Basel II Capital Requirement Sensitivity to the Definition of Default

Jiří Witzany*

The paper is motivated by a disturbing observation according to which the outcome of the regulatory formula significantly depends on the definition of default used to measure the Probability of Default (PD) and the Loss Given Default (LGD) parameters. Basel II regulatory capital should estimate, with certain probability level, the unexpected credit losses on banking portfolios, and so it should not depend on a particular definition of default that does not change real, historical and expected losses. This paper provides an explanation of the phenomenon based on the Merton default model and tests it using a Monte Carlo simulation. Moreover, it develops an analytical method to model LGD unexpected risk and to combine it with the PD unexpected risk. The developed formula and, in particular, its simplified version could be used to improve the current regulatory formula. The analysis, at the same time, provides a different insight into the issue of regulatory capital sensitivity to the definition of default. Finally, the paper performs a structural model-based simulation to test the hypothesis according to which scoring functions developed with a soft definition of default have weaker predictive power than the ones developed with a hard definition of default.

Introduction

The Basel II capital requirement formula (Basel, 2006, e.g., paragraph 328 for mortgages) has in principle the following form:

$$C = \textit{Unexpected Loss} - \textit{Expected Loss} = UDR.LGD - PD.LGD \quad \text{...(1)}$$

where *UDR* is a regulatory estimate of the unexpected default rate, *PD* is the estimated average default rate, and *LGD* is the loss given default. The resulting percentage of capital requirement (C) is multiplied by *EAD*, the exposure of the receivable at default (in most cases equal to the actual outstanding amount), to get the account level capital requirement in absolute terms. The calculation is done at account level, but the result should reflect only the account's contribution to unexpected risk within a well-diversified portfolio. The *UDR* is given by a regulatory formula that depends only on *PD* (other key parameters of the formula, i.e., correlation and probability level are set by the regulator as given constants, with a possible adjustment depending on *PD* in the case of correlation). In other words, the capital requirement C is a function of *PD* and *LGD,*

$$C = f(PD, LGD) = [UDR(PD) - PD].LGD \quad \text{...(2)}$$

The logic of Formula 2 is to give an estimation of unexpected systematic loss relative to the expected loss at 99.9% probability level. This goal is mainly achieved through the

* Assistant Professor, Faculty of Finance and Accounting, University of Economics, 13067 Prague 3, Czech Republic. E-mail: witzanyj@vse.cz

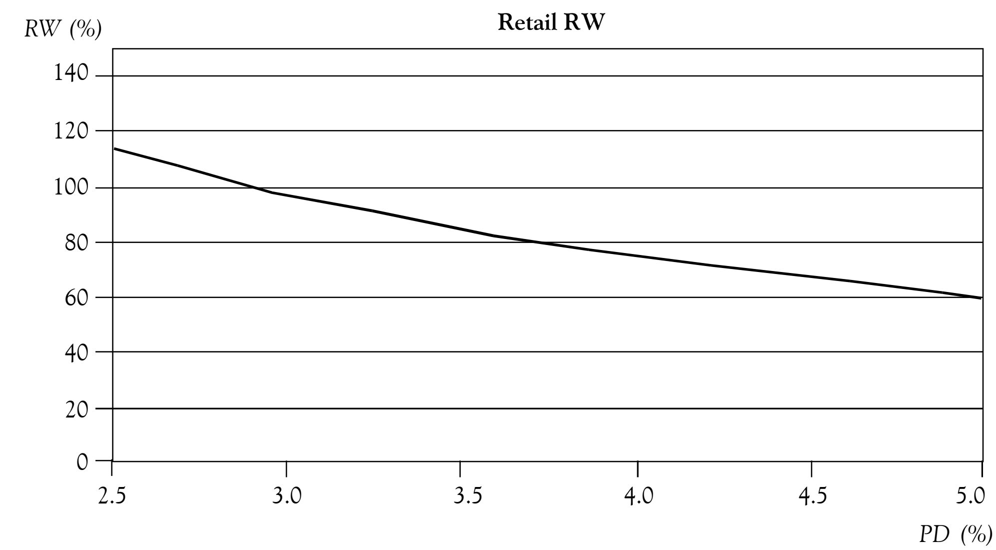
first part of the formula, $UDR(PD) - PD$, that models the difference between the unexpected default rate value and the expected mean default rate. On the other hand, the LGD parameter is defined in principle as a conservative estimate of the long-term average loss rates on defaulted loans. The regulator in addition, requires the banks to rather vaguely reflect the economic downturn conditions, dependence on PD or other factors to capture the relevant risks. Nevertheless under normal circumstances, the long-term average is considered as satisfactory. In that case, the possibility of an unexpectedly high loss rate is not captured by the formula at all.

In practice, the parameters PD and LGD are obtained from historical data on homogenous (in terms of product, segment and rating) sets of receivables. The values strongly depend on the definition of default, which determines the number of loans marked as defaulted. The definition of default must satisfy certain regulatory conditions, but the banks have a significant discretion in its implementation. Assume that a bank suffered a total loss L on its historical portfolio Π and recorded a number of defaults, N_D. Some of the defaulted receivables might have been recovered fully, while others contributed to the total loss L. If $\overline{E} = V/N$ (where V is the total portfolio Π volume and N is the total number of receivables) is the average exposure of the observed receivables and $\overline{L} = L/N_D$ is the average loss on defaulted receivables, then the loss given default parameter can be actuarially estimated as $LGD = \overline{L}/\overline{E}$. Similarly, PD can be estimated as N_D/N. Consequently, we get that the product $PD.LGD = L/V$ is the average percentage of observed loss rate, $LR = L/V$, on the portfolio, which does not depend on the definition of default. The definition just influences the decomposition of LR into the product $PD.LGD$, i.e., given PD (based on a definition of default), we calculate LGD as LR/PD and vice versa. This relationship holds as long as the LGD estimate is not stressed. However, the product, $PD.LGD$, serves as the bank's estimate of future losses which should be used for loan pricing, budgeting, etc. Hence, the stressing of $LR = PD.LGD$ should be quite consistent with the experienced losses, which might be moderately adjusted to certain negative macroeconomic or other developments. Hence, the historical loss rate LR should be replaced with a more conservative expected loss rate EL, which will be futuristic in approach. Then, we still have the identity, $PD.LGD = EL$. Thus, the capital requirement formula then takes the form,

$$C = f(PD, EL/PD) \qquad \text{...(3)}$$

where the expected loss EL is a fixed input parameter. One may ask why PD should be variable if observed from historical data. The point is, as already mentioned, there is the notion of default that strongly influences the counted number of defaults. According to the Basel II requirement, the receivables that are more than 90 days past due, where the debtor is legally proclaimed as a bankrupt, or that are considered probable not to be repaid in full, must be flagged as defaulted. The regulation leaves significant space for the banks to use a softer or harder definition of default, in particular due to the condition 'considered probable not to be repaid in full'. A bank may stick to the simple definition of 90 days past due plus legal events like bankruptcy, etc., or it may opt to use a soft definition of default, where many other clients are marked as defaulted earlier based on a number of additional conditions, just making sure that the overall probability of absorbing default (staying in default until maturity) remains above 50%, to be consistent with the common sense interpretation of the regulatory requirement. Another important parameter in the definition of default is the materiality threshold—only amounts above the threshold may cause a default. If the materiality threshold is too low, then many clients that, by mistake, left very small amounts unpaid may be marked as defaulted. There is also the issue of cross-cross default of retail clients which might be dealt with differently by different banks—does a default on credit card imply a default on a mortgage which is being paid? In the hard default approach, the bank may count on a historical dataset just 2.5% default rate, while in terms of the soft default definition, the rate may be as high as 5%. Of course, the different definitions of default should not and, according to our assumptions, do not change the observed and expected percentage loss. Since the underlying fundamental financial data remain the same, the capital requirement modeling the difference between the 99.9%-quantile of future loss and the mean (expected) loss should remain identical under different definitions of default. Nevertheless, Figure 1 shows that the risk weight (using the regulatory formula for retail loans) significantly declines from 115% to 62% when *PD* increases from 2.5% to 5%.

