

We used the same dataset as in (Gapko & Šmíd, 2012), ie, a historical dataset of mortgage delinquencies and started foreclosures, provided by the Mortgage Bankers Association. In our model we took the 90+ delinquency rate at the time t as the default rate, Q_t. Unfortunately, to our knowledge, there is no nationwide public database with banks' losses from mortgage portfolios that could be considered as our loss given default, L_t. Therefore we constructed its proxy by the rate of started foreclosures from the Mortgage Bankers Association and an index of median prices of new homes sold from the US Census Bureau. In particular, because the foreclosures dataset consists of all mortgage loans that fell into the foreclosure process and does not describe how successful the foreclosure process was, we discounted the foreclosures by estimated average values of the collaterals in the portfolio; even if, as we realized, our proxy of the LGD is apparently an ad hoc one, it reflects the fact that the LGD grows with decreasing prices of collaterals.

Formally, we put

$$Q_t = D_t,$$

where D_t is the 90+ delinquency rate at the time t and

$$L_t = \frac{F_t}{D_t J_t},$$

where F_t is the unadjusted rate of started foreclosures from the original dataset and J_t an estimated average value of collaterals in the portfolio calculated as

$$J_t = \sum_{s=t-r}^{t-1} \frac{N_{i,s}}{N_{t-r}} \cdot \frac{\Pi_t}{\Pi_s} \doteq \sum_{s=t-r}^{t-1} q_{t,s}(\omega_{t-1}) \cdot \frac{\Pi_t}{\Pi_s},$$

where $N_{i,s}$ is the number of individuals in the s-th generation at the time t, $\frac{N_{i,t}}{N_t}$ the proportion of individuals of the i-th generation in the whole portfolio at the time t, Π_s the value of the house price index at the time s (recall that we assume unit price of all the collaterals at the start of the mortgage and that $q_{t,s}$ is a function of the observed data).

Both datasets entering our calculations are depicted on the following chart (in percentage of the total outstanding balance).

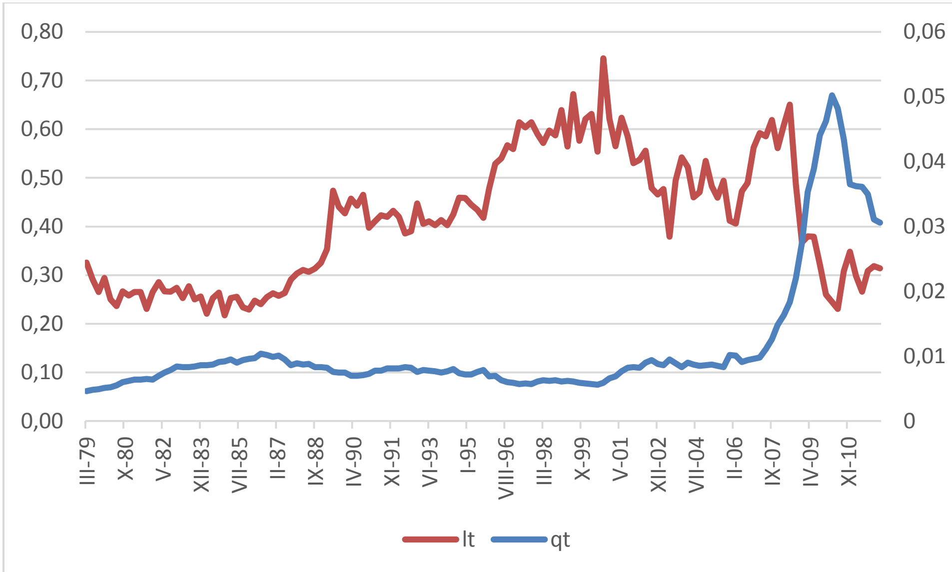

Figure 4.1: 90+ delinquency rate Q_t and the loss given default L_t

4.3.2 Choice of Parameters

In order to extract the rate of default and the loss given default, which is the first step in the estimation, we needed to restrict the number of parameters in the extracting functions given by (4.3) and (4.5). The parameters

$$c(\cdot), r, d, p(\cdot), \vartheta.(\cdot), \sigma$$

were further postulated as follows:

- The length of the mortgage, r was set to 120 quarters (30 years) based on the long-term average taken from the U.S. Housing Market Conditions survey published quarterly by the U.S. Department of Housing and Urban Development

- The variance of E (the individual factor driving the property price), ie, σ of the distribution with the c.d.f. equal to Φ was set at 0.12 because this value was found to be the one maximizing the log-likelihood in the single-generation model (Gapko & Šmíd, 2012)

- The size of the loan-to-value ratio d at the beginning of the loan is set to 1 (ie, the full mortgage nominal is collateralized by the borrower's property); this is a simplification and a possible point for the model enhancement.

- The quarterly interest rate, which determines the function p, is set to 1%; the function p uses the quarterly simple compounding interest to determine what amount of a mortgage remains to repay

- The standard deviation of each newcoming generation's wealth U_i is assumed to be normal with standard deviation equal to 5

- The parameter c - ie, the expected size of the mortgage, is assumed to be the same for all borrowers

Other parameters, eg, the split on individual generations in a given period, can be calculated directly or derived from our assumptions. For a better understanding of how the original datasets Q and L are translated into the common factors Y and I, resp., we include a comparison of Q and Y (*Figure 4.2*) and L and I (*Figure 4.3*). In the *Figures 4.2* and *4.3*, the values of the time series I and Q were adjusted to overlap the corresponding time series Y and L, resp. (i.e. Q multiplied by 100 and I multiplied by 10, so that the lines benefit from a single scale representation).

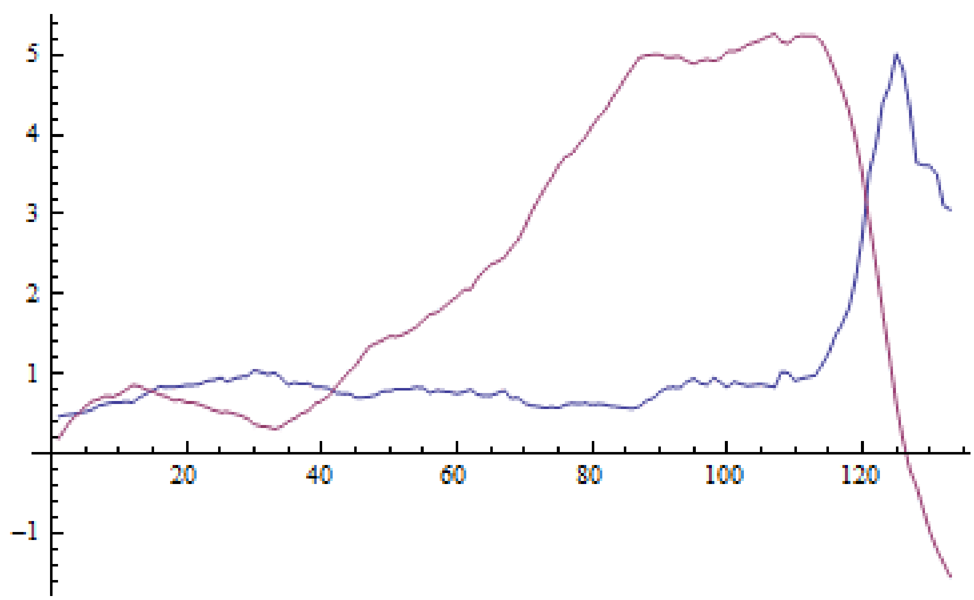

Figure 4.2: The comparison of Q (blue) and Y (violet)

