JEL Classification: G21, G28, C14
Keywords: credit risk, correlation, recovery rate, regulatory capital

On Deficiencies and Possible Improvements of the Basel II Unexpected Loss Single-Factor Model*

Jiří WITZANY – University of Economics, Prague (jiri.witzany@vse.cz)

Abstract
The goal of the Basel II regulatory formula is to model the unexpected loss on a loan portfolio. The regulatory formula is based on an asymptotic portfolio unexpected default rate estimation that is multiplied by an estimate of the loss given default parameter. This simplification leads to a surprising phenomenon where the resulting regulatory capital depends on a definition of default that plays the role of a frontier between the unexpected default rate estimate and the LGD parameter, whose unexpected development is not modeled at all or is modeled only partially. We study the phenomenon in the context of single-factor models where default and loss given default are driven by one systematic factor and by one or more idiosyncratic factors. In this theoretical framework we propose and analyze a relatively simple remedy of the problem requiring that the LGD parameter be estimated as an appropriate quantile on the required probability level.

1. Introduction

The Basel II regulatory formula (see Basel, 2006) aims to provide a sufficiently robust estimate of unexpected losses on banking credit exposures that should be covered by the capital. It is a compromise between the most advanced mathematical modeling techniques and the demand for a practical implementation. One of the most important simplifications is the decision to calculate unexpected losses (*UL*) using an estimate of the Unexpected Default Rate (*UDR*) multiplied through by the expected Loss Given Default (*LGD*) parameter, i.e., as $UL = UDR \cdot LGD$. The capital requirement (*C*) as a percentage of the exposure at default (*EAD*) is then set equal to the difference between the unexpected and expected loss (*EL*), $C = UL - EL = (UDR - PD) \cdot LGD$, where *PD* is the expected default rate, i.e., the probability of default.

The goal of this article is to study, theoretically and empirically in the context of single-factor models, certain inefficiencies of the regulatory formula $UL = UDR \cdot LGD$ caused mainly by the factorization of the unexpected loss into the unexpected default rate and the (rather expected) loss given default. Moreover, based on the analysis, we want to propose certain possible improvements to the formula.

While the expected default rate estimation based on the Vasicek (1987) approach is considered to be relatively robust, the resulting estimation of the unexpected loss has been criticized for neglecting the unexpected *LGD* (or equivalently recovery) risk. It has been empirically shown in a series of papers by Altman et al. (see, for example, 2004), Gupton et al. (2000), Frye (2000b, 2003), and Acharya et al. (2007) that there is not only a significant systematic variation of recovery rates,

* The research was supported by Czech Science Foundation grant nos. 402/06/0890 and 402/09/0732.

but also a negative correlation between frequencies of default and recovery rates, or equivalently a positive correlation between frequencies of default and losses given default. Consequently, the regulatory formula significantly underestimates the unexpected loss on the targeted confidence probability level (99.9%) and the time horizon (one year). Some authors have proposed alternative unexpected loss formulas incorporating the impact of recovery rate variation.

Frye (2000a, 2000b) used a single systematic factor model with an idiosyncratic factor driving the event of default and another independent idiosyncratic factor driving the recovery rate. The loading of the systematic factor for modeling default and recovery rates may differ. The recovery rate is modeled as a normal variable truncated at 100%. Frye does not provide an analytical formula but analyzes the robustness of the loss estimates using Monte Carlo simulation for different combinations of the input parameters. The parameters are also estimated using the maximum likelihood method from the Moody's Risk Service Default database. Alternatively, Dullmann and Trapp (2004) apply the logit transformation for recovery, modeling it in the same set up as Frye.

Pykhtin (2003) considers a single systematic factor model where default is driven by a systematic factor and an idiosyncratic factor while recovery is driven not only by the systematic factor and the independent idiosyncratic factor, but at the same time by another idiosyncratic factor driving the obligor's default. The collateral (recovery) value is set to have a lognormal distribution. Pykhtin arrives at an analytic formula which requires numerical approximations of the bivariate normal cumulative distribution values. The author admits that calibration of the model is difficult.

Tasche (2004) proposes a single-factor approach modeling the loss function directly. If there is no default the value of the loss function is zero and if there is a default (the systematic factor exceeds the default threshold) the value of the loss is drawn from a distribution as a function of the systematic factor. The obligor factor is decomposed as usual into the systematic and idiosyncratic factor. In other words, the single obligor factor is used to model the event of default and the loss given default as well. Tasche proposes to model *LGD* by a beta distribution. Quantiles of the loss function conditional on the systematic factor values may be expressed as an integral over a tail of the normally distributed factor. Tasche proposes to approximate the integral using Gauss quadrature and tests the model for different *PD*, mean/variance *LGD*, and correlation values. The approach is also elaborated in Kim (2006).

This study is motivated not only by the fact that the Basel II formula significantly underestimates the unexpected credit losses on the 99.9% confidence level, but also by the observation according to which the regulatory capital requirement depends on a definition of default which in a sense puts a border line between the *PD* and *LGD* parameters. This phenomenon has been analyzed in Witzany (2009) using a Merton model-based simulation. To give a more tractable analytical explanation we will apply the Tasche and Frye single-factor models as benchmarks against which we analyze the sensitivity of the regulatory formula. At the same time, we propose a simple modification of the regulatory formula in order to eliminate the problem. We propose to preserve the formula $UL = UDR \cdot LGD$ as well as the regulatory formula for unexpected default rate (*UDR*), but to reinterpret the parameter *LGD* as the 99.9% quantile of possible portfolio loss given default values. The Basel (2005) document goes in this direction, requiring *LGD* estimates to incorporate potential economic downturn conditions and adverse dependencies between default rates and recovery rates, but fails to specify the confidence probability level of those conservative estimations. We argue that any probability level below 99.9% preserves the problem of the definition of default sensitivity (and underestimation of the 99.9% loss function percentile), while the 99.9% *LGD* quantile solves the problem under reasonable modeling assumptions. We propose a single-factor beta distribution-based technique calibrated with account-level *LGD* mean, variance, cure rate, and a correlation to obtain robust estimates of the 99.9% *LGD* quantiles. As the reinterpretation of the formula leads to significantly higher capital requirements, we propose to reduce the probability level to the more realistic 99.5% currently used by the Solvency II proposal.