Figure 1: Dependence of Regulatory *RW*=12.5C on *PD*



This appears as a significant problem of the regulatory formula. On the one hand, the regulators might wish to motivate banks to use a conservative (soft) definition of default in order to start collection and work-out process soon. On the other hand, the sensitivity of the capital to the definition of default, as shown in Figure 1, is surprisingly dramatic. Soft definition of default may be desirable in terms of conservative classification of assets, but it could also pollute the quality of default statistics, if the definition is consistently used for development of scoring functions (the Basel II regulation does require banks to use the definition of default consistently throughout the credit process). Many clients marked as defaulted, using a soft definition, may return to ordinary repayment schedule later. The reason of the default in payment may not be real financial difficulties, but just disorganization on the client side, and repayments are put into order just after one or two reminders. Prediction of this type of default should not be mixed with prediction of real or absorbing default, i.e., default where the client does not return to ordinary repayments and the bank suffers a real loss. This notion of absorbing default (rather than soft default) should be captured by the hard definition of default.

In Section 2, we provide an explanation of the observed sensitivity to the definition of default and propose alternative approaches to modeling of *LGD* unexpected risk. In Section 3, we analyze the effects of the definition of default on the quality of scoring functions, developed using a standard technique and based on a structural credit default model simulation.

2. Unexpected Credit Losses

2.1 Economic vs. Regulatory Capital

Regulatory capital requirements for credit risk exposures aim to approximate the economic capital that captures the unexpected credit losses in a given time horizon and on a probability level. The two parameters in the context of Basel II are: 1 year and 99.9%. Thus in principle, we need to model the probability distribution of a random variable $V(1)$, the future value of the bank's credit portfolio, 1 year from the calculation time. The expected value of $V(l)$ should not only take into account the accrued interest but also the expected credit losses, so the 99.9%-quantile of $V(l) - E(V(1))$ is the value we need to estimate. The credit economic capital or credit Value at Risk (VaR) is in practice estimated by a number of techniques. For example, the well-known JP Morgan's CreditMetrics model (Gupton *et al.*, 1997) based on rating migration probabilities, reduced CreditRisk$^+$ model (Credit Suisse First Boston, 1997) and KMV Moody's PortfolioManager based on the structural Merton model.

The regulatory model must certainly accept a number of simplifications. Although the regulatory capital requirement is calculated at account level, it needs to take into account only the portfolio systematic risk and not the specific risk. Currently, the regulatory model does not reflect the level of diversification and specific correlation structure of the bank portfolio. It uses the overall average correlations differentiated for basic product classes and segments with certain dependence on the probability of default. The key simplification lies in the fact that the formula models principally unexpected default rate and not the unexpected loss rate. The Unexpected Loss Rate (*ULR)* according to Basel II is estimated as:

$$ULR = \mathrm{UDR}.LGD \qquad \ldots(4)$$

disregarding the maturity adjustment applied only for non-retail clients. Thus, the regulatory *ULR* formula does not take into account the unexpected risk of losses after default.

It is rather obvious and has been confirmed by a number of studies that the additional unexpected risk is quite significant (Altman *et al.*, 2004). This fact partially explains the empirical fact that the Basel II capital requirement depends on the definition of default. In this paper, we develop several alternative models, extending the unexpected loss calculation beyond the *UDR* to capture the (systematic or portfolio) risk of unexpected losses and to explain the sensitivity shown in Figure 1.

2.2 Basel II Unexpected Default Rate

The Basel II *UDR* formula can be expressed as follows:

$$UDR = N\left(\frac{N^{-1}(PD) + \sqrt{\rho}.N^{-1}(0.999)}{\sqrt{1-\rho}}\right) \qquad \text{...(5)}$$

where N is the cumulative standardized normal distribution function, N^{-1} is its inverse, ρ is the correlation set up by the regulator (15% for mortgage loans, 4% for revolving loans, and between 4% and 15% depending on *PD* for other retail loans.) It will be useful to recall here the principle of Formula 5 that was first discovered by Vasicek (1987).

For a client *j*, let T_j be the time to default on a client's debts. It is assumed that everyone will default once and as the time of the future event is unknown at present, the time $T_j < \infty$ is a random variable. If Q_j is the cumulative probability distribution of T_j then it can easily be verified that the transformed variable, $X_j = N^{-1}(Q_j(T_j))$, is a standardized normal (mean = 0, standard deviation = 1). The advantage is that after the transformation we can make the assumption that the variables are multivariate normal and given their mutual correlation ρ, properties of normal variables can be used to obtain an analytical result. This approach is called the Gaussian copula model. The following one-factor model is used,

$$X_j = \sqrt{\rho}.M + \sqrt{1-\rho}.Z_j \qquad \text{...(6)}$$

where M captures the systematic factor and Z_j the client-specific factors. All Z_j's and M have independent standard normal distributions.

The one-year probability of default *PD* of the client *j* can be expressed as:

$$Pr\left[T_j \le 1\right] = Pr\left[X_j \le N^{-1}(Q(1))\right]$$

$$= Pr\left[\sqrt{\rho}.M + \sqrt{1-\rho}.Z_j \le N^{-1}(Q(1))\right] = Pr\left[Z_j \le \frac{N^{-1}(Q(1)) - \sqrt{\rho}M}{\sqrt{1-\rho}}\right]$$

$$= N\left(\frac{N^{-1}(Q(1)) - \sqrt{\rho}M}{\sqrt{1-\rho}}\right) \qquad \text{...(7)}$$

The next step is to consider M as the systematic driver of portfolio default rates. The model can be used for a simulation as follows: First, generate randomly the value of M from a standardized normal distribution and then independently generate all Z_j. If the portfolio is large enough, then the simulated default rate on the portfolio will be given by Formula 7. If M is large, the simulated default rate will be low; if M is small, then the portfolio default rate will be high. For a given probability level x, the critical point of M is given by the quantile $N^{-1}(x)$. When M is replaced by $N^{-1}(x)$ and $Q(1)$ by the given average *PD*, we exactly get the regulatory Formula 5.

