A Note on the Vasicek's Model with the Logistic Distribution[1]

*Jiří WITZANY**

Abstract

The paper argues that it would be natural to replace the standard normal distribution function by the logistic function in the regulatory Basel II (Vasicek's) formula. In fact, such a model would be consistent with the standard logistic regression probability of default modeling approach. An empirical study based on the US commercial bank's loan historical delinquency rates from the period 1985 – 2012 re-estimates the default correlations and unexpected losses for the normal and logistic distribution models. The results indicate that the capital requirements could be up to 100% higher if the normal Vasicek's model was replaced by the logistic one.

Keywords: *credit risk, Basel II regulation, default rates*

JEL Classification: G20, G28, C51

1. Introduction

The stability of the global financial system depends on a good performance of the Basel II formula calculating the required regulatory capital (BCBS, 2006). The formula that is based on the seminal paper of Vasicek (1987) has been put into question in the context of the global financial crisis. The aim of this paper is to revisit the formula and propose an alternative based on logistic probability distribution.

The recently approved Basel III (BCBS, 2010) regulatory package tries to learn from the crisis. It aims to improve quality of the regulatory capital, to deal with the pro-cyclicality issue, to extend the risk coverage in the area of trading

* Jiří WITZANY, University of Economics in Prague, Faculty of Finance and Accounting, Department of Banking and Insurance, Winston Churchill Sq. 4, Prague 3, Czech Republic; e-mail: jiri.witzany@vse.cz

[1] This research has been supported by the Czech Science Foundation Grant P402/12/G097 *Dynamical Models in Economics*.

book products and to set global liquidity standards (see Šútorová and Teplý, 2013 or, Teplý, Vrábel and Černohorská, 2012). However, it does not change at all the key formula which is used to calculate the unexpected losses, i.e., the required capital. Our analysis shows that the formula does not have to cover extreme losses sufficiently well (see also Lall, 2012) and a stronger capital base could be achieved by its simple modification that is in fact consistent with bank credit risk modeling standards.

The capital requirement of a bank in the Internal Rating Based (IRB) approach is calculated as the unexpected loss (*UL*) less the expected loss (*EL*)

$$C = UL - EL = (UDR - PD) \times LGD \times EAD \tag{1}$$

decomposed into the unexpected default rate (*UDR*), probability of default (*PD*), loss given default (*LGD*), and the exposure at default (*EAD*) on the level of each individual exposure. The portfolio invariant capital requirements (see Gordy, 2003) are summed up across the bank's portfolio. The probability of default parameter is an output of an internal rating model. The loss given default and exposure at default parameters are set by the regulator in the Foundation (IRBF) approach, or estimated by an internal model of the bank in the Advanced (IRBA) approach. Our main focus is the unexpected default rate (*UDR*) that is calculated as a regulatory function of *PD*, asset correlation ρ (set by the regulation depending on the asset class and *PD*), and the probability level α (set at 0.999):

$$UDR_{\alpha}^{N} = \Phi\left(\frac{\Phi^{-1}(PD) + \sqrt{\rho} \cdot \Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right) \tag{2}$$

The formula, in detail explained in Section 0, is based on the assumption that the event of default is driven by a normally distributed variable. Moreover, the account level risk driving factor Y_i is decomposed into a single normally distributed systematic factor X and an independent normally distributed idiosyncratic factor

$$Y_i = \sqrt{\rho} X + \sqrt{1-\rho}\,\zeta_i \tag{3}$$

Since the idiosyncratic factors diversify away in a large and sufficiently granular portfolio, the unexpected losses are driven only by the systematic factor (Gordy, 2003). The choice of the normal distribution is certainly natural. In addition, its technical advantage is that the sum of two independent normal variables in (3) is again normally distributed. In fact, the model means that the correlation between defaults of different receivables is captured by a Gaussian copula (see, e.g., Cherubini, Luciano and Vecchiato, 2004).

On the other hand, the banking industry standard approach to internal ratings uses the technique of logistic regression based on the assumption that the event of default is driven by a logistically distributed variable. The logistic (or logit) regression can be compared to the probit binary choice regression where the events are driven by a normal variable. It is recognized that in case of probability of default modeling the logistic regression performs better, in terms of estimation results and in particular because of fatter tails of the logistic distribution. Gapko and Šmíd (2012) have investigated a complex dynamical version of the Vasicek's model based on the class of generalized hyperbolic distributions, however, according to our knowledge, a simple analogy of the Vasicek's formula using the logistic distribution has not been considered and compared to the normal distribution based formula (2). We argue that it is possible to replace simply the cumulative distribution function $\Phi(x)$ in (2) by the logistic function

$$\Lambda(x) = \frac{1}{1+e^{-x}}$$

The only technical complication is that the weighted sum of two independent logistically distributed variables is not exactly logistically distributed, but we will show that the empirical difference is negligible. The choice of the logistic distribution means that the correlation of defaults is captured in a way that should perform better in extreme situations. It is not just another, single systematic factor copula model, but in our view the most natural model consistent with the industry standard of probability of default modeling.