2. Sensitivity of Regulatory Capital to the Definition of Default

According to Basel II the contribution of a receivable to the unexpected loss of a well-diversified portfolio as a percentage of the exposure is estimated by the formula

$$UL = UDR \cdot LGD$$

where (1)

$$UDR = \Phi\left(\frac{\Phi^{-1}(PD) + \sqrt{\rho}\cdot\Phi^{-1}(0{,}999)}{\sqrt{1-\rho}}\right)$$

The correlation ρ is set by the regulator (e.g. 15% for mortgage loans, 4% for revolving loans, and somewhere between the two values depending on *PD* for other retail loans), while the parameters *PD* and *LG*D are estimated by the bank (in the IRBA approach).

The usual *LGD* estimation approach is based on a sufficiently large historical data set of a homogenous portfolio of receivables *A* in terms of product type, credit rating, and collateralization. The receivables have been observed over the period of one year and the defaulted cases subsequently for a sufficient period (usually one to three years) to have a record $l : A \to [0,1]$ of percentage losses $l(a)$ on the exposures at default if default occurred or 0 otherwise for every $a \in A$, and an indicator function $d : A \to \{0,1\}$ of default at the one-year horizon. It seems natural to require that $d(a) = 1$ if $l(a) > 0$ as in Tasche (2004), but in practice such a condition is difficult to achieve. According to section 452 of Basel (2006) obligors that are considered to be unlikely to pay their credit obligations in full and obligors more than 90 days overdue must be marked as defaulted. Some of the obligors marked as defaulted then naturally happen to pay all their obligations back; in particular, in the case of retail obligors days overdue may just be a result of payment indiscipline, not of a real lack of income to repay the loan. Moreover, some well-collateralized receivables (e.g. mortgages) are paid back in full even if the obligor itself is not able to repay the loan. Hence, we may require only that $l(a) > 0$ implies $d(a) = 1$ but not vice versa. The *PD* and *LGD* parameters of the Basel II formula (1) can be estimated in a simplified way from the given reference data set as

$$\overline{p} = \frac{|D|}{n}$$

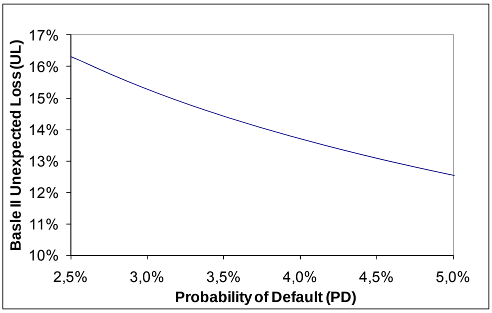
$$\overline{lgd} = \frac{\sum_{a \in A} l(a)}{|D|} = \frac{\bar{l}}{\overline{p}}$$

where $D = \{a \in A \mid d(a) = 1\}$, $n = |A|$, and $\bar{l} = \frac{1}{n}\sum_{a \in A} l(a)$.

Here we are using an equally weighted average loss given default that could be applied to a portfolio homogenous in terms of size. Let $D_H = \{a \in A \mid l(a) > 0\}$ be the set of receivables where we observed a positive loss, i.e., a hard default, and $\overline{p_H}$, $\overline{lgd_H}$ the averages as above. While the average (or expected) account-level percentage loss $\bar{l} = \overline{p} \cdot \overline{lgd} = \overline{p_H} \cdot \overline{lgd_H}$ remains unchanged, it is easy to see that $\overline{p_H} < \overline{p}$ and $\overline{lgd_H} > \overline{lgd}$ provided $D_H \subset D$. Because of the "unlikely to pay" condition in section 452 of Basel (2006) banks have certain freedom to set their own definition of default. However, if an obligor is marked as defaulted, then the probability of "not paying its obligation in full," i.e., of a hard default, should be naturally more than 50%. Consequently, the ratio $\overline{p} / \overline{p_H}$ may be in practice anywhere between 1 and 2. Banks may choose a lower days-past-due default threshold (e.g. 60 days), or a lower materiality condition (minimum amount past due implying the default), or apply different cross-default rules (default on one product implying defaults on other products with the same obligor), etc. More accounts with ultimate zero loss are then marked as defaulted. On the other hand, the definition of default must not be too soft, as noted above. Hence, given the same historical information (reference data set *A*) with the account-level average loss $EL = \bar{l}$ and choosing a different definition of default we obtain different possible values of $PD \in \left(\overline{p_H}, 2 \cdot \overline{p_H}\right)$ and $LGD = EL / PD \in \left(\overline{lgd_H}, 2 \cdot \overline{lgd_H}\right)$. Since the definition of default does not change the distribution of the total losses implied by the reference data set, the unexpected loss estimate given by (1) should remain essentially the same. However, *Figure 1* shows that this is not the case. When we set, for example, $EL = 2\%$ and let $PD \in (2.5\%, 5\%)$, then the $UL = UL(PD)$ parameter goes from 16.3% down to 12.5%. In other words, choosing the softest possible definition of default will reduce the capital requirement $C = UL - EL$ by almost 30% compared to the hard definition of default. The same effect can be observed for other combinations of *EL*, *PD*, and *LGD*. We will explain and prove the phenomenon in a general set up.

It could be argued that the problem is solved by the requirement (Basel, 2005) for *LGD* to reflect downturn economic conditions or *PD/LGD* correlation. However, this requirement given a sufficiently rich historical data set is normally implemented using only the data set $A' \subset A$ from years with economic downturn conditions and/or a high observed frequency of default. The *PD, LGD* parameters estimated from A' and *UL* calculated according to (1) will again depend on the definition of default in the same way as above.