2.3 Unexpected Loss Given Default – The Merton Model

The Merton model has been developed (Merton, 1974) in particular for corporate debtors. It is used as a theoretical foundation by the CreditMetrics methodology of J P Morgan. The idea is that a company is able to pay back its loans as long as the value of its assets is above the amount of the external liabilities. At the time of granting a new loan, the bank normally verifies that the assets are well above the external liabilities (i.e., the capital cushion is positive and sufficient). Nevertheless, later the value of the assets might stochastically go up or down, depending on a number of external and internal factors. To model the value of the assets, we use the geometrical Brownian motion equation, $dA = \mu A dt + \sigma A dz$, or in a simplified approach the Brownian motion stochastic equation is $dA = \mu dt + \sigma dz$, allowing negative values of A. The default at time t takes place, if $A(t) < L$, where L is the actual value of debtor's liabilities. The time of default can be defined as $T_{Def} = \inf\{t \geq 0;\ A(t) < L\}$. To model the unexpected default rate, *UDR*, we simply count the number of defaults, i.e., the probability *PD* that $T_{Def} \leq 1$ in a stressed scenario. The final loss in case of default, or its estimation at time 1, clearly depends on the terminal value $A(1)$ that has evolved randomly from the time T_{Def} to time 1. The model should also take into account the effect of a distressed sale or bankruptcy proceeding when assets are sold with a discount. Hence, setting $ULR = PD.LGD$ with an average (i.e., deterministic) *LGD*, we lose the variability of the development of the assets value, A, from the time T_{Def} to 1. This approach is equivalent to replacing the value $A(t)$ with $L^*\ (1 - LGD)$ for all $t \geq T_{Def}$.

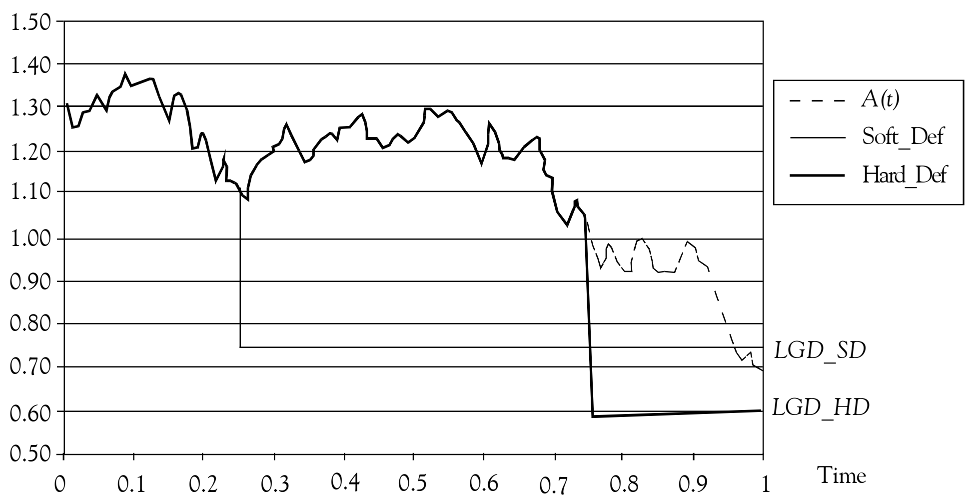
On the other hand, if at time 1 we have new information on the value of A, then the Best Estimate of Expected Losses (BEEL) generally differs from the initial estimate at the time of default. If A goes up again, then the new estimate would be better than the average *LGD*, and if the value of asset, A, deteriorates since the default, then the estimate would be lower. A possible estimation for BEEL is given by the expression, max $(0, L - H(A(1)))$, where $H(.)$ is a haircut function, reflecting the distress liquidation. In practice, the recovery process starts at the time of default (or sooner) and the first experiences soon reveal what the realistic recovery rate might be.

The structural model could be improved (in particular for retail loans and mortgages) as follows: Let $Q(t)$ model the liquid assets and let a partially or fully independent variable, $A(t)$, model the value the debtor's illiquid assets. The default in payment occurs when

Q is less than the amount to be paid P. It is then and usually with a delay that the bank starts to investigate the real value of the assets A. Once the assets are properly assessed, it is usually clear whether the value would be sufficient to cover the amount of the loan or not (though the process of liquidation may be time consuming). Consequently, BEEL at time 1 is normally a good account level estimate of the final loss, substantially better than the estimate of LGD set at time of default when the illiquid assets have not been inspected yet. Consequently, we infer that the losses recognized at time 1, normally a few months after default, are almost as precise as the losses calculated when the recovery process is completed (which may take a few years).

The above analysis shows that the regulatory formula can be interpreted as a model capturing the stochastic process A up to the time of default T_{Def}, then replacing the values of A by the average LGD, fully disregarding the element of unexpected LGD losses that can be almost fully recognized at the end of the year. Moreover, the sooner the defaults occur, the more the randomness of the after-default process is lost (Figure 2).

Figure 2: Stochastic Modeling of Assets up to the Time of Default



In this context, we can interpret the soft default as a larger threshold $L_1 > L$, arguing that some debtor's assets may not be liquid enough to cover the insufficient cash for loan repayment on time. If T_1 is defined using the threshold L_1, then generally $T_1 < T$. The two models yield $PD_1 > PD$. The loss given defaults then must be set to keep the expected loss $EL = PD \cdot LGD$ constant, i.e., $LGD_1 < LGD$. Since in the case of soft definition of default the model is less stochastic and more deterministic, one should not be surprised by the effect observed in Figure 2.

2.4 A Monte Carlo Simulation

We use the simple structural model, $dA = \mu dt + \sigma dz$, to test the phenomenon of the dependence of PD unexpected risk measures on the definition of default. The simulation is done on a large portfolio level with a correlation ρ involved (we use the regulatory level of ρ =10%). Analogous to Section 2.2, we decompose each random element dz_j into its systematic and debtor-specific parts, $dz_j = \sqrt{\rho}dw + \sqrt{1-\rho}dx_j$. We start with a large portfolio of old and new non-defaulted loans, and, for the sake of simplicity, set all outstanding amount to 1. We use a Monte Carlo simulation to plot the distribution of the loss rate on the portfolio level.

- Directly simulating the values of the loans at time 1 as min(1, $A_j(1)$) with a random haircut involved, if $A_j(1) < 1$;
- Simulating the values through the hard definition of default as $1 - LGD_H$ (where LGD_H is the average loss given hard default), if there is a default during the first year and 1 otherwise,
- Simulating similarly the values through a soft definition of default with a threshold, $S > 1$.

The steps of the simulation may be summarized as follows:

1. To simulate end-of-month $A_j(t)$ for each debtor $j = 1, ..., N$ (where N is the number of debtors or receivables in the portfolio), we need to set up the initial characteristics $A_j(0) > 1$, μ_j and σ_j. The values are randomly sampled with normal distributions such that the resulting mean portfolio loss and default rate correspond to normal values, e.g., 2% to 3%.

2. For $k = 1, ..., M$ (where M is the number of Monte Carlo portfolio simulations), we simulate the vector of monthly systematic factors $\Delta w(k, 0), ..., \Delta w(k,11)$ with the normal $N(0, \Delta t)$ distribution ($\Delta t = 1/12$).