In order to compare performance of the two formulas one could simply compare the outputs of (2) with the normal and logistic distributions for different segments and the probability of default values given the regulatory correlation and probability level. However, the correlation has been presumably calibrated with respect to the normal distribution and in order to compare meaningfully the two formulas we should recalibrate the correlation for the logistic distribution model in a way similar to the regulatory approach. Unfortunately, there is no public document that would disclose exactly the calculations underlying the regulatory asset correlation values. BCBS (2005) only indicates that the correlations have been chosen in order to fit historical default rate series. Nagpal and Bahar (2001) have used a non-parametric approach suggested by Lucas (1995) to estimate the correlation of US corporate defaults. Calem and Follain (2003, p. 27.) in a Federal Reserve System study on the asset correlation for the class of single-family mortgages propose to apply advanced portfolio unexpected loss models and "reverse-engineer" the asset correlation in (2). Many other studies like Hamerle, Liebig and Rösch (2003); Witzany (2011), or Crook and Belloti (2012) use a maximum likelihood estimation based on historical default data as outlined by BCBS (2005). Section 0 presents an empirical study on the US Federal Reserve statistical dataset of delinquency rates at commercial banks. We will use a simplified maximum likelihood estimation approach in order to calibrate the Vasicek's model with the normal and logistic distributions. The formulas, average PDs, and the estimated correlations will be used to calculate the unexpected default rates and compare the outcomes of the two models. Section 0 concludes, summarizes the results, and discusses further research opportunities.

2. Vasicek's Model with the Normal and Logistic Distribution

The Vasicek's formula (2) is obtained from a simple Merton-like structural model. The model assumes that there is a credit quality variable y_t developing over time and triggering default if it drops at or below certain level b. In case of corporate debtors the variable is interpreted as the asset value (or its relative change) and b as the indebtedness level. In case of private individuals the variable can be rather interpreted as a credit score capturing general repayment capacity of the debtor and b as a critical level where the debtor is not able to pay back his/her obligations any more. In both cases y_t is a latent, i.e. unobserved variable. In order to model the event of default at time t = 1we just need to model the probability distribution of $Y = y_t$. By definition, default happens, if and only if $Y \le b$. Since Y and b can be rescaled, we may assume without loss of generality that the mean of Y is 0 and that its variance is 1 (or another given constant). Moreover, we will assume that Y has a particular known distribution. For example, if Y has the standard normal distribution, and if we know that the probability of default is PD, then $PD = \Pr[Y \le b] = \Phi(b)$, and so $b = \Phi^{-1}(PD)$.

The classical Vasicek's model works with the normal distribution. Its goal is to find the distribution of future default rates (e.g. at t = 1) on a large (theoretically infinite) portfolio of receivables where the events of default might be partially correlated. It captures default correlation by decomposing the credit quality variable for each individual debtor i into a systematic part and an idiosyncratic part $Y_i = \sqrt{\rho}V + \sqrt{1-\rho}\zeta_i$, where V and all ζ_i 's are mutually independent variables with the standard normal distribution, and where $\rho \in (0,1)$ is a constant. Since the idiosyncratic factors ζ_i are independent and diversify away in a large (asymptotic) portfolio, the future default rate depends only on the systematic factor V:

$$\begin{aligned} DR(V) &= \Pr[Y_i \le b \mid V] = \Pr\left[\sqrt{\rho}V + \sqrt{1-\rho}\zeta_i \le \Phi^{-1}(PD) \mid V\right] = \\ &= \Pr\left[\zeta_i \le \frac{\Phi^{-1}(PD) - \sqrt{\rho}V}{\sqrt{1-\rho}} \mid V\right] = \Phi\left(\frac{\Phi^{-1}(PD) - \sqrt{\rho}V}{\sqrt{1-\rho}}\right) \end{aligned} \tag{4}$$

As the default rate increases with decreasing V, we can determine its quantiles simply plugging in the corresponding quantiles of V. The critical default rate that can be realized on a probability level $1-\alpha$, e.g. for $\alpha = 0.999$, is simply given by

$$V = \Phi^{-1}(1-\alpha) = -\Phi^{-1}(\alpha)$$

yielding the Vasicek's formula (2).

