Default and recovery correlations – a dynamic econometric approach

Integrating coherences between defaults and loss given default (LGD) is postulated by Basel II. If there is a positive correlation between the two, separate models for each lead to biased estimates for the LGD parameters, and the economic loss is overestimated. ***Alfred Hamerle***, ***Michael Knapp*** *and* ***Nicole Wildenauer*** *show that the bias vanishes if a simultaneous approach is used, leading to lower predicted LGDs and thus lower regulatory and economic capital requirements*

In credit risk models, the loss given default (LGD)[1] is either incorporated deterministically (as in Credit Risk+) or stochastically (as in CreditMetrics). In the latter case, the LGD may be drawn from a beta distribution.[2] In both cases, no correlation between default and LGD is considered.

In economic downturns, not only do probabilities of default (PDs) increase, but recovery rates also decrease. This pattern can be seen in historical data (see, for example, Frye, 2000, or Altman, *et al.*, 2003). To incorporate this relationship into credit risk models, several basic approaches have been proposed. Frye (2000) and Pykthin (2003) give first insights into a single-factor model for PD and recovery rate containing systematic and idiosyncratic risk. Düllmann & Trapp (2004) compare various transformations for the recovery rate. All of these authors use a single-factor model for PD as well as for recovery rate. They assume that the systematic risk that drives PDs influences recovery rates in the same way. Frye (2000) and Düllmann & Trapp (2004) estimate the PD model and use the realisations of the systematic risk factor as an input factor to the recovery rate model. Then the intercept, the parameter for the systematic and the idiosyncratic risk are estimated using a maximum-likelihood approach. Chava, Stefanescu & Turnbull (2006) use issuer-specific, time-dependent data to model default and recovery rates simultaneously. They also use the same systematic risk factor for defaults and recovery rates.

Our contribution includes four new aspects. First, we allow the systematic risk factors for PDs and recovery rates or LGDs to be different, leading to a more general approach. Second, we split the systematic risk in PD and LGD into an observable and an unobserved part. The observable part can be modelled using macroeconomic variables. This leads to a decreasing impact of the unobservable systematic risk factors. Third, the estimation of PD and LGD in separate models leads to biased estimates of the LGD parameters and, therefore, to biased predictions of the portfolio loss.[3] Risk measures such as value-at-risk (VAR) or conditional VAR (CVAR) tend to be overestimated. In this article, a simultaneous approach is developed to overcome this problem. Fourth, individual time-dependent data of issuers and bonds is used to estimate PD and LGD and the loss of a portfolio.[4]

This article is structured as follows. The next two sections show the modelling of PD and LGD. Then, we give the maximum likelihood equation to calculate the parameter estimates of the joint model, where PD, LGD and the correlation parameter of the PD model and the LGD model are estimated simultaneously. Then the simulation approach for the prediction of the loss distribution is shown. Then the data for the empirical analysis is shown. Following that, the results from the joint modelling of the PD and the LGD model are represented. Then the empirical loss distributions of the separate and the joint modelling of PD and LGD are interpreted. We then conclude.

PD model

The states 'default' and 'non-default' of obligor i in period t – in most applications, one year – are modelled using the indicator variable D_{it}, that is:

$$D_{it} = \begin{cases} 1 & \text{if borrower } i \text{ defaults in period } t \\ 0 & \text{otherwise} \end{cases}$$

where $i \in N_t$ and $t = 1, \ldots, T$. N_t denotes the 'risk set' consisting of all obligors of the portfolio who did not default at the beginning of period t.

For instance, in the model, the default event may be triggered when a metric variable S_{it} ($i \in N_t$, $t = 1, \ldots, T$) falls below a prescribed threshold c_{it} at a particular point in time (in the observed period) t:

$$S_{it} < c_{it} \Leftrightarrow D_{it} = 1 \quad (1)$$

S_{it} can be interpreted as the standardised return of a firm's assets. The random variable S_{it} is assumed to be latent and unobservable. The seminal works of Merton (1974, 1977) and Black & Scholes (1973) laid down the fundamentals for this approach. Here, the Basel II single-factor model is assumed for the random variable S_{it}, which triggers the default event, that is:

$$S_{it} = w F_t + \sqrt{1 - w^2} U_{it} \quad (2)$$

where $i \in N_t$ and $t = 1, \ldots, T$. F_t represent independent standard normally distributed, systematic risk components, that is, components that have an impact on all firms at a specified time and thus are not diversifiable. Idiosyncratic (and hence diversifiable) risk drivers U_{it} are also assumed to be standard normally