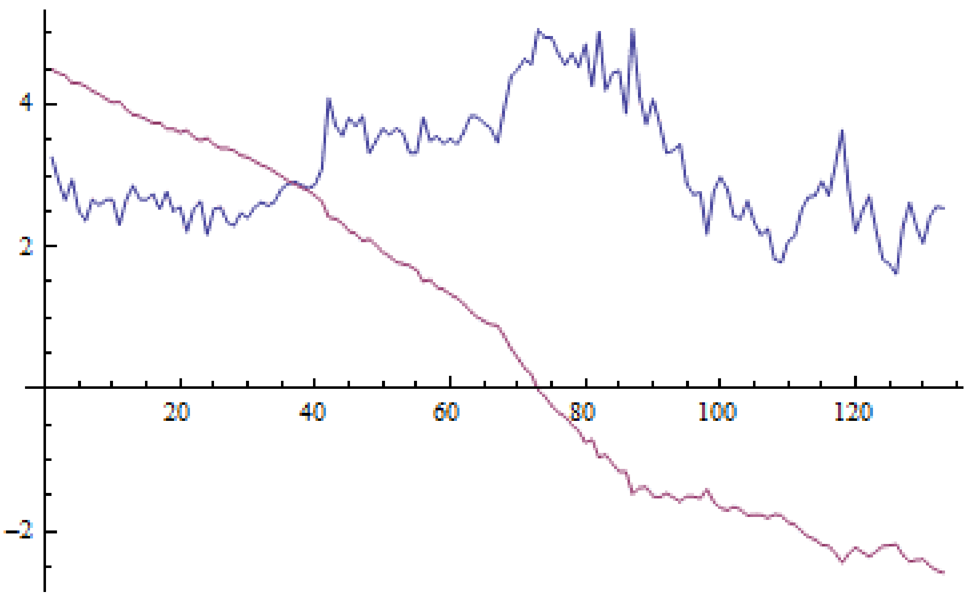

Figure 4.3: The comparison of *L* (blue) and *I* (violet)

From the beginning of the dataset, there was a sustained growth of house prices, which caused the collateral to exceed the mortgage outstanding amount and thus decreased the LGD. However, in 2007, there was a downturn in housing prices and this is reflected in the increase of the LGD. From the *Figures 4.2* and *4.3* we can graphically deduce that the evolution of both common factors might follow some trends, which suggests that there could be a dependence on several macroeconomic variables or stock market indexes. Thus, we chose a Vector Error Correction Model (VECM) with several exogenous macroeconomic variables, namely GDP, unemployment, interest rates, inflation, S&P 500 stock market index and the EUR/USD exchange rate, to capture the joint dynamics of the common factors Y and I. Note that we couldn't use any kind of real estate price index as the LGD values were adjusted by using such an index. Adding it would establish an unsought autocorrelation into the VECM error term.

4.3.3 Estimation and prediction

The VECM estimation was performed in the Gretl software. First, the stationarity tests of both VECM endogenous variables, ie Y and I, was performed and in both cases, the augmented Dickey-Fuller test rejected the stationarity. The Johansen's cointegration test rejects the absence of the first order cointegration between Y and I on the 10% probability level (see the Appendix

A.6 for detailed results of the Johansen's cointegration test and the corresponding cointegrating vectors). Moreover, the first VECM equation, explaining Y, shows that it strongly depends on the year-on-year GDP growth rate. No other macroeconomic variables considered were found significant in this equation, even after lagging them up to four quarters. The second VECM equation, explaining I, also shows dependency on one macroeconomic variable - unemployment rate. Therefore we left the two significant variables, ie, the GDP year-on-year growth rate and the unemployment rate in the model. The following table summarizes our findings. It is obvious that the model is able to explain Y with a much higher predictive power than I, which is probably caused by the fact that changes of I are based on a proxy instead of the actual LGD.

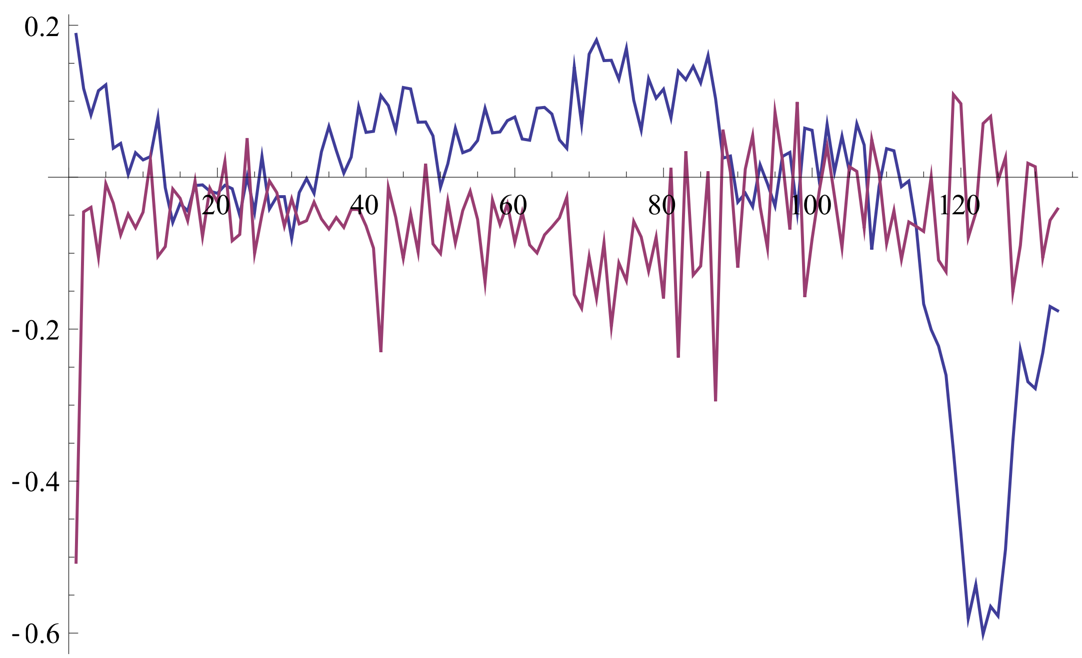

Figure 4.4: Returns of *Y* (blue) and *I* (violet)

Dependent variable	*Y (s.e.)*	*I (s.e.)*
constant	-0.0098 (0.03)	-0.14*** (0.04)
d1 PD common factor	0.96*** (0.04)	-0.17*** (0.05)
d1 LGD common factor	0.13* (0.07)	-0.24*** (0.09)
GDP year-on-year	0.72*** (0.23)	0.027 (0.3)
Unemployment rate	-0.05 (0.39)	1.07** (0.5)
Error correction term	-0.0067 (0.004)	0.016*** (0.006)
Adjusted R2	91%	15%

Table 4.1: results of the PD & LGD common factors VECM estimate

Thus the final pair of VECM equations is:

$$Y_t = -0.0098 + 0.96 \cdot dY_t + 0.13 \cdot dI_t + 0.72 \cdot GDP_t - 0.05 \cdot Unemployment_t - 0.0067 \cdot EC_t$$

$$I_t = -0.14 - 0.17 \cdot dY_t - 0.24 \cdot dI_t + 0.027 \cdot GDP_t + 1.07 \cdot Unemployment_t + 0.016 \cdot EC_t$$

We also performed tests of both normality and autocorrelation of residuals. All tests show that error terms of both equations are not autocorrelated and approximately normal.

After the model is estimated, we constructed a prediction of the common factors. To calculate the predicted Y and I, we needed a prediction of exogenous variables in the model, ie, the GDP y/y growth rate and the unemployment rate. As we measured the credit risk only, without an influence of deterioration in economic conditions, we assumed that the unemployment rate stayed for the prediction on its last value and the future GDP change is zero. The following two charts show the development of Y (*Figure 4.5*) and I (*Figure 4.6*), including the predicted value.

