Journal of the Operational Research Society (2012), 1–15

www.palgrave-journals.com/jors/

Forecasting probabilities of default and loss rates given default in the presence of selection

D Rösch[1]* and H Scheule[2]

[1]Leibniz Universität Hannover, Germany; and [2]University of Technology, Sydney, Australia

This paper offers a joint estimation approach for forecasting probabilities of default and loss rates given default in the presence of selection. The approach accommodates fixed and random risk factors. An empirical analysis identifies bond ratings, borrower characteristics and macroeconomic information as important risk factors. A portfolio-level analysis finds evidence that common risk measurement approaches may underestimate bank capital by up to 17% relative to the presented model.

Journal of the Operational Research Society advance online publication, 29 August 2012

doi:10.1057/jors.2012.82

Keywords: credit risk; correlation; probability of default; loss given default; recovery; tobit model

1. Introduction

Financial institutions were surprised that during the financial crisis, individual risk parameters deteriorated jointly. As a result, credit portfolio losses dramatically exceeded the predictions provided by internal risk models. Measuring credit portfolio losses is of great concern to fixed income investors. A large growth of investments in credit portfolios rather than single name credits has occurred through securitizations. The evaluation of credit portfolio risks requires the understanding of individual risk drivers as well as their dependence structure.

Credit portfolio risk is measured by various parameters such as default probabilities, loss rates given default, exposures at default and dependence parameters such as correlations and more general copulas. It is common practice to model these parameters independently and to introduce the dependence structure thereafter. This practice is supported by the implementation of isolated models provided by external vendors.

Various authors address the default likelihood. Important contributions are Merton (1974), Leland (1994), Jarrow and Turnbull (1995), Longstaff and Schwartz (1995), Madan and Unal (1995), Leland and Toft (1996), Jarrow *et al* (1997), Duffie and Singleton (1999), Gordy (2000, 2001), Shumway (2001), McNeil and Wendin (2007) and Duffie *et al* (2007).

Credit ratings are often used as aggregated explanations of financial risk. Ratings measure the financial risk of corporate bond issuers, corporate bond issues and structured finance securities. Fundamental issues relating to the general extent to which credit rating changes convey new information has a rich pedigree that is the subject of ongoing academic debate and investigation. For example, Radelet and Sachs (1998) find that rating changes are pro-cyclical which would suggest that they provide only a limited amount of new information to the market. Ederington and Goh (1993), Dichev and Piotroski (2001) and Purda (2007) find that corporate credit rating downgrades do provide news to the market, although most studies find that rating upgrades do not. Jorion *et al* (2005) show that after Regulation Fair Disclosure, the market impact of both downgrades and upgrades is significant and of greater magnitude compared to that observed in the pre-Regulation Fair Disclosure period. The relative roles of different CRAs have also been studied. For example, Morgan (2002) examine the effect of divergent Moody's and S&P ratings of banks.

Research on recoveries and loss rates given default (LGD) are quite recent. Pan and Singleton (2008) derive the implicit risk structure of recoveries from sovereign CDS spreads. Contributions which focus on recoveries from defaulted issuers include Carey (1998) and Pykhtin (2003). Empirical models for recoveries using explanatory co-variables which are economically motivated are provided by Dermine and Neto de Carvalho (2006), Acharya *et al* (2007), Qi and Yang (2009) and Grunert and Weber (2009). Most of the empirical results in the recent literature are based on common linear regression models analysing credit defaults.

Few exceptions exist: Pykhtin (2003) who accounts for this mortality bias and derives closed-form expressions for the Expected Loss and the Value-at-Risk. However, the

Correspondence: D Rösch, Institute of Banking & Finance, Faculty of Economics and Management, Leibniz Universität Hannover, Königsworther Platz 1, Hannover, Lower Saxony 30167, Germany.

paper does not provide an empirical solution for parameter estimation. Crook and Belotti (2011) estimate LGD models based on accounting information and macroeconomic information for credit card loans and compare OLS models with Tobit and decision tree models. Thomas *et al* (2012) and Matuszyk *et al* (2010) focus on the decision tree for a borrower's default.

Research on dependencies between risk parameters can be split into two categories. Firstly, dependencies between default events and asset value returns are modelled. Dietsch and Petey (2004) present a non-parametric approach and McNeil and Wendin (2007) apply a generalized mixed model approach using Maximum-Likelihood. Kiefer (2011) proposes Bayesian approaches for modelling correlated default events and infers regulatory capital. Boeker *et al* (2010) apply Bayesian approaches in conjunction with Markov-Chain-Monte-Carlo methods to model economic capital. Crook and Belotti (2012) provide evidence on asset correlations for credit card defaults. Secondly, Hu and Perraudin (2002), Tasche (2004), Miu and Ozdemir (2006) and Altman *et al* (2005) derive dependencies between default events and loss rates given default.

This paper extends the previous literature in various ways. Firstly, the paper develops a joint model for estimating and forecasting probabilities of default, recovery rates given default as well as asset correlations. The model explicitly takes into account that loss rates given default or recoveries can generally be observed after the occurrence of a default event and are censored otherwise. The sample of recoveries is therefore not representative of the population and is selected on the basis of observed losses or recoveries. Formulae for the unconditional and conditional probability of default, recovery rate given default, expected loss are developed. Conditional values are of greatest importance for credit risk modellers to comply with current modelling standards such as Basel II and Basel III, which require the specification of risk measures for economic downturns.

Secondly, this paper proposes an econometric estimation method for historic recovery rates. Under the censoring model simple estimators (from common linear OLS regression models) are inconsistent and we suggest an estimation method which yields consistent estimators.

Thirdly, the fixed effect model is extended by including a time-varying random effect, that is, unobservable systematic factors. The combined models include fixed effects given by explanatory control variables as well as random effects. The sensitivities of these random effects can be transformed into asset correlations which are a central parameter in Basel II and Basel III.

Fourthly, the models are applied to corporate bond issuers. The dynamics of recovery implied asset return volatilities, correlations and their determinants are analysed. Credit ratings have been highly criticized in the financial crisis due to their failure to predict corporate credit default risk. Using consistent estimators, the information content of credit ratings and their forecasting ability is augmented by additional bond issue and issuer characteristics as well as a macroeconomic variable. A portfolio-level analysis finds evidence that common risk measurement approaches may underestimate bank capital by up to 17% relative to the presented model.

The rest of the paper proceeds as follows. Section 2 defines a structural default process based on an obligor's asset value and an empirical version of the model. The model is extended to asset return correlations. Section 3 describes the data and presents the empirical results. In Section 4, the resulting Basel II capital is compared to a model with deterministic recoveries and the best practice US industry approach. Closed-end formulas for the Expected Loss, Value-at-Risk and Downturn Loss Given Default are presented. Section 5 concludes with a summary and a discussion of the model and the findings.

2. The basic models

2.1. Asset value dynamics and likelihood of a credit default

We derive the default probability and the recovery rate in an asset value model. Let V denote the value of a firm's assets as in Merton (1974). V is assumed to follow a stochastic process which can be described by

$$dV = \delta \cdot V \cdot dt + \sigma \cdot V \cdot dW, \tag{1}$$

where $\delta \in \Re$ is an exogenous parameter and $\sigma > 0$ is an exogenous volatility parameter, dt represents the passage of time and dW is a Brownian motion. The change in the logarithmic firm value $\ln V$ between time 0 and T for given value $v(0)$ of the firm at time 0 can be written as

$$S(T) = \ln V(T) - \ln v(0) = (\delta - 0.5\sigma^2)T + \sigma\sqrt{T} \cdot \varepsilon \tag{2}$$

where ε is a standard normally distributed random variable.

The firm is assumed to be financed by debt and equity. Debt consists of a zero coupon bond with notional k and maturity T. At maturity, the bondholders receive the lower of a payment k and the value of the firm's assets. In the case $V(T) < k$, the bond issue defaults and the bondholders receive a fraction of the notional which is also known as recovery. The default indicator is denoted by the random variable

$$D = \begin{cases} 1, & \text{credit default} \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

Hence, the probability of default is

$$\begin{aligned}\lambda &= P(D=1|v(0)) = P(V(T)<k|v(0))\\ &= P(S(T)<\ln k - \ln v(0))\\ &= P\left(R(T)<\frac{\ln\frac{k}{v(0)} - (\delta - 0.5\sigma^2)\cdot T}{\sigma\cdot\sqrt{T}}\right)\\ &= \Phi(-d(T)) \end{aligned} \quad (4)$$

where $\Phi(\cdot)$ is the standard normal cumulative density function, $R(T) = \left(S(T) - (\delta - 0.5\sigma^2)\cdot T\right)/\left(\sigma\cdot\sqrt{T}\right)$ is the normalized asset return and $d(T) = -\left(\ln\frac{k}{v(0)} - (\delta - 0.5\sigma^2)\cdot T\right)/(\sigma\cdot\sqrt{T})$ is the normalized default threshold which is also known as Distance-to-Default.

2.2. Severity of a bond default

In this setting, the repayment ratio RR is the minimum of the asset value to debt ratio and one

$$RR = \min\left\{\frac{V(T)}{k}, 1\right\} \quad (5)$$

Defining the default point c by

$$c = \ln k - \ln v_0 \quad (6)$$

gives the transformation

$$\begin{aligned}\ln RR &= \min\{\ln V(T) - \ln k, 0\}\\ &= \min\{\ln V(T) - \ln v(0) - (\ln k - \ln v(0)), 0\}\\ &= \min\{S(T) - c, 0\}\end{aligned} \quad (7)$$

Equation (7) shows that the natural logarithm (log) of the repayment ratio is normally distributed but censored at zero with non-zero values if a default event occurs.