[1] *The LGD is one minus the recovery rate.*
[2] *See Gupton, Finger & Bhatia (1997) and Credit Suisse First Boston (1997).*
[3] *The authors would like to thank an anonymous referee for helpful comments on this subject.*
[4] *Rösch & Scheule (2005) calculate empirical correlations in a simultaneous approach as well, but use aggregated data.*

distributed. Moreover, the unsystematic risks of different firms are assumed to be independent of each other and, of the systematic risk factors, F_t. w denotes the exposure to a common risk factor. w^2 denotes the correlation of S_{it} and S_{kt} of two borrowers i and k and is often referred to as the asset correlation.

Together with equation (1) and given threshold c_{it} we obtain the conditional PD in probit specification given the systematic risk factor F_t:

$$\begin{aligned}\lambda_{it}(F_t) &= P(S_{it} < c_{it} | F_t) \\ &= P\left(U_{it} < \frac{c_{it} - wF_t}{\sqrt{1-w^2}}\right) = \Phi\left(\frac{c_{it} - wF_t}{\sqrt{1-w^2}}\right)\end{aligned} \tag{3}$$

where $\Phi(\cdot)$ denotes the distribution function of the standard normal distribution.

The unconditional PD is obtained by integrating with respect to F_t, that is:

$$\lambda_{it} = E(\lambda_{it}(F_t)) = \int_{-\infty}^{\infty} \lambda_{it}(f_t)\varphi(f_t)df_t = \Phi(c_{it}) \tag{4}$$

where $\varphi(\cdot)$ denotes the density function of the standard normal distribution.

Next, observable obligor-specific and systematic risk factors can be integrated into the modelling approach. The observable components of systematic default risk capture changes in the macroeconomic environment, specifically cyclical developments, and are comprised in the vector $\mathbf{z}_{1,t-1}$ described below.[5] The key variables are macroeconomic indicators such as interest rates, the unemployment rate, the gross domestic product (GDP) growth rate, etc. As a result of this analysis, the major macroeconomic risk drivers affect PD with a time lag of at least one year. As the values of these risk factors are known when the prediction of PD is made, a further source of uncertainty is eliminated. Obligor-specific risk can be integrated into the form of rating information, size, legal form or age of a company. It is summarised in the vector $\mathbf{x}_{i,t-1}$. The corresponding parameter vectors for $\mathbf{x}_{i,t-1}$ and $\mathbf{z}_{1,t-1}$ are denoted by $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}_1$, respectively. Here, individual and time-dependent default thresholds c_{it} are determined for each obligor:

$$c_{it} = \beta_0 + \boldsymbol{\beta}'\mathbf{x}_{i,t-1} + \boldsymbol{\gamma}_1{}'\mathbf{z}_{1,t-1} \tag{5}$$

The special case considered by, for example, Düllmann & Trapp (2004) and others is given by the constant threshold:

$$\mathrm{c}_{\mathrm{it}} = \beta_0$$

for all i and t.

Taking (3) into account, this leads to the following specification for a certain risk segment, for example, a sector:

$$\lambda_{it}(f_t) = \Phi\left(\frac{\beta_0 + \boldsymbol{\beta}'\mathbf{x}_{i,t-1} + \boldsymbol{\gamma}_1{}'\mathbf{z}_{1,t-1} - wf_t}{\sqrt{1-w^2}}\right) \tag{6}$$

From (6) it can be seen that the PD model is a generalised linear mixed-effects model with probit link function. The parameters of the PD model can be estimated by the maximum likelihood method.[6]

LGD model

Here, we model the LGD rather than the recovery rate. More precisely, we consider the logit transformed LGD $Y_{t(j)}$ of the jth defaulted bond in period t:

$$Y_{t(j)} = \log\frac{LGD_{t(j)}}{1 - LGD_{t(j)}} \tag{7}$$

Written in terms of the recovery rate $R_{t(j)}$, we obtain:

$$Y_{t(j)} = \log\frac{1 - R_{t(j)}}{R_{t(j)}} = -\log\frac{R_{t(j)}}{1 - R_{t(j)}} \tag{8}$$

where $LGD_{t(j)} = 1 - R_{t(j)}$. The logit transformation of the recovery rate is also proposed by Schönbucher (2003) and Düllmann & Trapp (2004).[7]

We use the index $t(j)$ to indicate that the jth defaulted bond in period t is different from the jth defaulted bond in another period s ($t \neq s$, $j = 1, \ldots, m_t$). m_t denotes the number of defaulted bonds in period t.