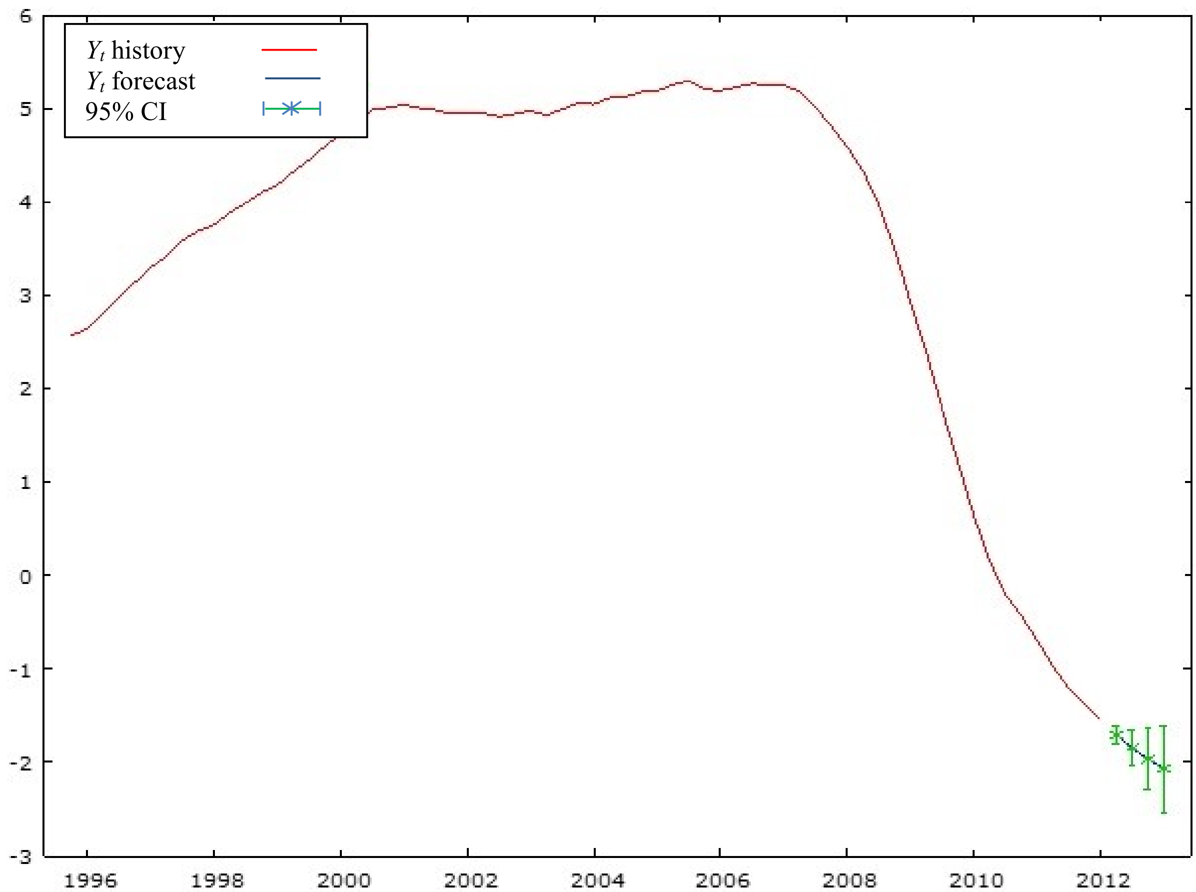

Figure 4.5: Development of *Y* with the predicted value (blue) and the prediction standard error (green)

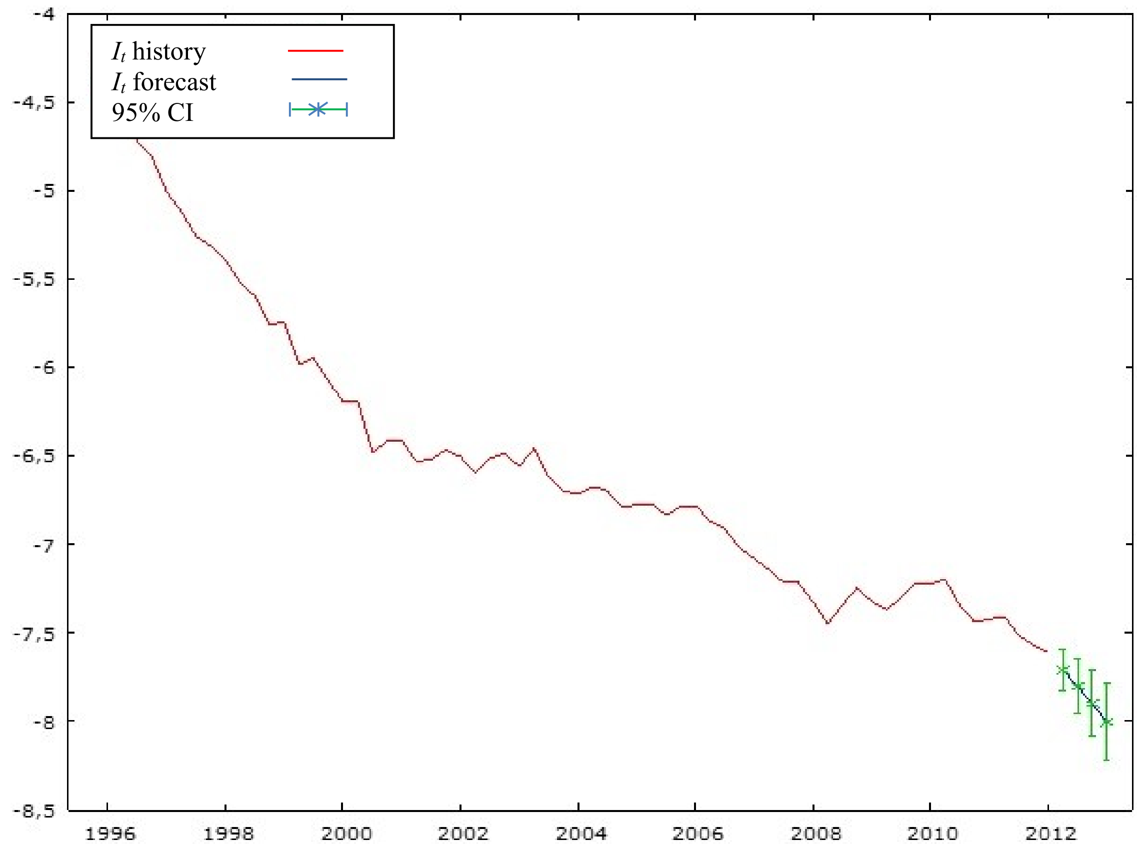

Figure 4.6: Development of I with the predicted value (blue) and the prediction standard error (green)

4.3.4 Prediction of losses

The remaining step was to predict a mean and a desired quantile losses. This was done by an inversion function to the factor extraction functions (see (4.3) and (4.5)) in the Mathematica software, by which we obtained predicted DR and LGD. These two values were then multiplied to get a loss. The mean loss prediction is quite straightforward as we already have the predicted values of both common factors. However, the quantile loss has to be calculated from the quantile value of both common factors. To be able to compare our quantile loss with the IRB model, we chose to simply calculated the 99.9th quantiles of Q and the 99.9th quantile of L and then multiply them[24]. The calculation of quantiles of Q and L from the quantiles of Y and I was done by the

[24] The 99.9th was chosen to reflect the IRB, which calculates the capital requirement for credit risk as a difference between the mean (expected) loss and the 99.9th quantile loss. Usually, the 99.9th quantile loss is interpreted as a multiplication of the 99.9th quantile of Q and a "downturn" LGD (usually calculated as a 95th quantile of L).

function (4.2) for Q and by (4.4) for L. Quantiles of common factors were obtained from their prediction standard error and the assumption that error terms of both VECM equations (see *Table 4.1*) are normally distributed. (Recall that we were not able to reject the normality). Thus,

$$Y_{q(0.999)} = Y_{t+i} + \sigma_Y \cdot N(0.999) \quad and$$
$$I_{q(0.999)} = I_{t+i} + \sigma_I \cdot N(0.95),$$

where $Y_{q(0.999)}$ and $I_{q(0.999)}$ are 99.9th quantiles of the factors Y and I, resp., Y_{t+i} and I_{t+i} are the common factors predictions, σ_Y and σ_I the regression standard errors and $N(0.999)$ and $N(0.95)$ the 99.9th and the 95th quantile of the standard normal distribution, resp. We constructed a one-quarter quantile loss prediction.

Because the Basel II IRB method calculates a twelve month forward quantile loss, to get a one quarter loss we divided the PD input (last DR value) by two (because the debtor's assets are assumed in the IRB model to be normally distributed, the quarterly PD is exactly one half of the one-year PD, according to the convolution of the normal distribution). We used just one quarter for all the predictions. Both the comparison of the predictions of mean losses calculated by our proposed model and the IRB, and the comparison of the predictions of quantile loss are summarized in the *Table 4.2*.

Model	***Our***	***IRB***
mean loss	0.84%	0.78%
99.9th quantile loss	1.23%	3.75%

Table 4.2: comparison of our model's and IRB losses

For the IRB model we have used the last value of default rate as an input for the PD and the last value of our adjusted LGD time series for an LGD. The difference between the IRB and our model computations is that the IRB treats LGD as a fixed variable, whereas in our proposed approach, we constructed a model for LGD predictions. As we can see from *Table 4.2*, our model predicts much lower quantile loss. This is due to the fact that the explanation of the development of default rates and LGD by our model is much neater than a crude ad-hoc approach of the IRB and thus the standard deviation of loss is lower.