The probability of default *PD* is one of the key inputs of the Vasicek's formula. In a standard banking credit rating model it is based on a credit scoring (regression) function that uses a vector of known explanatory variables **x**. The scoring model can be formulated in the same framework as the Vasicek's model: the event of default (at $t = 1$) is modeled by a latent variable Y_i that is decomposed into a known (or explained part) and an unknown part (reflecting an error and a future change of the credit quality variable), $Y_i = \boldsymbol{\beta}'\mathbf{x}_i + \epsilon_i$. In this case, we can assume without loss of generality that $b = 0$ since the default threshold can be incorporated into the constant term of $\boldsymbol{\beta}'\mathbf{x}_i$. Consequently, the probability of default is:

$$\Pr[Y_i \leq 0] = \Pr\left[\boldsymbol{\beta}'\mathbf{x}_i + \epsilon_i \leq 0\right] = \Pr\left[\epsilon_i \leq -\boldsymbol{\beta}'\mathbf{x}_i\right] = F\left(-\boldsymbol{\beta}'\mathbf{x}_i\right) \tag{5}$$

where F is the distribution function of ϵ_i. If F is a known distribution function independent on i then the coefficients $\boldsymbol{\beta}$ can be estimated maximizing the likelihood. I.e., given a set of historical observed default indicators $d_i \in \{0,1\}$, for $i = 1,...n$ (where $d_i = 1$ codes the event of default), and corresponding vectors $\mathbf{x}_i$ of explanatory variables, we maximize the likelihood function:

$$L = \prod_{i=1}^{n} F\left(-\boldsymbol{\beta}'\mathbf{x}_i\right)^{d_i} \left(1 - F\left(-\boldsymbol{\beta}'\mathbf{x}_i\right)\right)^{1-d_i}$$

If $F = \Phi$ is the normal cdf, then the model is known as the probit binary-response model (Greene, 2003, p. 1026). However, in case of credit scoring, researchers as well as practitioners prefer the logit model where $F = \Lambda$ is the logistic distribution. The model is considered to be more efficient, better analytically tractable, and compatible with the useful concepts of log-odds score and the Weight of Evidence (see, e.g., Witzany, 2010, p. 212).

The standard logistic distribution is defined by the cumulative distribution function

$$\Lambda(x) = \frac{1}{1+e^{-x}}$$

known as the *logistic function*, and equivalently by the probability density function

$$\lambda(x) = \Lambda'(x) = \frac{e^{-x}}{\left(1+e^{-x}\right)^2}$$

The mean of a variable with the logistic distribution is 0 and its variance is $\pi^2/3$. Note that the inverse of the logistic function assigning quantiles to given probabilities is the *logit function*:

$$\Lambda^{-1}(p) = \ln\left(\frac{p}{1-p}\right)$$

The general logistic distribution with mean μ and variance $s^2\pi^2/3$ can be defined by the following cdf:

$$\Lambda(x;\mu,s) = \frac{1}{1+e^{-(x-\mu)/s}}$$

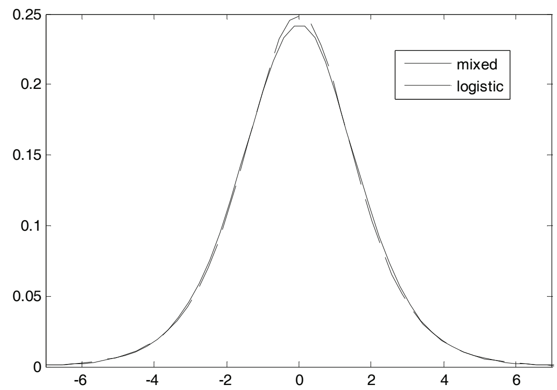
If the first two moments of a normal distribution, logistic distribution, and a t-distribution are matched, then it is easy to verify that the logistic and normal distributions diverge for quantiles beyond two standard deviations from the mean, with the t-distribution being much closer to the logistic one. Therefore, the logistic distribution has heavier tails and is more robust to inaccuracies in the underlying model or to errors in the data.