2.3. The empirical factor model

The subscript i is introduced for the respective borrower and the number of borrowers is denoted by n. A time-horizon of one year is considered. Thus, the transformed log-repayment ratio can be written as

$$\ln RR_i = \min\{S_i(1) - c_i, 0\} \quad (8)$$

$i=1,\ldots,n$. This representation assumes that the observed variables Y_i, that is, the log-repayment ratios, satisfy

$$Y_i = \ln RR_i = \min\{Y_i^*, 0\} \quad (9)$$

see Tobin (1958). Y^* is a latent variable generated by a classical regression model

$$Y_i^* = \boldsymbol{\beta}'\boldsymbol{x}_i + \sigma\cdot U_i \quad (10)$$

where $\boldsymbol{\beta}$ represents a vector of parameters, $\boldsymbol{x}_i$ a vector of covariates, which may include an intercept, and U_i a random error. Note that $y_i < 0$ implies an obligor default event. The errors are assumed to be independent and identically standard normally distributed.

Some useful model properties can be obtained following Maddala (1983). The conditional density of the log-repayment ratio, that is, the density of the log-recovery rate given default is

$$h(y_i|Y_i<0, x_i) = \frac{\phi(-(y_i - \boldsymbol{\beta}'\boldsymbol{x}_i)/\sigma)}{\sigma\cdot(1-\Phi(\boldsymbol{\beta}'\boldsymbol{x}_i/\sigma))} \quad (11)$$

for $y_i<0$, where $\phi(\cdot)$ is the density function of the standard normal distribution. A closed-form expression for the conditional expectation of the log-recoveries Y_i given $\boldsymbol{x}_i$ and $Y_i<0$ can be derived as

$$\begin{aligned}\mathbb{E}(Y_i|Y_i<0,\boldsymbol{x}_i) &= \frac{1}{1-\Phi(\boldsymbol{\beta}'\boldsymbol{x}_i/\sigma)}\int_{-\infty}^{0} z f(z)dz\\ &= \boldsymbol{\beta}'x_i - \sigma\frac{\phi(\boldsymbol{\beta}'\boldsymbol{x}_i/\sigma)}{1-\Phi(\boldsymbol{\beta}'x_i/\sigma)}\end{aligned} \quad (12)$$

where $f(\cdot)$ is the density of a normally distributed random variable with mean $\boldsymbol{\beta}'\boldsymbol{x}_i$ and variance σ^2. Note that the probability of default is

$$PD_i = P(D_i = 1|\boldsymbol{x}_i) = 1 - \Phi(\boldsymbol{\beta}'\boldsymbol{x}_i/\sigma) \quad (13)$$

The standardized linear predictor $\boldsymbol{\beta}'\boldsymbol{x}_i/\sigma$ is also known as the Distance-to-Default. Note that PD_i relates to the econometric model while λ relates to the theoretical asset value model.

Figure 1 shows a graphical interpretation of the relation between the linear predictor $\boldsymbol{\beta}'\boldsymbol{x}_i$, the probability of default (PD), and the volatility σ. Equation (13) shows that the PD

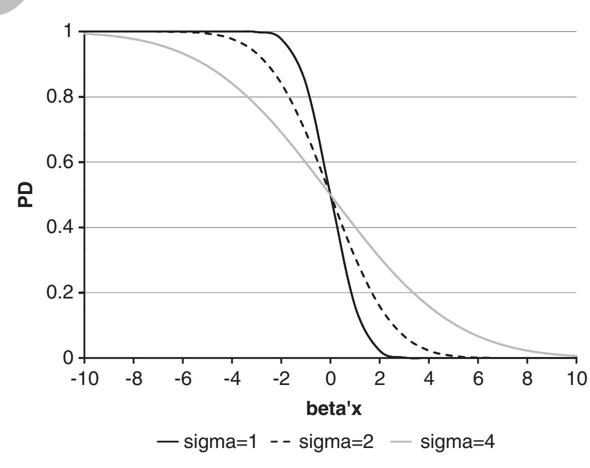

Figure 1 Relation between linear predictor $\boldsymbol{\beta}'\boldsymbol{x}$, volatility σ, and probability of default (PD).
Notes: This figure shows the relation between the linear predictor $\boldsymbol{\beta}'\boldsymbol{x}_i$, the probability of default (PD), and the volatility σ. Probabilities of default are calculated based on σ and $\beta'x$ according to Equation (13). PD is a non-linear decreasing function of the linear predictor and a non-linear increasing (decreasing) function of the volatility for low (high) linear predictors. For high σ, the relationship between $\beta'x$ and PD is linear and for low σ, a firm defaults with a high likelihood (ie, the PD is high) if $\beta'x< 0$ and a firm does not default with a high likelihood (ie, the PD is low) if $\beta'x< 0$.

is a non-linear decreasing function of the linear predictor and a non-linear increasing (decreasing) function of the volatility for low (high) linear predictors.

The conditional expectation of Y_i given $\boldsymbol{x}_i$ is

$$\mathbb{E}(Y_i|\boldsymbol{x}_i) = (\boldsymbol{\beta}'\boldsymbol{x}_i)(1 - \Phi(\boldsymbol{\beta}'\boldsymbol{x}_i/\sigma)) - \sigma\phi(\boldsymbol{\beta}'x_i/\sigma) \quad (14)$$

which is simply (12) times (13). Equations (12) and (14) have important consequences for the estimation of determinants for the recoveries using regression models. In both instances, the expectation of Y_i does not equal the linear predictor $\boldsymbol{\beta}'\boldsymbol{x}_i$. Thus, the estimates for $\boldsymbol{\beta}$ are biased and inconsistent if they are (i) estimated using non-zero observations of the Y_i (ie, using the recoveries of defaulted borrowers), or (ii) by treating the values of Y_i which are zero as regular dependent variables as in common linear regression models, see Goldberger (1972), Hausman and Wise (1977), Greene (1981), Bierens (2004).

The conditional variance of Y_i is given by

$$\begin{aligned}\mathbb{V}(Y_i|Y_i<0,\boldsymbol{x}_i) &= \sigma^2 - \sigma^2 \cdot \left(-\boldsymbol{\beta}'\boldsymbol{x}_i/\sigma + \frac{\phi(\boldsymbol{\beta}'x_i/\sigma)}{1-\Phi(\boldsymbol{\beta}'\boldsymbol{x}_i/\sigma)}\right) \\ &\quad \times \frac{\phi(\boldsymbol{\beta}'\boldsymbol{x}_i/\sigma)}{1-\Phi(\boldsymbol{\beta}'\boldsymbol{x}_i/\sigma)} \\ &= \sigma^2(1-\Delta(\Psi)) \quad (15)\end{aligned}$$

where $\Delta(\Psi) = [\lambda(\Psi) - \Psi]\,\lambda(\Psi)$ which is between 0 and 1, $\lambda(\Psi) = \frac{\phi(\Psi)}{1-\Phi(\Psi)}$ is the inverse Mills ratio, and $\Psi = \boldsymbol{\beta}'\boldsymbol{x}_i/\sigma$ · [1]

Finally, the expectation of the recovery rate given the firm's default is derived. First, we define the recovery rate given default as

$$RGD_i = \exp[Y_i^-] \quad (16)$$

that is, it is defined only if the borrower defaults. Then, the expected recovery rate given default is

$$\begin{aligned}ERGD_i &= \mathbb{E}(RGD_i) = \mathbb{E}(RR_i|D_i = 1, \boldsymbol{x}_i) \\ &= \int_{-\infty}^{0} \exp(y_i) \cdot h(y_i|Y_i<0,\boldsymbol{x}_i)\mathrm{d}y_i \\ &= \int_{-\infty}^{0} \exp(y_i) \cdot \frac{\phi(-(y_i - \boldsymbol{\beta}'\boldsymbol{x}_i)/\sigma)}{\sigma\cdot(1-\Phi(\boldsymbol{\beta}'\boldsymbol{x}_i/\sigma))}\mathrm{d}y_i \\ &= \frac{1}{1-\Phi(\boldsymbol{\beta}'\boldsymbol{x}_i/\sigma)} \cdot \exp(\boldsymbol{\beta}'\boldsymbol{x}_i + 0.5\sigma^2) \\ &\quad \times \Phi\left(-\frac{\boldsymbol{\beta}'\boldsymbol{x}_i + \sigma^2}{\sigma}\right) \quad (17)\end{aligned}$$

The derivation of the third equation is given in the Appendix. The expected loss rate given default (ELGD) is then defined as

$$ELGD_i = 1 - \mathbb{E}(RR_i|D_i = 1, \boldsymbol{x}_i) = 1 - ERGD_i \quad (18)$$

[1]We thank an anonymous referee for pointing this out.

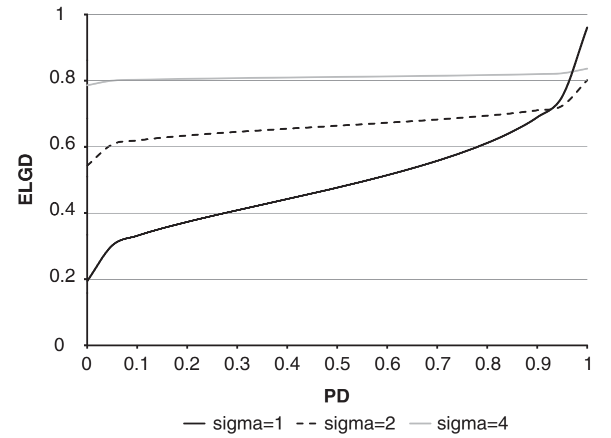

Figure 2 Relation between probability of default (PD), expected loss given default (ELGD), and volatility σ.
Notes: This figure shows the relation between PD, expected loss rate given default (ELGD), and the volatility σ. Given the volatility, the relationship between PD and ELGD is monotone: ELGD increases with the PD. The slope of the PD-ELGD curve depends on the volatility resulting in an approximately linear relation for higher values of the volatility. ELGD is calculated based on PD and σ according to Equations (17) and (18).