4.4 Conclusion

In the present paper, we suggested an estimable model of credit losses. The model is based on the assumption of underlying factors that are driving the probability of default and the loss given default. The two novelties of our approach are the multigenerational dimension of the model and the estimated relationship between underlying factors and a macroeconomic environment.

The empirical estimation shows that the model leads to more accurate predictions of future mean and quantile losses than in the Vasicek's framework. This might lead to a saving in the amount of capital that is needed to cover the quantile loss.

Even if the model is rather general and thus a bit more complicated to estimate due to the number of parameters, a bit less could be assumed if a user wished it, especially

- The distribution of the individual factors need not be the same in all periods but it might depend on the time and on the past of the common factor
- A dependence of the individual factors ΔE_t^i and ΔZ_t^i could be established

While the first generalization would not change our formulas much (some indexes would have to be added to the present notation) the second one would bring the necessity to work with a conditional distribution of ΔE given not defaulting, for which no analytical formula exists, even in the simple case of normal factors.

Appendix

A.1 Definitions and Auxiliary Results

First, we have to take into account that the borrowers have to be renumbered in each period in order to remove those who defaulted or fully repaid their mortgage and add those who came newly. Let us assume that the renumbering at t is done as follows: once the indexes $1, 2, \ldots, i-1$ are assigned, a random variable D_t^i is drawn from the Bernoulli distribution with parameter π_t. The index i is consequently given to a newcomer, if $D_t^i = 1$ or to the first unindexed borrower who did not default at t and does not repay fully his mortgage at t, if $D_t^j = 0$. Let us denote S_t^i the starting time of the debtor, indexed by i at t.

Now, denote,

$$\Omega_0 = (\dot{a}_0^1, S_0^1, \dot{a}_0^2, S_0^2, \ldots)$$

and

$$\Omega_t = (Y_\tau, I_\tau, Z_\tau^1, E_\tau^1, Z_\tau^2, E_\tau^2, \ldots)_{\tau \leq t}$$

for $t > 0$ and note that, as the distribution of D_t^i depends only on π_t, which itself is a part of the vector ω_t, we have that D_t^i is conditionally independent of $\Omega_t, D_t^1, D_t^2, \ldots, D_t^{i-1}, D_t^{i+1}, \ldots$ given ω_t.

Further, we have to formulate rigorously the assumptions concerning the distribution of the initial wealth and the property price. In particular, we assume that, for each i, $\left(C^i, A_{S_i}^i\right) = \left(C_{i,S_i}, U_{i,S_i}\right)$, where

C

for any i and t, $\left(C_{i,t}, U_{i,t}\right)$ is conditionally independent of $\Omega_t, \left(C_{j,t}, U_{j,t}\right)_{j \neq i}$ given ω_t, and the conditional distribution of $\left(C_{i,t}, U_{i,t}\right)$ given ω_t equals for all i.

Finally, denote $\omega_\infty = (\Delta Y_\tau, \Delta I_\tau)_{\tau \in \mathbb{N}}$ and assume that

$\boldsymbol{A(0)}$

variables $\dot{a}_0^1, S_0^1, \dot{a}_0^2, S_0^2, \ldots$ are mutually independent and independent of ω_t, Ω_t for any $t > 0$, such that a_0^i has the same strictly increasing continuous conditional c.d.f. given S_i for each i.

Now, let us prove that

Lemma 6

For each $t > 0$ the following is true:

$\boldsymbol{A(t)}$

For any i, $\dot{a}_{t-1}^i, \ldots$ is conditionally independent of $\omega_\infty, \left(S_{t-1}^j, a_{t-1}^j\right)_{j\neq i}$ given S_{t-1}^i, ω_{t-1}, such that a_{t-1}^i has the same strictly increasing continuous conditional c.d.f. for each i.

Proof. Let us proceed by induction: For $t = 0$, the assertion follows from $\boldsymbol{A(0)}$. Now, assume $\boldsymbol{A(t)}$ and try to prove $\boldsymbol{A(t+1)}$. Let $i \in \mathbb{N}$. From the basic properties of conditional expectations, we have

$$\mathbb{P}\left(a_t^i < x \middle| D_t^i, \omega_\infty, \left(S_t^k\right)_{k\in\mathbb{N}}, \left(a_t^k\right)_{k\neq i}\right)$$
$$= \begin{cases} \mathbb{P}\left(U_t^i < x \middle| D_t^i, \omega_\infty, \left(S_t^k\right)_{k\in\mathbb{N}}, \left(a_t^k\right)_{k\neq i}\right) = \mathbb{P}\left(U_t^i < x \middle| \omega_t\right) & \text{on } \left[D_t^i = 1\right] \\ \mathbb{P}\left(a_{t-}^{J_i} < x \middle| D_t^i, \omega_\infty, \left(S_t^k\right)_{k\in\mathbb{N}}, \left(a_t^k\right)_{k\neq i}\right) = \mathbb{E}\left(\pi(x) \middle| D_t^i, \omega_\infty, \left(S_t^k\right)_{k\in\mathbb{N}}, \left(a_t^k\right)_{k\neq i}\right) & \text{on } \left[D_t^i = 0\right] \end{cases}$$

where

$$\pi(x) = \mathbb{P}\left(a_{t-}^{J_i} < x \middle| \left(D_t^k, S_{t-1}^k, S_t^k\right)_{k\in\mathbb{N}}, \omega_\infty, J^i, \left(a_t^k\right)_{k\neq i}\right)$$

and J^i is the index of the borrower indexed by i at t given the numbering from $t-1$. On the set $[J_i = j]$, we get

$$\pi(x) = \mathbb{P}\left(a_{t-1}^j + \Delta Z_t^j + \Delta Y_t < x \middle| \left(D_t^k, S_{t-1}^k, S_t^k\right)_{k\in\mathbb{N}}, \omega_\infty, J^i, \left(a_t^k\right)_{k\neq i}\right)$$

$$= \mathbb{E}\left(\mathbb{P}\left(a_{t-1}^{j} + \Delta Z_t^j + \Delta Y_t < x \middle| Q_t^j, \left(D_t^k, S_{t-1}^k, \Delta Z_t^k\right)_{k\in\mathbb{N}}, \omega_\infty, J^i, \left(a_{t-1}^k\right)_{k\neq j}\right) \middle| \left(D_t^k, S_{t-1}^k, S_t^k\right)_{k\in\mathbb{N}}, \omega_\infty, \left(a_t^k\right)_{k\neq i}\right)$$

$$= \mathbb{E}\left(\rho(x) \middle| \left(D_t^k, S_{t-1}^k, S_t^k\right)_{k\in\mathbb{N}}, \omega_\infty, J^i, \left(a_t^k\right)_{k\neq i}\right),$$

where

$$\rho(x) = \mathbb{P}\left(a_{t-1}^j + \Delta Z_t^j + \Delta Y_t < x \middle| \mathbf{1}\left[a_{t-1}^j + \Delta Z_t^j + \Delta Y_t < 0\right], S_{t-1}^j, \omega_t\right)$$

(the last "=" is due to $\boldsymbol{A(t)}$) where, by the textbook calculation

$$\rho(x) = \psi(x, S_{t-1}, \omega_{t-1}), \qquad \psi(x, s, \omega) = \frac{\Psi_t^s(x - \Delta Y_t | s, \omega) - \Psi_t^s(-\Delta Y_t | s, \omega)}{1 - \Psi_t^s(-\Delta Y_t | s, \omega)}$$

on the set $M = \left[Q_t^j = 0, S_t^i = S_{t-1}^i\right]$. Now, because $[J_i = j] \subset M$ and $[J_i = j]_{j\in\mathbb{N}}$ cover the set $\left[D_t^i = 0\right]$, we have by Local Property ((Kallenberg, 2002), Lemma 6.2) that

$$\pi\,(x) = \psi\left(x, S_t^i, \omega_{t-1}\right)$$

on $\left[D_t^i = 0\right]$ finally giving

$$\begin{aligned}
&\mathbb{P}\left(a_t^i < x \middle| \omega_\infty, \left(S_t^k\right)_{k\in\mathbb{N}}, \left(a_t^k\right)_{k\neq i}\right) \\
&= \mathbb{E}\left(\mathbb{P}\left(U_t^i < x \middle| \omega_t\right)\mathbf{1}_{[\boldsymbol{D_1=1}]} + \psi\left(x, S_t^i, \omega_{t-1}\right)\mathbf{1}_{[\boldsymbol{D_1=0}]} \middle| \omega_\infty, \left(S_t^k\right)_{k\in\mathbb{N}}, \left(a_t^k\right)_{k\neq i}\right) \\
&= \mathbb{E}\left(\mathbb{P}\left(U_t^1 < x \middle| \omega_t\right)\mathbf{1}_{[\boldsymbol{D_t^1=1}]} + \psi\left(x, S_t^i, \omega_{t-1}\right)\mathbf{1}_{[\boldsymbol{D_t^1=0}]} \middle| \omega_t\right),
\end{aligned} \tag{8}$$

where the last "=" is due to the conditional independence of D_t of Ω_t, hence $\boldsymbol{A(t+1)}$ is proved.

