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

Modeling a Distribution of Mortgage Credit Losses

Petr Gapko
Martin Šmíd

IES Working Paper: 23/2010

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

[UK FSV – IES]

Opletalova 26
CZ-110 00, Prague
E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Institut ekonomických studií
Fakulta sociálních věd
Univerzita Karlova v Praze

Opletalova 26
110 00 Praha 1

E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Disclaimer: The IES Working Papers is an online paper series for works by the faculty and students of the Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Czech Republic. The papers are peer reviewed, but they are *not* edited or formatted by the editors. The views expressed in documents served by this site do not reflect the views of the IES or any other Charles University Department. They are the sole property of the respective authors. Additional info at: ies@fsv.cuni.cz

Copyright Notice: Although all documents published by the IES are provided without charge, they are licensed for personal, academic or educational use. All rights are reserved by the authors.

Citations: All references to documents served by this site must be appropriately cited.

Bibliographic information:
Gapko, P., Šmíd, M. (2010). "Modeling a Distribution of Mortgage Credit Losses" IES Working Paper 23/2010. IES FSV. Charles University.

This paper can be downloaded at: http://ies.fsv.cuni.cz

Modeling a Distribution of Mortgage Credit Losses

Petr Gapko*
Martin Šmíd#

* Institute of Information Theory and Automation,
Academy of Sciences of the Czech Republic and
IES, Charles University Prague
E-mail: petr.gapko@seznam.cz

Institute of Information Theory and Automation,
Academy of Sciences of the Czech Republic

September 2010

Abstract:
One of the biggest risks arising from financial operations is the risk of counterparty default, commonly known as a “credit risk”. Leaving unmanaged, the credit risk would, with a high probability, result in a crash of a bank. In our paper, we will focus on the credit risk quantification methodology. We will demonstrate that the current regulatory standards for credit risk management are at least not perfect, despite the fact that the regulatory framework for credit risk measurement is more developed than systems for measuring other risks, e.g. market risks or operational risk. Generalizing the well known KMV model, standing behind Basel II, we build a model of a loan portfolio involving a dynamics of the common factor, influencing the borrowers' assets, which we allow to be non-normal. We show how the parameters of our model may be estimated by means of past mortgage deliquency rates. We give a statistical evidence that the non-normal model is much more suitable than the one assuming the normal distribution of the risk factors. We point out how the assumption that risk factors follow a normal distribution can be dangerous. Especially during volatile periods comparable to the current crisis, the normal distribution based methodology can underestimate the impact of change in tail losses caused by underlying risk factors.

Keywords: Credit Risk, Mortgage, Delinquency Rate, Generalized Hyperbolic Distribution, Normal Distribution

JEL: G21

Acknowledgements:
Support from the Czech Science Foundation under grants 402/09/H045 and 402/09/0965, and GAUK 46108 is gratefully acknowledged.

I. Introduction

Credit risk – the risk of counterparty default – accompanies every business idea, but is most visible in the banking sector. Banking business is based on the simple idea of reallocation of capital. A typical bank collects deposits and offers them on in the form of loans. Therefore, banking business is a high leverage (and risky) business with a high proportion of external financing.

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). Basel II is a document describing minimum principles for risk management in the banking sector. It is applicable all over the world, and in the European Union it is implemented into European law by the Capital Requirements Directive (CRD) (European Commission, 2006). However, the regulatory standards need to be simple enough so that every bank subject to regulation is capable of using them. This leads to a trade-off between simplicity and accuracy.

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 (e.g. JP Morgan CreditMetrics, etc.), 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.

[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.

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 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.

II. 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).

II.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 method, and because regulatory 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

[2] A supervisor is a regulator of 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.

each separate exposure.[4] The average loss that could occur in the following 12 months is calculated as follows:

$$EL = E(PD) \cdot E(LGD) \cdot EAD, \quad (1)$$

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 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.

II.2. Expected and unexpected loss for a portfolio

---

[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…).

[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.