Figure 2 shows the relation between PD, expected loss rate given default (ELGD), and the volatility σ. Given the volatility, the relationship between PD and ELGD is monotone: ELGD increases with the PD. The slope of the PD-ELGD curve depends on the volatility resulting in an approximately linear relation for higher values of the volatility. In other words, the positive correlation between the likelihood and severity of credit risk is driven by the random asset value and therefore embedded in the causal model. Note that actual defaults and recoveries (or losses) given default are realizations of random variables in Equations (3) and (16) and will take on values different from their expectations shown in Figure 2.

The Tobit model parameters are estimated conditional on default using the Maximum-Likelihood method. The probability that obligor i has not defaulted conditional on x_i is

$$1 - PD_i = \Phi(\boldsymbol{\beta}'\boldsymbol{x}_i/\sigma) \quad (19)$$

The density of the log-recovery is

$$h(y_i|x_i) \cdot (1 - \Phi(\boldsymbol{\beta}'\boldsymbol{x}_i/\sigma)) = \frac{\phi(-(y_i - \boldsymbol{\beta}'\boldsymbol{x}_i)/\sigma)}{\sigma} \quad (20)$$

and therefore the likelihood for an observed pattern of non-defaults and log-recoveries is

$$\mathfrak{L} = \prod_{i\in\{y_i=0\}} (\Phi(\boldsymbol{\beta}'\boldsymbol{x}_i/\sigma)) \cdot \prod_{i\in\{y_i<0\}} \left(\frac{\phi(-(y_i - \boldsymbol{\beta}'\boldsymbol{x}_i)/\sigma)}{\sigma}\right) \quad (21)$$

It may be more convenient to calculate the log-likelihood

$$\ell = \sum_{i\in\{y_i=0\}} \ln(\Phi(\boldsymbol{\beta}'\boldsymbol{x}_i/\sigma)) + \sum_{i\in\{y_i<0\}} \ln\left(\frac{\phi(-(y_i-\boldsymbol{\beta}'\boldsymbol{x}_i)/\sigma)}{\sigma}\right) \quad (22)$$

which is then maximized with regard to the parameters $\boldsymbol{\beta}$ and σ. The estimator is consistent and asymptotically normal, see Amemiya (1973), Davidson and MacKinnon (1993).

2.4. Extension of the model to asset return correlations

The framework which has been presented thus far incorporates the residual volatilities but does not take into account that the firms' asset returns may be cross-sectionally correlated. Correlations are an important input into modern credit portfolio risk models. Small changes of the correlation between asset returns may have a high impact on the portfolio loss distribution and related measures.

The random error U_i of Equation (10) is decomposed into

$$U_i = \omega \cdot F + \tilde{\sigma} \cdot V_i \quad (23)$$

where F is a systematic error component which simultaneously affects all assets (which is also known as a systematic random effect), and V_i is an idiosyncratic error affecting only asset i, $i=1,\ldots,n$. All errors are standard normally distributed and independent from each other, ω and $\tilde{\sigma}$ are parameters which express the exposure to the systematic and idiosyncratic factors. Note that the total variance is $\mathbb{V}(U_i)=\sigma^2=\omega^2+\tilde{\sigma}^2$. Thus, the correlation between two latent variables Y_i^* and Y_j^* of asset i and j is given by

$$\rho = \frac{\mathbb{C}(Y_i^*, Y_j^*)}{\sigma\cdot\sigma} = \frac{\omega^2}{\sigma^2} = \frac{\omega^2}{\omega^2+\tilde{\sigma}^2} \quad (24)$$

where $\mathbb{C}(\cdot)$ denotes the covariance. This parameter plays a crucial role in most commercial credit risk models as well as Basel II which will be discussed in Section 4.

The latent variable Y_i^* extends to

$$Y_i^* = \boldsymbol{\beta}'\boldsymbol{x}_i + \omega\cdot F + \tilde{\sigma}\cdot V_i \quad (25)$$

F is an annual realization and ω can be estimated using the econometric specification

$$Y_{it}^* = \boldsymbol{\beta}'\boldsymbol{x}_{it} + \omega\cdot F_t + \tilde{\sigma}\cdot V_{it} \quad (26)$$

where $i\in n_t$, $t=1,\ldots,T$. T is the number of time series observations available (eg, the number of years) and n_t is the set of borrowers in period t. Given this notation the parameters can be estimated by the Maximum-Likelihood method as shown below.

Consider a given realization of the systematic factor $F_t=f_t$. Conditional on f_t the Likelihood for each period is

$$\mathfrak{L}_t = \prod_{i\in\{y_{it}=0\}} (\Phi((\boldsymbol{\beta}'\boldsymbol{x}_{it}+\omega\cdot f_t)/\tilde{\sigma})) \times \prod_{i\in\{y_{it}<0\}} \left(\frac{\phi(-(y_{it}-\boldsymbol{\beta}'\boldsymbol{x}_{it}-\omega\cdot f_t)/\tilde{\sigma})}{\tilde{\sigma}}\right) \quad (27)$$

Note that f_t is not observable and that the expectation is calculated with respect to F_t

$$\mathbb{E}(\mathfrak{L}_t) = \int_{-\infty}^{\infty} \prod_{i\in\{y_{it}=0\}} (\Phi((\boldsymbol{\beta}'\boldsymbol{x}_{it}+\omega\cdot f_t)/\tilde{\sigma})) \quad (28)$$

$$\times \prod_{i\in\{y_{it}<0\}} \left(\frac{\phi(-(y_{it}-\boldsymbol{\beta}'\boldsymbol{x}_{it}-\omega\cdot f_t)/\tilde{\sigma})}{\tilde{\sigma}}\right)\phi(f_t)\mathrm{d}f_t \quad (29)$$

Finally, using a time series of T observations, the Log-Likelihood is

$$\ell = \ln\mathfrak{L} = \mathrm{In}\left(\prod_{t=1}^{T}\mathbb{E}(\mathfrak{L}_t)\right) = \sum_{t=1}^{T}\ln\mathbb{E}((\mathfrak{L}_t) \quad (30)$$

which is then maximized with regard to the parameters β, ω and $\tilde{\sigma}$. This operation can be solved numerically using Adaptive Gaussian Quadrature (see Pinheiro and Bates, 1995; Rabe-Hesketh *et al*, 2002).[2]

3. Empirical study

3.1. Data

The empirical analysis is based on recoveries provided by the rating agency Moody's.[3] Moody's measures the recovery of a bond issue upon occurrence of a default event, that is, if

- Interest and/or principal payments are missed or delayed,

[2] A simulation study was conducted to ensure the consistency of the estimators.

[3] Moody's collects general information on bond issuer and issues as well as default and market price information given default events. Note that the only information this paper uses which is uniquely generated by Moody's is the credit rating. This credit rating is very similar to the ones published by other rating agencies such as Standard & Poor's and Fitch: we have hand-collected from the Bloomberg database 63 151 ratings at origination which were rated by Moody's and Standard & Poor's, 38 346 bond ratings at origination which were rated by Moody's and Fitch and 34 578 bond ratings at origination which were rated by Standard & Poor's and Fitch. The Spearman correlation coefficient is 0.9819 for Moody's and Standard and Poor's, 0.9738 for Moody's and Fitch and 0.9702 for Standard and Poor's and Fitch. This analysis of these rating pairs suggests that credit ratings are similar for the three rating agencies as the correlation coefficients are very high. These findings are consistent with Guettler and Wahrenburg (2007) as well as interviews with employees of the three agencies.

- Chapter 11 or Chapter 7 bankruptcy is filed, or
- Distressed exchange such as a reduction of the financial obligation occurs.

In order to guarantee a homogeneous risk segment, the data set was restricted to regular US bond issues. The observation period includes the years 1982–2009. This data set includes 473 951 observations with 1653 default and recovery events. A recovery rate is defined as the ratio of the price of defaulted debt obligations after 30 days of the occurrence of a default event and the par value.

Table 1 shows the number of observations, default rate and mean recovery per year.

Figure 3 shows that the ratio of non-investment grade issues to total issues co-moves with the default rate which demonstrates the ability of Moody's ratings to predict defaults.

The grey bars indicate years which include a period of economic downturn as indicated by the National Bureau of Economic Research (NBER) and show three downturns for the US economy, a first one in 1991 during the First Gulf War, a second one in 2001 during the downturn in the internet industry and the terrorist attack in the US and a third one from 2008 onwards also known as the Global Financial Crisis.

Generally speaking, the default rate decreases and the recovery rate increases with increases in credit quality. This negative relationship between default and recovery rates is displayed in Figure 4. Again, the grey bars indicate years which include a period of economic downturn as indicated by the NBER.

Figures 5 and 6 show histograms for the absolute recoveries and recoveries which are transformed by the natural logarithm. The distribution of the log-recoveries confirms the assumption of a censored standard normal distribution of Y_i^* in Equation (10).