♣

Lemma 7

For any $i \in N$, S_t^i is conditionally independent of ω_∞, $(S_t^j)_{j\neq i}$, given ω_t.

Proof. For $t = 0$ the Lemma follows from $\boldsymbol{A(0)}$. Let $\tau > 0$ and let the Lemma holds for $t = \tau - 1$. ie,

$$\mathbb{P}\left(S_{\tau-1}^{k} \middle| \left(S_{t}^{j}\right)_{j \neq k}, \dot{\omega}_{\infty}\right) = \mathbb{P}\left(S_{\tau-1}^{k} \middle| \dot{\omega}_{\tau-1}\right).$$

By our construction, S_{τ}^{i} is a function of $S_{\tau-1}^{J_i}$ where J_i is defined by the previous proof. Similarly to the previous proof we show that, on $[J_i = j]$ the probability that $S_{\tau}^{i} = s$ given all the variables $a_{\tau-1}^{\cdot}$ depends only on $q_{\tau-1}(\omega_{\tau-1})$ and on π_{τ}.

♣

Lemma 8

$Q_t^1, Q_t^2, \ldots$ *are mutually conditionally independent given* ω_t.

Proof. It follows from Lemma 6 that Q_t^i is conditionally independent of $\left(Q_t^k, S_{t-1}^k\right)_{k \neq i}$ given $\left(\omega_t, S_{t-1}^i\right)$. Thanks to Lemma 7 and independence of variables $\Delta Z_t^{\cdot}$ we get that S_{t-1}^i is conditionally independent of $\left(Q_t^k, S_{t-1}^k\right)_{k \neq i}$ given ω_t which gives the Lemma by the Chain rule for conditional independence ((Kallenberg, 2002), Proposition 6.8).

A.2 Proof of Proposition 3

By (Kallenberg, 2002), Corollary 4.5.

$$L_t = 1 - \frac{\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} Q_t^i G_t^i}{\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} Q_t^i H_t^i}.$$

Further, by Lemma 8 and by the independence of variables E, the summands in both sums are conditionally independent given ω_t, hence, by the Law of large numbers,

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} Q_t^i G^i = \mathbb{E}(Q_t^1 G_t^1 | \omega_t) = \mathbb{E}(\mathbb{E}(Q_t^1 G_t^1 | \omega_t, S_{t-1}^1) | \omega_t)$$

$$= \sum_{s=t-r}^{t-1} \mathbb{E}(Q_t^1 G_t^1 | \omega_t, S_{t-1}^1 = s) q_{t-1,s}(\omega_{t-1})$$

$$= \sum_{s=t-r}^{t-1} \mathbb{E}(G_t^1 | \omega_t, S_{t-1}^1 = s) \mathbb{E}(Q_t^1 | \omega_t, S_{t-1}^1 = s) q_{t-1,s}(\omega_{t-1})$$

$$= \sum_{s=t-r}^{t-1} \mathbb{E}(G_t^1 | \omega_{t-1}, S_t^1 = s) \mathbb{E}(Q_t^1 | \omega_{t-1}, S_t^1 = s) q_{t-1,s}(\omega_{t-1})$$

$$= \sum_{s=t-r}^{t-1} \upsilon_{t,s,\omega_{t-1}} h_{t-s}\left(\Delta I_{s,t}\right)$$

$$h_r(\iota) = d\mathbb{E}(\exp\{min(\iota + e_r, w_r)\}) \qquad e_r \sim \mathcal{N}(0, r\sigma^2)$$

and analogously,

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} Q_t^i H^i = \sum_{s=t-r}^{t-1} \upsilon_{t,s,\omega_{t-1}}$$

As to h, we are getting

$$\begin{aligned} h_r(\iota) &= d\mathbb{E}(\exp\{\iota\} \exp\{\min(e_r, w_r - \iota)\}) \\ &= de^{\iota}\left[\int_{-\infty}^{w_r-\iota} e^x d\Phi_{(r)}(x) + e^{w_r-\iota}\left(1 - \Phi_{(r)}(w_r - \iota)\right)\right] \\ &= de^{\iota}\int_{-\infty}^{w_r-\iota} e^x d\Phi_{(r)}(x) + p(t-s)\left(1 - \Phi_{(r)}(w_r - \iota)\right), \end{aligned}$$

where $\Phi_{(\nu)}$ is a c.d.f. $\mathcal{N}(0, \nu\sigma^2)$ - when we put $\varsigma = \sqrt{r}\sigma$, we get

$$\int_{-\infty}^{w_r-\iota} e^x d\Phi_{(r)}(x) = \int_{-\infty}^{w_r-\iota} \frac{1}{\sqrt{2\pi}\varsigma} e^{-\frac{x^2}{2\varsigma^2}} e^x dx$$

$$= \frac{1}{\sqrt{2\pi}\varsigma} \int_{-\infty}^{w_r-\iota} \exp\left\{-\frac{x^2 - 2\varsigma^2 x + \varsigma^4}{2\varsigma^2} + \frac{1}{2}\varsigma^2\right\} dx$$

$$= \exp\left\{\frac{1}{2}\varsigma^2\right\} \int_{-\infty}^{w_r-\iota} \frac{1}{\sqrt{2\pi}\varsigma} \exp\left\{-\frac{(x-\varsigma^2)^2}{2\varsigma^2}\right\} dx$$

$$= \exp\left\{\frac{1}{2}\varsigma^2\right\}\mathbb{P}[N(\varsigma^2,\varsigma^2) < w_r - \iota] = \exp\left\{\frac{1}{2}\varsigma^2\right\}\varphi\left(\frac{w_r - \iota}{\varsigma} - \varsigma\right)$$

hence

$$h_r(\iota) = d\exp\left\{\frac{1}{2}r\sigma^2 + \iota\right\}\varphi\left(\frac{w_r - \iota}{\sqrt{r}\sigma} - \sqrt{r}\sigma\right) + p(r)\left[1 - \varphi\left(\frac{w_r - \iota}{\sqrt{r}\sigma}\right)\right].$$

The monotonicity is proved by the fact that

$$\begin{aligned}
\frac{\partial}{\partial\iota}h_r(\iota) &= de^{\iota}\left(\Phi_{(r)}(w_r - \iota)e^{(w_r-\iota)} - \int_{-\infty}^{w_r-\iota}\Phi_{(r)}(x)e^x dx\right)\\
&= de^{\iota}\left(\Phi_{(r)}(w_r - \iota)\int_{-\infty}^{w_r-\iota}e^x dx - \int_{-\infty}^{w_r-\iota}\Phi_{(r)}(x)e^x dx\right)\\
&= de^{\iota}\left(\int_{-\infty}^{w_r-\iota}\Phi_{(r)}(w_r - \iota)e^x dx - \int_{-\infty}^{w_r-\iota}\Phi_{(r)}(x)e^x dx\right)\\
&= de^{\iota}\left(\int_{-\infty}^{w_r-\iota}\left[\Phi_{(r)}(w_r - \iota) - \Phi_{(r)}(x)\right]e^x dx\right) > 0.
\end{aligned}$$

A.3 Proof of Proposition 5

The fact that $q_{t,t} = \pi_t$ follows from the definition, as well as the fact that $q_{t,s} = 0$ for $s \le t - r$.