Figure 1 Unexpected Loss According to the Basel II Formula if *EL* = *PD*·*LGD* = 2% is Fixed and *PD* Varies from 2.5% to 5% ($\rho = 15\%$)



3. Alternative Single-Factor Models

The single-factor models of Frye (2000a, 2000b), Pykhtin (2003), Tasche (2004), and others can be generally described as follows. Let the (percentage) loss of a receivable in the given time horizon be an increasing function of one systematic factor X and of a vector $\vec{\zeta}$ of idiosyncratic factors $L = L(X, \vec{\zeta})$. The factor X captures macroeconomic or other systematic influences that may develop in time, while $\vec{\zeta}$ reflects specificities of each individual obligor in a portfolio. Hence, the impact of $\vec{\zeta}$ is diversified away in a large (asymptotic) portfolio, while X remains as a risk factor. Consequently, the future unknown loss on a large portfolio can be modeled as $E[L \mid X]$ (see Gordy, 2003, for details). Since we assume that L is increasing in X the problem to find the quantiles of $E[L \mid X]$ reduces to a calculation of the quantiles of X. If x is the desired (e.g. for 99.9%) quantile of X then $UL = E[L \mid X = x]$. This is a clear advantage of the single-factor approach compared to the multi-factor approach, where we work with a vector $\vec{X}$ of systematic factors instead of one factor X and the determination of the quantiles of $E[L \mid \vec{X}]$ becomes complex.

The expression for the unexpected loss may be decomposed into two parts corresponding to the unexpected default rate and loss given default:

$$E[L \mid X = x] = P[L > 0 \mid X = x] \cdot E[L \mid L > 0, X = x]$$

Here we use the hard definition of default $D_H = 1 \Leftrightarrow L > 0$, while as explained above in practice we usually need to work with a softer definition of default. We will say that $D = D(X, \vec{\zeta}) \in \{0,1\}$ is a consistent notion of default provided $L > 0 \Rightarrow D = 1$. Then the unexpected loss may be in general decomposed as

$$UL = P[D = 1 \mid X = x] \cdot E[L \mid D = 1, X = x] \tag{2}$$

The simplest version of the single-factor model is probably the model proposed by Tasche (2004). The loss function $L = L(X, \zeta)$ is driven by one standard-normally distributed factor $Y = \sqrt{\rho} X + \sqrt{1-\rho}\,\zeta$, where X and ζ are independent standard-normally distributed, and ρ is their correlation. If L is assumed to have a cumulative probability distribution function $F_L:[0,1]\to[0,1]$ then we may express the loss function in the form $L(X,\zeta)=F_L^*\left(\Phi\left(\sqrt{\rho}X+\sqrt{1-\rho}\zeta\right)\right)$ or just $L(Y)=F_L^*\left(\Phi(Y)\right)$, where $F_L^*(z)=\inf\{l:F_L(l)\geq z\}$ is the generalized inverse of F_L. Note that if p is the probability of default, then

$$F_L(0)=P[L=0]=1-p \tag{3}$$

In a sense a more natural model has been proposed by Frye (2000a,2000b), which may in generalized form be described as follows. Let $Y_1=\sqrt{\rho_1}X+\sqrt{1-\rho_1}\zeta_1$ and $Y_2=\sqrt{\rho_2}X+\sqrt{1-\rho_2}\zeta_2$ be two standard-normally distributed factors with one systematic and two independent idiosyncratic factors. The correlations ρ_1 and ρ_2 may in general be different. The first factor Y_1 drives defaults in the model while the second Y_2 is assumed to drive losses in case of default, i.e., there is a default threshold y_D and a non-negative non-decreasing function G so that the loss function can be expressed as follows:

$$L(X,\zeta_1,\zeta_2)=\begin{cases}0 \text{ if } Y_1\leq y_D,\\ G(Y_2) \text{ otherwise}\end{cases} \tag{4}$$

If F_G is the distribution function of the random variable $G(Y_2)$, then the loss function may be again expressed as $G(Y_2)=F_G^*\left(\Phi(Y_2)\right)$.

The Pykhtin (2003) model in a sense unifies the two models. In generalized form let Y_1 be the driver of default as above; on the other hand let

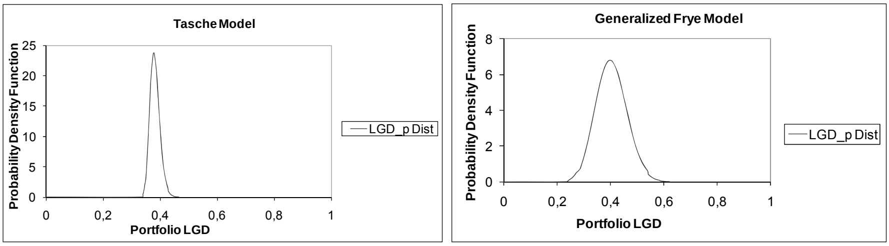
$$Y_2=\sqrt{\rho_2}X+\sqrt{1-\rho_2}\left(\sqrt{\omega}\zeta_1+\sqrt{1-\omega}\zeta_2\right) \tag{5}$$

be the driver of loss given default incorporating not only the systematic factor and a new idiosyncratic factor, but also the idiosyncratic factor from Y_1. The loss function $L(X,\zeta_1,\zeta_2)$ is expressed by (4) as in Frye's model. The approach enables us to model the fact that loss in the case of obligors default is determined by the value of assets and the specific financial situation at the time of default as well as by the workout/bankruptcy specific development.