Table 1 Number of observations, default rates and mean recoveries per year

Year	*Total observations*	*Default rate (%)*	*Mean recovery (%)*
1984	805	0.25	34.78
1985	1224	0.49	53.88
1986	1900	1.26	48.42
1987	2637	1.29	67.41
1988	3059	0.92	38.52
1989	3611	0.89	39.41
1990	4003	1.87	30.22
1991	3972	1.59	37.51
1992	4132	0.92	49.28
1993	4666	0.41	39.78
1994	5474	0.29	48.72
1995	6548	0.49	51.13
1996	8433	0.20	44.82
1997	13 046	0.17	50.02
1998	18 176	0.19	45.08
1999	23 473	0.34	35.53
2000	27 182	0.34	22.16
2001	27 656	0.62	37.84
2002	27 740	0.52	41.71
2003	29 140	0.36	33.77
2004	33 689	0.09	59.26
2005	44 511	0.05	57.25
2006	48 789	0.07	63.67
2007	48 190	0.02	74.20
2008	45 454	0.78	17.57
2009	36 441	0.46	28.83
Sum/average	473 951	0.57	44.26

Notes: This table shows the number of observations, default rate and mean recovery per year. Default rate is the ratio between the number of defaulted issuers and the total number of issuers. Recovery rate is the ratio of the price of defaulted debt obligations after 30 days of the occurrence of a default event and the par value.

3.2. Fixed effects models

The data set which includes default events, recoveries and credit ratings is merged with bond issuer characteristics (from Compustat) and macroeconomic information and the following models are estimated:

- Model 1: bond ratings;
- Model 2: bond issue characteristics and bond issuer characteristics;
- Model 3: bond ratings, bond issue characteristics and bond issuer characteristics;
- Model 4: bond ratings, bond issue characteristics, bond issuer characteristics and macroeconomic variable.

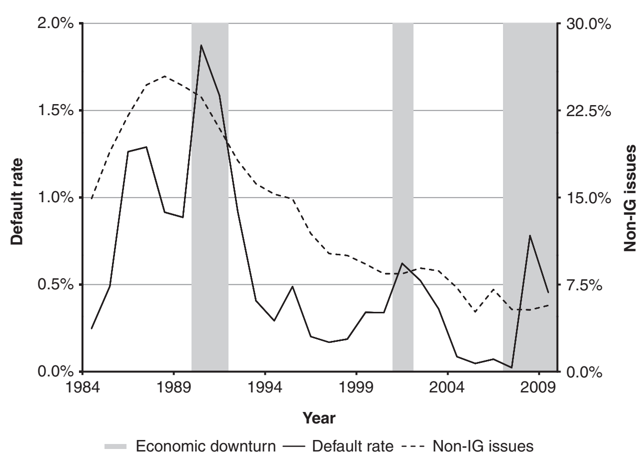

Figure 3 Default rate and non-investment grade rate.
Notes: This figure shows that the ratio of non-investment grade issues to total issues co-moves with the default rate which demonstrates the ability of Moody's ratings to predict defaults. Default rate is the ratio between the number of defaulted issues and the total number of issues. The non-investment grade rate is the number of non-investment grade issues to the total number of issues. The grey bars indicate years which include a period of economic downturn as indicated by the National Bureau of Economic Research (NBER).

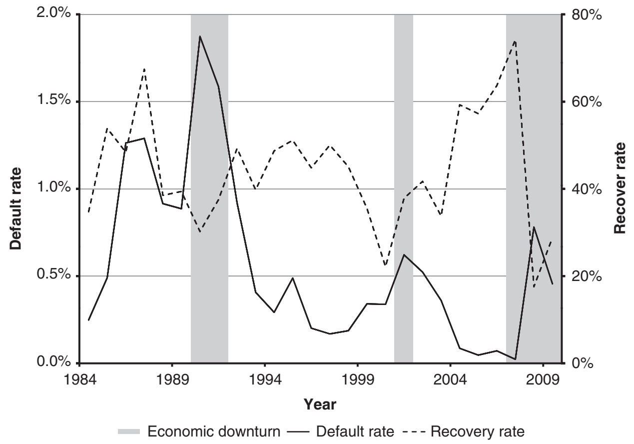

Figure 4 Default rates for all issues and recovery rates for all issues.
Notes: This figure shows the econometric properties of default rates and recovery rates and the negative relationship between these two variables. Default rate is the ratio between the number of defaulted issuers and the total number of issuers. Recovery rate is the ratio of the price of defaulted debt obligations after 30 days of the occurrence of a default event and the par value. The grey bars indicate years which include a period of economic downturn as indicated by the National Bureau of Economic Research (NBER).

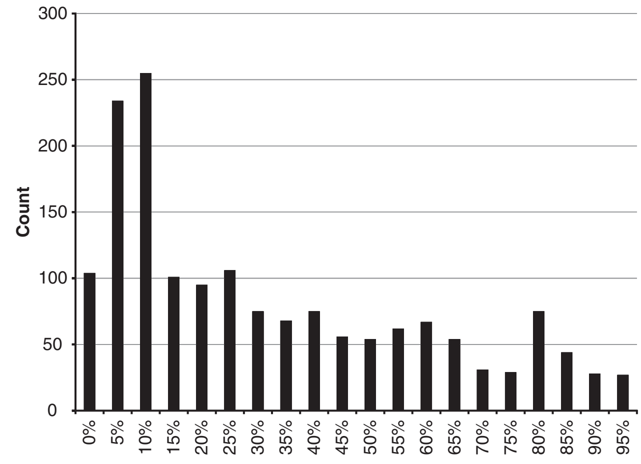

Figure 5 Absolute frequencies for recoveries.
Notes: This shows a histogram for the absolute recoveries. Recovery rate is the ratio of the price of defaulted debt obligations after 30 days of the occurrence of a default event and the par value.

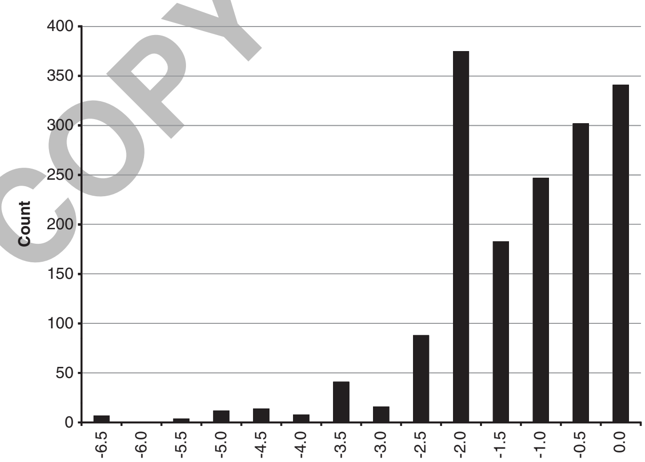

Figure 6 Absolute frequencies for log-recoveries.
Notes: This figure shows a histogram for the recoveries which are transformed by the natural logarithm. The distribution of the log-recoveries confirms the assumption of a censored standard normal distribution of Y_i^*. Recovery rate is the ratio of the price of defaulted debt obligations after 30 days of the occurrence of a default event and the par value.

Bond ratings are categorized into the classes investment grade, Ba, B and Caa-C to ensure that a meaningful number of default events and therefore recoveries are available. The ratings are then dummy-coded as follows:

$$x_{it}^j = \begin{cases} 1, & \text{issue } i \text{ has assigned rating grade } j \\ & \text{at the beginning of year } t \\ 0, & \text{otherwise} \end{cases} \tag{31}$$

for grades j = Ba, B, C.

Bond issue characteristics are the subordination and seniority level (dummy coded categories senior secured, senior unsecured and subordinated). Bond issuer characteristics are the age of a firm (measured by the number of years from the first observation in the Compustat data set), the size (measured by the natural logarithm of total assets), Tobin's Q (ie, the market to book value of assets), net worth (ie, the equity less cash and short-term investments to total assets, the profitability (ie, earnings before interest, tax and depreciation to assets). These variables are commonly used in the finance literature to capture financial risk and control for firm heterogeneity. All financial variables are winsorised at the 5th and 95th percentile to limit the effect of possibly spurious outliers. Financial institutions are major bond issuers and are dummy-coded. We use real GDP growth as a macroeconomic variable. We have tested other variables, such as inflation and interest rates. These variables have a lower explanatory power and we exclude multiple macroeconomic variables from the analysis due to the limited amount of time period for the data sample.