Let $t - r < s < t$, and let J_i be the previous index of the borrower indexed by i at t (it can be eg, a zero if the borrower is a newcomer). Clearly, $S_t^i = s \Leftrightarrow \mathrm{D}_t^i = 0 \wedge \mathrm{S}_{t-1}^{J_i} = \mathrm{s}$ which implies

$$\begin{aligned}
&\mathbb{P}\left(S_t^i = s \middle| \omega_t = \omega\right) = \mathbb{P}\left(D_t^i = 0, S_{t-1}^{J_i} = s \middle| \omega_t = \omega\right)\\
&= (1 - \pi_t)\mathbb{P}\left(S_{t-1}^{J_i} = s \middle| D_t^i = 0, \omega_t = \omega\right)\\
&= (1 - \pi_t)\sum_j \mathbb{P}\left(S_{t-1}^{j} = s \middle| J_i = j, D_t^i = 0, \omega_t = \omega\right)\mathbb{P}\left(J_i = j \middle| D_t^i = 0, \omega_t = \omega\right).
\end{aligned} \tag{9}$$

Further, as

$$J_i = j \Leftrightarrow \sum_{k=1}^{i-1} \left(1 - D_t^k\right) = \sum_{k=1}^{i-1} \mathbf{1}[Q_{t-1}^k = 0, S_{t-1}^k \neq t - r] \wedge Q_t^j = 0 \wedge S_{t-1}^j \neq t - r \wedge D_t^i = 0$$

we have, from the conditional independence

$$\mathbb{P}\left(S_{t-1}^j = s \middle| J_i = j, D_t^i = 0, \omega_t = \omega\right) = \mathbb{P}\left(S_{t-1}^j = s \middle| Q_t^j = 0, S_{t-1}^j \neq t - r, \omega_t = \omega\right)$$

$$= \frac{\mathbb{P}\left(S_{t-1}^j = s, Q_t^j = 0, S_{t-1}^j \neq t - r \middle| \omega_t = \omega\right)}{\mathbb{P}\left(Q_t^j = 0, S_{t-1}^j \neq t - r \middle| \omega_t = \omega\right)}$$

$$= \frac{\mathbb{P}\left(S_{t-1}^j = s, Q_t^j = 0 \middle| \omega_t = \omega\right)}{\mathbb{P}\left(Q_t^j = 0 \middle| \omega_t = \omega\right) - \mathbb{P}\left(Q_t^j = 0, S_{t-1}^j = t - r \middle| \omega_t = \omega\right)}$$

$$= \frac{\mathbb{P}\left(Q_t^j = 0 \middle| S_{t-1}^j = s, \omega_t = \omega\right)\mathbb{P}\left(S_{t-1}^j = s \middle| \omega_t = \omega\right)}{\mathbb{P}\left(Q_t^j = 0 \middle| \omega_t = \omega\right) - \mathbb{P}\left(Q_t^j = 0 \middle| S_{t-1}^j = t - r, \omega_t = \omega\right)\mathbb{P}\left(S_{t-1}^j = t - r \middle| \omega_t = \omega\right)}$$

$$= \frac{\left(1 - \Psi_t^s(-\Delta Y_t | s, \omega_{t-1})\right) q_{t-1,s}(\omega_{t-1})}{1 - \Psi_t(-\Delta Y_t | \omega_{t-1}) - \left(1 - \Psi_t^s(-\Delta Y_t | t - r, \omega_{t-1})\right) q_{t-1,t-r}(\omega_{t-1})},$$

which, not being dependent on j, may be pulled out from the sum in (9).

The formula for a_t is proved similarly to (8).

A.4 Calculated I and Y factors values

I_t	Y_t
-0.5085704994306764	0.1750314224323941
-0.5652179040619789	0.29499719362208887
-0.6003286726860134	0.39726324690304315
-0.7142846940288936	0.4831112931300562
-0.7089909630863085	0.5663748188399981
-0.7296087627792128	0.62177922672209
-0.8315862085305106	0.6441704838440533
-0.8747250743449888	0.682030325836802
-0.9321022932466507	0.6858068582489915
-0.9860740928505496	0.7405781071561675
-0.9858812516808483	0.7577703785643948
-1.0546798131712412	0.8184914895224069
-1.145392089822586	0.7888497569859749
-1.1751966741339186	0.775126515775665

I_t	Y_t
-1.204994385192282	0.7232628751690667
-1.2648677941577025	0.6632669319563341
-1.260374426704848	0.6583394306796603
-1.3500983104660855	0.6403376031522052
-1.3369241183442202	0.6451641723491223
-1.3959845393225698	0.6232469136022897
-1.3720450761896632	0.5865533336011293
-1.4661322775399819	0.5940264737705111
-1.521125870894923	0.5672149289185254
-1.470950517947169	0.5263431815334215
-1.5679739732370455	0.49789457212455956
-1.6234246412146365	0.5019103505517739
-1.621815068122503	0.4875356085556648
-1.6518621741666402	0.4396282008924923
-1.7151686515242597	0.4109841958469174
-1.7429709850430082	0.32515617728602303
-1.7967922300621437	0.3100144978177349
-1.8625266685037791	0.3081501781065848
-1.8919278860619082	0.2787210204651825
-1.9500296887890578	0.3255752965436441
-2.0254666334766185	0.3950295288920353
-2.089506961283927	0.41605229267203253
-2.1316710987302976	0.47216029243711627
-2.1788597478638367	0.4983019007057647
-2.226520388009039	0.5844420997329862
-2.2777508058634357	0.6292435689470355
-2.367188234052972	0.6857891494125798
-2.594356033029133	0.793265394561697
-2.623541508469589	0.8896939710333649
-2.655889675010422	0.9614458230550806
-2.7849465612772697	1.0832541224488552
-2.834434509674115	1.1663248530853085
-2.9258144719173815	1.2419643894700658
-2.9017781740620467	1.3040355528285077
-2.9780728286101437	1.3666716897090283
-3.082126572614423	1.3937441423074184
-3.141079492999823	1.4168865758521243
-3.215099626032445	1.4378472491515306
-3.2488734100415506	1.4924037581671599
-3.2641663842983104	1.5085738460642748
-3.323310264895081	1.5682211183943218
-3.461740546955769	1.6670219549897594

I_t	Y_t
-3.4919983930611553	1.7178038176817647
-3.559792024846867	1.7745752374061303
-3.593055705493035	1.8551937470012327
-3.6759506164820945	1.9295322483218422
-3.7324532575324296	2.0119527618635393
-3.810294058415235	2.0369920679564513
-3.916957096881624	2.122115778729583
-3.9727811065858822	2.2144681272152598
-4.040798354571718	2.329177803058618
-4.09948038181535	2.381594273869628
-4.126705632853109	2.4230376958972113
-4.282073231913425	2.5690930936642893
-4.45807091967635	2.6627834002645407
-4.554001558418366	2.822242326907804
-4.700553586639593	2.983432728717481
-4.800448068721535	3.1096777136749996
-5.025636371994059	3.2483920342861388
-5.144405586880597	3.396950555671169
-5.232449146312347	3.563173763135644
-5.315362276622163	3.6438185605119937
-5.391198274264772	3.6859169963916907
-5.488498722219545	3.8107749091389342
-5.589986709233622	3.9388572504064294
-5.755159332299376	4.042711954406166
-5.721204447655005	4.144359576488124
-5.982769682007945	4.247016201093456
-5.934641473210264	4.40610974416346
-6.067801131025043	4.533408337128003
-6.165165793387324	4.695116831740003
-6.130527783445086	4.821349614109967
-6.442817423082921	4.920074694177565
-6.37748590441058	4.942613452907782
-6.357131532218663	4.962394123435942
-6.468428808140153	4.929488297901919
-6.485003801424921	4.895025598134301
-6.430964814163771	4.884997791830792
-6.496787410453988	4.923402339718606
-6.571742935480333	4.877755342046011
-6.496945467286127	4.844222462613068
-6.493495277547879	4.871753707722411
-6.530422790657934	4.906942159577974
-6.409450781122464	4.876018307094994

I_t	Y_t
-6.586804210438476	4.891653353971619
-6.660044635243046	4.969424922364519
-6.680176973574653	4.95397406620541
-6.635464917782348	4.998234060406817
-6.686936649990608	5.056231411087124
-6.769848471880518	5.092698314368622
-6.753934617729041	5.116752373834525
-6.747612136694832	5.172631548920785
-6.7972248093633505	5.204304681275314
-6.739863311297612	5.122303099560624
-6.749001028313339	5.102472539314004
-6.821725233325911	5.13486442761028
-6.884861152879483	5.129154855664604
-6.98646857435743	5.1461208293027285
-7.057653114118144	5.146289380443744
-7.109802025416156	5.0757816626116465
-7.183184631816278	4.930827382292277
-7.191606185092257	4.721869858041339
-7.2955414604882405	4.502402636615213
-7.419561716076927	4.238439841274942
-7.309811733979121	3.8990319206010104
-7.215072218263964	3.438252929058797
-7.284428929963496	2.854060530507241
-7.34812392952542	2.2972577707675708
-7.271114630884305	1.701483175928924
-7.195429609139502	1.1330434995862353
-7.189447577803872	0.5500622267900723
-7.178559216759887	0.07173414915160903
-7.3228487609341855	-0.29043797485787093
-7.412594594693715	-0.5096603926497952
-7.385587164982854	-0.7829475659868338
-7.374518201029718	-1.0557827419142134
-7.477512757618346	-1.293237953380052
-7.538693072005547	-1.4512192390021252
-7.573463355480406	-1.612371543292268