In analogy to the model for the PD, the following approach for the transformed LGD $Y_{t(j)}$ is specified:[8]

$$Y_{t(j)} = \mu + b_1 G_t + b_2 E_{t(j)} \tag{9}$$

where μ denotes the mean of the transformed LGD $Y_{t(j)}$.[9] Now the systematic risk factor is denoted with G_t, which is assumed to be independent standard normally distributed. The idiosyncratic risk drivers $E_{t(j)}$ are assumed to be standard normally distributed and independent of each other and of the systematic risk factor G_t. b_1 denotes the impact of the systematic risk factor G_t, and b_2 denotes the impact of the idiosyncratic risk factor $E_{t(j)}$. As an extension to Frye (2000), Pykthin (2003) and Düllmann & Trapp (2004), we assume there may be different unobservable systematic risk factors F_t and G_t influencing PD and LGD.

We use an extended version where the (transformed) LGD is explained by observable issuer- and debt-specific factors $\mathbf{v}_{t-1(j)}$ such as debt rating, seniority, guarantee of a second entity, maturity, etc. The economic conditions can be incorporated by macroeconomic factors $\mathbf{z}_{2,t-1}$, which may be different from the ones used in the PD model. Hence, μ transforms to $\mu_{t(j)}$ and can be written as:

$$\mu_{t(j)} = \alpha_0 + \mathbf{a}'\mathbf{v}_{t-1(j)} + \tilde{\mathbf{a}}_2{}'\mathbf{z}_{2,t-1} \tag{10}$$

Taking (9) into account, we obtain:

$$Y_{t(j)} = \alpha_0 + \boldsymbol{\alpha}'\mathbf{v}_{t-1(j)} + \boldsymbol{\gamma}_2{}'\mathbf{z}_{2,t-1} + b_1 G_t + b_2 E_{t(j)} \tag{11}$$

The corresponding parameter vectors for $\mathbf{v}_{t-1(j)}$ and $\mathbf{z}_{2,t-1}$ are denoted by $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}_2$, respectively. α_0 is an intercept. The LGD model is a linear mixed-effects model for the logit transformed LGD (see McNeil & Wendin, 2005).

Joint model estimation

Dependence between default rates and LGD can be found in empirical data.[10] The question is: how to insert this coherence?

[5] *The index one at $z_{1,t-1}$ is to distinguish the vector of macroeconomic values of the PD model from the one in the LGD model.*

[6] *See, for example, Hamerle,* et al. *(2005) or McNeil & Wendin (2005).*

[7] *This transformation ensures that estimated and predicted recovery rates and LGD are between zero and one. We model LGD instead of recovery rates. (8) shows that modelling the recovery rate means modelling $-Y_{t(j)}$ instead of $Y_{t(j)}$. The special cases of an LGD of zero and one cannot be incorporated in (7) and should be treated separately because they would lead to a transformed LGD of $-\infty$ and an LGD that is not defined, respectively.*

[8] *This model is equivalent to the logit specification in Düllmann & Trapp (2004), although another parameterisation is used here.*

[9] *The LGD is always conditional on default. Therefore, only data of defaulted bonds is used to estimate the model. Non-defaulted bonds are assigned a loss of zero.*

[10] *See Frye (2000).*

Frye (2000) introduces a model where the same systematic risk that drives the default rate also drives the recovery rate. He assumes large portfolios that are perfectly diversified. As this is not always true in the credit portfolios of typical banks, another method to introduce correlation between default and LGD is needed.

In this article, correlation between defaults and LGDs is introduced via the unobservable systematic risk factors F_t and G_t, which may be correlated. However, a serious difficulty arises if the PD model and the LGD model are estimated separately. Assuming that F_t and G_t are correlated leads to biased estimates of the LGD model when using only model (11) for estimation. In particular, the average LGD is overestimated.[11] The bias vanishes if F_t and G_t are independent normal random variables.

Several simulation studies were carried out.[12] They show that the bias may be considerable and should not be neglected. It should be mentioned that this is in contrast to usual econometric multi-equation models where estimators based on single-equation models are usually still consistent but, in general, not efficient.