3. For each k, we simulate the vectors of account-specific monthly systematic changes $\Delta x(k, j, 0), ..., \Delta x(k, j, 11)$ for $j = 1, ..., N$. Again with the normal $N(0, \Delta t)$ distribution, all generated random numbers are sampled independently. We set,

   $$\Delta z(k,j,m) = \sqrt{\rho}\,\Delta w(k,m) + \sqrt{1-\rho}\,\Delta x(k,j,m) \text{ for } m = 0, ..., 11$$

4. We calculate inductively starting with $A(k, j, 0) = A_j(0)$ the simulated asset values,

   $$A(k,j,m/12) = A(k,j,(m-1)/12) + \mu_j\,\Delta t + \sigma_j\,\Delta z(k,j,m-1) \text{ for } m = 1, ..., 12.$$

5. For each k and j, we calculate the account level loss rate as $LR(k, j) = 0$ if $A(k, j, 1) \geq 1$ and as $LR(k, j) = (1 - A(k, j, 1))(1 - haircut)$ if $A(k, j, 1) < 1$, where the randomly generated haircut partially reflects the fact that assets must be sold in a distress to cover the loan at least partially. The portfolio loss rate is then,

$$LR^k = \frac{1}{N}\sum_{j=1}^{N} LR(k,j) \qquad \ldots(8)$$

and the mean loss rate, i.e., expected loss, is $EL = \frac{1}{M}\sum_{k=1}^{M} LR^k$. The values of LR^k can be plotted into a histogram (Figure 3). The 99.9%-quantile, $q_{99.9\%}(LR^k)$ (unexpected loss rate) and the VaR loss rate, $LR_{VaR} = q_{99.9\%}(LR^k) - EL$ are then estimated from the obtained distribution (Table 1).

6. For each k, we count the number N_H^k of hard defaults, i.e., of debtors j for which $A(k, j, m) < 1$ for some m. The observed (simulated) portfolio default rate is then, $PD_H^k = N_H^k / N$. Then, we calculate the mean default rate as $PD_H = \frac{1}{M}\sum_{k=1}^{M} PD_H^k$ and $LGD_H = EL/PD_H$. The regulatory unexpected loss formula is based on estimation of the unexpected default rate. We estimate $q_{99.9\%}$ (PD_H^k) from the distribution of PD_H^k and set $LR_{Var,H} = (q_{99.9\%}(PD_H^k) - PD_H).LGD_H$. This calculation is equivalent to resetting the $A(k, j, m) = LGD_H$ for all defaulted accounts from the month of default. The estimated unexpected loss would then be the same if we proceed as in Step 5. The distribution of the losses can also be depicted by the histogram of $PD_H^k.LGD_H$.
7. We repeat the calculations for a soft default threshold, $S > 1$.

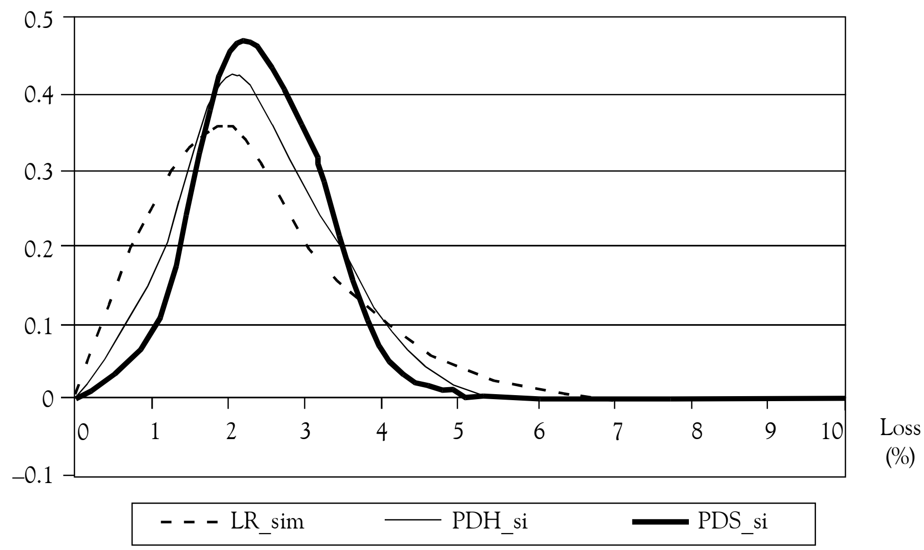
Table 1: Characteristics of the Simulated Loss Rate Distributions

(in %)

	***EL* (Mean)**	***PD***	***LGD***	***LR* Std**	***LR_Var* (99.9%)**
Loss Rate Simulation	1.90	NA	NA	1.25	6.96
PD Hard Def. Simulation	1.90	7.43	25.63	0.95	4.50
PD Soft Def. Simulation	1.90	13.70	13.90	0.75	3.03

Figure 3 and Table 1 clearly demonstrate that unexpected losses based on *PD* simulation significantly underestimate the unexpected losses based on full loss rate simulation. Moreover, the softer definition of default reduces variability of the simulated loss distribution measured, e.g., by the standard deviation or by the 99.9% quantile (4.5% in case of hard default simulation, and 3.03% in case of soft default simulation). The difference would be even larger if we simulate separately the liquid and illiquid assets (with a correlation factor). The value of illiquid assets recognized shortly after

Figure 3: Histograms of Simulated Losses (N = 200, M = 1,000)



default and which may be significantly below 1 mostly determines the final loss on the account. In the above simulation, the account level loss rate depends on the simulated value of $A(k, j, 1)$ which would not be too far from the level where the assets have been a few months before.

2.5 Gaussian Copula and Beta Distribution-Based Model

Now, we propose, in a spirit similar to the Basel II approach, an analytic formula for the unexpected loss on a homogenous portfolio of defaulted receivables due to lower than expected recoveries. The resulting *LGD* unexpected risk is an interesting measure per se, which can be used, for example, for a consistent definition of *LGD,* economic capital, etc. Subsequently, we combine the resulting distribution of *LGD* losses with the (regulatory) distribution of probabilities of default to get an integral measure of unexpected total credit losses.

Let LR_j denote the percentage loss rate (i.e., 1 – the recovery rate) on a defaulted receivable j ($j = 1, \ldots, n$) observed from the perspective of a completed recovery process. The completion may take up to 3-5 years; however, as we argued in the previous section, the loss given default recognition takes much shorter time (e.g., revaluation of collateral, negotiations with the client, etc.), while implementation of the chosen recovery strategy takes a long time. Hence, by modeling the distribution of LR_j (on account or portfolio level) in full recovery time horizon, we get a realistic estimation of the *LGD* unexpected risk in 1 year (or time of default to 1 year) horizon.