All models include an intercept which represents the reference category, that is, for Model (1) bonds issued by investment grade rated firms, for Model (2) senior secured bonds issued by non-financial institutions and for

Table 2 Parameter estimates for the Tobit models

	Model 1	*Model* 2	*Model* 3	*Model* 4
Intercept	12.1750***	−4.6692***	4.4124***	4.2426***
	(0.2839)	(0.8212)	(0.9952)	(0.9911)
Rating Ba	−1.9831***		−1.8699***	−2.0451***
	(0.1776)		(0.4212)	(0.4234)
Rating B	−4.3996***		−3.6187***	−3.8009***
	(0.1401)		(0.3293)	(0.3351)
Rating Caa-C	−7.9420***		−7.0786***	−7.1722***
	(0.1972)		(0.3926)	(0.3959)
Financial institution		2.4335***	0.3206	0.1389
		(0.1951)	(0.2373)	(0.2401)
Senior unsecured		−2.2169***	−1.5841***	−1.6320***
		(0.4418)	(0.4873)	(0.4827)
Subordinated		−2.7214***	−1.8490***	−1.8749***
		(0.4637)	(0.5004)	(0.4965)
Age		0.0544***	0.0666***	0.0724***
		(0.0057)	(0.0059)	(0.0061)
Size		1.5464***	0.6270***	0.6326***
		(0.0909)	(0.0937)	(0.0936)
Tobin's Q		3.0645***	3.0454***	2.9505***
		(0.2557)	(0.2591)	(0.2571)
Net worth		12.7333***	10.7734***	10.5525***
		(0.5931)	(0.5572)	(0.5527)
Profitability		0.9846	1.7785	0.3337
		(1.8624)	(1.8905)	(1.8646)
GDP growth				15.1641***
				(2.8516)
σ	3.8590***	4.5307***	4.3765***	4.3630***
	(0.0867)	(0.1580)	(0.1520)	(0.1514)
AIC	19 679	9402	8850	8824
Obs.	473 951	237 437	237 437	237 437

Notes: This table shows the results of Tobit models for the logarithm of the recovery rate with rating grades and seniority status as explanatory variables; standard deviations are in parentheses; ***indicates significance at the 1% level, **indicates significance at the 5% level, *indicates significance at the 10% level. AIC is Akaike's Information Criterion and measures the goodness-of-fit.

Models (3) and (4) senior secured bonds issued by investment grade rated non-financial institutions.

An unreported model without co-variates results in the constant (or mean transformed asset return) of 13.1729 and the volatility is 4.8789. This implies a Distance-to-Default of 2.7 (ie, 13.1729/4.8789) and an average probability of default of 0.9% (ie, $\Phi(-2.37)$) and the expected recovery from Equation (17) is 39.00%. Note that the averages presented in Table 1 average over period averages.

Table 2 shows the results of the parameter estimates for Model (1)–Model (4). The standard errors are reported in parentheses in each row below the parameter estimates and both estimates are significantly different from zero. It is interesting that information which explains traditionally the likelihood of default is also able to explain the severity of default (ie recovery rates or LGDs). Most variables are significant and bond issue characteristics, bond issuer characteristics and macroeconomic variable augment the information included in ratings provided by rating agencies.

With regard to Model (1), the results show that all three rating effects are significantly different from zero, which indicates significant differences for the three grades. For instance, the constant for a grade Ba borrower is $12.1750-1.9831=10.1920$, which yields a lower distance to default than an investment grade rating (intercept 12.1750) and therefore a higher PD and a lower expected recovery compared to an IG grade borrower. Similarly, the effects for the other grades can be interpreted where the highest default probability and lowest recovery is assigned to the riskiest grade Caa-C. In Model (2), financial institutions have a lower probability of default and a higher recovery. Senior unsecured loans and subordinated loans have a lower recovery than senior secured loans (the reference category). All financial variables have the expected sign. For example, the probability of default decreases (and recovery increases) with net worth. The interpretation is the same for age, size, Tobin's Q and profitability. Model (3) confirms that the results hold if bond ratings, bond issue characteristics and bond issuer characteristics are included. Model (4) shows that also the macro economy has a significant impact on default probabilities and recovery rates: an increase in the growth rates of real GDP results in lower probabilities of default and a higher recovery.

Financial information for the borrower is not observable for every rated credit and the sample size is higher for Model (1) than for Model (2) to Model (3). Therefore, we focus on the comparison of Model (2) to Model (3), which shows that the inclusion of additional information into the model reduces the volatility from 4.5307 to 4.3630. This demonstrates that the models capture valuable information regarding the idiosyncratic error in the process of the asset returns. Moreover, the last row shows that the Akaike Information Criterion (AIC) declines from 9402 in Model (2) to 8824 in Model (4).

3.3. Random effects model

Table 3 shows the estimation results for models which include a systematic random effect in addition. In Model (5), which include only ratings, the coefficients for the rating grades are close to the coefficients which we obtained without a random effect (ie closer than two standard deviations away). Moreover, we can calculate the total volatility as $\sqrt{\omega^2+\tilde{\sigma}^2}=\sqrt{1.3439^2+3.5595^2}=3.8047$ which is also very close to the volatility from the model without a systematic risk component. However, some part of the total volatility can now be attributed to the systematic variation (ie, the asset correlation). The asset correlation given in the last row is then calculated as $\rho=((\omega^2)/(\omega^2+\tilde{\sigma}^2))=((1.3439^2)/(1.3439^2+3.5595^2))=0.1248$. In other words, 12.5% of total return variance relates to systematic risk and 87.5% to idiosyncratic risk. The

Table 3 Parameter estimates for the Tobit models with fixed and random effects

	Model 5	*Model* 6	*Model* 7	*Model* 8
Intercept	12.0685***	−4.9221***	5.9980***	5.2284***
	(0.3892)	(0.8236)	(0.9647)	(1.0629)
Rating Ba	−2.0006***		−3.2155***	−3.2248***
	(0.1796)		(0.3947)	(0.3948)
Rating B	−4.6298***		−4.873***	−4.8813***
	(0.1500)		(0.3342)	(0.3344)
Rating Caa-C	−8.2223***		−7.9772***	−7.9736***
	(0.2057)		(0.3934)	(0.3933)
Financial institution		2.2147***	−0.08288	−0.08844
		(0.1802)	(0.2138)	(0.2139)
Senior unsecured		−2.5079***	−1.6983***	−1.6996***
		(0.3919)	(0.4201)	(0.4201)
Subordinated		−3.0495***	−1.8684***	−1.8719***
		(0.4206)	(0.4387)	(0.4386)
Age		0.0611***	0.07832***	0.07845***
		(0.0056)	(0.0059)	(0.0059)
Size		1.6635***	0.5597***	0.5639***
		(0.0907)	(0.0851)	(0.0852)
Tobin's Q		2.3664***	2.3484***	2.3477***
		(0.2190)	(0.2194)	(0.2194)
Net worth		11.1376***	9.3533***	9.3556***
		(0.5244)	(0.4904)	(0.4903)
Profitability		−1.2928	0.0666	0.0628
		(1.6807)	(1.6637)	(1.6628)
GDP growth				27.9906
				(17.5909)
ω	1.3439***	1.4983***	1.6585***	1.5466***
	(0.1976)	(0.2575)	(0.2815)	(0.2667)
$\tilde{\sigma}$	3.5595***	3.9476***	3.6910***	3.6907***
	(0.0867)	(0.1355)	(0.1256)	(0.1256)
σ	3.8047	4.2224	4.0465	4.0017
ρ	0.1248	0.1259	0.1680	0.1494
AIC	18 267	8559	7782	7782
Obs.	473 951	237 437	237 437	237 437

Notes: This table shows the results of Tobit models with random effects for the logarithm of the recovery rate with rating grades and seniority status as explanatory variables; standard deviations are in parentheses; ***indicates significance at the 1% level, **indicates significance at the 5% level, *indicates significance at the 10% level. AIC is Akaike's Information Criterion and measures the goodness-of-fit.

inclusion of issuer and issue specific characteristics and the GDP also leads to similar coefficients as in the models without random effect. The asset correlation slightly increases from Model (6) to Model (7) when ratings are additionally included. The dummy for financial institutions is no longer significant indicating that the information about the difference in recoveries between non-financials and financials is already captured by the rating. However, the other coefficients (with the exception of profitability in every model) remain significant and add explanatory power in addition to the rating. When the GDP change is included in Model (8) the correlation is slightly reduced as some information about the macroeconomy which was captured by the random factor in Models (5) to (7) is now explained by the GDP. In this regression the observable risk factor (GDP change) and the random risk factor are somewhat interchangeable, as the GDP change is not statistically significant and correlation is reduced only a little after including GDP change.

4. Implications for portfolio credit risk

4.1. Measurement of portfolio credit risk

This section focuses on the determination of economic and regulatory capital under the Basel II rules. Note that general and specific provisions by the financial institutions should be sufficient to cover the expected losses, while (Tier I and Tier II) capital should be sufficient to cover the difference between the 99.9th percentile of the future loss and the Expected Loss, which is also known as the Credit-Value-at-Risk.

Thus, the probability distribution of the future loss of a credit portfolio and risk figures derived thereof, such as the Expected Loss or the Value-at-Risk are of a central concern to financial institutions. This generally requires the forecast of the loss distribution for a future time period such as one year. In the following, the time subscript is dropped for efficiency of exposition. We denote the exposure of loan i in the portfolio by a_i, which is assumed to be known. Then, the total exposure of the portfolio is $a=\sum_i^n a_i$ and the proportion of loan exposure i in the entire portfolio is denned as $\eta_i = a_i/a$.

The random loss of borrower $i, i=1,\dots,n$ as a fraction of its total exposure is denoted by

$$L_i = (1 - RGD_i)\cdot D_i \tag{32}$$

where RGD_i is the recovery rate given default.