Now, let us revisit the Vasicek's model. Since the latent credit variable Y_i is assumed to be logistically distributed in the standard PD estimation approaches, we should also keep this assumption consistently in the Vasicek's model whose aim is to estimate unexpected default rate. However, the technical disadvantage of the logistic distribution is that, if we assume $Y_i = \sqrt{\rho}V + \sqrt{1-\rho}\zeta_i$ is a weighted sum of two independent logistically distributed variables V and ζ_i, then the result is not exactly logistically distributed variable. The cumulative distribution function of Y_i can be written as the double integral

$$\Lambda_\rho(x) = \iint\limits_{\sqrt{\rho}v+\sqrt{1-\rho}z\le x} \lambda(v)\lambda(z)dvdz = \iint\limits_{\sqrt{\rho}v+\sqrt{1-\rho}z\le x} \frac{e^{-v-z}}{\left(1+e^{-v}\right)^2\left(1+e^{-z}\right)^2}dvdz \qquad (6)$$

which unfortunately does not yield the logistic function again. A possible approach to approximate is a mixture of normal distributions (Stefanski, 1991). However, Figure 1 indicates that the differences between the logistic and mixed logistic distribution are almost negligible for a reasonably low value of ρ.

F i g u r e 1
Comparison of the Logistic and Mixed Logistic Distributions ($\rho = 0.1$)



Source: Own calculations.

Table 1 lists 95%, 99%, and 99.9% quantiles of the normal, logistic, and mixed logistic distributions (6) for several sample correlations. The differences between the logistic and mixed logistic distributions are more significant only for $\alpha = 99.9\%$ and for higher correlation values, yet still not large when compared to the difference between the normal and logistic distributions.

T a b l e 1
Comparison of Normal, Logistic, and Mixed Logistic Distribution Quantiles

α	**Normal**	**Logistic**	**Mixed log., ρ = 0.05**	**Mixed log., ρ = 0.10**	**Mixed log., ρ = 0.15**	**Mixed log., ρ = 0.20**	**Mixed log., ρ = 0.30**
0.95	2.983	2.943	2.943	2.944	2.945	2.951	2.956
0.99	4.219	4.600	4.554	4.519	4.495	4.471	4.441
0.999	5.605	6.933	6.842	6.759	6.678	6.594	6.500

Source: Own calculations.

Therefore, let us assume that the credit variable $Y_i = \sqrt{\rho}V + \sqrt{1-\rho}\zeta_i$ is decomposed into a systematic and idiosyncratic part where V and all ζ_i are mutually independent variables with the logistic distribution and $\rho \in (0,1)$ being a constant. Consequently, Y_i has the "almost-logistic" distribution given by (6). Repeating the argument (4) we obtain the following formula for the default rate on the confidence level α:

$$UDR_\alpha^{L0} = \Pr[Y_i \le b \,|\, V = \Lambda^{-1}(1-\alpha)] = \Pr\left[\sqrt{\rho}V + \sqrt{1-\rho}\zeta_i \le \Lambda_\rho^{-1}(PD) \,|\, V = \Lambda^{-1}(1-\alpha)\right] =$$
$$= \Pr\left[\zeta_i \le \frac{\Lambda_\rho^{-1}(PD) + \sqrt{\rho}\Lambda^{-1}(\alpha)}{\sqrt{1-\rho}}\right] = \Lambda\left(\frac{\Lambda_\rho^{-1}(PD) + \sqrt{\rho}\Lambda^{-1}(\alpha)}{\sqrt{1-\rho}}\right) \quad (7)$$

Since $\Lambda_\rho^{-1}(PD) \cong \Lambda^{-1}(PD)$ we propose to use the formula

$$UDR_\alpha^L = \Lambda\left(\frac{\Lambda^{-1}(PD) + \sqrt{\rho}\Lambda^{-1}(\alpha)}{\sqrt{1-\rho}}\right) \tag{8}$$

as an alternative to the Vasicek's formula (2). We will call the formula "*logistic Vasicek*". The advantage of this simplified formula is that it is fully analytical while in (7) the value $\Lambda_\rho^{-1}(PD)$ needs to be estimated numerically. Table 2 shows that the differences between the precise formula (7) and its approximation (8) are very small. Moreover, the correlation ρ will be calibrated with respect to the formula (8) rather than (7) further diminishing the error of this approximation.

Table 2

A Comparison of the Logistic Vasicek's UDR Calculated According to (7) and its Approximation (8) with $\rho = 0.1$ and $\alpha = 0.999$

PD	**0.2%**	**0.5%**	**1%**	**2%**	**4%**	**6%**
$\Lambda_\rho^{-1}(PD)$	–6.0731	–5.1940	–4.5311	–3.8552	–3.1774	–2.7659
$\Lambda^{-1}(PD)$	–6.2126	–5.2933	–4.5951	–3.8918	–3.1781	–2.7515
UDR_α^{L0}	1.63%	4.02%	7.77%	14.66%	25.98%	35.13%
UDR_α^{L}	1.41%	3.64%	7.30%	14.18%	25.97%	35.48%

Source: Own calculations.

