International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:8, No:12, 2014

Modeling Default Probabilities of the Chosen Czech Banks in the Time of the Financial Crisis

Petr Gurný

***Abstract*—**One of the most important tasks in the risk management is the correct determination of probability of default (PD) of particular financial subjects. In this paper a possibility of determination of financial institution's PD according to the credit-scoring models is discussed. The paper is divided into the two parts. The first part is devoted to the estimation of the three different models (based on the linear discriminant analysis, logit regression and probit regression) from the sample of almost three hundred US commercial banks. Afterwards these models are compared and verified on the control sample with the view to choose the best one. The second part of the paper is aimed at the application of the chosen model on the portfolio of three key Czech banks to estimate their present financial stability. However, it is not less important to be able to estimate the evolution of PD in the future. For this reason, the second task in this paper is to estimate the probability distribution of the future PD for the Czech banks. So, there are sampled randomly the values of particular indicators and estimated the PDs' distribution, while it's assumed that the indicators are distributed according to the multidimensional subordinated Lévy model (Variance Gamma model and Normal Inverse Gaussian model, particularly). Although the obtained results show that all banks are relatively healthy, there is still high chance that "a financial crisis" will occur, at least in terms of probability. This is indicated by estimation of the various quantiles in the estimated distributions. Finally, it should be noted that the applicability of the estimated model (with respect to the used data) is limited to the recessionary phase of the financial market.

***Keywords*—**Credit-scoring Models, Multidimensional Subordinated Lévy Model, Probability of Default.

I. Introduction

THE internationalization of financial markets has significantly expanded investment opportunities and risk. No doubt, the probability of default (PD) is one of the key input factors of credit risk modeling and measuring. Estimating the borrower's risk level by assigning a different PD to each borrower is now widely employed in many banks (especially since 2008, when most of the world has been going through a period of financial and economic turmoil). The false estimation of PD leads to unreasonable rating; incorrect pricing of financial instruments and thereby it was one of the causes of the financial crisis.

There are a lot of models to estimate the probability of default in financial world. Among the most widely used ones are models, generally known as credit-scoring models. These are multivariate models which use the main financial indicators of a company as input, attributing a weight to each of them that reflects its relative importance in forecasting default. This paper is related to a number of other studies focused on the credit-scoring models. Decisive boost to the development and spread of these models came in the 1960s, with studies by [1], [2]. Other important contributions in that field are [3]-[8]. The vast majority of already proposed credit scoring models were derived on a sample of non-financial institutions, mainly due to the fact that defaults of financial institutions occur relatively scarcely and not all the data are publicly available. Nevertheless, there were several more or less sufficient attempts to identify the key factors for healthy financial institutions, originating from financial statements, see e.g. [9] and references therein.

The goal of the paper is the application and verification of possibility usage of the chosen credit-scoring models (linear discriminant analysis, logit regression and probit regression) to modeling default probabilities of the commercial banks during the financial crisis period. Within these methods they are arranged three models for determination of probability of default of financial institutions on the basis of sample of almost 300 American commercial banks and consequently the chosen model is applied on three Czech commercial banks with a view to estimate their probability distribution of PD with one year prediction. Finally, there is calculated risk of the PD on the various level of probability.

To estimate PDs' prediction it is necessary to simulate particular financial indicators. While till lately the geometric Brownian motion was the most used model for the description of the progress of the particular underlying factor, presently it is tendency to analyze and to simulate the time series of the particular variables by means of Lévy processes as it could be helpful to better predictor of the models ability, see [10] or [11].

The paper is organized as follows. In the theoretical part of the paper we explain shortly the linear discriminant analysis and regression models (logit and probit approach) at first. Next, Lévy processes and copula function will be more detailed presented. Further we will apply the theoretical knowledge on the chosen sample of the US banks, we will determine the model and consequently we will apply the model on the portfolio of three Czech commercial banks to determine their distributions of the PDs.

II. Methodology and Literature Review

First, there will be briefly introduced the credit scoring model in this part of the paper, then we will give more detail the Lévy processes (focusing on subordinated Lévy processes)

The paper is based on research activities sponsored through GACR Project No. P403/14-15175P and ESF Project No. CZ.1.07/2.3.00/20.0296. The support is greatly acknowledged.

Petr Gurný is with the Department of Finance, VSB – Technical University of Ostrava, 72100 Czech Republic (e-mail: petr.gurny@vsb.cz).

and copula functions (focusing on Gaussian copula function).