To avoid the bias in the LGD model, we suggest a joint estimation procedure using a maximum likelihood procedure that allows the joint estimation of all coefficients, including those of models (6) and (11), with observable risk factors.

We assume that the random vector $(F_t, G_t)'$ has a standard normal distribution with correlation parameter ρ, that is:

$$\begin{pmatrix} F_t \\ G_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix};\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right) \tag{12}$$

For the joint distribution of F_t and $Y_{t(j)}$, we obtain:

$$\begin{pmatrix} F_t \\ Y_{t(j)} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ \mu_{t(j)} \end{pmatrix};\begin{pmatrix} 1 & b_1\rho \\ b_1\rho & b_1^2+b_2^2 \end{pmatrix}\right) \tag{13}$$

Furthermore, $Cov(y_{t(j)}, y_{t(l)}) = b_1^2$ for two bonds j and l, where $j \neq l$.

Since the random vector $(Y_{t(1)}, \dots, Y_{t(m_t)}, F_t)'$ has a multivariate normal distribution, the conditional distribution of $\mathbf{y}_t = (Y_{t(1)}, \dots, Y_{t(m_t)})'$ given F_t is also multivariate normal (see, for example, Searle (1971), page 47). Its mean vector and covariance matrix can be calculated as:

$$\boldsymbol{\mu}_t(F_t) = \boldsymbol{\mu}_t + \begin{pmatrix} b_1\rho \\ \vdots \\ b_1\rho \end{pmatrix} F_t \tag{14}$$

where $\boldsymbol{\mu}_t = (\mu_{t(1)}, \dots, \mu_{t(m_t)})'$ and:

$$\boldsymbol{\Sigma}_{\mathbf{y}_t|F_t} = \begin{pmatrix} b_1^2\left(1-\rho^2\right)+b_2^2 & & b_1^2\left(1-\rho^2\right) \\ & \ddots & \\ b_1^2\left(1-\rho^2\right) & & b_1^2\left(1-\rho^2\right)+b_2^2 \end{pmatrix} \tag{15}$$

Then, the conditional joint density function of defaults and LGDs given the unobservable risk factor f_t can be written as:

$$\begin{aligned} h\left(d_{it}, i \in N_t, \mathbf{y}_t \middle| f_t\right) &= (2\pi)^{-\frac{m_t}{2}} \left|\boldsymbol{\Sigma}_{\mathbf{y}_t|f_t}\right|^{-\frac{1}{2}} \\ &\times \exp\left\{-\frac{1}{2}\left[\mathbf{y}_t - \boldsymbol{\mu}_t(f_t)\right]' \boldsymbol{\Sigma}^{-1}_{\mathbf{y}_t|f_t}\left[\mathbf{y}_t - \boldsymbol{\mu}_t(f_t)\right]\right\} \\ &\times \prod_{i\in N_t} \lambda_{it}(f_t)^{d_{it}}\left[1-\lambda_{it}(f_t)\right]^{(1-d_{it})} \end{aligned} \tag{16}$$

The unconditional joint density is given by:

$$h\left(d_{it}, i \in N_t, \mathbf{y}_t\right) = \int_{-\infty}^{\infty} h\left(d_{it}, i \in N_t, \mathbf{y}_t \middle| f_t\right)\varphi(f_t)\,df_t \tag{17}$$

Observing a time series with T periods leads to the final unconditional log-likelihood function:

$$\begin{aligned} & l(\beta_0, \hat{\mathbf{a}}, \tilde{\mathbf{a}}_1, w, \alpha_0, \mathbf{a}, \tilde{\mathbf{a}}_2, b_1, b_2, \rho) \\ &= \sum_{t=1}^{T} \ln\left(\int_{-\infty}^{\infty} h\left(d_{it}, i \in N_t, \mathbf{y}_t \middle| f_t\right)\varphi(f_t)\,df_t\right) \end{aligned} \tag{18}$$

The integral on the right-hand side of (18) can be approximated by the adaptive Gaussian quadrature as described in Pinheiro & Bates (1995). Usually, the log-likelihood function (18) is numerically optimised with respect to the unknown parameters for which several algorithms, such as the Newton-Raphson or the (dual) quasi-Newton optimisation method, exist.[13]

Simulating loss distributions

We have to forecast the loss distribution for the next year $T+1$. The portfolio loss L_{T+1} for this period is a random variable given by:

$$L_{T+1} = \sum_{i\in N_{T+1}} D_{i,T+1} Loss_{i,\,T+1} \tag{19}$$

The loss of each bond j in year $T+1$ is modelled as the product of exposures at default (EAD) and LGD for each bond. The sum of the losses of the bonds for each borrower i leads to the borrower's loss, $Loss_{i,\,T+1}$.