We have argued that the logistic Vasicek's formula is compatible with the logistic regression widely used to estimate and model probabilities of defaults. In order to compare the impact of the two formulas we can as a first approximation use the regulatory correlations and calculate the differences. Figure 2 and Table 3 compare the unexpected default rates calculated according to the normal and logistic Vasicek's formulas with $\rho = 0.1$ and $\alpha = 0.999$. The logistic *Unexpected Default Rate* (UDR) estimate is clearly much more conservative than the normal Vasicek's UDR calculation.

Table 3

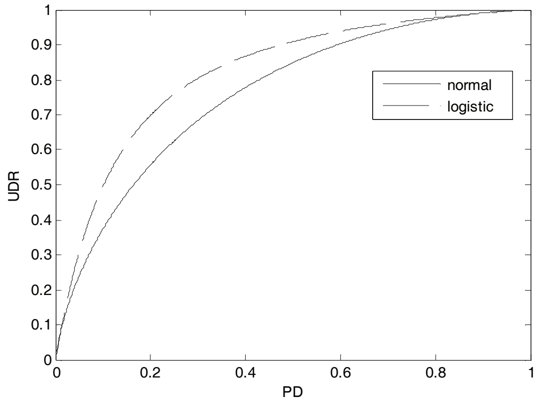
Normal and Logistic Vasicek's Unexpected Default Rates (UDR) for Selected PD Values ($\rho = 0.1$, $\alpha = 0.999$)

PD	0.0100	0.0200	0.0300	0.0400	0.0500	0.0700	0.1000	0.1500	0.2000
Normal UDR	0.0775	0.1282	0.1704	0.2074	0.2408	0.2996	0.3742	0.4751	0.5568
Logistic UDR	0.0730	0.1418	0.2039	0.2597	0.3097	0.3955	0.4965	0.6163	0.6987
Difference	–0.0045	0.0136	0.0335	0.0522	0.0689	0.0959	0.1224	0.1412	0.1418

Source: Own calculations.

F i g u r e 2

Comparison of the Normal and Logistic Vasicek's Unexpected Default Rate (UDR) as a Function of PD ($\rho = 0.1$, $\alpha = 0.999$)

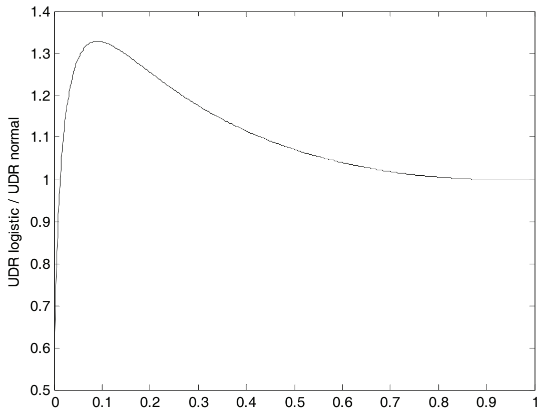


Source: Own calculations.

Figure 3 shows that the relative differences are more than 20% for PD values in the range from 3% to 25%.

F i g u r e 3

The Ratio between the Logistic and the Normal Vasicek's UDRs



Source: Own calculations.

However, such a comparison is not really consistent since the regulatory correlations have been calibrated so that the model (2) fits a representative dataset. Unfortunately, the regulator (i.e. BCBS – the Basel Committee for Banking Supervision) does not disclose the exact method or the dataset used for the calibration. BCBS (2005) only indicates that the correlation is chosen so that it fits the variability of a historical series of default rates.

Let us consider a series of historical default rates $p_{1,}...p_n$ observed over one-year, or possibly quarterly, or even over shorter time-horizons. Assume, that the default rate realizations are given by the model

$$p_t = g(v_t;\rho) = F\left(\frac{F^{-1}(PD) - \sqrt{\rho}v_t}{\sqrt{1-\rho}}\right)$$

where $F = \Phi$ or $F = \Lambda$, $PD = \frac{1}{n}\sum_{t=1}^{n} p_t$ is just the average observed default rate (i.e. a long term PD estimate), and the systematic factors v_t are iid $N(0,1)$ or logistically distributed, respectively. In order to fit this simple model we just need to estimate the correlation ρ maximizing the likelihood function (or rather its logarithm)