A. Credit Scoring Models

Credit scoring models are statistical models derived by means of econometric methods, such as regression analysis, discriminant analysis, logit and probit models, or even neural networks or panel models – see any econometric textbook or credit risk handbook, such as [12] or [13]. They work on the principle of assigning weights to financial and economic indicators. Weights express the significance of these indicators in the estimation of the borrower's default.

1. Logit and Probit Models

Logit and probit regression analysis are the multivariate techniques which allow for estimating the probability that an event occurs or not, by predicting a binary dependent outcome from a set of independent variables. Thus the goal is to model probability P_i that default occurs, for the given borrower i, by specifying the model

$$P_i = f\left(\alpha + \sum_{j=1}^{n} \beta_j x_{i,j}\right),$$

where x_j are particular financial indicators for *i*-th borrower and α, β_j are estimated parameters.

There are a lot of ways of specifying P_i, but in this paper we will focus on the logit and probit transformation, thus logit and probit model.

In logit model we use logistic transformation:

$$P_i = 1/\left[1 + \exp(-\alpha - \sum_{j=1}^{n} \beta_j x_{i,j})\right]. \tag{1}$$

In the case of probit model we use the cumulative distribution function of normal distribution:

$$P_i = \int_{-\infty}^{\alpha + \sum_{j=1}^{n} \beta_j x_{i,j}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt. \tag{2}$$

Due to nonlinear features of these models it is necessary to use maximum likelihood method for parameters estimation.

2. Linear Discriminant Analysis (LDA)

A basic idea of LDA is to maximize the difference between the two groups, while the differences among particular members of the same group are minimized. Within credit risk models, one group consists of good borrowers (non-defaulted – *group A*), while the other includes bad ones (already defaulted – *group B*). The differences are measured by means of the discriminant variable – *score z*. For a given borrower i, we calculate the score as follows:

$$z_i = \sum_{j=1}^{n} \gamma_j x_{j,i}, \tag{3}$$

where x_i are particular financial indicators and γ_j is the vector of estimated parameters.

Linear discriminant analysis can be used to produce a direct estimate of the probability of default. See [2] or [14] for calculation and more details.

B. Multidimensional Subordinated Lévy Processes

The first focus at Lévy models with jumps goes back to 1930's. The most recent and complete monographs on the theory behind and/or application of Lévy models are [10] and [15]-[17]. In this section, we first describe the marginal distributions of subordinated Lévy models. Then, we will show, how to obtain multidimensional distribution from marginal distributions by means of copula functions.

1. Marginal Distribution

We can define a Lévy process $\{X(t)\}_{t\geq 0}$ as cádlág real value stochastic process with $X(0)=0$ which is stochastically continuous and has stationary independent increments. For a given infinitely divisible distribution, we can define the triplet of Lévy characteristics, $\{\gamma, \sigma^2, \nu(dx)\}$, where the former two define the drift of the process (deterministic part) and its diffusion. The latter is a Lévy measure. If it can be formulated as $\nu(dx) = u(x)dx$, it is a Lévy density. It is similar to the probability density, with the exception that it need not be integrable and zero at origin.

Let X be a Brownian motion. If we replace standard time t in Brownian motion X, $X(t;\mu,\sigma) = \mu dt + \sigma Z(t)$, by its suitable function $l(t)$ as follows:

$$X(l(t);\theta,\nu) = \theta l(t) + \nu Z(l(t)) = \theta l(t) + \nu\varepsilon\sqrt{l(t)}, \tag{4}$$

we get a subordinated Lévy model. Due to the simplicity one of the most suitable candidates for the function $l(t)$ seem to be either the variance gamma model (VG) – the overall process is driven by a gamma process from the gamma distribution with shape *a* and scale *b* depending solely on variance *k*, $G[a,b]$, or normal inverse Gaussian model (NIG) – the subordinator is given by an inverse Gaussian process based on the inverse Gaussian distribution, $IG[a,b]$.

The final step to get a model for a marginal distribution depends on the issue we are going to solve. For example, if the task is to model the prices of a financial asset, i.e. strictly positive value, we should evaluate the Lévy model (4) in the exponential part:

$$S(t) = S_0 \exp(\mu t + x(l(t)) + \omega t), \tag{5}$$

where μ states a long-term drift of the price (average return) and is the mean correcting parameter. By contrast, if we model a variable, which can be both positive and negative (e.g. price returns), we can proceed as follows:

$$x(t) = \mu t + X(l(t)) - \theta t, \tag{6}$$

so that the long-term drift is fit again.