The expected loss of borrower i as a fraction of its total exposure can be calculated as

$$\begin{aligned}
\mathbb{L}_i &= \mathbb{E}(L_i) = \mathbb{E}(D_i|\boldsymbol{x}_i) - \mathbb{E}(RGD_i\cdot D_i|\boldsymbol{x}_i)\\
&= \Phi(-\boldsymbol{\beta}'\boldsymbol{x}_i/\sigma) - \frac{1}{1-\Phi(\boldsymbol{\beta}'\boldsymbol{x}_i/\sigma)}\cdot\exp(\boldsymbol{\beta}'\boldsymbol{x}_i + 0.5\sigma^2)\\
&\quad\times\Phi\left(-\frac{\boldsymbol{\beta}'\boldsymbol{x}_i+\sigma^2}{\sigma}\right)\cdot\Phi(-\boldsymbol{\beta}'\boldsymbol{x}_i/\sigma)\\
&= \Phi(-\boldsymbol{\beta}'\boldsymbol{x}_i/\sigma) - \exp(\boldsymbol{\beta}'\boldsymbol{x}_i + 0.5\sigma^2)\cdot\Phi\left(-\frac{\boldsymbol{\beta}'\boldsymbol{x}_i+\sigma^2}{\sigma}\right)\\
&= PD_i - ERGD_i\cdot PD_i\\
&= PD_i\cdot ELGD_i
\end{aligned} \tag{33}$$

where the second line follows from the fact that the recovery is different from zero only if the borrower defaults and $PD_i = P(D_i = 1|\boldsymbol{x}_i)$ is the probability of default from Equation (13). The loss rate of a portfolio of loans is the weighted average of the individual loan loss rates given by

$$L = \sum_i^n \eta_i(1 - RR_i)\cdot D_i \tag{34}$$

The expected portfolio loss is obtained as

$$\begin{aligned}
\mathbb{L} &= \mathbb{E}\left(\sum_{i=1}^n \eta_i L_i\right) = \sum_{i=1}^n \eta_i\mathbb{E}(L_i) =\\
&= \sum_{i=1}^n \eta_i\cdot[PD_i - ERGD_i\cdot PD_i]\\
&= \sum_{i=1}^n \eta_i\cdot PD_i\cdot ELGD_i
\end{aligned} \tag{35}$$

The dependency structure of the loans is crucial for the probability distribution of the portfolio loss and risk measures such as the Value-at-Risk. Generally speaking, the density of Equation (34) cannot be expressed analytically but can be obtained by Monte-Carlo simulation. Gordy (2003) and Pykhtin (2003) show that an analytical solution for the percentiles of the distribution can be given in the special case of a single stochastic risk factor (which is the case in our model) and an infinitely granular portfolio. The expected loss rate for borrower i is expressed conditional on the systematic risk factor:

$$\begin{aligned}
\mathbb{L}_i(F) &= \mathbb{E}(L_i|F) = \mathbb{E}(D_i|\boldsymbol{x}_i, F) - \mathbb{E}(RR_i\cdot D_i|\boldsymbol{x}_i, F)\\
&= \Phi(-(\boldsymbol{\beta}'\boldsymbol{x}_i + \omega\cdot F)/\tilde{\sigma}) - \frac{1}{1-\Phi((\boldsymbol{\beta}'\boldsymbol{x}_i+\omega\cdot F)/\tilde{\sigma})}\cdot\exp(\boldsymbol{\beta}'\boldsymbol{x}_i + \omega\cdot F + 0.5\tilde{\sigma}^2)\cdot\Phi\left(-\frac{\boldsymbol{\beta}'\boldsymbol{x}_i+\omega\cdot F+\tilde{\sigma}^2}{\tilde{\sigma}}\right)\\
&\quad\times\Phi(-(\boldsymbol{\beta}'\boldsymbol{x}_i + \omega\cdot F)/\tilde{\sigma})\\
&= \Phi(-(\boldsymbol{\beta}'\boldsymbol{x}_i + \omega\cdot F)/\tilde{\sigma}) - \exp(\boldsymbol{\beta}'\boldsymbol{x}_i + \omega\cdot F + 0.5\tilde{\sigma}^2)\cdot\Phi\left(-\frac{\boldsymbol{\beta}'\boldsymbol{x}_i+\omega\cdot F+\tilde{\sigma}^2}{\tilde{\sigma}}\right) = CPD_i(F) - CERGD_i(F)\cdot CPD_i(F)\\
&= CPD_i(F)\cdot CELGD_i(F)
\end{aligned} \tag{36}$$

where

$$CPD_i(F) = \Phi(-(\boldsymbol{\beta}'\boldsymbol{x}_i + \omega \cdot F)/\tilde{\sigma}) \quad (37)$$

is the conditional default probability, while

$$\begin{aligned} CERGD_i(F) = & \frac{1}{1 - \Phi((\boldsymbol{\beta}'\boldsymbol{x}_i + \omega \cdot F)/\tilde{\sigma})} \\ & \times \exp(\boldsymbol{\beta}'\boldsymbol{x}_i + \omega \cdot F + 0.5\tilde{\sigma}^2) \\ & \times \Phi\left(-\frac{\boldsymbol{\beta}'\boldsymbol{x}_i + \omega \cdot F + \tilde{\sigma}^2}{\tilde{\sigma}}\right) \end{aligned} \quad (38)$$

and $CELGD_i(F) = 1 - CERGD_i(F)$ are the conditional expected recovery rate given default and expected loss given default given the systematic factor. The random loss of a granular portfolio is given by

$$L^{\infty} = \sum_i^n \eta_i \mathbb{L}_i(F) \quad (39)$$

and is therefore a monotonically increasing function of the systematic factor. Thus, the α-percentile of the future loss, referred to as Value-at-Risk, is obtained as

$$L^{\alpha} = \sum_i^n \eta_i \mathbb{L}_i\left(F = \Phi^{-1}(1-\alpha)\right) \quad (40)$$

for $0 < \alpha < 1$. Note that this expression reduces to the core of IRB Basel II formula after a simple reparameterization if the recovery is not modelled via the asset value model, and instead, is assumed to be deterministic. In Equation (24), the asset correlation was defined as $\rho = \omega^2/\sigma^2$ with $\sigma^2 = \omega^2 + \tilde{\sigma}^2$. Noting that $1 - \rho = \tilde{\sigma}^2/\sigma^2$ and rewriting the conditional probability of default results in

$$\begin{aligned} CPD_i(F) &= \Phi(-(\boldsymbol{\beta}'\boldsymbol{x}_i + \omega \cdot F)/\tilde{\sigma}) \\ &= \Phi\left(-\frac{\boldsymbol{\beta}'\boldsymbol{x}_i \cdot \sigma}{\tilde{\sigma} \cdot \sigma} - \frac{\omega \cdot F \cdot \sigma}{\tilde{\sigma} \cdot \sigma}\right) \\ &= \Phi\left(-\frac{\boldsymbol{\beta}'\boldsymbol{x}_i}{\sigma} \cdot \frac{\sigma}{\tilde{\sigma}} - \frac{\omega \cdot F}{\sigma} \cdot \frac{\sigma}{\tilde{\sigma}}\right) \\ &= \Phi\left(-\frac{\boldsymbol{\beta}'\boldsymbol{x}_i}{\sigma} \cdot \frac{1}{\sqrt{1-\rho}} - \sqrt{\rho} \cdot F \cdot \frac{1}{\sqrt{1-\rho}}\right) \\ &= \Phi\left(\frac{\Phi^{-1}(PD_i) - \sqrt{\rho} \cdot F}{\sqrt{1-\rho}}\right) \end{aligned} \quad (41)$$

which is the conditional default probability in the Basel II IRB approach in terms of asset correlation where (i) the systematic factor is fixed to the 99.9th percentile of a standard normally distributed variable and (ii) the asset correlation is expressed as a function of the default probability.

Finally, the model allows for a straightforward definition of so-called 'Downturn Loss Given Default' for the Basel II model. The downturn probability of default can be defined by the conditional default probability of Equation (41). A similar interpretation is possible for the recovery (or the loss given default) and the individual or portfolio loss rate. To see this, note that Equations (36), (38) and (40) depend only on the systematic factor. Therefore a 'downturn recovery' is defined as the conditional expected recovery given an adverse realization of the systematic factor according to Equation (38)

$$\begin{aligned} & CERGD_i(F = \Phi^{-1}(1-\alpha)) \\ & = \frac{1}{1 - \Phi\left((\boldsymbol{\beta}'\boldsymbol{x}_i + \omega \cdot \Phi^{-1}(1-\alpha))/\tilde{\sigma}\right)} \\ & \quad \times \exp\left(\boldsymbol{\beta}'\boldsymbol{x}_i + \omega \cdot \Phi^{-1}(1-\alpha) + 0.5\tilde{\sigma}^2\right) \\ & \quad \times \Phi\left(-\frac{\boldsymbol{\beta}'\boldsymbol{x}_i + \omega \cdot \Phi^{-1}(1-\alpha) + \tilde{\sigma}^2}{\tilde{\sigma}}\right) \end{aligned} \quad (42)$$

with a downturn loss given default given as $CELGD_i(F = \Phi^{-1}(1-\alpha)) = 1 - CERGD_i(F = \Phi^{-1}(1-\alpha))$.

In the granular portfolio the Downturn LGD is then given as in Equation (40) where α can be set to 0.999 as proposed by Basel II. In other words, the downturn LGD is then based on the same economic stress as the probability of default.

In summary, given the estimation of a single credit risk model all common credit risk measures may be calculated. This is shown exemplary for the random effect model, Model (1). Table 4 shows in the first three rows the unstressed measures probability of default, loss given default and expected loss for different credit ratings categories. Rows four to seven show the stressed credit measures conditional probability of default, conditional

Table 4 Summary of credit risk measures derived from random effects model

	Rating IG	*Rating Ba*	*Rating B*	*Rating C*
PD	0.0008	0.0041	0.0253	0.1560
ELGD	0.5160	0.5520	0.6061	0.6937
Expected loss	0.0004	0.0022	0.0153	0.1082
Empirical asset correlation	0.1248	0.1248	0.1248	0.1248
CPD	0.0131	0.0483	0.1780	0.5343
CELGD	0.5674	0.6154	0.6877	0.8004
Value-at-risk ($a = 0.999$)	0.0074	0.0297	0.1224	0.4277

Notes: PD is calculated according to Equation (13), ELGD is calculated according to Equation (18), Expected loss is calculated according to Equation (33), CPD is calculated according to Equation (37), CELGD is calculated as one minus ERGD which is calculated according to Equation (38), value-at-risk is calculated according to Equation (36).

expected loss given default and Value-at-Risk based on the 99.9th percentile of the random systematic risk factor.