A.5 Mathematica code

Function, calculation of common factors I and Y from PD and LGD

```
fstepM[vec_,spv_,r_,n_,sStd_,tol_,sE_,ld_,pts_]:=Module[{cn,cj,ndf,mu,
cc,cf,mun,rnv,sc,scg,scf,td,qd,in,is,inW,denS,br},
```

```
cn=Transpose[{Transpose[vec[[1]]][[1]]+RandomReal[NormalDistribution[]
,Length[vec[[1]]]],Transpose[vec[[1]]][[2]]+1}];

cj=Sort[cn,#1[[1]]<#2[[1]]& ];

ndf=Max[1,Round[Length[cj]spv[[1]]]];

mu=cj[[ndf,1]];

td=Take[Transpose[cj][[2]],ndf];

qd=Table[Count[td,i]/Length[cj],{i,r}];



in=0;

denS=Sum[pts[[i]] qd[[i]],{i,r}];

br=0;

While[1-1/denS Sum[h[in,If[i==1,0,Sum[vec[[2,-j,2]],{j,1,i-
1}]],sE,ld,i] qd[[i]],{i,r}]>spv[[2]]+tol ||1-1/denS
Sum[h[in,If[i==1,0,Sum[vec[[2,-j,2]],{j,1,i-1}]],sE,ld,i]
qd[[i]],{i,r}]<spv[[2]]-tol, If[(-Sum[dh[in,If[i==1,0,Sum[vec[[2,-
j,2]],{j,1,i-1}]] ,sE,ld,i] qd[[i]],{i,r}]/denS)==0,If[br==0,inW=-
2;br=1;, Break[]],inW=in-(1-1/denS Sum[h[in,If[i==1,0,Sum[vec[[2,-
j,2]],{j,1,i-1}]],sE,ld,i] qd[[i]],{i,r}]-spv[[2]])/(-
Sum[dh[in,If[i==1,0,Sum[vec[[2,-j,2]],{j,1,i-1}]],sE,ld,i]
qd[[i]],{i,r}]/denS)];in=inW;];



cc={Drop[Transpose[cj][[1]],ndf]-
mu,Drop[Transpose[cj][[2]],ndf]}//Transpose;

cf=Select[cc,#[[2]]<r&];

mun=-mu;



rnv=Sort[RandomReal[NormalDistribution[mun,sStd],n ]];

sc=Select[rnv,#>0&];

scg=Transpose[{sc,Table[0,{Length[sc]}]}];

scf=Join[cf,scg];
```

```
{scf,Append[vec[[2]],{mun,in}]}

];

h[\[Iota]_,is_,\[Sigma]E_,ldf_,ts_]:=Exp[\[Iota]+is+ldf+\[Sigma]E^2/2
ts]CDF[NormalDistribution[0,1],(-ldf-\[Iota]-is)/(Sqrt[ts]\[Sigma]E)-
Sqrt[ts]\[Sigma]E]+1-CDF[NormalDistribution[0,1],(-ldf-\[Iota]-
is)/(Sqrt[ts]\[Sigma]E)]

dh[\[Iota]_,is_,\[Sigma]E_,ldf_,ts_]:=E^(\[Iota]+is+ldf+ts
\[Sigma]E^2/2) CDF[NormalDistribution[0,1],(-ldf-\[Iota]-
is)/(Sqrt[ts]\[Sigma]E)-Sqrt[ts]\[Sigma]E]

sdens[n_,r_,sStd_,spv_,tol_,sE_,ld_,pts_]:=Module[{rnvs,scs,scgs,ndf},

rnvs=Sort[RandomReal[NormalDistribution[0,sStd],n ]];

ndf=Max[1,Round[n spv[[1]]]];

scs=Drop[rnvs,ndf]-rnvs[[ndf]];

scgs=Transpose[{scs,Table[0,{Length[scs]}]}];


Nest[ fstepM[#,spv,r,n,sStd,tol,sE,ld,pts]&
,{scgs,Table[{0,0},{r}]},2r]

]

fytM[n_,r_,sStd_,lt_,gt_,sd_,sm_,tol_,sE_,ld_,pts_]:=Fold[
fstepM[#1,#2,r,n,sStd,tol,sE,ld,pts]& ,{sd,sm},{lt,gt}//Transpose]
```

Function, calculation of PD and LGD from common factors

```
fstepMinv[vec_,spvInv_,r_,n_,sStd_,tol_,sE_,ld_,pts_]:=Module[{cn,cj,c
c,cf,ndf,rnv,sc,scg,scf,td,qd,denS,Lt,Gt},

cn=Transpose[{Transpose[vec[[1]]][[1]]+RandomReal[NormalDistribution[s
pvInv[[1]],1],Length[vec[[1]]]],Transpose[vec[[1]]][[2]]+1}];

cj=Sort[cn,#1[[1]]<#2[[1]]& ];

ndf=Count[Negative[Transpose[cn][[1]]],True];

Lt=ndf/Length[cj];

td=Take[Transpose[cj][[2]],ndf];
```

```
qd=Table[Count[td,i]/Length[cj],{i,r}];


denS=Sum[pts[[i]] qd[[i]],{i,r}];

Gt=1-1/denS Sum[h[spvInv[[2]],Sum[vec[[2,-j,2]],{j,1,i-1}],sE,ld,i]
qd[[i]],{i,r}];


cc={Drop[Transpose[cj][[1]],ndf],Drop[Transpose[cj][[2]],ndf]}//Transp
ose;

cf=Select[cc,#[[2]]<r&];


rnv=Sort[RandomReal[NormalDistribution[spvInv[[1]],sStd],n ]];

sc=Select[rnv,#>0&];

scg=Transpose[{sc,Table[0,{Length[sc]}]}];

scf=Join[cf,scg];


{scf,Append[vec[[2]],{spvInv[[1]],spvInv[[2]]}],Append[vec[[3]],{Lt,Gt
}]}

];

h[\[Iota]_,is_,\[Sigma]E_,ldf_,ts_]:=Exp[\[Iota]+is+ldf+\[Sigma]E^2/2
ts]CDF[NormalDistribution[0,1],(-ldf-\[Iota]-is)/(Sqrt[ts]\[Sigma]E)-
Sqrt[ts]\[Sigma]E]+1-CDF[NormalDistribution[0,1],(-ldf-\[Iota]-
is)/(Sqrt[ts]\[Sigma]E)]

sdensInv[n_,r_,sStd_,spvInv_,tol_,sE_,ld_,pts_]:=Module[{rnvs,scs,scgs
,ndf},

rnvs=Sort[RandomReal[NormalDistribution[spvInv[[1]],sStd],n ]];

ndf=Count[Negative[rnvs],True];

scs=Drop[rnvs,ndf]-rnvs[[ndf]];

scgs=Transpose[{scs,Table[0,{Length[scs]}]}];
```

```
Nest[ fstepMinv[#,spvInv,r,n,sStd,tol,sE,ld,pts]& ,{scgs,Table[{-rnvs[[ndf]],0},{r}]},2r]

]

fytMinv[LG_,\[CapitalDelta]Yt_,\[CapitalDelta]It_,sd_,sm_,n_,r_,sStd_,tol_,sE_,ld_,pts_]:=Fold[ fstepMinv[#1,#2,r,n,sStd,tol,sE,ld,pts]& ,{sd,sm,{LG}},{\[CapitalDelta]Yt,\[CapitalDelta]It}//Transpose]
```