2. Copula Function

A useful tool of dependency modeling is the copula function. In this paper, we restricted ourselves to ordinary copula functions. Basic reference for the theory of copula functions is [18], while [19] and [20] target mainly on the application issues in finance.

First we will define a *n*-variate copula *C* as the joint distribution function of *n* Uniform (0,1) random variables. If we label the *n* random variables as $(U_1, U_2, ..., U)$ then we can write down the copula *C* as $C(u_1, ...u_n) = \Pr(U_1 \leq u_1, ..., U_n \leq u_n$. Well known Sklar's theorem states that for any joint distribution function F, there is a unique copula *C* that satisfies:

$$F(x_1, ..., x_n) = C(F_1(x_1), ..., F_n(x_n)). \qquad (7)$$

Sklar's theorem proves that in examining multivariate distributions, we can separate the dependence structure from the marginal distributions. Conversely, we can construct a multivariate joint distribution from a set of the marginal distributions, and a selected copula. The dependence structure is captured in the copula function and is independent of the form of the marginal distributions.

Let *R* be a symmetric, positive definite matrix with $diag\ (R) = 1$ and let Φ_R be a standardized multivariate normal distribution with correlated matrix *R*. Then the multivariate Gaussian copula is defined as

$$C(u_1, ..., u_n; R) = \Phi_R\left(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_n)\right), \qquad (8)$$

where $\Phi^{-1}(u)$ denotes the inverse of the normal cumulative distribution function.

3. Parameter Estimation of Multidimensional Subordinated Lévy Models

There exist three main approaches to parameter estimation for copula function based dependency modeling: exact maximum likelihood method (EMLM), inference for margins (IFM), and canonical maximum likelihood (CML). While for the former all parameters are estimated within one step, which might be very time consuming, mainly for high dimensional problems or complicated marginal distributions, the latter two methods are based on estimating the parameters for the marginal distribution and parameters for the copula function separately. While assuming IFM, marginal distributions are estimated in the first step and the copula function in the second one, for CML instead of parametric margins empirical distributions are used. For more details, see any of the empirically oriented literature such as [20].

In this paper we will assume IFM approach. In order to estimate the parameters of marginal distributions, generalized method of moments will be used. Table I shows the basic characteristics of the VG distribution, NIG distribution and GBM (Geometric Brownian Motion) for comparison.

TABLE I
FIRST FOUR MOMENTS OF THE VG, NIG AND GBM DISTRIBUTION

Moments	VG (θ, σ, ν)	NIG (θ, σ, ν)	GBM (μ, σ^2)
Mean	θ	θ	μ
Variance	$\sigma^2 + \nu\theta^2$	$\sigma^2 + \nu\theta^2$	σ^2
Skewness	$\theta\nu\left(3\sigma^2 + 2\nu\theta^2\right)\left(\sigma^2 + \nu\theta^2\right)^{-3/2}$	$3\theta\nu\left(\sigma^2 + \nu\theta^2\right)^{-1/2}$	0
Kurtosis	$3\left(1 + 2\nu - \dfrac{\nu\sigma^4}{\left(\sigma^2 + \nu\theta^2\right)^2}\right)$	$3\left(\dfrac{\theta^2\nu(1+5\nu) + \sigma^2(1+\nu)}{\sigma^2 + \nu\theta^2}\right)$	3

III. CREDIT MODELS ESTIMATION AND VERIFICATION

First, there will be the models from empirical sample derived in this part of the paper. Estimated models will be verified on the control sample and finally the best fitted model will be chosen.

A. Data Description

In this paper we will work with a sample of 298 financial institutions. We can observe basic characteristics of the empirical sample of the banks in the Table II. The defaulted banks are thought the financial institutions which have gone into liquidation or undergone financial restructuring processes (e.g. take-over by another company or by government). The samples of the financial institutions were chosen randomly pursuant to the publicly available information (www.federalreserve.gov, www.failedbankreporter.com). As a second step the financial indicators from financial statements were identified, see [9]. In addition to financial indicators of particular banks, three macroeconomic indicators (particularly the growth of the GDP, long-term interest rate and unemployment rate) were included among the independent variables to be better reflected the economic situation on the market. It's necessary to note that all the data were collected during the financial crisis during the years 2008 - 2010. Due to large sensitivity of the models on the input data, it's necessary to understand the limitations of the estimated models and reached results only on this phase of the market evolution.