The portfolio loss can also be expressed as a portion of the total exposure:

$$L^*_{T+1} = \frac{L_{T+1}}{\sum_{i\in N_{T+1}} EAD_{i,\,T+1}} \tag{20}$$

Here, the EAD are assumed to be fixed. N_{T+1} is the set of all borrowers who are in the portfolio at the beginning of year $T+1$. The PDs and the LGDs are modelled by (6) and (11), respectively. The unknown parameters in the models are replaced by the corresponding maximum-likelihood estimates.

The predicted loss distribution for year $T+1$ is determined by simulation. We proceed as follows:

- **Step 1.** A random realisation f_{T+1} of the unobservable systematic risk factor is drawn from a standard normal distribution.
- **Step 2.** Using the values of $\mathbf{x}_{iT}$ and $\mathbf{z}_{1,\,T}$, we calculate the predictions of the conditional PD $\hat{\lambda}_{i,\,T+1}(f_{T+1})$, $i \in N_{T+1}$. According to the conditional independence, N_{T+1} Bernoulli events are generated with conditional PDs $\hat{\lambda}_{i,\,T+1}(f_{T+1})$.
- **Step 3.** The predicted conditional transformed LGD $\hat{y}_{T+1(j)}(f_{T+1})$, $j = 1, \dots, m_{T+1}$ is calculated for all m_{T+1} defaulted borrowers using the observed values of $\mathbf{v}_{T(j)}$ and $\mathbf{z}_{2,\,T}$ and m_{T+1}-variate normal distribution with mean vector (14) and covariance matrix (15). Then $\hat{y}_{T+1(j)}(f_{T+1})$ is retransformed to obtain the conditional $\hat{LGD}_{T+1(j)}(f_{T+1})$.
- **Step 4.** Taking into account the relevant EAD and the predicted LGDs, the portfolio loss realisation is calculated.

[11] *The bias can been seen if there is a positive correlation between defaults and LGD. During a recession, PDs increase, leading to more defaults. LGDs are also higher in a recession if defaults and LGD are correlated. In economic booms, PDs decrease, leading to fewer defaults. LGDs also decrease, but occur only in the event of a default. These lower LGDs are therefore under-represented, leading to an upward bias in the separate estimation of LGD.*

[12] *The results of the simulation studies can be obtained from the authors.*

[13] *The estimation of the parameters of (18) can be done in SAS using PROC IML and the CALL NLPQN and CALL NLPFDD routines to calculate parameter estimates and the Hesse matrix.*

- **Steps 1** through **4** are repeated five million times, and the loss distribution of the portfolio is calculated.

For the empirical loss distribution, expected loss and risk measures such as unexpected loss, VAR or CVAR are easily obtained.[14]

The data

We use data from Moody's Default Risk Service database. The data set contains information about issuers and debt rated by Moody's since 1970. If an issuer defaults several times, only the first default event is included in the PD and LGD data set.

Since Moody's changed its rating methodology in 1982, we use data from 1982–2003 for the PD and the LGD data set.[15] To obtain a homogeneous data set, the data is restricted to issuers domiciled in the US. This restriction is necessary because economic development is not the same in different countries. Thus we would either have to use different systematic risk factors for each country or restrict the data set to only one country.

Obligors are grouped into different sectors (referenced as 'broad industries' by Moody's). As default rate, and therefore recovery observation, is scarce in financial sectors such as 'banking', 'finance', etc., and in sectors such as 'sovereign' or 'public utility', only data from the 'industrial', 'transportation' and 'other non-bank' sectors is used. These last three sectors are referred to as the aggregated sector 'industry' in this article. The data set contains about 3,800 rated issuers with about 27,500 observations. It mainly contains rating information and other issuer-specific information such as finer sectoral classification, default rate per year, etc.

