Modeling a distribution of mortgage credit losses

Petr Gapko[1], Martin Šmíd[2]

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. 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.

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

JEL Classification: G21

1 Introduction

In our paper, we will focus on credit risk quantification methodology. The minimum standards for credit risk quantification are often heavily regulated. The current recommended system of financial regulation is 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). The regulation is designed in the way that banks are required to cover with a stock of capital a certain quantile loss from a certain risk (i.e. from a risk that counterparty wouldn’t pay back its liabilities).

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).

In this paper, we will introduce a new approach to quantifying credit risk with a focus on a mortgage portfolio which can be classed with the Value-at-Risk models. Our approach is different from the regulatory method in the assumption of the loss distribution. In the general version of our model, we assume that risk factors that drive losses can be distributed not only standard normal assumed by the regulatory framework 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). To test our model, we will demonstrate its goodness-of-fit on a nationwide mortgage portfolio. Moreover, we will compare our results with the regulatory approach.

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 regulatory approach. At the end we summarize our findings and offer recommendations for further research.

2 Credit risk measurement methodology

The Basel II allows two possible quantification methods for the credit risk: the “Standardized Approach” (STA) and the “Internal Rating Based Approach” (IRB). The IRB approach is more advanced than STA and is based on

[1] Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Institute of Economic Studies, Faculty of Social Sciences, Charles University, Prague
[2] Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic

a Vasicek-Merton credit risk model (Vasicek, 1987) 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 PD 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.

PD is usually obtained by one of the following methods: from a scoring model (Moody's KMV, JP Morgan CreditMetrics, etc.), from a Merton-based distance-to-default model (mainly used for commercial loans; Merton, 1973 and 1974) or as a long-term stable average of past 90+ delinquencies.

Two basic measures of credit risk are expected and unexpected losses. 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 the loss distribution. The expected (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[3] and EL is the abbreviation for "Expected Loss."

For calculations of unexpected losses, it is usually assumed that losses follow a certain distribution in time. The regulatory IRB framework uses for this purpose a mix of distributions of two risk factors, one individual for each borrower and one common for all borrowers. Both factors are assumed to follow a standard normal distribution and to be correlated with a certain assigned value of the correlation coefficient.

3 Our approach

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

$$\log A_{i,1} = \log A_{i,0} + \eta + \gamma X_i \quad (2)$$

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).

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

$$\log A_{i,t} = \log A_{i,t-1} + Y_t + Z_{i,t} \quad (3)$$

where $A_{i,t}$ is the wealth of the i -th borrower at the time $t \in \mathbf{N}_0$, $Z_{i,t}$ is a random variable specific to the borrower and Y_t is the common factor following a general (adapted) stochastic process (such an assumption makes sense, for instance, if Y_t models a macroeconomic variable or a price on a capital market).

We assume all $(Z_{i,t})_{i \leq n, t \in \mathbf{N}}$ to be mutually independent and idependent of $(Y_t)_{t \in \mathbf{N}}$, such that all $Z_{i,t}$, $i \leq n$, $t \in \mathbf{N}$, are identically distributied with $\mathbf{E} Z_{1,1} = 0$, $\mathrm{var}(Z_{1,1}) = \sigma$, $\sigma > 0$, 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).

[3] Exposure-at-default is a Basel II expression for the amount that is (at the moment of the calculation) exposed to default.

Percentage loss in the generalized model

Denote $\overline{Y}_{t-1} = (Y_\tau)_{\tau \le t-1}$ the history of the common factor up to the time $t-1$. Analogously to the original model, the conditional probability of the banrkupcy of the i -th borrower at the time t given $\overline{Y}_t$ equals to

$$\mathbf{P}\left(A_{i,t} \le B_{i,t} \middle| \overline{Y}_t\right) = \mathbb{P}\left(Z_t \le c_{i,t} - Y_t \middle| \overline{Y}_t\right) = \Psi(c_{i,t} - Y_t),$$

$$c_{i,t} = \log B_{i,t} - \log A_{i,t-1}$$

where $B_{i,t}$ are the borrower's debts.

Denote L_t the percentage loss of all the portfolio of the borrowers 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 with homogeneous (in the sense that $c_{i,t} = c_t$, $i \le n$) borrowers that

$$L_t \doteq \Psi(c_t - Y_t), \qquad t \in \mathbb{N},$$

which furter implies that

$$\varepsilon_t = Y_t - Y_{t-1} \doteq \Psi^{-1}(L_t) - \Psi^{-1}(L_{t-1}) \quad (4)$$

hence

$$L_t \doteq \Psi(\Psi^{-1}(L_{t-1}) - \varepsilon_t) \qquad (5)$$

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 to be normally distributed and the common factor to be multiplicatively defined by

$$Y_t = (1 + \varsigma_t) Y_{t-1}$$

where $\varsigma_1, \varsigma_2, \cdots$ are i.i.d. (possibly non-normal) variables (note that our choice of the dynamics corresponds to the assumption of i.i.d. returns if the common factor stands for prices of a financial instrument). 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. By (4), we get that

$$\varsigma_t = \frac{Y_t - Y_{t-1}}{Y_{t-1}} \doteq \dot{\varsigma}_t \qquad \dot{\varsigma}_t = \frac{\psi^{-1}(L_t) - \psi^{-1}(L_{t-1})}{\psi^{-1}(L_{t-1}) - Y_0} \qquad (6)$$

which allows us to use the sample $\dot{\varsigma}_1, \dot{\varsigma}_2, \cdots$ to estimate parameters of ς_1 (and consequently the distribution of the losses).