TABLE II
BASICS CHARACTERISTIC OF THE EMPIRICAL SAMPLE

	number of observation	percent	cumulative percent
non-defaulted banks	137	45.97	45.97
defaulted banks	161	54.03	100
total	298		

Table II illustrates that the ratio between defaulted and non-defaulted banks in the input sample is roughly the same (i.e. certainly not empirical). This also implies that the resulting estimated models will not suitable for direct estimation of real PD, since they will be obviously biased. For the estimation of the real PD would be necessary to perform a calibration. However, this does not affect the conclusions reached in this paper.

B. Model Estimation

The results of the logit analysis specified in the previous methodology section are summarized in Table III, where the variables are as defined in Table X (Appendix A).

TABLE III
ESTIMATION RESULTS FOR LOGIT MODEL

Status	Coefficient	Std. Error	z	P>z
x17_GDP	-399.0495	182.8138	-2.18	0.029
x5_ROAA	-82.79004	36.47131	-2.27	0.023
x10_PL GL	111.3344	44.07752	2.53	0.012
x14_EQ TA	-237.7291	99.90404	-2.38	0.017
_cons	18.39238	8.6913	2.12	0.034

where x5, x10, x14 and x17 denotes ROAA (Return on Average Assets), PL GL (Problem Loans on Gross Loans), EQ TA (Shareholders' Equity / Total Assets) and GDP (year growth rate of GDP), respectively. These indicators were selected on the basis of stepwise method (forward selection followed by backward elimination) with significance levels 5% for adding variable into the model and 10% for removing variable from the model. The value of the Log likelihood is -7.7047 and the pseudo R2 is equal to 0.9637. The value of the likelihood-ratio test statistic (χ^2 distributed with 4 degrees of freedom) is 395.74 and $\Pr > \chi^2 = 0.0000$.

The results of the probit analysis are summarized in Table IV.

TABLE IV
ESTIMATION RESULTS FOR PROBIT MODEL

Status	Coefficient	Std. Error	z	P>z
x17_GDP	-225.2364	100.365	-2.24	0.025
x5_ROAA	-42.95094	17.47756	-2.46	0.014
x10_PL GL	58.81323	21.78227	2.70	0.007
x14_EQ TA	-129.9351	55.06857	-2.36	0.018
_cons	10.3022	4.944014	2.08	0.037

The value of the Log likelihood is -7.7896 and the pseudo R2 is equal to 0.9621. The value of the likelihood-ration test statistic is 395.56 and $\Pr > \chi^2 = 0.0000$.

Table V presents the results of the linear discriminant analysis.

TABLE V
ESTIMATION RESULTS FOR LINEAR DISCRIMINANT ANALYSIS

Status	Coefficient	Wilks' Lambda	F	Sig.
x17_GDP	34.635	0.481	315.428	0.000
x5_ROAA	6.967	0.609	187.829	0.003
x10_PL GL	-4.881	0.616	181.795	0.001
x14_EQ TA	19.615	0.520	269.615	0.009

The value of the test statistic for model is 360.356 (χ^2 distributed with 4 degrees of freedom) and $\Pr > \chi^2 = 0.0000$.

C. Verification and Comparison of the Estimated Models on the Control Sample

For verification of the success rate of the estimated models we will apply these models on the control sample of 100 American commercial banks. These banks were chosen randomly in such way that the control sample contained 50 non-default and 50 default banks. For the verification of the predicting abilities of the estimated models, they were applied on the data one year before decisive day. ROC analysis is used for measure of the model's quality prediction in this study, see Table VI for results. For more detailed description of ROC curve see [21].

TABLE VI
STATISTICAL DESCRIPTION OF THE AREA UNDER THE CURVE (AUC) FOR ESTIMATED MODELS

Test Results Variables	AUC	Std. Error	Asymptotic Sig. b
logit model	0.988	0.012	0.000
probit model	0.987	0.012	0.000
LDA	0.964	0.011	0.001

The asymptotic significance of each model is less than 0.05, so all are doing better than guessing. Nevertheless, from the results it is clear that logit model shows the best results in application on the control sample (regardless to the probit model) and so we can say this model is the most appropriate model for prediction of the banks default. On the other hand, the LDA model appears as the least appropriate for the prediction of the banks' failure. For the above reason, we will use for subsequent application the logit model estimated in Table III.