Since we are interested in particular in unexpected loss given default modeling let us compare the three models in this respect. In the Tasche model, the unexpected loss given (hard) default conditional on the value of the systematic factor can be expressed as

$$E[L\mid L>0,X=x]=\frac{1}{1-\Phi\left(\frac{y-\sqrt{\rho}x}{\sqrt{1-\rho}}\right)}\int_{\frac{y-\sqrt{\rho}x}{\sqrt{1-\rho}}}^{\infty}L\left(\sqrt{\rho}x+\sqrt{1-\rho}\zeta\right)\phi(\zeta)d\zeta \tag{6}$$

Figure 2 Portfolio Loss Distribution in the Tasche and Generalized Frye Model
($\mu = 0.4$, $\sigma = 0.15$, $\rho = 0.15$, $p_H = 0.01$)



where $y = \Phi^{-1}\left(P[L=0]\right)$. On the other hand, for the Frye model we get a nicer formula

$$E[L \mid L>0, X=x] = E[G(Y_2) \mid X=x] = \int_{-\infty}^{+\infty} G\left(\sqrt{\rho_2}x + \sqrt{1-\rho_2}\zeta\right)\phi(\zeta)d\zeta \qquad (7)$$

since the value of $G(Y_2)$ does not depend on the idiosyncratic factor driving the default conditional on $X = x$. Regarding the Pykhtin model we get the following double integral:

$$E[L \mid L>0, X=x] = E\left[G(Y_2) \mid Y_1 > y_D, X=x\right] =$$

$$= \frac{1}{1-\Phi\left(\dfrac{y-\sqrt{\rho_1}x}{\sqrt{1-\rho_1}}\right)} \int_{\frac{y-\sqrt{\rho_1}x}{\sqrt{1-\rho_1}}}^{\infty} \int_{-\infty}^{+\infty} G\left(\sqrt{\rho_2}X + \sqrt{1-\rho_2}\left(\sqrt{\omega}\zeta_1 + \sqrt{1-\omega}\zeta_2\right)\right)\phi(\zeta_2)\phi(\zeta_1)d\zeta_2 d\zeta_1 \qquad (8)$$

This approach, if properly calibrated, represents an economically more faithful model compared to the other two. In fact, the three correlation parameters of the model may be linked to the default correlation, the loss given default correlation, and the default – loss given default correlation. Nevertheless, since the model is difficult to calibrate and as it is computationally complex we will focus on the Tasche and Frye models. In fact, the two models are special cases of the Pykhtin model: set in (5) $\omega = 1$ and $\rho_1 = \rho_2$ for the Tasche model and $\omega = 0$ for the Frye model.

Hence, given an account level distribution of *LGD* modeled by the function *G*, formulas (6)–(8) give the transformed portfolio level (average) $LGD = E[L \mid L>0, X]$ driven only by the systematic factor *X*. Since *X* is assumed to follow a standard normal distribution we obtain a distribution of the average (asymptotic) portfolio *LGD*. To model account level loss given default we will use the beta distribution with minimum 0 and maximum 1 determined by its mean μ and standard deviation σ. *Figure 2* shows the transformed distributions of the portfolio *LGD* in the two models given that $\mu = 0.4$, $\sigma = 0.15$, and $\rho = 0.15$. In the Tasche model we used the probability of default $p_H = 0.01$. The distribution of the portfolio *LGD* does not depend on the probability of default in the case of the Frye model; it does in the case of the Tasche model, but the shape appears different but low if we test different values of p_H. It is obvious that the variance of the portfolio *LGD* is much lower in the case of Tasche model than in the case of the Frye model. In fact, the standard deviation of the former is approximately 4.5% while the standard deviation of the latter is 9.7%.

The Tasche model in spite of its appealing simplicity turns out to be inappropriate if the unexpected loss is to be factorized according to (2). If the correlation is calibrated for the unexpected default rate calculation (i.e., fitting the correlation parameter to a time series of observed rates of default) then the portfolio *LGD* variance is too low compared to the empirical observations. This follows, for example, from the study of Frye (2003) showing that *LGD* in bad years is almost twice the *LGD* in good years, or Frye (2000b) where the Frye model correlation coefficients ρ_1 and ρ_2 calibrated to a Moody's database appear to be almost equal. Another disadvantage of the Tasche model is that *PD* estimations cannot be separated from *LGD* estimations. On the other hand, the Frye model can be calibrated separately according to the volatility of frequencies of default over a number of years and according to the volatility of portfolio *LGD* observed in a time series.

4. An Analysis of the Sensitivity of the Regulatory Capital Formula

The phenomenon described in Section 2 has been partially explained in Witzany (2009) using a Merton (1974) model-based simulation where we argued that a softer definition of default terminates the asset value stochastic process sooner than a hard definition of default, thus reducing the variance of the total losses due to the average LGD set at the time of default.

To provide a better analytical explanation of the difference between the real loss quantile and the regulatory loss quantile estimation (and its dependence on the definition of default) we will use the Frye and Tasche one-factor models as benchmarks against which we compare the regulatory unexpected loss estimation. In both cases the unexpected loss we need to estimate is given by

$$UL = E[L \mid X = x] = P[D = 1 \mid X = x] \cdot E[L \mid D = 1, X = x] \tag{9}$$

where $x = \Phi^{-1}(\alpha)$ and α is the regulatory probability level 0.999.

Let us consider the Tasche model first. Let $p_H = P[L > 0] = P[D_H = 1]$ be the probability of "hard default" ($D_H = 1 \Leftrightarrow L > 0$). Note that according to (3) $F_L(0) = 1 - p_H$, where F_L is the distribution function of L. Consequently, $L = L(Y) = F_L^*\left(\Phi(Y)\right) = 0$ for $Y \le \Phi^{-1}\left(1 - p_H\right)$ and $L = F_L^*\left(\Phi(Y)\right) > 0$ for $Y > \Phi^{-1}(1 - p_H)$. Hence $y_h = \Phi^{-1}(1 - p_h)$ is the hard default critical point for the factor Y. As already explained, banks naturally use a softer definition of default. Let us assume that such a definition of default is represented by another critical point $y < y_H$. With this new definition of default $D = 1 \Leftrightarrow Y > y$ the loss (given default) may be zero with a positive probability, $P[L = 0 \mid D = 1] = P\left[y < Y \le y_H\right] > 0$. This new definition of default D does not change anything on the unexpected loss $UL = E[L \mid X = x]$, where the notion of default is irrelevant. On the other hand, the regulatory estimation of unexpected loss turns out to be different for the two definitions of default:

$$
\begin{aligned}
UL_{reg}^{H} &= P\left[Y > y_H \mid X = x\right] \cdot E\left[L \mid Y > y_H\right] \\
UL_{reg}(y) &= P[Y > y \mid X = x] \cdot E[L \mid Y > y]
\end{aligned} \tag{10}
$$