The recovery rate, and thus LGD, can be observed directly from the price of the defaulted debt approximately one month after default. This is Moody's definition of the recovery rate, which is known as market recovery rate in the literature. The LGD in per cent is calculated as 100 minus the price of the defaulted bond, that is, 100 minus the recovery. Information about debt includes loans, bonds and preferred stock from issuers of different countries. The aggregated sector 'industry' contains about 85% of all debt obligations, so – for the same reason as in the PD data set – the data was restricted to this sector. Obligors or bonds within this sector are assumed to be exposed to the same systematic risk factor. For the same reason, only bonds were used, and all loans or preferred stocks were eliminated from the data set. This is not a severe constraint as 90% of the debt obligations were classified as bonds.[16]

Empirical analysis

We use two different models to calculate and forecast PD and LGD. Model A contains explanatory variables in the separate PD and LGD models (see (6) and (11)). Model B stands for the simultaneous approach including explanatory variables (see (16)). The resulting estimates for the PD and LGD model are presented in table A. As the bias is only present in the LGD parameters – especially in the constant term – the parameters of model A are only shown for the LGD model.

The results for the PD model can be interpreted as follows. Borrowers with a better issuer rating – for example with a rating of Aaa to Aa3 – have smaller estimated PDs than issuers with a rating of Caa or worse. The same relationship holds for the other rating classes. The better the rating class, the smaller the PD. If the interest rate of the federal funds with a time lag of one year increases, the estimated PD also increases. The same relationship holds for the default rate with a time lag of one year.

A. Parameter estimates for models A and B for the PD and LGD model and the correlation parameter

Effect	Model A	Model B
Parameters for the PD model		
Constant		–1.7945*
Issuer rating Aaa to A3 (t – 1)		–2.5870*
Issuer rating Baa1 to Baa3 (t – 1)		–1.8141*
Issuer rating Ba1 to Ba3 (t – 1)		–1.2678*
Issuer rating B1 to B3 (t – 1)		–0.5770*
Interest rate Fed Fund (t – 1)		0.09333*
Default rate (%) (t – 1)		0.09769*
w		0.1334*
Parameters for the LGD model		
Constant	–0.8349*	–1.7240*
Debt rating Ba3 to B3 (t – 1)	–0.2039**	–0.1912**
Seniority: senior unsecured	0.6041*	0.5806*
Seniority: senior subordinated	0.6926*	0.6648*
Seniority: subordinated and junior subordinated	1.0262*	1.0262*
Relative seniority: 2 and 3	0.4979*	0.4963*
Additional backing by a third party	–0.2676**	–0.2728**
Bond maturity (in years)	0.03402*	0.03607*
Volume of defaulted bonds ($ million)	0.001096*	0.001059*
Default rate (%) (t – 1)	0.1558**	0.1496**
b_1	0.5358*	0.4849*
b_2	1.3139*	1.3081*
□		0.6088**
* significant at 1% level, ** significant at 5% level		

The parameter estimates of the LGD models A and B can be interpreted in the following way. A bond with a debt rating of Ba3 to B3 with a time lag of one year has a lower (transformed) LGD than a bond with debt rating of Caa to C. Senior unsecured bonds have a higher (transformed) LGD than senior secured bonds. The other seniority classes are interpreted equivalently. Bonds with a relative seniority of one have higher LGDs than those with a lower relative seniority.[17] The longer the maturity, the higher the (transformed) LGD. The uncertainty of future cash flows may cause this coherence. The higher the volume of the defaulted bonds, the higher the (transformed) LGD. A higher volume of a defaulted bond leads to a higher supply in the market for defaulted bonds. This leads to a lower price and thus a higher LGD for this bond. The higher the default rate with a time lag of one year, the higher the LGD. This result can also be drawn back to supply and demand in the market for defaulted bonds.[18]

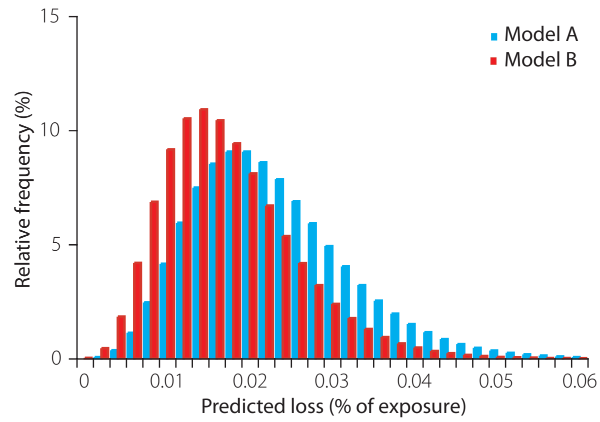
Macroeconomic variables such as the lagged default rate try to explain cyclical variations of the LGD or the default rate. Here,