4.2. Application: Basel II regulatory capital

Table 5 shows the key risk parameters for the calculation of bank capital for the various rating classes. The risk parameters include the unstressed parameters Basel asset correlation, probability of default (PD) and loss rate given default (ELGD) as well as the stressed parameters conditional probability of default (CPD) and downturn loss rate given default. Two approaches are compared for the latter: the downturn LGD may firstly be calculated according to the empirical derivation presented by Equation (42): CELGD or secondly by a proposal by the Federal Reserve System (2006): US CELGD which applies a linear relationship of the downturns LGD on ELGD:

$$US\,CELGD = 8\% + 92\% \times ELGD \qquad (43)$$

Note that under the Basel IRB approach (see Basel Committee, 2006), the regulatory capital is equal to the difference between the Value-at-Risk (ie, the product of Basel CPD and loss given default) and the Expected Loss (ie, the product of PD and ELGD). The Value-at-Risk is based on the 99.9th percentile of the random systematic risk factor and pre-specified asset correlations. The two last rows of Table 5 show that both a deterministic recovery rate (ELGD) as well as the US proposal lead to an underestimation of the regulatory capital which increases with the credit risk in a rating category. For rating C, this underestimation is up to almost 17% in the instance of deterministic recoveries (ELGD) and almost 13% for Equation (43) (US ELGD).

Figure 7 confirms these observations by comparing the capital requirement which results from ELGD, US CELGD and CELGD for various probabilities of default. For high PDs (x-axis) the underestimation can be even higher (up to 25% for US ELGD) than given the values of our empirical study.

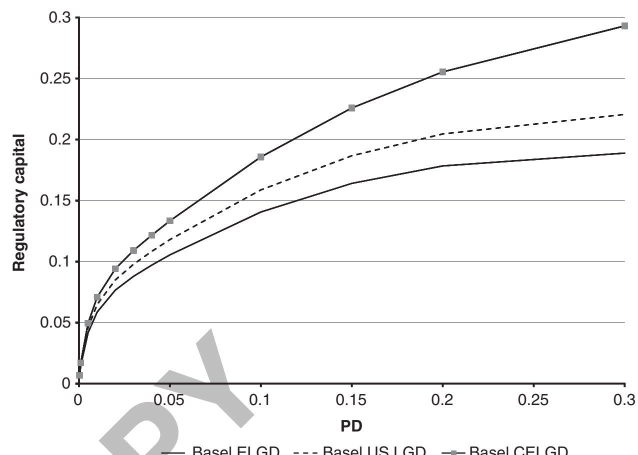

Figure 7 Comparison of Basel II capital for ELGD, US CELGD and CELGD.

Notes: The regulatory capital is the difference between the Value-at-Risk and Expected Loss, that is the difference between the product of (i) CPD and ELGD and the product of PD and ELGD, (ii) CPD and US CELGD and the product of PD and ELGD, as well as (iii) CPD and CELGD and the product of PD and ELGD. No maturity adjustment is included.

Table 5 Summary of Basel II credit risk measures derived from Random Effects model

	Rating IG	*Rating Ba*	*Rating B*	*Rating C*
Basel asset correlation	0.2355	0.2179	0.1539	0.1200
PD	0.0008	0.0041	0.0253	0.1560
CPD	0.0279	0.0868	0.2097	0.5254
ELGD	0.5160	0.5520	0.6061	0.6937
CELGD	0.5674	0.6154	0.6877	0.8004
US CELGD	0.5547	0.5879	0.6376	0.7182
Credit value-at-risk (ELGD)	0.0140	0.0458	0.1140	0.2840
Credit value-at-risk (CELGD)	0.0154	0.0513	0.1311	0.3400
Credit value-at-risk (US CELGD)	0.0151	0.0489	0.1206	0.2969
Underestimation (ELGD)	9.28%	10.72%	13.05%	16.48%
Underestimation (US CELGD)	2.29%	4.65%	8.01%	12.69%

Notes: This table shows in the first panel the unstressed measures probability of default, loss given default and expected loss for different credit ratings categories. The second panel shows the stressed credit measures conditional probability of default, conditional expected loss given default and value-at-risk based on the 99.9th percentile of the random systematic risk factor. The (Basel) asset correlation is calculated by inserting the estimated probability of default into the Internal Ratings-based Approach formula for the asset correlation. PD is calculated according to Equation (13). (Basel) CPD is calculated according to Equation (41) and the Basel asset correlation. ELGD is calculated according to Equation (18). CELGD is calculated according to Equation (42). US CELGD is calculated according to a proposal by US regulators (see Federal Reserve System, 2006) according to Equation (43). This proposal applies a linear formula for economic downturns for the LGD: $8\% + 92\% \times$ ELGD. Credit value-at-risk is equal to the difference between the value-at-risk (ie, the product of Basel CPD and loss given default) and the expected loss (ie, the product of PD and ELGD). The underestimation compares CELGD with the regulatory capital based on (i) ELGD and (ii) US CELGD.

5. Discussion

Financial institutions have implemented in their risk measurement and management frameworks separate models for credit risk parameters which are often provided by external vendors. This practice results in independent and often constant recovery rates. Owing to the model independence, financial institutions were surprised that during the current financial crisis, individual risk parameters deteriorated jointly.

The current risk measurement approach has multiple drawbacks. Firstly, default probabilities, recovery rates and correlations are often modelled as constant over time. Secondly, credit risk parameters are modelled independently and possibly inconsistently. Thus, dependencies between parameters are not included. Thirdly, conditional parameters such as recoveries which are conditional upon the occurrence of default are modelled by (ordinary least square) regression models, which do not take the conditionality into account and lead to inconsistency of the estimated parameters.

In response to these shortcomings, this paper provides a top down approach in which individual credit risk parameters are derived in a closed formula from a single model. This model allows for a dynamic and consistent modelling of credit portfolio risks. This framework is regression based and requires the observation of past recoveries or losses but does not require market prices. A causal relationship between credit quality, recovery rate, volatility, and correlation is established. Formulae for the unconditional and conditional probability of default, recovery rate given default, expected loss are developed. This approach allows financial institutions to have a consistent approach across different credit risk measures used to derive provisions, economic and regulatory capital as well as other applications such as credit pricing.

An empirical analysis provides evidence for the inferred relationship between credit quality, recoveries and correlation. Credit ratings have been highly criticized in the current financial crisis due to their failure to predict corporate credit default risk. Using the consistent estimation technique, the information content of credit ratings and their forecasting ability is augmented by additional borrower characteristics and a macroeconomic variable. The study finds that bond ratings, bond issue characteristics, bond issuer characteristics and macroeconomic variable explain both default probabilities and recovery rates or LGDs. These results are also important in light of the current discussion of the accuracy of credit ratings. The study confirms both views: credit ratings contain useful information but additional variables such as bond issue characteristics, bond issuer characteristics and macroeconomic variable add value.

In addition, the study analysed portfolio credit risk both for economic and regulatory capital allocation and identified an underestimation of the regulatory capital if downturn loss rates given defaults are estimated applying current best practice approaches.

In relation to the current financial crisis, the paper may facilitate changes to best practice in credit portfolio risk modelling and formulation of minimum modelling standards.

Acknowledgements—The authors would like to thank two anonymous referees, the editor of this journal and the participants of the EFA Annual Meeting in Athens, the EFMA Annual Meeting in Milan, the International Conference on Price, Liquidity, and Credit Risks in Konstanz, the 11th Symposium on Finance, Banking, and Insurance, Karlsruhe, the Workshop Risk Management, Obergurgl as well as the financial seminars at the University of Bristol, Deutsche Bundesbank, University of Edinburgh, Global Association of Risk Professionals, Leibniz University Hannover, University of Ulm, Hong Kong Institute of Monetary Research, Melbourne Centre for Financial Studies and The University of Melbourne for valuable comments. The support of the Australian Centre for Financial Studies, the Australian Prudential Regulation Authority, the Centre for International Finance and Regulation and the Hong Kong Institute for Monetary Research is gratefully acknowledged.

References

Acharya VV, Bharath ST and Srinivasan A (2007). Does industry-wide distress affect defaulted firms? Evidence from creditor recoveries. *Journal of Financial Economics* **85**(3): 787–821.

Altman E, Brady B, Resti A and Sironi A (2005). The link between default and recovery rates: Theory, empirical evidence and implications. *Journal of Business* **78**(6): 2203–2227.

Amemiya T (1973). Regression analysis when the dependent variable is truncated normal. *Econometrica* **41**(6): 997–1116.

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

Bierens HJ (2004). *The tobit model*. Working Paper.

Boeker K, Crimmi A and Fink H (2010). Bayesian risk aggregation: Correlation uncertainty and expert judgement. In: Boeker K (ed). *Rethinking Risk Measurement and Reporting—Volumes I*. Risk Books: London.

Carey M (1998). Credit risk in private debt portfolios. *Journal of Finance* **53**(4): 1363–1387.

Crook J and Belotti T (2011). Loss given default models incorporating macroeconomic variables for credit cards. *International Journal of Forecasting* **28**(1): 171–182.