Since the portfolio is *LGD* homogenous, we can assume that the distribution of all LR_j is the same with certain cumulative probability distribution function Q . LR_j can be transformed, as in Section 2.2, to a standardized normal variable $Y_j = N^{-1}(Q(LR_j))$. Standardized loss (recovery) rates are used, for example, by the KMV LossCalc methodology (Gupton, 2005; and Kim and Kim, 2006). Let us again use the one factor model for Y_j,

$$Y_j = \sqrt{\rho}.V + \sqrt{1-\rho}.W_j \qquad \text{...(9)}$$

with independent standardized normal V and W_j. A number of studies (Altman and Fanjul, 2004) have not only confirmed that there is a correlation between the rates of default and the recovery rates, but also that the two variables are driven by a common economic factor. This, in particular, is intuitive in the case of mortgages, when a poor state of economy drives not only more clients to default but also reduces the value of the collaterals. The correlation can be estimated from the historical data. However, if there are only limited data (which is mostly the case), it makes sense to use the same regulatory correlation coefficients (with an average taken for the Basel II class of 'other receivables', where the correlation coefficient depends on *PD*). For a given probability level x (e.g., 99.9%), similarly as above, the losses at portfolio level are generated just by the value of V as the independent values of W_j diversify away for a large n. The unexpected portfolio loss rate takes place if V is at the high level expressed by the quantile $N^{-1}(x)$. The unexpected loss rate is as follows:

$$ULR = E\left[Q^{-1}\left(N\left(\sqrt{\rho}N^{-1}(x) + \sqrt{1-\rho}.W\right)\right)\right] \qquad \text{...(10)}$$

where the expectation has taken over all the values of a standardized normal variable, W. It might be tempting to use instead the median,

$$ULR \cong Q^{-1}(N\left(\sqrt{\rho}\,N^{-1}(x)\right)$$

but since the function Q^{-1} is generally strongly nonlinear, the error (difference between the median and mean) is too significant as confirmed by our empirical testing. Consequently, we strongly recommend the use of Equation 10, where the expected value is calculated numerically using the standardized normal density of W, i.e.,

$$ULR = \int_{-\infty}^{+\infty} Q^{-1}\left(N\left(\sqrt{\rho}\,N^{-1}(x) + \sqrt{1-\rho}.w\right)\frac{1}{2\pi}e^{-\frac{w^2}{2}}dw\right) \qquad \text{...(11)}$$

Compared to the Vasicek formula, we have not eliminated the probability distribution function, Q. Note that there is one difference: While in the case of default rates we model at account level a variable, taking only two values (0 if no default and 1 if default); in the case of *LGD*, we model a variable, taking in general any value in the range [0, 1] and so the distribution Q does matter even at the portfolio level. Given the correlation and the probability level, x, we still need to model the Q distribution of *LGD*.

Empirical studies (Gupton, 2005; and Schuermann, 2004) show that the beta distribution is relatively appropriate for *LGD* modeling. The probability density function of beta distribution with minimum 0, maximum 1, and parameters α and β is:

$$Beta(x,\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1} \qquad ...(12)$$

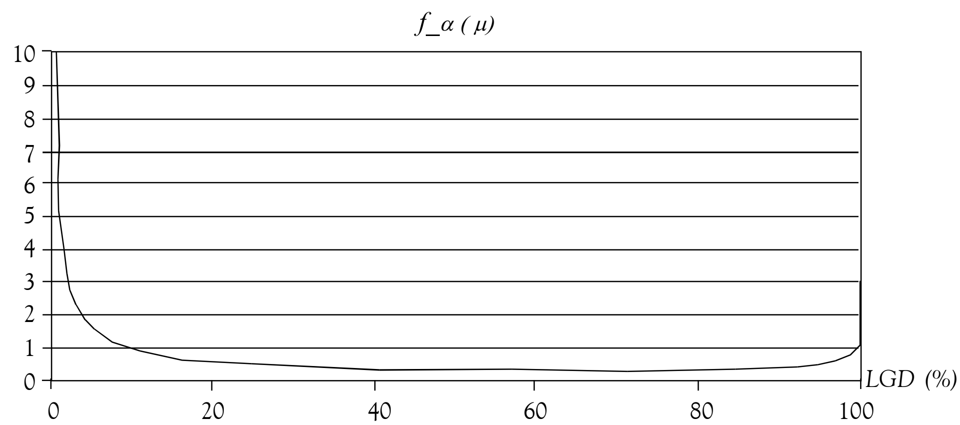
where $\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}e^{-x}\,dx$ is the standard gamma function. The parameters α and β can be calculated from the mean μ and standard deviation σ of the modeled variable,

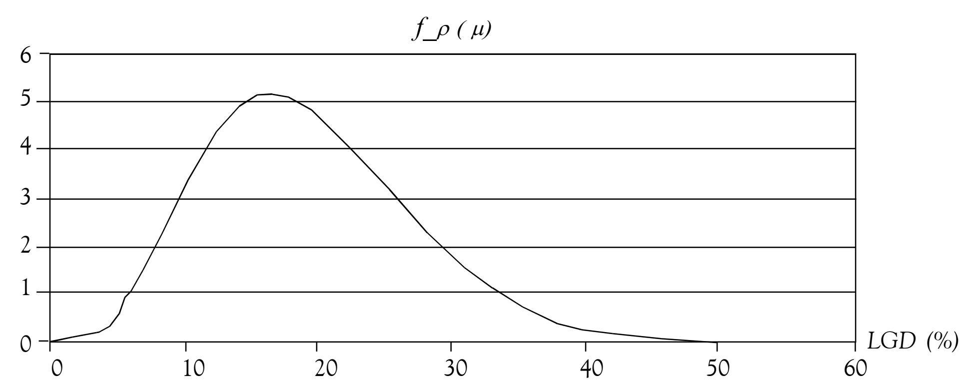
$$\alpha=\mu\left(\frac{\mu(1-\mu)}{\sigma^2}-1\right)$$

$$...(13)$$

$$\beta=(1-\mu)\left(\frac{\mu(1-\mu)}{\sigma^2}-1\right)$$

Figure 4: Account Level Beta Distribution (μ = 20% and σ = 30%) and Its Transformation into a Portfolio Distribution (ρ = 10%)





To calibrate the distribution function, we can either use the mean and the standard deviation from historical *LGD* data on the product, or if not available, use the public data which are, unfortunately, available only for public corporate bonds and bank loans (Altman *et al.*, 2004)). Figure 4 shows how a bimodal account level beta distribution is transformed into a portfolio beta distribution.

2.6 Lognormal Distribution-Based Model – *PD* and *LGD* Convolution

Portfolio level total loss rate, expressed as *LR* = *PD.LGD,* can be modeled through a distribution of the two variables *PD* and *LGD* and their (log) convolution. *LGD* here is not taken as an average but as a random variable defined as the (in advance unknown) loss rate achieved on the defaulted part of the loan portfolio. The regulatory (Vasicek) model provides a reasonable model for *PD* and the above technique (in Section 2.5) provides a model for the *LGD* variable.