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 (i) 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 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]

III. Our approach

III.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

$$\log A_{i,1} = \log A_{i,0} + \eta + \gamma \mathrm{X}_i \qquad (2)$$

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).

III.1.1. The generalization

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

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

$$\log A_{i,t} = \log A_{i,t-1} + Y_t + U_{i,t} \tag{3}$$

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

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

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 $\mathbb{E}Z_{1,1} = 0$, $\mathrm{var}(Z_{1,1}) = \sigma$, $\sigma > 0$, 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)).

III.1.2. 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

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

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

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

furter implying that

$$Y_t \doteq \Psi^{-1}(L_{t-1}) - \Psi^{-1}(L_t) \tag{4}$$

and

$$L_t \doteq \Psi(\Psi^{-1}(L_{t-1}) - Y_t) \tag{5}$$

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 (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.

III.2. 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:

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

where

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

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:

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

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

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

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

IV. Data and results

III.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 1. We observe an unprecedentedly huge increase in the 90+ delinquency rate beginning with the second quarter of 2007.

[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.).

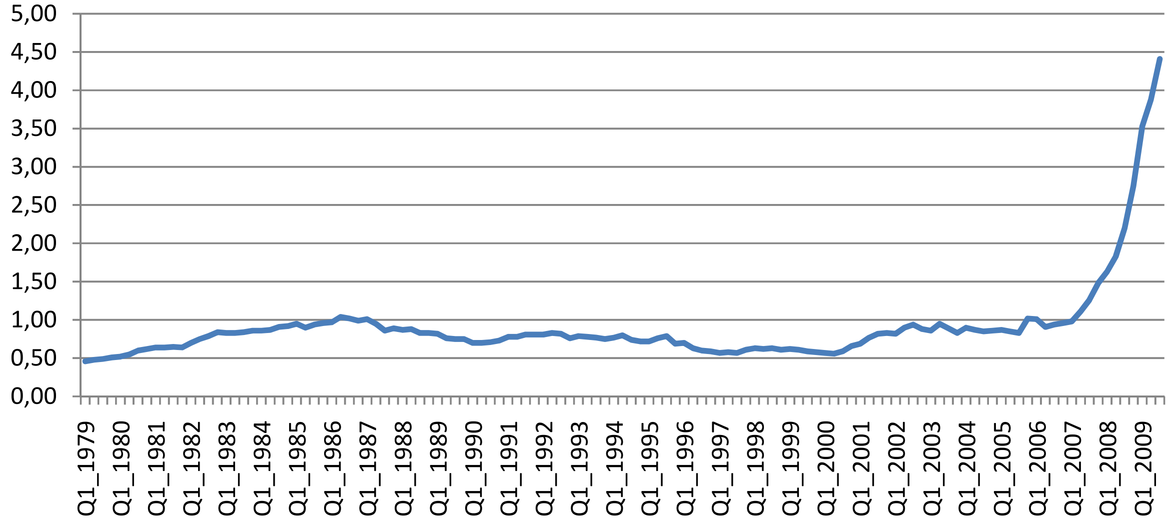

Figure 1: Development of US 90+ delinquency rate

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, displaying the common factor (left axis), adjusted for inflation, against the S&P 500 stock index. The 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%).