Crook J and Belotti T (2012). Asset correlations for credit card defaults. *Applied Financial Economics* **22**(2): 87–95.

Davidson R and MacKinnon JG (1993). *Estimation and Inference in Econometrics*. Oxford University Press: New York.

Department of the Treasury, Federal Reserve System and Federal Deposit Insurance Corporation (2006). Basel II capital accord—Notice of proposed rulemaking. September.

Dermine J and Neto de Carvalho C (2006). Bank loan losses-given-default: A case study. *Journal of Banking and Finance* **30**(4): 1219–1243.

Dichev I and Piotroski J (2001). The long-run stock returns following bond ratings changes. *Journal of Finance* **56**(1): 173–203.

Dietsch M and Petey J (2004). Should SME exposures be treated as retail or corporate exposures? A comparative analysis of default

probabilities and asset correlations in French and German SMEs. *Journal of Banking and Finance* **28**(4): 773–788.
Duffie D and Singleton K (1999). Modeling term structures of defaultable bonds. *Review of Financial Studies* **12**(4): 687–720.
Duffie D, Saita L and Wang K (2007). Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics* **83**(3): 635–665.
Ederington LH and Goh JC (1993). Is a bond rating-downgrade bad news, good news, or no news for stockholders? *Journal of Finance* **48**(5): 2001–2008.
Goldberger AH (1972). *Selection bias in evaluating treatment effects: Some formal illustrations*. Institute for Research on Poverty Discussion Paper 123–172. University of Wisconsin: Madison.
Gordy MB (2000). A comparative anatomy of credit risk models. *Journal of Banking and Finance* **24**(1–2): 119–149.
Gordy MB (2001). What wags the tail? Identifying the key assumptions in models of portfolio credit risk. In: Alexander C (ed). *Mastering Risk: Volume 2: Applications*, Pearson Education, FT Press: New Jersey. pp 122–140.
Gordy MB (2003). A risk-factor model foundation for ratings-based bank capital rules. *Journal of Financial Intermediation* **12**(3): 199–232.
Greene W (1981). On the asymptotic bias of the ordinary least squares estimator of the tobit model. *Econometrica* **49**(2): 505–513.
Grunert J and Weber M (2009). Recovery rates of commercial lending: Empirical evidence for German companies. *Journal of Banking and Finance* **33**(3): 505–513.
Guettler A and Wahrenburg M (2007). The adjustment of credit ratings in advance of defaults. *Journal of Banking and Finance* **31**(3): 751–767.
Hausman JA and Wise DA (1977). Social experimentation, truncated distributions and efficient estimation. *Econometrica* **45**(4): 919–939.
Hu YT and Perraudin W (2002). *The dependence of recovery rates and defaults*. Mimeo.
Jarrow RA and Turnbull SM (1995). Pricing derivatives on financial securities subject to credit risk. *Journal of Finance* **50**(1): 53–85.
Jarrow RA, Lando D and Turnbull SM (1997). A Markov model for the term structure of credit risk spreads. *Review of Financial Studies* **10**(2): 481–523.
Jorion P, Liu Z and Shi S (2005). Informational effects of regulation FD: Evidence from rating agencies. *Journal of Financial Economics* **76**(2): 309–330.
Kiefer KM (2011). Default estimation, correlated defaults, and expert information. *Journal of Applied Econometrics* **26**(2): 173–192.
Leland HE (1994). Corporate debt value, bond covenants and optimal capital structure. *Journal of Finance* **49**(4): 1213–1252.
Leland HE and Toft K (1996). Optimal capital structure, endogeneous bankruptcy, and the term structure of credit spreads. *Journal of Finance* **51**(3): 987–1019.
Longstaff FA and Schwartz E (1995). A simple approach to valuing risky fixed and floating rate debt. *Journal of Finance* **50**(3): 789–819.
Madan D and Unal H (1995). Pricing the risk of recovery in default with absolute priority rule violation. *Journal of Banking and Finance* **27**(6): 1001–1218.
Maddala GD (1983). *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge University Press: Cambridge.
Matuszyk A, Mues A and Thomas LC (2010). Modelling LGD for unsecured personal loans: Decision tree approach. *Journal of the Operational Research Society* **61**(3): 393–398.
McNeil AJ and Wendin JP (2007). Bayesian inference for generalized linear mixed models of portfolio credit risk. *Journal of Empirical Finance* **14**(2): 131–149.
Merton RC (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance* **29**(2): 449–470.
Morgan DP (2002). Rating banks, risk and uncertainty in an opaque industry. *American Economic Review* **92**(4): 874–888.
Miu P and Ozdemir B (2006). Basel Requirement of Downturn LGD: Modeling and Estimating PD & LGD Correlations. *Journal of Credit Risk* **2**(2): 43–68.
Pan J and Singleton KJ (2008). Default and recovery implicit in the term structure of sovereign CDS spreads. *Journal of Finance* **68**(5): 2345–2384.
Pinheiro JC and Bates DM (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics* **4**(1): 12–35.
Purda L (2007). Stock market reaction to anticipated versus surprise rating changes. *Journal of Financial Research* **30**(2): 301–320.
Pykhtin M (2003). Unexpected recovery risk. *Risk* **16**(8): 74–78.
Qi M and Yang X (2009). Loss given default of high loan-to-value residential mortgages. *Journal of Banking and Finance* **33**(5): 788–799.
Rabe-Hesketh S, Skrondal A and Pickles A (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *Stata Journal* **2**(1): 1–21.
Radelet S and Sachs J (1998). The East Asian financial crisis: Diagnosis, remedies, prospects. *Brookings Papers* **28**(1): 1–90.
Shumway T (2001). Forecasting bankruptcy more accurately: A simple hazard model. *Journal of Business* **74**(1): 101–124.
Tasche D (2004). *The Single Risk Factor Approach to Capital Charges in Case of Correlated Loss Given Default Rates*. Working Paper.
Thomas LC, Matuszyk A and Moore A (2012). Comparing debt characteristics and LGD models for different collections policies. *International Journal of Forecasting* **28**(3): 196–203.
Tobin J (1958). Estimation of relationships for limited dependent variables. *Econometrica* **26**(1): 24–36.

Appendix

Derivation of the expected recovery rate given default

$$ERGD_i = \mathbb{E}(RR_i | D_i = 1, \boldsymbol{x}_i) = \frac{1}{1 - \Phi(\boldsymbol{\beta}' \boldsymbol{x}_i / \sigma)} \cdot \exp(\boldsymbol{\beta}' \boldsymbol{x}_i + 0.5\sigma^2) \cdot \Phi\left(-\frac{\boldsymbol{\beta}' \boldsymbol{x}_i + \sigma^2}{\sigma}\right)$$

Substitute $\mu_i = \beta' x_i$ and $PD_i = 1 - \Phi(\mu_i/\sigma)$ and write

$$\begin{aligned}
ERGD_i &= \int_{-\infty}^{0} \exp(y_i) \cdot h(y_i | Y_i < 0, \boldsymbol{x}_i) \mathrm{d}y_i \\
&= \int_{-\infty}^{0} \exp(y_i) \cdot \frac{\phi(-(y_i - \mu_i)/\sigma)}{\sigma \cdot (1 - \Phi(\mu_i/\sigma))} \mathrm{d}y_i \\
&= \frac{1}{\sigma \cdot PD_i} \int_{-\infty}^{0} \exp(y_i) \cdot \frac{1}{\sqrt{2\pi}} \cdot \exp\left(\frac{-(-y_i + \mu_i)^2}{2\sigma^2}\right) \mathrm{d}y_i \\
&= \frac{1}{\sigma \cdot PD_i} \cdot \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{0} \exp\left(y_i - \frac{\mu_i^2 - 2y_i\mu_i + y_i^2}{2\sigma^2}\right) \mathrm{d}y_i \\
&= \frac{1}{\sigma \cdot PD_i} \cdot \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{0} \exp\left(\frac{2\sigma^2 y_i - \mu_i^2 + 2y_i\mu_i - y_i^2}{2\sigma^2}\right) \mathrm{d}y_i \\
&= \frac{1}{\sigma \cdot PD_i} \cdot \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{0} \exp\left(\frac{-y_i^2 + 2y_i(\mu_i + \sigma^2) - (\mu_i + \sigma^2)^2 - \mu_i + (\mu_i + \sigma^2)^2}{2\sigma^2}\right) \mathrm{d}y_i \\
&= \frac{1}{\sigma \cdot PD_i} \cdot \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{0} \exp\left(\frac{-\left(y_i - (\mu_i + \sigma^2)\right)^2 + 2\mu_i\sigma^2 + \sigma^4}{2\sigma^2}\right) \mathrm{d}y_i \\
&= \frac{1}{\sigma \cdot PD_i} \cdot \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{0} \exp\left(\frac{-\left(y_i - (\mu_i + \sigma^2)\right)^2}{2\sigma^2}\right) \cdot \exp(\mu_i + 0.5\sigma^2) \mathrm{d}y_i \\
&= \frac{1}{PD_i} \cdot \exp\left(\mu_i + 0.5\sigma^2\right) \cdot \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-\left(y_i - (\mu_i + \sigma^2)\right)^2}{2\sigma^2}\right) \mathrm{d}y_i \\
&= \frac{1}{PD_i} \cdot \exp\left(\mu_i + 0.5\sigma^2\right) \cdot \Phi\left(-\frac{\mu_i + \sigma^2}{\sigma}\right)
\end{aligned}$$

Received January 2012;
accepted May 2012 after one revision