Data a parameters

```
data=Import[""][[1]];

lt=Transpose[Drop[data,1]][[1]]/100;

gt=Transpose[Drop[data,1]][[4]];

tol=10^-8;

ld=Log[1];

sE=0.12;

((1+urok)^r urok)/((1+urok)^r-1) ;

spv={0.005,0.004};

n=10000;

sStd=5;

r=120;

urok=0.01;

Table[(urok-1/(1+urok)^r+1/(1+urok)^i)/(1-1/(1+urok)^r),{i,r}];

pts=Prepend[Table[(-1/(1+urok)^r+1/(1+urok)^i)/(1-1/(1+urok)^r),{i,r-1}],1];
```

Results

```
Timing[fis=sdens[n,r,sStd,spv,tol,sE,ld,pts];]

sd=fis[[1]];

sm=fis[[2]];
```

```
ListLinePlot[{sm[[All,1]],sm[[All,2]]},PlotRange->All]

Timing[vytS=fytM[n,r,sStd,lt,gt,sd,sm,tol,sE,ld,pts];]

ListLinePlot[{Drop[vytS[[2,All,1]],3r],Drop[vytS[[2,All,2]],3r]},PlotR
ange->All]

ListLinePlot[Drop[vytS[[2,All,2]],3r]]

ListLinePlot[{Accumulate[Drop[vytS[[2,All,1]],3r]],Accumulate[Drop[vyt
S[[2,All,2]],3r]]}]

ListLinePlot[{lt,gt},PlotRange->All]

sdInv=vytS[[1]];

smInv=Take[vytS[[2]],-r];

ListLinePlot[{Take[Transpose[smInv][[1]],-
10],Take[Transpose[smInv][[2]],-10]}]

vytSinv=fytMinv[{lt[[1]],gt[[1]]},Take[Transpose[smInv][[1]],-
10],Take[Transpose[smInv][[2]],-
10],sdInv,smInv,n,r,sStd,tol,sE,ld,pts];

ListLinePlot[{vytSinv[[2,All,1]],vytSinv[[2,All,2]]},PlotRange->All]

ListLinePlot[{Join[Take[lt,133],Drop[vytSinv[[3,All,1]],1]],Join[Take[
gt,133],Drop[vytSinv[[3,All,2]],1]]}]
```

A.6 The Johansen test of cointegration for Y and I and corresponding cointegrating vectors

Rank	Eigenvalue	Trace test	[p-value]	Lmax test	[p-value]
0	0.093990	13.766	[0.0890]	12.733	[0.0852]
1	0.0079771	1.0332	[0.3094]	1.0332	[0.3094]

Beta (cointegrating vectors)

Y	-1.1699	-0.25296
I	-0.78858	-0.65458

Renormalized beta coefficients

Y	1.0000	0.38645
I	0.67404	1.0000

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

Carling, K., Jacobson, T., Lindé, J., & Roszbach, K. (2007). Corporate credit risk modeling and the macroeconomy. Journal of Banking & Finance, Volume 31, Issue 3, pp. 845–868.

d'Ecclesia, R. (2008). Estimation of credit default probabilities. *XLIII EWGFM meeting.* London.

Duffie, D. (2005). Credit risk modeling with affine processes. Journal of Banking & Finance, Volume 29, Issue 11, pp. 2751–2802.

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

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

Gapko, P., & Šmíd, M. (2012a). Modeling a Distribution of Mortgage Credit Losses. *Ekonomicky casopis*, 60(10): 1005-1023.

Gupton, G. M., Finger, C. C., & Bhatia, M. (1997). *CreditMetrics™ - Technical Document.* J.P. Morgan & Co. Incorporated.

Kallenberg, O. (2002). *Foundations of Modern Probability, 2nd edition.* New York: Springer.

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

Vasicek, O. A. (1987). *Probability of Loss on Loan Portfolio.* KMV.

Vasicek, O. A. (2002). The Distribution of Loan Portfolio Value. *Risk.*

Witzany, J. (2010). On deficiencies and possible improvements of the Basel II unexpected loss single-factor model. *Czech Journal of Economics and Finance*, 60(3): 252-268.

Yang, B. H., & Tkachenko, M. (2012). Modeling exposure at default and loss given default: empirical approaches and technical implementation. *The Journal of Credit Risk,* 8(2): 81-102.

Report on Opponents' comments

Rita D'Ecclesia:

The following changes were applied according to the specific comments:

Essay 1:

- The table with descriptive statistics of the used dataset was added and the surrounding text adjusted accordingly

- The relationship between the S&P stock index and the Y common factor was estimated by an autoregressive model and results briefly commented in the text – this relationship was used only to show that there might be some dependence of the common factor on macroeconomic environment; this dependency is then examined more accurately in the Essay 3

Essay 2:

- All equations in the paper were numbered

- A short explanation was added to the page 8, where Rt is defined (the difference to RD)

- Equation 3.4 is taken from Vasicek, where the default probability is an input – it assumes that it is constant at a given time, but dependent on the two risk factors in its evolution; as this part of the essay serves only as a description of existing framework, the dissertation wasn't changed at this place.

- Non-random LGD at page 5 corrected to random LGD

- The repetition at the bottom of the page 5 serves as an explanation that Frye and Pykhtin started from the same point, thus the equation was kept.

- Equation 3.9 (= 3.7 in the opponent's report) is taken from Pykhtin so it was kept

- Footnote on the page 8 added to explain that the assumption of loans lasting only one period is very restrictive and thus is a focus of our future research.

- The error in the notation of Rt was fixed in the section 3.4
- The relationship between Y and S&P 500 was estimated by the autoregressive model and a short note on the results was added; we are aware that this simple model of the correlation is far from being perfect, however, we use this simplification because modeling the correlation is not the main focus of this essay
- Descriptive statistics was added for the time series Rt and Dt; also, a short notice on the stationarity test of the factors Y and I was shortly commented
- In the first version of the essay, we calculated the Johansen test additionally to the Engle-Granger, because Engle-Granger, despite confirming nonstationarity of both tested series, did not reject the nonstationarity of residuals from the cointegrating regression; however, we decided to run the VECM and our results show that there is a dependence structure between Y and I; a discussion was added to the section 3.4, Johansen test removed
- Discussion of stationarity tests of residuals added
- A note on the difference of Y on I and I on Y dependence added
- A note explaining how the simulation was done in practice added; the reference was fixed
- We consider our model dynamic because for LGD and PD we are able to estimate the dynamic relationship between underlying risk factors and macroeconomic environment and to predict dynamic LGD and PD

Essay 3:

- A note added to the section 4.2.5, explaining that values of all parameters except of $P((X,Y,\pi)\in \cdot)$ were chosen based on empirical observations or expert judgment
- The explanation of parameters choice in the section 4.3.2. was improved
- Figure 4.1 switched to double range
- Better explanation of the adjustment of I and Q at figures 4.2 and 4.3 added

- A short note that stationarity of both endogenous VECM variables was rejected added (at the beginning of the section 4.3.3)

- Results of the Johansen test for cointegration and the cointegrating vectors added to the appendix; no additional normalizations (except the one provided in the appendix, which was done automatically by the Gretl software) were used for Y and I

- Scales present at Figures 4.5 and 4.6, legend added to describe better the visualizations

- σ_Y and σ_I are standard errors of the cointegration regression (newly mentioned in the text)

- As the evolution of borrower's assets, and thus the quarterly increments of the underlying risk factors in the IRB framework are assumed to be normally distributed, the yearly increment of the underlying risk factors will be distributed normally with a quadruple variance

Tomáš Tichý:

- The introduction of the first paper was updated and a paragraph, which describes the evolution of the banking regulation since the paper was published, was added.

- The references in the third paper were extended.

- References to Frye and Pykhtin added to the introduction – the text referred to their work, but the correct reference was missing.

- References of Eberlein (2001, 2002) we fixed.

- Systemic replaced by systematic.

- Misspellings and other suggestions fixed and accepted.

Jiří Witzany:

The following changes were applied according to the specific comments:

Essay 1:

- A sentence explaining that $Z_{i,t}$ are identically distributed only in the case when the loans last only one period added. As discussed in the essay, this is the biggest shortcoming of the first and the second essays, which we dealt with in the last essay.

- A note that we consider all loans as lasting only one period added to the sub-chapter 2.4.2. This allows all loans to enter and exit the calculation each period

Essay 2:

- A discussion why we used the rate of foreclosures started on the defaulted accounts was added to the data description. Unfortunately, for the overall US mortgage market, there does not exist (to our knowledge) a better publicly available dataset of LGD

Essay 3:

- The results were recalculated using 95th quantile of L instead the 99.9th