Note that there is a difference between the second part of the "real" unexpected loss (9), where the expected value is conditional upon the systematic factor $X = x$ and the second part of the regulatory formula (10), where the expected value is conditional only upon the event of default. It is shown in the *Technical Appendix* that the regulatory unexpected loss (10) is indeed less than the "real" unexpected loss (9) and moreover that the unexpected loss estimate $UL_{reg}(y)$ is an increasing function of y, i.e., it gets smaller with a softer definition of default:

$$
UL_{reg}(y) < UL_{reg}^{H} < UL, \text{ for } y < y_H \tag{11}
$$

In case of the Frye model we have

$$
\begin{aligned}
UL_{reg}^{H} &= P\left[Y_1 > y_H \mid X = x\right] \cdot E\left[G\left(Y_2\right) \mid Y_1 > y_H\right] \\
UL_{reg}(y) &= P\left[Y_1 > y \mid X = x\right] \cdot E\left[G\left(Y_2\right) \mid Y_1 > y\right]
\end{aligned} \tag{12}
$$

and the *Technical Appendix* again shows that the property (11) holds. Hence, we have demonstrated that the numerically observed deficiency of the Basel II formula (*Figure 1*) generally holds in the context of the Tasche or Frye model.

5. Improved Regulatory Formula

In Section 3 we gave a general definition of the one (systematic) factor model.

We have seen that if D is a consistent notion of default, then the loss may be decomposed to $UL = P[D = 1 \mid X = x] \cdot E[L \mid D = 1, X = x]$. It is not in general obvious that the conditional default rate $P[D = 1 \mid X = x]$ as well as the conditional loss given default $E[L \mid D = 1, X = x]$ are increasing functions of x. However, this is a property of the aforementioned one-factor models (Tasche, Frye, Pykhtin). Consequently, it is correct in the context of one-factor models where both the conditional PD and conditional LGD are increasing functions of the systematic factor X to state that

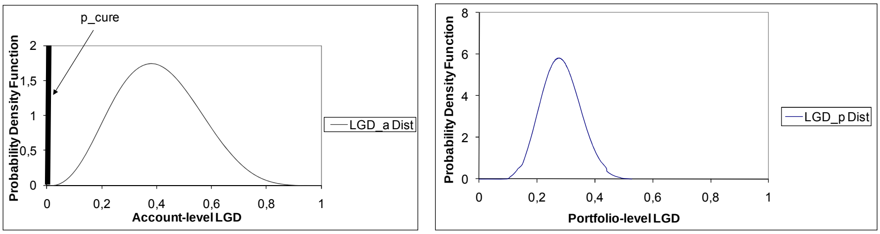
$$
UL = UDR \cdot ULGD \tag{13}
$$

where $UDR = P[D = 1 \mid X = x]$ is the α-quantile of possible default rates and $ULGD =$ $= E[L \mid D = 1, X = x]$ the α-quantile of possible LGDs, with x being the α-quantile of X. The unexpected default rate is underestimated by the regulatory formula (1) as shown in Section 4, but we can improve it significantly by requiring that LGD be not the expected loss given default but the unexpected portfolio level loss given default ($ULGD$) on the same 99.9% probability level.

For practical applications we propose to use the generalized Frye model. In the notation of Section 3 we just need to estimate

$$
\begin{aligned}
ULGD &= E\left[L \mid Y_1 > y_D, X = x\right] = E\left[G\left(Y_2\right) \mid X = x\right] = \\
&= \int_{-\infty}^{+\infty} G\left(\sqrt{\rho_2}\, x + \sqrt{1-\rho_2}\, \zeta\right) \phi(\zeta)\, d\zeta
\end{aligned} \tag{14}
$$

Figure 3 Account and Transformed Portfolio Loss Given Default Distribution ($p_{cure} = 0.3$, $\mu = 0.4$, $\sigma = 0.15$, $\rho = 0.1$)



To complete our model we need to propose an appropriate loss given default function *G*. We may follow Witzany (2009) by specifying that $G(Y)$ has a beta distribution calibrated to the empirical mean and standard deviation. However, since we consider that the default definition may be in a practice a softer one with a non-negligible percentage p_{cure} of receivables marked as defaulted being cured, i.e., ultimately ending up with zero loss, we extend the model as follows. Let μ and σ be the mean and standard deviation of the observed positive losses assumed to have a beta distribution with minimum 0 and maximum 1. Let $B(t,\mu,\sigma)$ be the corresponding cumulative beta distribution function on [0,1], then the mixed distribution function incorporating the possibility of cures is defined by

$$F(t) = p_{cure} + \left(1 - p_{cure}\right) \cdot B(t,\mu,\sigma) \tag{15}$$

Finally setting $G(Y) = F^{*}\left(\Phi(Y)\right)$ we see that *G* has the distribution given by *F*. To estimate the 99.9% *ULGD* we just need to evaluate (14) numerically for $x = \Phi^{-1}(0.999)$. *Figure 3* illustrates the account-level *LGD* density function given by *F* for a given set of parameters (with mass weight p_{cure} at 0) and the transformed portfolio level *LGD* density function of $E[L \mid X]$ derived from (14) .