IV. ESTIMATION OF PD DISTRIBUTION

In this section we will proceed to the main task of the paper. First, it will be describe the data set – financial indicators. Then, we estimate their future marginal distribution and put them together by a given copula function and estimate the distribution of future PD as based on logit model estimated in previous section. Finally, knowing the PD distribution, we will focus on the tail of these distributions with a view to estimate the extreme value of the PD on the various levels of significance for every particular bank from the portfolio.

A. Data

We collected four financial indicators of three key Czech banks on the quarterly basis over the last thirteen years. In particular, ČSOB (Československá obchodní banka), KB (Komerční banka), and GE (GE Money Bank) were studied on the basis of financial indicators identified in Table III.

The indicators, which have been already identified as significant, are annual growth rate of GDP (GDP), Return on average assets (ROAA), Problem loans relative to gross loans (PL GL) and Shareholders' Equity to Total Assets (EQ TA). We can observe their evolution during the last thirteen years in Fig. 1.

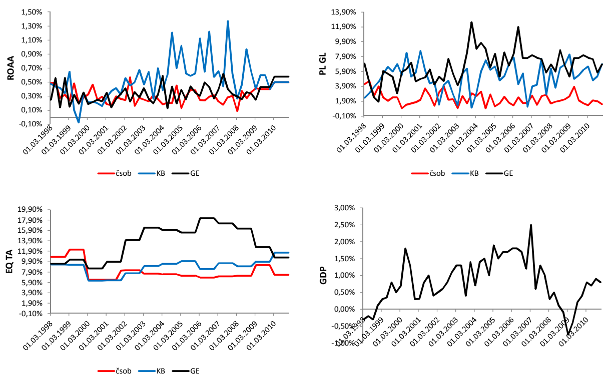

Fig. 1 Evolution of financial indicators over 1998 – 2010

We can observe that most of the indicators were very instable over the time. This is especially apparent in the case of KB and GE bank, while indicators of ČSOB seem to be relatively stable. The evolution of the GDP growth is well known, including recession around years 1998 and 2009. These facts also results into variability of PD over the time, see Fig. 2. According to the model, all analyzed banks had relatively close to default stage during 1998-1999 (PDs were 40%, except ČSOB) and during 2009 (their PDs were around 60%). This was caused by the recession stage of Czech economic. Nevertheless, these high PDs may also be consequence of the estimation of the model from the set of US banks (because of the lack of defaulted banks data for original sample in Central Europe). GE had a problem also over 2000 a 2001, which was probably caused by higher proportion of problem loans relative to gross loans and by lower proportion of equity on total assets.

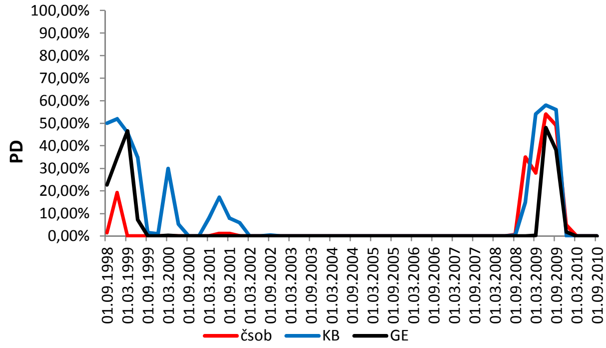

Fig. 2 Evolution of PD over 1998 – 2010

B. Marginal Distribution

In order to get the future distribution of particular financial indicators x_i, we will assume $n = 100000$ independent scenarios for *(i)* Geometric Brownian motion, $x_i \in N(\mu, \sigma^2)$, *(ii)* variance gamma model, $x_i \in VG(\theta, \sigma, \nu)$ and *(iii)* normal inverse Gaussian model, $x_i \in NIG(\theta, \sigma, \nu)$. It is generally recommended to use maximal likelihood approach for estimation of the model parameters. However, since our data set consists of only few values, we will follow the generalized method of moments. In this paper we are interested solely in the distribution of PD for the next time moment. We therefore sample just once from the estimated distribution to get one scenario, i.e. the future value of a given indicator x_i.