$$L = \prod_{t=1}^{n} L(p_t;\rho), \text{ where } L(p_t;\rho) = \frac{F'(v_t)}{g'(v_t;\rho)} \tag{9}$$

$$v_t = g^{-1}(p_t;\rho) = \frac{F^{-1}(p_t)\sqrt{1-\rho} - F^{-1}(PD)}{\sqrt{\rho}}$$

and

$$g'(v_t;\rho) = \frac{\partial g}{\partial v_t} = \frac{\sqrt{\rho}}{\sqrt{1-\rho}} F'\left(\frac{F^{-1}(PD) - \sqrt{\rho}v_t}{\sqrt{1-\rho}}\right) = \frac{\sqrt{\rho}}{\sqrt{1-\rho}} F'\left(F^{-1}(p_t)\right)$$

A more sophisticated model used by Hamerle, Liebig and Rösch (2003) or Hamerle and Rösch (2006) takes into consideration all the exposure-level and general economic information available at the time $t-1$. Let $\mathbf{x}_{i,t-1}$ denote all the covariates available for the receivable *i* at the time $t-1$. In line with the credit scoring model (5) the credit variable $Y_i = \boldsymbol{\beta}'\mathbf{x}_{i,t-1} + \sqrt{\rho}V_t + \sqrt{1-\rho}\zeta_{it}$ is decomposed into a known predictable part, a latent systematic factor, and an idiosyncratic factor. The probability of default conditional on $\mathbf{x}_{i,t-1}$ and conditional on a realization of the systematic factor then is

$$\Pr[Y_i \le 0 \mid \mathbf{x}_{i,t-1}, V_t] = \Pr\left[\boldsymbol{\beta}'\mathbf{x}_{i,t-1} + \sqrt{\rho}V_t + \sqrt{1-\rho}\zeta_{it} \le 0\right] = F\left(\frac{-\boldsymbol{\beta}'\mathbf{x}_i - \sqrt{\rho}V_t}{\sqrt{1-\rho}}\right)$$

Hamerle Liebig and Rösch (2003) or Hamerle and Rösch (2006) use the normal distribution $F = \Phi$, but we argue that it is more appropriate to use the logistic distribution $F = \Lambda$. In order to estimate the parameters $\boldsymbol{\beta}$ and ρ we need to know the default indicators $d_{it} \in \{0,1\}$ and the covariates $\mathbf{x}_{i,t-1}$ on the level of each receivable $i = 1,...N_t$ and for all $t = 1,...n$. The likelihood function in this case involves integrations over the latent factors v_t:

$$L = \prod_{t=1}^{n} \int_{-\infty}^{+\infty} \prod_{i=1}^{N_t} g(v_t; \mathbf{x}_i, \rho)^{d_{it}} \left(1 - g(v_t;\ \mathbf{x}_i, \rho)\right)^{1-d_{it}} dF(v_t)$$

where

$$g(v_t;\ \mathbf{x}_i, \rho) = F\left(\frac{-\boldsymbol{\beta}'\mathbf{x}_i - \sqrt{\rho} v_t}{\sqrt{1-\rho}}\right)$$

According to Hamerle and Rosch (2006) the integral can be approximated by a Gaussian quadrature and the likelihood needs to be maximized numerically. The confidence intervals can be estimates using the Fisher information matrix that is evaluated numerically.

3. Empirical Study

In order to compare the potential impact of the normal and logistic Vasicek's formulas we use the data provided by the U.S. Federal Reserve.[2] The dataset contains quarterly delinquency rates on loans and leases at all U.S. commercial banks from 1985(Q1) until 2012(Q1). Delinquent loans are defined as those past-due 30 days or more and still accruing interest as well as those in nonaccrual status. They are measured as a percentage of the end-of-period loans. Since delinquent loans are regularly charged-off the delinquency rate can be considered to be a proxy of an ordinary default rate. The exposures are segmented as mortgages (residential, commercial, farmland), consumer loans, credit cards, leases, commercial and industrial (corporate) loans, and agricultural loans. The delinquency rates are given as seasonally adjusted and non-adjusted. Figure 4 shows the volatile development of the overall delinquency rate. A similar pattern is followed by the delinquency rates of the individual segments.

Although the notion of delinquency is not fully compatible with the Basel concept of default (ninety or more days past due), we will use the seasonally adjusted delinquency rate in order to indicatively compare the two alternative Vasicek's formulas.