The unexpected loss estimated using the described technique is, however, still sensitive to the definition of default, although *Figure 5* shows that the sensitivity is moderate and opposite compared to the regulatory capital (the unexpected loss estimation increases with softer definition of default). Another applicable solution is to adjust the probability of (conventional soft) default *p* with the observed probability of cures p_{cure} and then apply the hard default-based formula

$$UL = UDR\left(p\,(1 - p_{cure})\right) \cdot ULGD(\mu,\sigma) \tag{16}$$

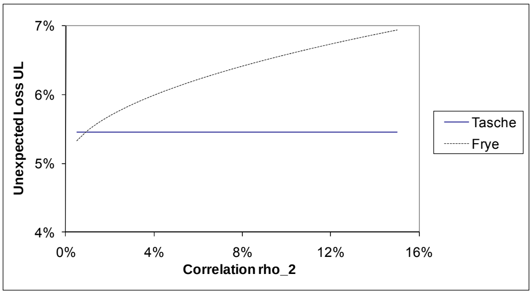
where the unexpected loss given (hard default) is estimated according to (14) using the beta distribution with mean μ and standard deviation σ .

6. Numerical Study

We are going to compare the values of regulatory unexpected loss in different scenarios: unexpected loss in the Tasche model and unexpected loss in the Frye model with and without the cure rate. The scenarios are specified by the probability

Figure 4 Comparison of the 99.9% Unexpected Loss in the Tasche Model and Frye Model Depending on ρ_2

(with $p_H = 1\%$, $\mu = 45\%$, $\sigma = 15\%$, $\rho = 15\%$, $p_{cure} = 0\%$)



of hard default p_H, loss given (hard) default mean μ, and standard deviation σ, correlations $\rho = \rho_1$ and ρ_2, and the cure rate p_{cure}. Formally, we have two definitions of default, D_H and D_S, so that $P[D_H = 1 \mid D_S = 1] = 1 - p_{cure}$. Consequently, the probability of soft default is recalculated as $p_S = p_H / (1 - p_{cure})$. Moreover, $D_S = 1 \& D_H = 0 \Rightarrow L = 0$ and so the loss given soft default mean can be expressed as

$$\mu_S = E[L \mid D_S = 1] = E[L \mid D_H = 1](1 - p_{cure}) = \mu(1 - p_{cure})$$

and the *LGD* standard deviation as

$$\sigma_S = \sqrt{E\left[L^2 \mid D_H = 1\right](1 - p_{cure}) - \mu^2 (1 - p_{cure})^2} = $$
$$= \sqrt{(1 - p_{cure})(\sigma^2 + p_{cure}\mu^2)}$$

since $E\left[L^2 \mid D_H = 1\right] = \sigma^2 + \mu^2$.

Figure 4 shows how it is difficult to align the Tasche and Frye models. If we fix $\rho = 15\%$ as the correlation related to the unexpected default rate and the other parameters as specified below, then in order to obtain the same unexpected loss in the Frye model as in the Tasche model the *LGD* correlation must be reduced to 1% or even less. Such a calibration is in contradiction with empirical studies such as Frye (2000b, 2003). Thus, we focus rather on the Frye modeling approach.

Figure 5 compares the regulatory unexpected loss estimate with different estimation approaches based on the Frye model explained at the end of the previous section. While the UL_reg curve shows the regulatory unexpected loss declining with the cure rate going up, the UL_Frye_S curve based on the beta distribution calibrated to μ_S and σ_S turns out to be increasing. The dependence is weaker if we use the mixed beta distribution (15) (UL_Frye_S_c) and logically it is fully eliminated (UL_Frye_H) when we use (16). We consider the positive sensitivity of UL_Frye_S_c to the cure rate, motivating banks to use a harder definition of default, to be much more acceptable than the negative sensitivity of UL_reg, which motivates banks to use

Figure 5 Regulatory and Frye Model-Based 99.9% Unexpected Loss Estimations Depending on p_{cure}

(with $p_H = 2.5\%$, $\mu = 70\%$, $\sigma = 15\%$, $\rho = \rho_1 = 15\%$, $\rho_2 = 8\%$)

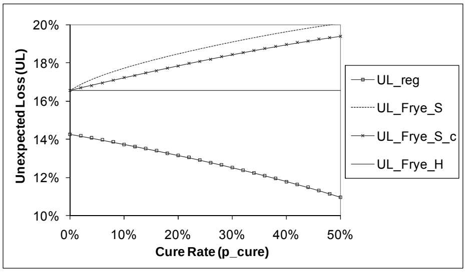
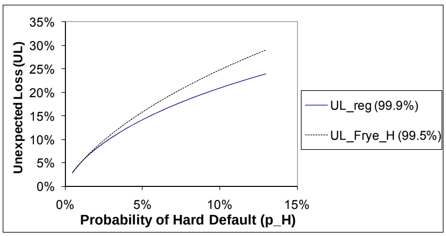
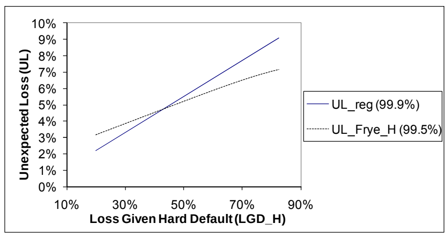


Figure 6 Sensitivity of the 99.9% Regulatory and 99.5% Frye Model Unexpected Loss to p_H and μ

(with, $\sigma = 22\%$, $\rho = \rho_1 = 15\%$, $\rho_2 = 13\%$, $\mu = 45\%$ in the first graph, and $p_H = 1\%$ in the second)





a softer definition of default, which is not usually ideal for credit risk modeling as pointed out in Witzany (2009). The problem is fully solved by recalculating the probability of soft default to the probability of hard default, which, however, might be rather difficult to communicate in practice.

Incorporation of the unexpected loss given default into the unexpected loss calculation significantly increases the value compared to the regulatory unexpected loss. If we wanted to set up the model in line with the current regulatory capital values we could consider reducing the (artificially high) regulatory level. It turns out that the level of 99.5% (proposed, for example, for Solvency II) leads to comparable values of the regulatory UL on the 99.9% level and the Frye model UL on the 99.5% level. The relationship nevertheless depends on the σ and ρ_2 values. *Figure 6* compares the sensitivity of the 99.9% regulatory UL and 99.5% Frye model UL (16) to the probability of default and expected loss given default, other parameters being fixed. The 99.5% Frye UL turns out to be more sensitive than the regulatory UL with respect to the probability of default but less than the expected loss given default. Hence, by appropriate recalibration of the confidence level we do not obtain the same unexpected loss estimations in all scenarios, but using the proposed model we obtain a better correspondence between the real risk and the economic capital, more robust calculations, and at the same time an overall comparable average level of capital.