[14] *The VAR is defined as the quantile of the loss distribution at a certain level. The unexpected loss is defined as the difference between the VAR and the expected loss. The CVAR is defined as the conditional mean of losses greater than the respective VAR.*
[15] *In contrast to Frye (2000), we do not truncate our data after 1997 but treat the rating grades Caa and Caa1, Caa2 and Caa3 as a single (pooled) grade.*
[16] *For a more detailed description of the data set, see Hamerle, Knapp & Wildenauer (2006).*
[17] *A subordinated bond of one borrower that also has senior unsecured bonds has a relative seniority of two, whereas a subordinated bond of another borrower that has a junior subordinated bond besides the subordinated one has a relative seniority of one.*
[18] *A detailed interpretation of the LGD parameters can be found in Hamerle, Knapp & Wildenauer (2006).*

B. Simulation results for models A and B

	Model A	Model B
Expected loss	2.26472%	1.73245%
VAR (99%)	5.01583%	4.13341%
VAR (99.9%)	6.33643%	5.34340%
Unexpected loss (99%)	2.75111%	2.40096%
Unexpected loss (99.9%)	4.07171%	3.61096%
CVAR (99%)	5.59054%	4.66208%
CVAR (99.9%)	6.88703%	5.85396%
Standard deviation	0.94414%	0.80622%

1 Loss distribution for simulations of models A and B



the default rate can be seen as a proxy for the cyclical movements. The lagged default rate can explain PD and LGD better than other macroeconomic variables such as GDP growth or the unemployment rate.[19] The bias in the LGD parameters can be seen when comparing the constants of models A and B. The estimate of the constant is smaller in model B.[20] b_1 denotes the impact of the systematic risk factor and b_2 denotes the impact of the remaining idiosyncratic risk factor.

The estimated correlation parameter ρ indicates a positive relationship between default and LGD.

Simulation results

Two different simulations are used to illustrate the effect using separate or simultaneous models to predict the portfolio loss. In simulation A, the separate models are used to predict defaults and LGDs, and the empirical correlation of the realisations of the systematic risk factors is taken into account.[21] In simulation B, the simultaneous estimation procedure for PD and LGD is used.

We use a data set of 586 obligors that did not default at the end of 2003 in the adjusted Moody's data set to predict the loss distribution for 2004.[22] A portfolio is used where each obligor holds one bond that is worth its face value at default, and thus an EAD of $100.

We obtain the results shown in table B from simulations A and B. They are given as a percentage of the total exposure (see (20)).

A comparison of the values of the loss distribution in table B shows that the expected loss is higher in model A. This result can be attributed to the overestimation of LGD parameters, which finally leads to the overestimation of the portfolio loss. The VAR and the unexpected loss are also smaller for model B both for the 99% and the 99.9% level. The CVAR is also smaller in model B. Hence the overestimation of the parameters of the LGD model leads to higher predicted LGDs and thus to higher predicted

References

Altman E, B Brady, A Resti and A Sironi, 2003
The link between default and recovery rates: theory, empirical evidence and implications
Working paper

Basel Committee on Banking Supervision, 2004
International convergence of capital measurement and capital standards – a revised framework

Black F and M Scholes, 1973
The pricing of options and corporate liabilities
Journal of Political Economy 81, May–June, pages 637–659

Chava S, C Stefanescu and S Turnbull, 2006
Modeling expected loss with unobservable heterogeneity
Working paper

Credit Suisse First Boston, 1997
CreditRisk+: a credit risk management framework

Düllmann K and M Trapp, 2004
Systematic risk in recovery rates – an empirical analysis of US corporate credit exposures
Deutsche Bundesbank, discussion paper, 02/2004, Series 2: Banking and Financial Studies

Frye J, 2000
Depressing recoveries
Risk November, pages 108–111

Gupton G, C Finger and M Bhatia, 1997
CreditMetrics – technical document

Hamerle A, M Knapp, T Liebig and N Wildenauer, 2005
Incorporating prediction and estimation risk in point-in-time credit portfolio models
Deutsche Bundesbank, discussion paper, 13/2005, Series 2: Banking and Financial Studies