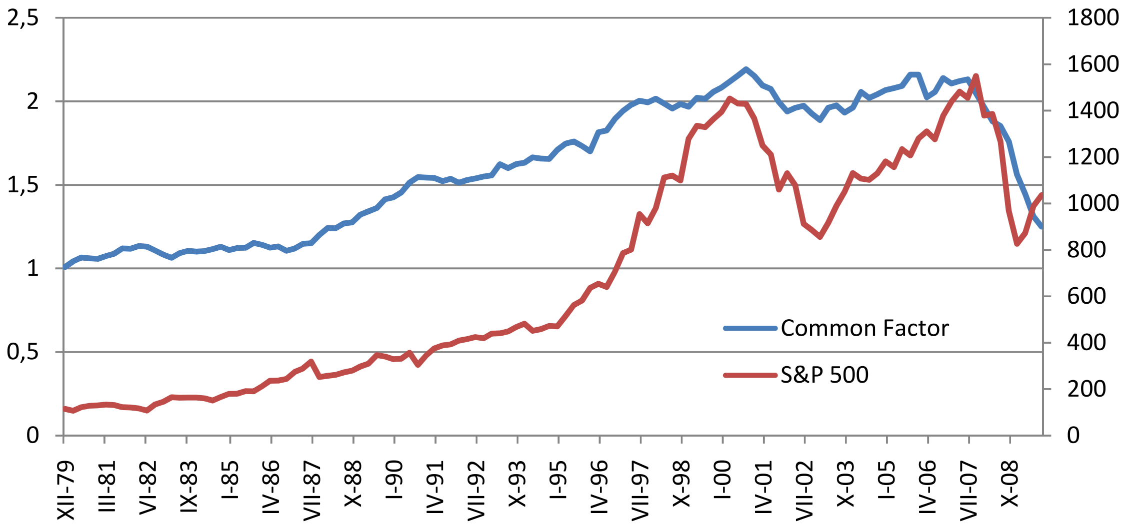

Figure 2: Comparison of the development of the common factor and lagged S&P 500 returns

III.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:

$$\chi^2 = \sum_{i=1}^{t} (O_i - E_i)^2 / E_i, \tag{8}$$

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 1 shows graphically the difference between the estimated generalized hyperbolic and normal distributions. From Figure 1 we can see that the GHD is able to describe better both the skewness and the kurtosis of the dataset.

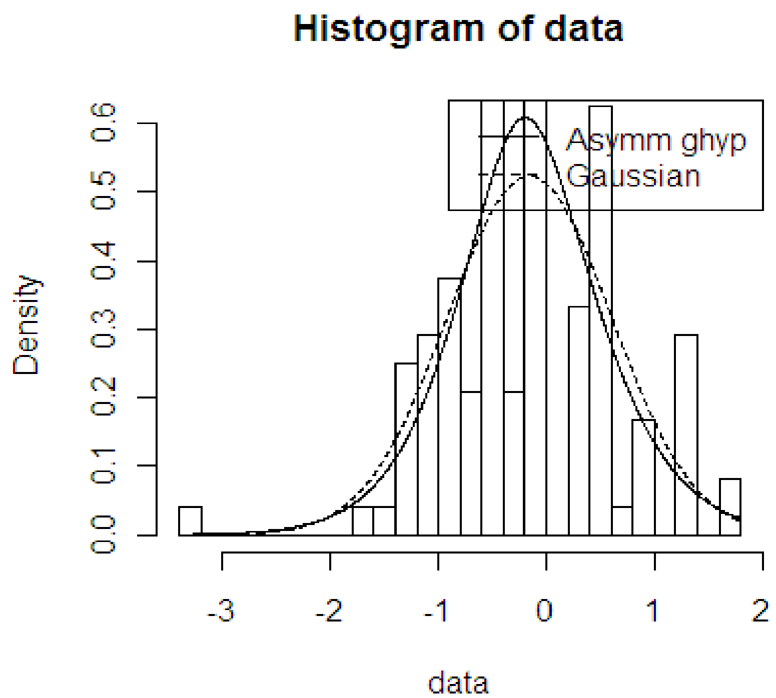

Figure 3: Compared histograms: GHD vs. Normal vs. dataset

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]).

III.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 (4) to get

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

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.

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

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

The first column in the Table 2 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 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.