We propose a practical approach based on lognormal distributions to combine the two factors. If we assume that *PD* and *LGD* are jointly lognormal, then *LR* is also lognormal, ln *LR* = ln(*PD.LGD*) = ln *PD* + ln *LGD*. Given the means and standard deviations of *PD* and *LGD* (or rather of ln *PD* and ln *LGD*) and assuming that the logs of the two variables have a correlation ρ , we can calculate the mean and standard deviation of *LR* and all quantiles needed. It can be empirically verified that it is feasible to approximate the distribution of *PD* with the lognormal one. The approximation is more questionable for *LGD* as it takes values in the range [0, 1], with mean somewhere in the middle. Nevertheless, as *LR* again has a distribution empirically close to the lognormal one, the approach may be considered as acceptable. Recall that if *X* is a lognormal random variable such that $\ln(X/X_0)$ is $N(m, s^2)$, i.e., normal with mean *m* and standard deviation *s*, then

$$E(X)=X_0 e^{m+\frac{s^2}{2}}$$
$$\sigma(X)=X_0 e^m \sqrt{e^{s^2}-1} \qquad \text{...(14)}$$

If we are given $E(X) = X_0$ and $\sigma(X) = \sigma_0 < X_0$, then Equation 14 is solved by setting

$$s=\sqrt{-\ln\left(1-(\sigma_0/X_0)^2\right)} \text{ and } m=-s^2/2 \qquad \text{...(15)}$$

So, given the key parameters of the distributions of *PD* and *LGD*, $E(PD) = p_0$, $\sigma(PD) = \sigma_{PD}$, $E(LGD)= l_0$, $\sigma(LGD)= \sigma_{LGD}$, and a correlation ρ of ln *PD* and ln *LGD*, we calculate the standard deviations s_{PD} and s_{LGD}, and the means m_{PD} and m_{LGD} of ln *PD* and ln *LGD* from Equation 15. If the variable $\ln LR/(p_0 l_0) = \ln PD/p_0 + \ln LGD/l_0$, then the mean is $m = m_{PD} + m_{LGD}$ and the standard deviation is:

$$s=\sqrt{s_{PD}^2+2\rho\sigma_{PD}\,\sigma_{LGD}+s_{LGD}^2} \qquad \text{...(16)}$$

It is then easy to verify that

$$E(LR) = p_0 l_0 e^{\rho s_{PD} s_{LGD}}, \text{ and} \quad \text{...(17)}$$

$$\sigma(LR) = E(LR)\sqrt{1 - Ae^{-2\rho s_{PD} s_{LGD}}}, \text{ where } A = \left(1-(\sigma_{PD}/p_0)^2\right)\left(1-(\sigma_{LGD}/l_0)^2\right) < 1 \quad \text{...(18)}$$

Table 2 shows the results for different input parameters of *PD* and *LGD*. The columns titled, '*PDLR* Std' and 'VaR_*PDLR*', show the standard deviation and 99.9% Var of the variable $LR = PD.l_0$ with the given parameters, i.e., of the loss rate modeled based on the *PD* distribution in the regulatory manner. On the other hand, the columns titled '*LR* Std' and 'VaR_*LR*' show the measures obtained, based on the convolution *PD.LGD*, using Equations 17 and 18. It again shows how significantly the regulatory *PD*-based modeling approach underestimates the total unexpected loss, if *LGD* variability is taken into account.

Table 2: The Mean, Standard Deviation and 99.9% VaR of *LR=PD.LGD*

(in %)

PD Mean	*PD* Std	*LGD* Mean	*LGD* Std	ρ	*LR* Mean	*PDLR* Std	*LR* Std	VaR_ *PDLR*	VaR_ *LR*
2.0	1.50	60.00	15.0	15.0	1,24	0.90	0.98	12.0	14.6
4.0	2.50	30.00	10.0	10.0	1.23	0.75	0.86	7.0	9.7
4.0	2.50	30.00	10.0	25.0	1.27	0.75	0.92	7.0	11.2
6.0	4.00	60.00	15.0	15.0	3.71	2.40	2.64	25.1	31.5
10.0	5.00	50.00	15.0	15.0	5.13	2.50	3.03	17.7	26.3

We need to keep in mind that these are portfolio level distribution parameters. Hence, we do need to use the copula approach, as explained in Section 2.5, to transform an account-specific beta distribution with observed parameters into a portfolio distribution with a mean and standard deviation used in the lognormal model.

The model also provides another explanation of the regulatory capital sensitivity to the definition of default. The idea is to fix parameters of the loss distribution *LR* and recalculate the 'regulatory' *PD*-based economic capital. That is, for a given *PD* mean p_0, we have to find the standard deviation σ_{PD} and estimate the regulatory-like *PD*-based capital based on the 99.9%-quantile of the distribution PD^*l_0.

To do that we have to solve two equations for three unknowns l_0, σ_{LGD}, and σ_{PD}. To get a unique solution, we need to add a relationship between l_0 and σ_{LGD} that could be based, for example, on empirical observations presented in Table 3.

Let us apply simple linear regression to the data to get the approximate relationship for the portfolio *LGD* standard deviation

$$\sigma_{LGD} = 9.84\% - 2.4\% l_0 \quad \text{...(19)}$$

Table 3: Loss Given Default on Public Corporate Bonds (1974-2003) and Bank Loans (1989(Q3)-2003)				
Loan/Bond Seniority	**Median (%)**	**Mean (%)**	**Standard Dev. (%)**	**Portfolio Std. Dev. (%)**
Senior-Secured Loans	27.00	31.50	24.40	8.52
Senior-Unsecured Loans	49.50	45.00	28.40	9.57
Senior-Secured Bonds	45.51	47.16	23.05	8.20
Senior-Unsecured Bonds	57.73	65.11	26.62	9.08
Senior-Subordinated Bonds	67.65	69.83	24.97	8.65
Subordinated Bonds	68.04	70.97	22.53	8.07
Discount Bonds	81.75	79.07	17.64	6.96
Total Sample Bonds	59.95	65.69	24.87	8.65

Note: Based on prices just after default on bonds and 30 days after default on loans. The last column shows standard deviations of transformed portfolio loss distribution with 10% correlation based on Equation 11.

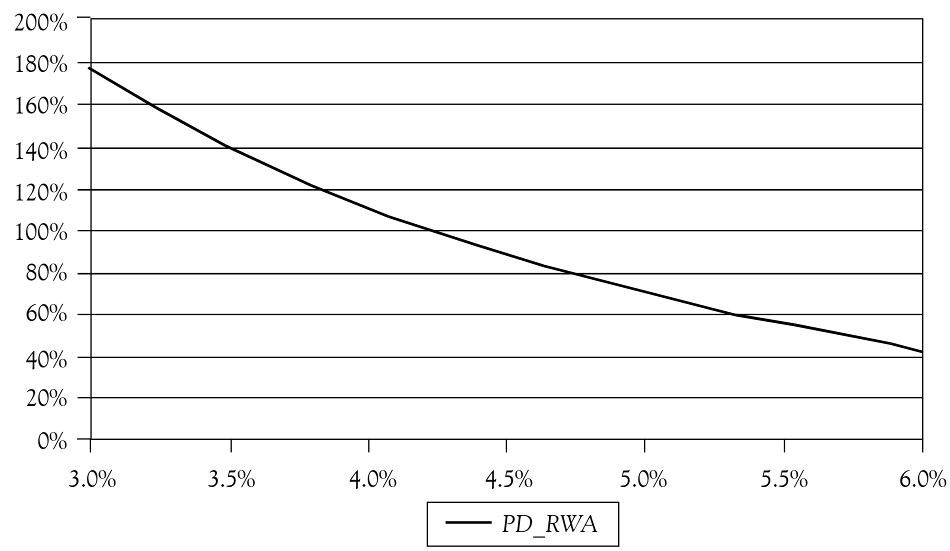
Source: Emery (2003) and Altaian and Fanjul (2004).