In the Table VII we can observe descriptive statistics of returns of modeled distributions for each particular indicator. It is very important to compare these characteristics with the empirical ones. In particular, we have to notice insufficient skewness and kurtosis within GBM. Gaussian process is not able to model skewness and kurtosis as compared to Lévy processes, which allow modeling of the higher moments of probability distribution.

TABLE VII
COMPARISON OF THE PROBABILITY DISTRIBUTION MOMENTS OF THE EMPIRICAL AND MODELED DISTRIBUTION

		ROAA				PL GL				EQ TA			
		empiric.	*VG*	*NIG*	*GBM*	*empiric.*	*VG*	*NIG*	*GBM*	*empiric.*	*VG*	*NIG*	*GBM*
CSOB	MEAN	***0.0049***	0.0040	0.0042	0.0049	***-0.0265***	-0.0280	-0.0263	-0.0244	***-0.0064***	-0.0061	-0.0066	-0.0060
	VAR	***0.2562***	0.2586	0.2597	0.2552	***0.2542***	0.2506	0.2534	0.2567	***0.0118***	0.0114	0.0116	0.0119
	STDEV	***0.5061***	0.5085	0.5096	0.5052	***0.5042***	0.5006	0.5034	0.5066	***0.1085***	0.1070	0.1078	0.1090
	SKEW	***-0.2810***	-0.2617	-0.3572	-0.0021	***-0.0574***	-0.0978	-0.0341	-0.0067	***-3.6901***	-3.6335	-3.6135	-0.0067
	KURT	***4.5224***	4.5536	4.5475	3.0283	***2.7235***	3.0893	2.9491	2.9811	***24.8456***	23.6038	25.3044	2.9611
KB	MEAN	***-0.0167***	-0.0104	-0.0166	-0.0199	***0.0049***	0.0031	0.0048	0.0046	***0.0045***	0.0049	0.0038	0.0035
	VAR	***0.3996***	0.4050	0.4057	0.4013	***0.3906***	0.3852	0.3950	0.3953	***0.0059***	0.0059	0.0062	0.0061
	STDEV	***0.6322***	0.6364	0.6370	0.6335	***0.6250***	0.6206	0.6285	0.6287	***0.0767***	0.0771	0.0789	0.0778
	SKEW	***-0.3192***	-0.3349	-0.2822	0.0057	***-0.5024***	-0.4386	-0.4842	0.0011	***-2.2180***	-2.1100	-2.1065	-0.0144
	KURT	***4.4106***	4.7158	4.6250	2.9911	***4.3197***	4.1315	4.7742	2.9194	***15.4215***	15.8136	13.9136	3.0086
GE	MEAN	***-0.0016***	-0.0021	-0.0025	-0.0062	***0.0046***	0.0041	0.0042	0.0034	***0.0045***	0.0045	0.0044	0.0043
	VAR	***0.3686***	0.3682	0.3667	0.3762	***0.1190***	0.1206	0.1169	0.1183	***0.0066***	0.0070	0.0066	0.0066
	STDEV	***0.6071***	0.6068	0.6056	0.6133	***0.3449***	0.3472	0.3419	0.3439	***0.0813***	0.0834	0.0810	0.0815
	SKEW	***-0.3564***	-0.3477	-0.3615	0.0498	***0.7187***	0.7447	0.6073	0.0015	***0.9433***	1.0049	0.8507	0.0015
	KURT	***3.3182***	3.2605	3.2869	2.9760	***4.2595***	5.0449	4.2559	2.9810	***10.7351***	11.4827	10.8115	2.9510

From observable results we have clearly stated that the Lévy process models empirical distribution better than the traditional GBM, while the differences between the VG and NIG are negligible.

C. Linear Dependency among Variables (Gaussian Copula)

In order to estimate the probability of default of any of the banks, the dependency among particular indicators must be taken into account. We consider the Gaussian copula function, i.e. the inputs are marginal distributions and empirical correlation matrix. Applying this copula on the particular marginal distribution we get the progress of the indicators with required dependencies. Afterwards by applying the simulated indicators into the estimated logit model we easily get the probability distribution of PDs of analyzed banks. The resulting descriptive statistics is depicted in Table VIII.