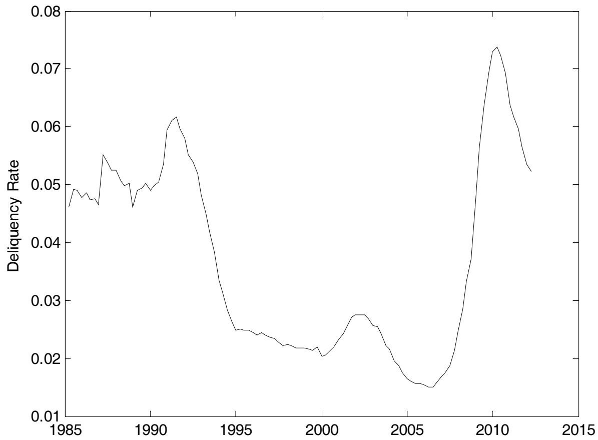
Since Basel II is based on segmentation we estimate the correlation parameters separately for a few most important segments, namely: residential mortgages, credit cards, consumer loans, and corporate loans. Because we do not have transaction level data, we implement just the simple estimation method outlined in Section 0.

---

[2] See <http://www.federalreserve.gov/releases/chargeoff/>.

F i g u r e 4

Seasonally Adjusted Delinquency Rates of U.S. Commercial Bank Loans 1987 – 2012 (All Products)



Source: Own calculations; <www.federalreserve.gov/releases/chargeoff>.

T a b l e 4

Estimation Results (Correlations and Unexpected Default Rates; Estimation Errors of the Correlation Parameter are Shown in Parentheses)

	Mortgages	Consumer Loans	Credit Cards	Corporate Loans
Exposure (2012 Q1, USD mil)	2 071 042	580 497	595 894	1 326 836
Average PD	0.0357	0.0350	0.0449	0.0309
Correlation (normal)	0.0751 (0.0101)	0.0034 (0.0005)	0.0057 (0.0008)	0.0470
Log-likelihood (normal)	219.4560	402.8462	299.9271	289.02
Correlation (logistic)	0.1209 (0.0158)	0.0056 (0.0009)	0.0086 (0.0015)	0.0899
Log-likelihood (logistic)	225.47	402.83	300.53	284.68
Likelihood ratio (normal/ logistic)	12.03	–0.03	1.21	–8.69
UDR-PD (normal)	0.1243	0.0161	0.0263	0.0791
UDR-PD (logistic)	0.2423	0.0219	0.0362	0.1606
Relative difference	94.9%	36.0%	37.6%	103.0%

Source: Own calculations.

Table 4 shows the estimation results. For each of the four segments we have estimated the correlation maximizing (9) with $F = \Phi$ or $F = \Lambda$ (normal and logistic). The estimation error is given by the inversion of the second derivative of the log-likelihood function at the estimated correlation parameter. The two models can be compared, for example, by using the ordinary likelihood ratio

$$D = -2\ln\left(\frac{L_{\text{normal}}}{L_{\text{logistic}}}\right) = -2\ln\left(L_{\text{normal}}\right) + 2\ln\left(L_{\text{logistic}}\right)$$

The ratio shows a significantly better fit of the logistic model in case of the most important mortgage loans segment. The normal distribution model performs better for the corporate segments, and the results are mixed for credit cards and consumer loans. The estimated correlation of the logistic model is higher than the one of the normal model in case of all the segments. While the correlations for the mortgage and corporate segments are at expected levels, the credit cards and consumer loans correlations are unusually low and would be probably replaced by a minimum threshold. The differences between the unexpected default rate increments ($UDR - PD$) for the two models are even higher due to the effect indicated by Figure 2. Since the unexpected default rate increment $UDR - PD$ is a multiplicative factor of the Basel II formula (1) we may conclude that the capital requirement would be almost doubled for the mortgage and corporate segments if the normal Vasicek's formula was replaced by the logistic one. The exposure weighted difference between the normal Vasicek's and logistic Vasicek's formula capital requirement turns out to be 93.4%.

Conclusion

We have argued that the Vasicek's formula presenting a quantitative core of the Basel II regulation should be based rather on the logistic than on the normal distribution in line with standard logistic regression PD estimation approach. Based on an empirical analysis, the standard normal cumulative distribution function in the Vasicek's formula can be simply replaced by the logistic function. If the correlation parameters were kept on the current regulatory level then the capital requirement could be 20 – 30% higher compared to the current formula results. Nevertheless, if the correlation parameters were re-estimated for the logistic distribution model capturing better extreme events (like the recent financial crisis) than, according to our empirical study, the differences are even more dramatic reaching 90 – 100%.