7. Conclusion

We have demonstrated and analytically explained that regulatory capital according to the Basel II formula is sensitive to the definition of default. We have seen in Section 5 that the problem may be relatively simply theoretically solved in the context of general single-factor models requiring that the LGD parameter be reinterpreted as the 99.9% percentile of possible losses given default (or generally using the same percentile as for the unexpected default rate). We have considered three particular one (systematic) factor models and concluded that the one with two idiosyncratic factors proposed by Frye is the most appropriate to implement in practice. The best results are provided by the model where the observed probability of soft default is adjusted using the cure rate to obtain the probability of hard default (which can be fully determined only ex post). Since the extended model gives higher unexpected loss values the confidence level can be recalibrated to a lower value (e.g. 99.5%) to achieve comparable capital levels. The resulting formula, compared to the regulatory formula, provides more robust and economically more faithful estimates of unexpected credit losses.

Technical Appendix

The goal of the appendix is to demonstrate property (11) stated in Section 4 in the case of the Tasche and Frye models.

Let us start with the Tasche model and show that $UL_{reg}(y) < UL$. Let $p = 1 - \Phi(y)$ be the probability of soft default. Note that

$$P[Y > y \mid X = x] = P\left[\sqrt{\rho}x + \sqrt{1-\rho}\zeta > y\right] = \Phi\left(\frac{\Phi^{-1}(p) + \sqrt{\rho}x}{\sqrt{1-\rho}}\right) \quad (17)$$

coincides with the regulatory formula for the unexpected default rate. The difference between (10) and (9) lies in the second part of formula (10), i.e., in $E[L \mid Y > y]$, where the regulation in general requires an average loss given default in the sense of the discussion above while the "real" unexpected loss (9) can be decomposed as

$$E[L \mid X = x] = P[Y > y \mid X = x] \cdot E[L \mid X = x, Y > y]$$

It appears obvious that

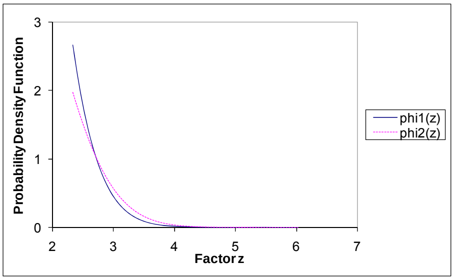
$$E[L \mid X = x, Y > y] > E[L \mid Y > y] \quad (18)$$

although the full proof is unfortunately rather technical. The right-hand side of the inequality (18) can be written as

$$E[L \mid Y > y] = \frac{1}{1-\Phi(y)} \int_y^{\infty} L(z)\phi(z)\,dz = \int_y^{\infty} L(z)\phi_1(z)\,dz$$

where $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$ is the standard normal distribution density function and $\phi_1(z) = \frac{\phi(z)}{1-\Phi(y)}$. The left-hand side of (18) equals

Figure 7 Density Functions ϕ_1 and ϕ_2



$$E[L \mid X = x, Y > y] = \frac{1}{1-\Phi\left(\frac{y-\sqrt{\rho}x}{\sqrt{1-\rho}}\right)} \int_{\frac{y-\sqrt{\rho}x}{\sqrt{1-\rho}}}^{\infty} L\left(\sqrt{\rho}x + \sqrt{1-\rho}\zeta\right)\phi(\zeta)\,d\zeta =$$

$$= \frac{1}{1-\Phi\left(\frac{y-\sqrt{\rho}x}{\sqrt{1-\rho}}\right)} \int_{y}^{\infty} L(z)\phi\left(\frac{z-\sqrt{\rho}x}{\sqrt{1-\rho}}\right)\frac{dz}{\sqrt{1-\rho}} = \int_{y}^{\infty} L(z)\phi_2(z)\,dz \tag{19}$$

where

$$\phi_2(z) = \frac{\phi\left(\frac{z-\sqrt{\rho}x}{\sqrt{1-\rho}}\right)}{\sqrt{1-\rho}\cdot\left(1-\Phi\left(\frac{y-\sqrt{\rho}x}{\sqrt{1-\rho}}\right)\right)}$$

Both densities $\phi_1(z)$ and $\phi_2(z)$ are normalized over the interval $[y,+\infty)$, hence to show that

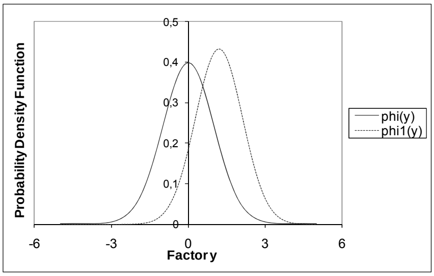
$$\int_{y}^{\infty} L(z)\phi_1(z)\,dz < \int_{y}^{\infty} L(z)\phi_2(z)\,dz \tag{20}$$

we need to analyze the relationship between the two densities. It follows from the properties of the normal distribution density that (provided $x > 0$ and $\rho > 0$) there is an $\tilde{y} > y$ so that $\phi_1(z) > \phi_2(z)$ on $[y,\tilde{y})$ and $\phi_1(z) < \phi_2(z)$ on $(\tilde{y},+\infty)$. See *Figure 7* for an illustration with $y = \Phi^{-1}(0.99)$, $x = \Phi^{-1}(0.999)$, and $\rho = 0.1$. Provided $L(z)$ is an increasing function (and not constant on $[y,+\infty)$) the inequality (18) follows immediately.