TABLE VIII
ESTIMATION OF THE FUTURE MEANS VALUES AND STDEV OF PDS

	E (PD)			stdev (PD)		
	VG	*NIG*	*GBM*	*VG*	*NIG*	*GBM*
ČSOB	***4.83%***	***4.70%***	***4.28%***	0.1481	0.1467	0.1149
KB	***4.02%***	***4.51%***	***5.38%***	0.1703	0.1815	0.2000
GE	***2.98%***	***2.80%***	***2.74%***	0.1392	0.1331	0.1247

It is obvious from results that risk of default of the analyzed banks is on a relatively low level and all these three banks show a financial safety for all the considered models (see expected PD from 2.74% to 5.38%). Taking into account the conclusion from previous section, we can say that utilization of the GBM for default modeling shows a slight undervaluation of the default risk within analyzed banks. GE appears to be relatively most stable, which may be caused by a high proportion of equity on total assets. Here is difference between results reached on the same portfolio in [22], where it was working with the model, in which macroeconomic indicators were not included.

D. Analysis of Extreme Quantiles within Modeled Distributions

Another interesting issue is, what is the maximum loss on the probability level 90%, 95%, 99% or even 99.99% for every from analyzed banks? To answer this question, we use the cumulative distribution function (CDF) of modeled PDs. Because we are now interested in PD values close to 100%, we will work with the function (1-CDF) for a better interpretation. The results are apparent in Table IX.

TABLE IX
ESTIMATED VALUES OF PD ON THE VARIOUS LEVEL OF PROBABILITY

	0.001			0.01			0.05			0.1		
	VG	NIG	GBM	VG	NIG	GBM	VG	NIG	GBM	VG	NIG	GBM
ČSOB	99.81%	99.72%	86.71%	86.97%	86.01%	59.49%	31.00%	30.10%	27.89%	11.87%	11.31%	13.21%
KB	100.00%	100.00%	100.00%	99.83%	99.96%	99.98%	21.49%	22.68%	29.35%	1.38%	1.78%	3.29%
GE	100.00%	100.00%	99.80%	96.10%	95.47%	84.43%	10.03%	9.28%	10.95%	1.68%	1.58%	1.96%

For example, we can say that with probability 5% it will be the future PD of ČSOB greater than 31.00% within VG model. From the results it is clear, that GBM undervalues the PD values especially for the low quantiles (see difference on the 0.01 quantile within ČSOB for e.g. VG and GBM, 86.97% and 59.49%), which can be very critical in risk management. The PD values for VG and NIG model are again similar. Generally we can say that it is much more convenient to use more sophisticated models such as Lévy processes to model probability of PD distribution, at least in time of financial crisis.

V. CONCLUSION

The main reason for estimating the probability of default is use in risk management, valuation of the credit derivatives, estimation of the creditworthiness of borrowers and estimation of a bank's capital adequacy. Incorrect estimation of the PD can lead to an incorrect valuation of risk and financial problems of a company. In this paper, we attempted to estimate the distributions of PD for the next stages of three Czech banks assuming that the PD was determined by the evolution of financial indicators in estimated credit-scoring models.

In the first section of the application portion of the paper, we estimated three credit-scoring models (logit model, probit model and linear discriminant analysis model) from a set of 298 US commercial banks. After verification and comparison with the control sample, the logit model was chosen as the most appropriate model for the prediction of commercial banks' PDs in stages of market depression. Similar statistical significance was also produced by the probit model. LDA appeared to be the least appropriate model. It is important to note the assumptions of the estimated models, and hence, the possibilities and limitations of their utilization. All three prediction models were estimated from a dataset obtained during a time of financial crisis; therefore, the utilization of these models is also restricted to this particular phase of market evolution. Another limitation is attributable to the lag between the calculation of relevant indicators and the date of each bank's default. From this perspective, the estimated models are one-period prediction models, with a rather short time of prediction of default (1-2 years). Taking into account these limitations and the obtained results, the estimated logit model was chosen as the most appropriate model for predicting a bank's default.

The estimated logit model was then applied to a portfolio of three key Czech banks with the intention to estimate their future PD probability distributions. We assumed three candidates for the evolution of financial indicators, a Gaussian distribution, VG distribution and NIG distribution. It was found that the VG and NIG processes fit the empirical statistics better than the traditional GBM.

From the results, according to the estimated model, all of the analysed Czech banks appear to be relatively financially stable, with a slight undervaluation of PD using the GBM. Finally, analysis of the extreme quantiles of modelled distributions was performed. Based on the results, we can state that in terms of risk management, it seems safer to use more sophisticated models, such as Lévy processes, to model PD probability distributions to avoid a possible underestimation of the risk, at least in times of financial crisis.