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 attributed: 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 99.9th 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 99.9th 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. 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:

$$M(u) = e^{u\mu}\left(\frac{\alpha^2-\beta^2}{\alpha^2-(\beta+u)^2}\right)^{\lambda/2}\frac{K_\lambda(\delta\sqrt{\alpha^2-(\beta+u)^2}}{K_\lambda(\alpha^2-\beta^2)}, \quad (1)$$

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

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

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

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

$$+ \frac{(\beta\delta)^2}{\alpha^2 - \beta^2} \left(\frac{K_{\lambda+2}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}{K_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)} - \left(\frac{K_{\lambda+1}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}{K_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)} \right)^2 \right) \quad (3)$$

By substituting from equations (2) and (3) into equation (1) 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

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

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

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

and that

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

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

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

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

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

and, consequently,

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

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

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

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

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

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

Bibliography

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. *Lévy Processes: Theory and Applications*, 319–337.

Eberlein, E. (2002). The Generalized Hyperbolic Model: Financial Derivatives and Risk Measures. *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*

IES Working Paper Series

2010

1. Petra Benešová, Petr Teplý : *Main Flaws of The Collateralized Debt Obligation's Valuation Before And During The 2008/2009 Global Turmoil*
2. Jiří Witzany, Michal Rychnovský, Pavel Charamza : *Survival Analysis in LGD Modeling*
3. Ladislav Kristoufek : *Long-range dependence in returns and volatility of Central European Stock Indices*
4. Jozef Barunik, Lukas Vacha, Miloslav Vosvrda : *Tail Behavior of the Central European Stock Markets during the Financial Crisis*
5. Onřej Lopušník : *Různá pojetí endogenity peněz v postkeynesovské ekonomii: Reinterpretace do obecnější teorie*
6. Jozef Barunik, Lukas Vacha : *Monte Carlo-Based Tail Exponent Estimator*
7. Karel Báťa : *Equity Home Bias in the Czech Republic*
8. Petra Kolouchová : *Cost of Equity Estimation Techniques Used by Valuation Experts*
9. Michael Princ : *Relationship between Czech and European developed stock markets: DCC MVGARCH analysis*
10. Juraj Kopecsni : *Improving Service Performance in Banking using Quality Adjusted Data Envelopment Analysis*
11. Jana Chvalkovská, Jiří Skuhrovec : *Measuring transparency in public spending: Case of Czech Public e-Procurement Information System*
12. Adam Geršl, Jakub Seidler : *Conservative Stress Testing: The Role of Regular Verification*
13. Zuzana Iršová : *Bank Efficiency in Transitional Countries: Sensitivity to Stochastic Frontier Design*
14. Adam Geršl, Petr Jakubík : *Adverse Feedback Loop in the Bank-Based Financial Systems*
15. Adam Geršl, Jan Zápal : *Economic Volatility and Institutional Reforms in Macroeconomic Policymaking: The Case of Fiscal Policy*

16. Tomáš Havránek, Zuzana Iršová : *Which Foreigners Are Worth Wooing? A Meta-Analysis of Vertical Spillovers from FDI*
17. Jakub Seidler, Boril Šopov : *Yield Curve Dynamics: Regional Common Factor Model*
18. Pavel Vacek : *Productivity Gains From Exporting: Do Export Destinations Matter?*
19. Pavel Vacek : *Panel Data Evidence on Productivity Spillovers from Foreign Direct Investment: Firm-Level Measures of Backward and Forward Linkages*
20. Štefan Lyócsa, Svatopluk Svoboda, Tomáš Výrost : *Industry Concentration Dynamics and Structural Changes: The Case of Aerospace & Defence*
21. Kristýna Ivanková : *Isobars and the Efficient Market Hypothesis*
22. Adam Geršl, Petr Jakubík : *Relationship Lending, Firms' Behaviour and Credit Risk: Evidence from the Czech Republic*
23. Petr Gapko, Martin Šmíd : *Modeling a Distribution of Mortgage Credit Losses*

All papers can be downloaded at: http://ies.fsv.cuni.cz

Univerzita Karlova v Praze, Fakulta sociálních věd
Institut ekonomických studií [UK FSV – IES] Praha 1, Opletalova 26
E-mail : ies@fsv.cuni.cz http://ies.fsv.cuni.cz