It is in fact much easier to show that the regulatory unexpected loss is less than the Frye model unexpected loss. In this case we just need to prove that

$$E\left[G(Y_2)\right] < E\left[G\left(\sqrt{\rho_2}X + \sqrt{1-\rho_2}\zeta\right) \mid X = x\right] \tag{21}$$

Figure 8 Density Functions ϕ and ϕ_1



with the notation from Section 3. The left-hand side simply equals $\int_{-\infty}^{\infty} G(y)\phi(y)\,dy$, while the right-hand side can after substitution be written as $\int_{-\infty}^{\infty} G(y)\phi_1(y)\,dy$, where $\phi_1(y) = \phi\left(\frac{y-\sqrt{\rho}x}{\sqrt{1-\rho}}\right) / \sqrt{1-\rho}$.

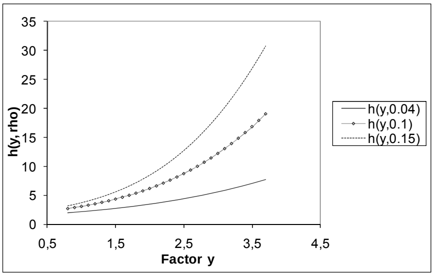
For $\rho > 0$ it can be verified that the function $\phi_1(y) < \phi(y)$ on interval $(-\infty, y_0)$ and $\phi_1(y) > \phi(y)$ on $(y_0, +\infty)$ (see *Figure 8*). Consequently, it again holds provided G is non-decreasing and strictly increasing on a non-trivial interval.

Next we want to show that the function $UL_{reg}(y) = P[Y > y \mid X = x] \cdot E[L \mid Y > y]$ defined according to (10) is an increasing function of $y \le y_H$ for a certain range of feasible values for y and ρ. Since $E[L \mid Y > y] = \frac{E[L]}{P[Y > y]}$ we just need to show that the ratio between the unexpected loss $UL_{reg}(y)$ and the expected loss $E[L]$ not depending on y

$$h(y,\rho) = \frac{UL_{reg}(y)}{E[L]} = \frac{P[Y > y \mid X = x]}{P[Y > y]} = \frac{1-\Phi\left(\frac{y-\sqrt{\rho}x}{\sqrt{1-\rho}}\right)}{1-\Phi(y)} \tag{22}$$

is an increasing function of y. Note that the equation (22) is identical for the Tasche and Frye models. Unfortunately, we cannot prove generally that the function $h(y,\rho)$ is increasing in $y > 0$ for any given correlation $\rho > 0$. In fact, it is not increasing on $(0,\infty)$ with $\rho > 0$, since $\frac{y-\sqrt{\rho}x}{\sqrt{1-\rho}} > y$ and so $h(y,\rho) < 1$ for large y while clearly $h(y,\rho) > 1$ for smaller values of $y > 0$. However, it can be shown using numerical approximations that the function is increasing over a range of admissible values for y and ρ. *Figure 9* shows the function (22) strongly increasing with the values

Figure 9 The Ratio between the Unexpected Loss $UL_{reg}(y)$ and the Expected Loss $E[L]$ increases with y



$x = \Phi^{-1}(0.999)$ and $\rho = 0.4, \rho = 0.1, \rho = 0.15$, and for $y = \Phi^{-1}\left(1 - p_D\right)$ over the range [0.8, 3.7] corresponding to admissible *PD* values in the interval [0.01%, 21.2%].

Consequently, we have demonstrated that indeed:

$$UL_{reg}(y) < UL_{reg}^{H} < UL, \text{ for } y<y_H$$

for both models.

REFERENCES

Acharya VV, Bharath S, Srinivasan A (2007): Does Industry-wide Distress Affect Defaulted Firms? – Evidence from Creditor Recoveries. *Journal of Financial Economics*, 85(3):787–821.

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

Altman E, Resti A, Sironi A (2004): Default Recovery Rates in Credit Risk Modeling: A Review of the Literature and Empirical Evidence. *Economic Notes by Banca dei Paschi di Siena SpA*, 33(2):183–208.

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

Basel Committee on Banking Supervision Paper (2005): *Guidance on Paragraph 468 of the Framework Document.* Bank for International Settlements.

Brady B, Chang P, Miu P, Ozdemir B, Schwartz D (2006): *Discount Rate for Workout Recoveries: An Empirical Study.* SSRN working paper.

Dullman K, Trapp M (2004): Systematic Risk in Recovery Rates – An Empirical Analysis of U.S. Corporate Credit Exposures. *EFWA Basel Paper*.

Emery K (2003): *Moody's Loan Default Database as of November 2003*. Moody's Investors Service.

Frye J (2000a): Collateral Damage. *Risk,* 13(4):91–94.

Frye J (2000b): Depressing recoveries. *Risk,* 13(11):106–111.

Frye J (2003): A false sense of security, *Risk,* 16(8):63–67.

Gordy MB (2003): A risk-factor model foundation for ratings-based bank capital rule. *Journal of Financial Intermediation*, 12:199–232.

Gupton GM (2005): Advancing Loss Given Default Prediction Models: How the Quiet Have Quickened. *Economic Notes by Banca dei Paschi di Siena SpA,* 34(2):185–230.

Gupton GM, Gates D, Carty L (2000): Bank loan losses given default. *Moody's Global Credit Research*, Special Comment.

Kim J, Kim H (2006): Loss Given Default Modelling under the Asymptotic Single Risk Factor Assumption. *MPRA, Working Paper.*

Merton RC (1974): On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. The *Journal of Finance*, 29:449–470.

Pykhtin M (2003): Unexpected recovery risk. *Risk*, 16(8):74–78.

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

Tasche D (2004): The single risk factor approach to capital charges in case of correlated loss given default rates. *Deutsche Bundesbank, Working paper*.

Vasicek O (1987): Probability of Loss on a Loan Portfolio. *KMV Working Paper*, February.

Witzany J (2009): Basel II Capital Requirement Sensitivity to the Definition of Default. *Icfai University Journal of Financial Risk Management*, 6(1):55–75.