APPENDIX A

TABLE X
DESCRIPTION OF FINANCIAL INDICATORS

Indicator	Description of the indicator	Indicator's group
LTA	Logarithm of total assets	Size
YAEA	Interest income / Average Interest Earning Assets	Profitability
CIBL	Interest Expense / Average Interest Bearing Liabilities	Profitability
NIM	Net Interest Margin	Profitability
ROAA	Return on Average Assets	Profitability
ROAE	Return on Equity	Profitability
IE	Interest Expense / Interest Income	Profitability
CIR	Cost to Income Ratio	Efficiency
PE OI	Personnel Expenses / Operating Income	Efficiency
PL GL	Problem Loans / Gross Loans	Assets Quality
LLR GL	Loan Loss Reserve / Gross Loans	Assets Quality
PL EQ LLR	Problem Loans / (Shareholders' Equity + Loan Loss Reserve)	Assets Quality
T1	Tier 1 ratio	Capital adequacy
EQ TA	Shareholders' Equity / Total Assets	Capital adequacy
CAR	Capital Adequacy	Capital adequacy
D EQ	Total Deposits / Shareholders' Equity	Capital adequacy

REFERENCES

[1] W. Beaver, "Financial ratios as predictors of failures. Empirical Research in Accounting: Selected Studies – 1966", supplement to *Journal of Accounting Research*, 4, 1966, pp. 71-111.
[2] E. I. Altman, "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy". *Journal of Finance*, September, 1968, pp. 589-609.
[3] D. A. McFadden, "Comment on discriminant analysis versus logit analysis". *Annals of Economic and Social Measurement*, 1976, pp. 511–523.
[4] A. Santomero and J. D. Visno, "Estimating the probability of failure for commercial banks and the banking system". *Journal of Banking and Finance 1*, 1977, pp. 185–215.
[5] J. A. Ohlson, "Financial ratios and the probabilistic prediction of bankruptcy". *Journal of Accounting Research 18*, 1980, pp. 109–131.
[6] M. E. Zmijeski, "Methodological issues related to the estimation of financial mistress prediction models". *Journal of Accounting Research 22*, 1984, pp. 59–82.
[7] M. Queen, and R. Roll, "Firm mortality: using market indicators to predict survival". *Financial Analysts Journal 3*, 1987, pp. 9–26.
[8] T. Shumway, "Forecasting bankruptcy more accurately: a simple hazard rate model". *Journal of Business 74*, 2001, pp. 101–124.
[9] A. Peretsky and A. Karminsky, "Models for Moody's bank rating", *BOFIT Discussion Papers 17*, 2008.
[10] R. Cont and P. Tankov, *Financial Modelling with Jump Processes*. Chapman & Hall press, 2004.
[11] W. Schoutens, *Lévy Processes in Finance: Pricing Financial Derivates*. Wiley, 2003.
[12] W. Green, *Econometric Analysis*. 6th ed. Prentice Hall, 2008.
[13] B. Engelmann and R. Rauhmeier, *The Basel II Risk Parameters*. Springer Verlag, 2006.
[14] A. Resti and A. Sironi, *Risk Management and Shareholders' Value in Banking: From Risk Measurement Models to Capital Allocation Policies*. Wiley, 2007.
[15] A. Kyprianou, W. Schoutens and P. Wilmott, *Exotic Options Pricing and Advanced Lévy Models*. Chichester: Wiley, 2005.
[16] D. Applebaum, *Lévy Processes and Stochastic Calculus*. Cambridge University Press, 2004.
[17] O.E. Barndorff-Nielsen, T. Mikosch and S.I. Resnick, *Lévy processes. Theory and Applications*. Boston: Birkhäuser. 2001.
[18] R.B. Nelsen, *An Introduction to Copulas*. 2nd ed. Springer, 2006.
[19] J. Rank, *Copulas. From theory to application in finance.* Risk books, 2007.
[20] G.¨Cherubini, E. Luciano and W. Vecchiato, *Copula Methods in Finance*. Wiley, 2004.
[21] J. A. Hanley and B. J. McNeil, "The Meaning and Use of the Area under a Receiver Operating Characteristics (ROC) Curve". *Radiology*, 1982, pp. 561-557.
[22] P. Gurný and T. Tichý, "Estimation of PD implied by scoring model for Czech banks". *Proceedings of the 12th International Conference on Finance and Banking*. Silesian University in Opava, 2010, pp. 79-88.