Intuitively speaking, since σ_{LGD} remains almost constant and the standard deviation of the total losses remains constant as well, the standard deviation of PD, σ_{PD}, also remains constant. Consequently, the ratio σ_{PD}/p_0 decreases as we move from the hard to the soft definition of default. This implies that in the *PD*-based approach, where unexpected losses are modeled through the variable $PD.l_0$, the standard deviation, $\sigma(PD_S.l_S) = \sigma_{PD,S}.l_s \approx \frac{\sigma_{PD,S}}{p_S}EL$, for the soft definition default is lower than the standard deviation, $\sigma(PD_H.l_H) \approx \frac{\sigma_{PD,H}}{p_H}EL$, in the hard default case.

To perform a numerical simulation, we may proceed as follows: Let us set the mean of *LR* as $EL = \mathrm{E}(LR)$ and standard deviation as $\sigma(LR)$ at a level corresponding to the observed data. Fixing the two parameters of the *LR* distribution, let p_0 range from an initial (hard default value) PD_H to PD_H*2 due to gradual softening of the default. Next, we need to solve the Equations 17, 18 and 19 for the unknowns l_0, σ_{LGD} and σ_{PD}. Initially we set, $s_{PD} = s_{LGD} = 0$, then we calculate l_0 from Equation 17, and σ_{LGD} and s_{LGD} from Equations 19 and 15, and finally s_{PD} from Equation 16. The calculation is repeated until s_{PD} converges to a value with a given precision.

Our numerical simulation (Figure 5) indeed shows a strong dependence of the $PD.l_0$ 99.9% VaR (i.e., a proxy for the regulatory capital) based on the described model, where we start with the hard default probability 3% and change the definition of default so that PD goes up to 6% (the other parameters are: $E(LR) = 1.25\%$, $\sigma(LR) = 1\%$, and $\rho = 15\%$). The standard deviation of LGD follows Equation 19 for changing definition of defaults, while the standard deviation of PD and the mean of LGD are recalculated according to Equations 17 and 18 in order to keep the total loss rate mean and standard deviation, $\sigma(LR)$, constant. It is obvious that the dependence based on the convolution of lognormal distributions is quite similar to the initially observed dependence (Figure 1).

Figure 5: 99.9% *PD*-Based Var Capital Dependence on the Definition of Default



2.7 Binomial *LGD* Model

According to some empirical studies, unimodal beta distribution does not necessarily faithfully model the distribution of recovery rates (Schuermann, 2004). The recovery rates can be for some products bimodal—the recovery rates are either rather low or rather high. This is not surprising, in particular, in the case of collateralized products like mortgages. The collateral is either successfully sold and the defaulted receivables are more or less paid back, or there is an unexpected problem with the receivable and the recovery is low. Taking a simplified approach, we can assume that the variable LGD is binomial, i.e., there are only two possible recovery rates (and LGD) values: 0 and 1. Then (looking backward or into the future) we can distinguish just two types of defaults—full-loss-defaults and zero-loss-defaults. As there is no loss on zero-loss-defaults, those cases may be forgotten and all we need to model are the full-loss defaults.

The probability of full-loss-defaults is *PD**LGD*, as *LGD* is the probability of full loss conditioned by a default and *PD* is the probability of default. The loss conditioned by a full-loss-default is certainly 100%, so unexpected default rate is equal to the unexpected loss in this case. The event of a full-loss-default can be modeled using the Gaussian copula Vasicek model:

$$UEL_2 = N\left(\frac{N^{-1}(PD.LGD) + \sqrt{\rho}.N^{-1}(0.999)}{\sqrt{1-\rho}}\right).EAD \qquad \ldots(20)$$

UEL_2 captures both the unexpected default rate and *LGD*, given a systematic correlation ρ. To get the contribution of *LGD*, we need to deduct the unexpected loss capturing only the unexpected default rate:

$$UEL_1 = N\left(\frac{N^{-1}(PD) + \sqrt{\rho}.N^{-1}(0.999)}{\sqrt{1-\rho}}\right).LGD.EAD \qquad \ldots(21)$$

Unexpected loss is measured by Equations 20 and 21 as the loss amount out of the total portfolio outstanding before default. Hence, the *LGD* economic capital as a percentage of the defaulted notional EAD**PD* can be expressed as:

$$LGD_{VaR} = UEL_{LGD} = \frac{UEL_2 - UEL_1}{EAD.PD} \qquad \ldots(22)$$

One may note that there is a dependence of the *LGD* economic capital on *PD* (the economic capital is allocated to a portfolio of already defaulted receivables). This problem can be solved letting *PD* converge to 1. Then we get a model for a fully defaulted portfolio and the Equation 22 converges to:

$$LGD_{VaR} = \left(N\left(\frac{N^{-1}(LGD) + \sqrt{\rho}.N^{-1}(0.999)}{\sqrt{1-\rho}}\right) - LGD\right) \qquad \ldots(23)$$

There is a connection between the binomial approach and modeling of *LGD* via the bimodal beta distribution with two extreme values. It can be easily empirically verified that the binomial approach is rather conservative compared to the previous models.

3. Development of Scoring Functions and the Definition of Default

In the previous part of the paper, we analyzed the question whether the definition of default impacts the quality of scoring functions developed using ordinary methods (like logistic regression), modeling the probability of default. Banks could certainly use one definition of default for regulatory purposes (i.e., regulatory *PD* and *LGD* calculation) and the other, more appropriate, for development of scoring function. Nevertheless, according to Basel II, the definition of default should be used consistently throughout the organization (Basel, 2006, paragraph 456), hence the regulatory definition should also be used for the development of scoring function. As we have analyzed in the previous sections, banks are in a sense (probably unintentionally) motivated to use a softer definition of default. Our hypothesis on the other hand, is that softer definition of default leads to a worse scoring function. The reasoning is that banks need to model real losses, i.e., number of clients that really do not pay their obligations to the bank. Soft definition of default (e.g., defined as just 30 overdue or with a very low materiality threshold) may capture a large number of clients (in a particular retail) that just forgot to pay an exact amount on time but do not really intend not to pay and do have enough financial resources. It may be argued that poor payment discipline is already correlated to higher probability of default, but separation of the two types of the events (poor payment discipline and real lack of resources/intention not to pay the loan back) is difficult. Hence, it appears preferable to use an absorbing (hard) definition of default where clients, once marked as defaulted do not return (of course with some exceptions) to the ordinary non-defaulted status and there is a real problem. Application of a soft definition of default on the other hand, implies that many clients marked as defaulted become current, i.e., non-defaulted with observed $LGD = 0$. Moreover, this situation may repeat several times for a single client.