As it was already said, we assume the distribution of ς_1 to be generalized hyperbolic and we use the ML estimation to get its parameters. Moreover, we compare our choice to several other classes of distributions.

4 Data and results

4.1 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). To verify that 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[4] The rate used is the best substitute for losses that banks faced from their mortgage portfolios, relaxing the LGD variable. 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.

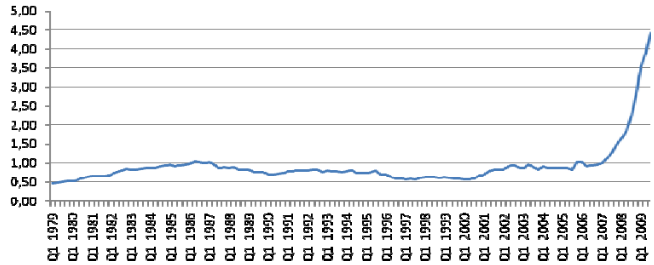

Figure 1 Development of US 90+ delinquency rate

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, gamma, extreme value, beta and the class of generalized hyperbolic distributions. The dataset used for distribution fitting was constructed from the above described data by using formula (6) from the previous part. 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 the fitting procedure we used the R package "ghyp". We used the chi-square goodness-of-fit test. In general, only five from the considered distributions were not rejected to describe the dataset based on the chi-square statistic (on a 95% level).

Beside the chi-square statistic, we used a different statistic to compare all the tested distributions and sort them by their values: the Anderson-Darling statistic (Anderson & Darling, 1952) and the Wasserstein distance. All the statistics are measures of the distance between the original sample and the tested distribution. The following table summarizes our results. It includes distributions that were not rejected based on the chi-square statistic. The table is sorted by the Anderson-Darling statistic:

Distribution	Wasserstein metric	Anderson Darling	Chi-square statistic	P-value of chi-square
Generalized hyperbolic	0.0080	0.193	4.19	0.96
LogLogistic	0.0089	0.278	5.05	0.93
Logistic	0.0099	0.309	6.64	0.83
Inverse Gaussian	0.0141	0.849	17.35	0.10
Normal	0.0156	0.896	15.96	0.14

Table 1 Comparison of goodness-of-fit of tested distributions

According to the Table 1, both the chi-square and Anderson-Darling statistics show that the generalized hyperbolic distribution (GHD) has the best fit. Our calculations show that the class of generalized hyperbolic distributions is able to describe the behavior of delinquencies much better than the other distributions widely used for risk assessment (normal, lognormal, logistic, gamma), 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[5]).

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

In this section, we compare the capital requirement calculated by the IRB regulatory approach (assuming that both risk factors are driven by the standard normal distribution) and our dynamic framework with a generalized

[4] 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.).

[5] The correlation 15% is a benchmark set for the mortgage exposures in the Basel II framework and thus we will use this benchmark for our computations.

hyperbolic distribution. 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 faced the following problem: we estimated the generalized hyperbolic distributions on quarterly observations and thus we needed to transform the quarterly changes obtained to yearly figures. In particular, to forecast a yearly loss, we may repeatedly use (4) to get

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

which leads to 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 standard normal and the normal distributions were 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%	19.7605%	22.9078%

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 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 much higher quantile losses in both cases. However, heavy tails of the GDH distribution further evoke higher quantile losses, which at the end lead to a higher capital requirement.

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 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 uses the standard normal distribution as an underlying distribution that drives risk factors for credit risk.

Our results show that the mix of standard normal distributions used in the Basel II regulatory framework is, at the 99.9% level of probability, underestimating the potential unexpected loss on the one-year horizon. Therefore, introducing the dynamics may lead to a better capturing of tail losses. Within our dynamic model we have further compared the predictions based on the normal and the class of generalized hyperbolic distributions. Our results show that the heavy-tailed generalized hyperbolic distribution predicts the biggest tail loss.

Acknowledgements

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

References

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

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

[3] 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.

[4] 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

[5] Barndorff-Nielsen, O. E., Blæsild, P., & Jensen, J. L. (1985). *The Fascination of Sand.* A Celebration of Statistics, 57–87.

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

[7] Eberlein, E. (2001). *Application of Generalized Hyperbolic Lévy Motions to Finance.* Lévy Processes: Theory and Applications, 319–337.

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

[9] 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.

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

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

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

[13] JP Morgan. (n.d.). *CreditMetrics.*

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

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

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

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

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

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

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

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

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