Hamerle A, M Knapp and N Wildenauer, 2006
Modelling loss given default: a "point in time" approach
The Basel II Risk Parameters – Estimation, Validation and Stress Testing by B Engelmann and R Rauhmeier, Springer, pages 127–142

McNeil A and J Wendin, 2005
Bayesian inference for generalized linear mixed models of portfolio credit risk
Working paper, forthcoming in Journal of Empirical Finance

Merton R, 1974
On the pricing of corporate debt: the risk structure of interest rates
Journal of Finance 29, pages 449–470

Merton R, 1977
On the pricing of contingent claims and the Modigliani-Miller theorem
Journal of Financial Economics 5, pages 241–249

Pinheiro J and D Bates, 1995
Approximations to the log-likelihood function in the nonlinear mixed-effects model
Journal of Computational and Graphical Statistics 4 (1), March, pages 12–35

Pykthin M, 2003
Unexpected recovery risk
Risk April, pages 74–78

Rösch D and H Scheule, 2005
A multifactor approach for systematic default and recovery risk
Journal of Fixed Income 15 (2), September, pages 63–75

Schönbucher P, 2003
Credit derivatives pricing models: models, pricing and implementation
Wiley, Chichester

Searle S, 1971
Linear models
Wiley, New York

losses. If the estimation bias is ignored, the economic loss is overestimated.

The loss distributions of simulations A and B are illustrated in figure 1.

Conclusion

In the past, only a few studies have carried out PD and LGD at the same time (see, for example, Frye (2000), Pykthin (2003), Düllmann & Trapp (2004) or Chava, Stefanescu & Turnbull, (2006)). They assume that PD and LGD are driven by a systematic risk factor that is the same for all issuers in the same year, and idiosyncratic factors that are different for all issuers and debt obligations.

In most empirical analysis concerning the link between PD and LGD, defaults and LGDs are positively correlated. This conclusion is drawn from the fact that, in the past, in an economic downturn, default and loss rates increased. Most of the authors therefore assume that the same systematic risk that affects PD also affects LGD. We generalise this approach by allowing different systematic risk factors to affect PD and LGD. The correlation between these systematic risk factors can be estimated empirically from historical data. Modelling PD and LGD separately leads to biased estimates in the LGD model. A simultaneous model is used here that generalises the approach of Rösch & Scheule (2005) for individual time-dependent data.

The systematic risk factors are split into an observable and an unobserved part by taking macroeconomic variables into account. We use a dynamic individual approach to model PD and LGD, including issuer- and bond-specific variables as well as the macroeconomic variables. Using this approach, we obtain more precise PD and LGD predictions, leading to more exact predictions of economic capital.

PD and LGD are estimated using US data from Moody's Default Risk Service in the aggregated sector 'industry'. The time-dependent issuer- and bond-specific and the macroeconomic variables are incorporated with a time lag of one year. Therefore, PD and LGD predictions for the next year can be made on the basis of values that are known at the time the prediction is made.

We use Monte Carlo simulation to calculate the impact of the dynamic individual approach for PD and LGD on economic capital. Using the parameter estimates from the simultaneous model, the parameters of the portfolio loss distribution, such as expected loss, VAR and CVAR, are smaller than those resulting from the separate PD and LGD model. Thus, neglecting this prediction bias using the separate models leads to the overestimation of economic loss, which may be costly for banks. ■

Alfred Hamerle is head of, Michael Knapp is academic adviser and Nicole Wildenauer (now at Commerzbank) was research assistant at the department of statistics at the Faculty of Business, Economic and Information Systems at the University of Regensburg, Germany.
Email: Alfred.Hamerle@wiwi.uni-regensburg.de
Michael.Knapp@wiwi.uni-regensburg.de
Nicole.Wildenauer@wiwi.uni-regensburg.de

[19] Another way to incorporate cyclical effects might be to use a vector autoregressive model for F_t and G_t, (see McNeil & Wendin, 2005).
[20] This bias has been confirmed in a simulation study where the true values of the PD and LGD model are given to simulate a data set. In a next step, the parameters are estimated using the simulated data set.
[21] For simplicity, the PD parameters of the simultaneous model were used. This can be done as the PD parameters are not biased like the LGD parameters.
[22] The data set is adjusted to contain only US borrowers in the sector industry, as in the data set used for the model estimation.