It would be optimal to test our hypothesis on real data. Due to the lack of real data, we use for a simulation the structural Merton-like model applied to a retail portfolio. We model the disposable income, A, as a Wiener process, $dA = \mu dt + \sigma dz$, with different initial parameters, $A(0)$, μ and σ. The soft default event is defined by the condition, Days Past Due (DPD) ≥ 60 and hard default event by DPD ≥ 180 (relaxing the regulatory 90 DPD condition). The amount of a monthly installment is constant at 1 and the event missing a payment at time t occurs if $A(t) < 1$, or when the client forgets to pay. The second independent event $Ind(t)$ (with values 0 or 1) is parameterized by a probability p characterizing payment discipline of a client set at the beginning of the simulation. We assume that the events 'forgetting to pay in a month' are independent. Hence, for example, if $p = 10\%$, then the client may with the probability 1%=10%*10%, forget to send two consecutive payments and be marked as (soft) defaulted, even though there is no problem with his ability to pay. The probability of forgetting four consecutive payments is in this situation only 0.01%, hence almost negligible. Consequently, the simulated default data contain much more 'noise' when the soft definition of default is used rather than when the hard definition of default is applied.

To simulate a history of days past due, it is sufficient to generate only end-of-month $A(t)$ like in Section 2.4. The initial attributes of a client, $A(0)$, μ, σ and p, can be interpreted as the initial disposable income, income growth potential (e.g., due to age, education, etc.), income stability (e.g., due to marital status, number of children, etc.), and payment discipline. We consider the triple, $A(0)$, μ and σ, as a proxy for full demographic data provided by a client applying for a loan. The three variables are used as the explanatory variables of our logistic regression. Banks normally approve only the applications promising sufficient probability of repayment. Hence generating the initial data, we estimate (by a separate simulation) the probability of default and accept only those with probability of default less than 15%. On the other hand, we randomly modify the initial triple to capture the fact that clients are always not completely precise in their applications, and moreover, there is generally a time lag between the time of application and the start of our simulation.

We generate, in the way described above, one set of clients to be used for the development of scoring function and another set of clients for testing. We first generate the monthly values, $A(t)$ and $Ind(t)$, for each client. The data are then used to select the soft and hard defaulted clients. The first set of data is used for a logistic regression function development with the explanatory variables, $A(0)$, μ and σ, to estimate the probability of: (1) Soft default and (2) Hard default. The regression is performed using the application STATA, and the resulting coefficients are presented in Table 4. The first scoring function would be used by a bank relying on the soft definition of default, with the second relying

Table 4: Soft and Hard Default Logistic Regression Functions

Soft Default-Based Logistic Regression

Logit Estimates	Number of Obs.	=	100.0000
	LR $\chi^2(3)$	=	685.7200
	Prob > χ^2	=	0.0000
Log Likelihood = –2,767.2986	Pseudo R^2	=	0.1102

Def_0_1_	Coef.	Std. Err.	z	P > |z|	(95% Conf. Interval)	
initial_	3.514344	0.2114122	16.623	0.000	3.099983	3.928704
drift	5.95629	0.7287103	8.174	0.000	4.528045	7.384536
std	–2.656194	0.1629243	–16.303	0.000	–2.97552	–2.336869
_cons	–1.680595	0.2891526	–5.853	0.000	–2.243404	–1.117786

Hard Default-Based Logistic Regression

Logit Estimates	Number of Obs.	=	100.0000
	LR χ^2 (3)	=	521.4000
	Prob > χ^2	=	0.0000
Log Likelihood = –1,641.1698	Pseudo R^2	=	0.1371

Def_0_1_	Coef.	Std. Err.	z	P > |z|	(95% Conf. Interval)	
initial_	4.456413	0.3213701	13.867	0.000	3.826539	5.086287
drift	10.83548	1.013854	10.687	0.000	8.848358	12.82259
std	–2.594334	0.2060633	–12.590	0.000	–2.998211	–2.190458
_cons	–2.172141	0.4258281	–5.101	0.000	–3.006749	–1.337533

on the hard definition of default. The most important output of a scoring function is how it orders (ranks) the different clients—this allows assigning ratings from bad to good to different sets of clients, and estimating their probability of default. Notice that the relative differences between the soft default and hard default scoring functions coefficients (in particular comparing the coefficients of μ). Consequently, the functions may provide quite different ranking to identical clients. A better quality of the hard default scoring function is already indicated by the R^2 coefficient (13.7% compared to 11% of the soft default scoring function). By calculating the Gini coefficient on the testing dataset, we again get a better result for the hard scoring function (72.4% compared to 71.4%).

Conclusion

In this paper, we offer several explanations of the observed dependence of the regulatory capital on the definition of default. It is clear that regulators should seek a remedy for the issue. One possibility is to strictly control the definition of default used by banks, another is to modify the regulatory formula to capture the *LGD* unexpected risk. The latter approach is preferable as a number of studies, including this one, indicate that the *LGD* risk component is significant. The fact that the current formula does not capture the *LGD* unexpected risk is a reason behind the observed sensitivity to the definition of default. Our paper provides two alternative formulas that could be used for the proposed new regulation: a complex one decomposing and combining the *PD* and *LGD* unexpected risk, and a simplified one based on a straightforward generalization of the Vasicek formula.

The current regulatory formula (probably unintentionally) motivates banks to use a softer definition of default. Our simulation indicates that such a definition may not serve well to the quality of credit risk management tools, in particular for development of scoring functions. Further research to confirm the hypothesis should be however done on real banking data. ❖

Bibliography

1. Altman E and Fanjul G (2004), "Defaults and Returns in the High Yield Bond Market: Analysis Through 2003", NYU Salomon Center Working Paper.

2. Altman E, Resti A and Sironi A (2004), "Default Recovery Rates in Credit Risk Modelling: A Review of the Literature and Empirical Evidence", *Economic Notes by Banca dei Paschi di Siena SpA*, Vol. 33, No. 2, pp. 183-208.

3. Basel Committee on Banking Supervision (1996), "Amendment to the Capital Accord to Incorporate Market Risks", Bank for International Settlements.

4. Basel Committee on Banking Supervision (1998), "The Framework for Internal Controls in Banking Organizations", Bank for International Settlements.

5. Basel Committee on Banking Supervision (2005), "Guidance on Paragraph 468 of the Framework Document", Bank for International Settlements.

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

7. Brady B, Chang P, Miu P, Ozdemir B and Schwartz D (2006), "Discount Rate for Workout Recoveries: An Empirical Study", Working Paper.

8. Credit Suisse First Boston (1997), *CreditRisk$^+$: Technical Documentation*, London.

9. Emery K (2003), Moody's Loan Default Database as of November 2003, Moody's Investors Service, December.

10. Gupton G M (2005), "Advancing Loss Given Default Prediction Models: How the Quiet Have Quickened", *Economic Notes by Banca dei Paschi di Siena SpA*, Vol. 34, No. 2, pp. 185-230.

11. Kim J and Kim H (2006), "Loss Given Default Modelling Under the Asymptotic Single Risk Factor Assumption", Working Paper, MPRA.

12. Merton Robert C (1974), "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates", *The Journal of Finance*, Vol. 29, pp. 449-470.

13. Gupton G M, Finger C C and Bhatia M (1997), *CreditMetrics: Technical Document*, J P Morgan and Co., New York.

14. Schuermann T (2004), "What Do We Know About Loss Given Default", in *Credit Risk Models and Management*, 2nd Edition, Risk Books, London.

15. Vasicek O (1987), "Probability of Loss on a Loan Portfolio", KMV Working Paper.

Reference # 37J-2009-03-03-01