The result shows a significant model risk of the current Basel II regulation. The proposed logistic Vasicek's formula is not just a formula based on a special distribution or a copula function. In our view, it is the most natural choice in the context of current PD modeling standards. The differences indicate that the global banking system might be seriously undercapitalized and vulnerable in times of a financial crisis. The alarming differences between the normal and logistic Vasicek's formula capital requirement calculated on the U.S. commercial banks' delinquency dataset call for further research and testing. For example, Gapko and Šmíd (2012) applied a dynamical version of the Vasicek's model with of the generalized hyperbolic distribution and also estimated increased capital requirements, but only in the range of 20 – 30% on a mortgage delinquency rate time series.

Further investigation of the proposed and other competing models should be based on larger and better quality datasets containing exposure level information and using a fully Basel II compliant definition of default. The estimation of the key correlation parameter should take into account all the available exposure level and macroeconomic information according to the methodology outlined in Section 0. This extension could, possibly, reduce the estimated correlation and the variability of the default rates implied by the model.

Literature

BCBS (2005): An Explanatory Note on the Basel II IRB Risk Weight Functions. Basel: Bank for International Settlements.

BCBS (2006): Basel Committee on Banking Supervision: International Convergence of Capital Measurement and Capital Standards, A Revised Framework – Comprehensive Version. Basel: Bank for International Settlements.

BCBS (2010): Basel III: A Global Regulatory Framework for More Resilient Banks and Banking Systems. Basel: Bank for International Settlements.

CALEM, P. S. – FOLLAIN, J. R. (2003): The Asset-Correlation Parameter in Basel II for Mortgages on Single-Family Residences. Washington, DC: Board of Governors of the Federal Reserve System.

CHERUBINI, U. – LUCIANO, E. – VECCHIATO, W. (2004): Copula Methods in Finance. Chichester: John Willey & Sons Finance.

CROOK, J. – BELLOTI, T. (2012): Asset Correlations for Credit Card Defaults. Applied Financial Economics, *22*, No. 2, pp. 87 – 95.

DOUGLAS, W. D. (2007): The Distribution of Defaults and Bayesian Model Validation. Journal of Risk Model Validation, *1*, No. 1, pp. 23 – 53.

GAPKO, P. – ŠMÍD, M. (2012): Modeling a Distribution of Mortgage Credit Losses. Ekonomický časopis/Journal of Economics, *60*, No. 10, pp. 1005 – 1023.

GORDY, M. (2003): A Risk Factor Foundation for Ratings Based Bank Capital Rules. Journal of Financial Intermediation, *12*, No. 3, pp.199 – 232.

GREENE, W.H. (2003): Econometric Analysis. 5th edition. New Jersey: Prentice Hall.

HAMERLE, A. – LIEBIG, T. – RÖSCH, D. (2003): Benchmarking Asset Correlations. Risk, *16,* No. 11, pp. 77 – 81.

HAMERLE, A. – RÖSCH, D. (2007): Parameterizing Credit Risk Models. Journal of Credit Risk, *2,* No. 4, pp. 101 – 122.

LALL, R. (2012): From Failure to Failure: The Politics of International Banking Regulation. Review of International Political Economy, *19*, No. 4, pp. 609 – 638.

LUCAS, D. J. (1995): Default Correlation and Credit Analysis. Journal of Fixed Income, *4*, No. 4, pp. 76 – 87.

NAGPAL, K. – BAHAR, R. (2001): Measuring Default Correlation. Risk, *14*, No. 3, pp. 129 – 132.

STEFANSKI L. A. (1991): A Normal Scale Mixture Representation of the Logistic Distribution. Statistics & Probability Letters, *11*, No. 1, pp. 69 – 70.

ŠÚTOROVÁ, B. – TEPLÝ, P. (2013): The Impact of Basel III on Lending Rates of EU Banks. Czech Journal of Economics and Finance, *63*, No. 3, pp. 226 – 243.

TEPLÝ, P. – VRÁBEL, M. – ČERNOHORSKÁ, L. (2012): The VT Index as an Indicator of Market Liquidity Risk in Slovakia. Ekonomický časopis/Journal of Economics, *60*, No. 3, pp. 223 – 238.

VASICEK, O. (1987): Probability of Loss on a Loan Portfolio. [KMV Working Paper.] San Francisco: KMV Corporation.

WITZANY, J. (2010): Credit Risk Management and Modeling. Praha: Nakladatelství Oeconomica.

WITZANY, J. (2011): A Two Factor Model for PD and LGD Correlation. Bulletin of the Czech Econometric Society, *18*, No. 28, pp. 1 – 19.