DELFT UNIVERSITY OF TECHNOLOGY

REPORT 08-10

Generalized Beta Regression Models for Random Loss-Given-Default

X. Huang and C.W. Oosterlee

ISSN 1389-6520

Reports of the Department of Applied Mathematical Analysis

Delft 2008

Generalized Beta Regression Models for Random Loss-Given-Default

Xinzheng Huang[a,b,*], Cornelis. W. Oosterlee[a,c]

[a] *Delft Institute of Applied Mathematics, Delft University of Technology, Mekelweg 4, 2628CD, Delft, the Netherlands*

[b] *Group Risk Management, Rabobank, Croeselaan 18, 3521CB, Utrecht, the Netherlands*

[c] *CWI - National Research Institute for Mathematics and Computer Science, Kruislaan 413, 1098 SJ, Amsterdam, the Netherlands*

May 23, 2008

Abstract

We propose a new framework for modeling systematic risk in Loss-Given-Default (LGD) in the context of credit portfolio losses. The class of models is very flexible and accommodates well skewness and heteroscedastic errors. The quantities in the models have simple economic interpretation. Inference of models in this framework can be unified. Moreover, it allows efficient numerical procedures, such as the normal approximation and the saddlepoint approximation, to calculate the portfolio loss distribution, Value at Risk (VaR) and Expected Shortfall (ES).

1 Introduction

In the context of credit portfolio losses, the quantity Loss-Given-Default (LGD) is the proportion of the exposure that will be lost if a default occurs. The uncertainty about the actual LGD constitutes an important source of the credit portfolio risk in addition to the default risk. In practice, e.g., in both CreditMetrics (Gupton et al. 1997) and KMV Portfolio Manager (Gupton & Stein 2002), the uncertainty in the LGD rates of defaulted obligors is assumed to be a Beta random variable independent for each obligor. The Beta distribution is well-known to be very flexible, modeling quantities constrained in the interval

*Corresponding author; E-mail: X.Huang@ewi.tudelft.nl. This author would like to thank Dr. Mâcé Mesters from Rabobank for helpful discussions.

$[0, 1]$. Depending on the choice of parameters, the probability density function can be unimodal, U-shaped, J-shaped or uniform.

However, extensive empirical evidence, see e.g., Hu & Perraudin (2002), Altman et al. (2005), shows that this simple approach is insufficient. It is now well understood that LGD is positively correlated to the default rate, in other words, LGD is high when the default rate is high, which suggests that there is also systematic risk in LGD, just like in the default rates. A heuristic justification is that the LGD is determined by the collateral value which is sensitive to the state of the economy.

Based on results of a non-parametric estimation procedure, Hu & Perraudin (2002) further showed that without taking into account the PD/LGD correlation the economic capital, or Value at Risk (VaR), of a loan portfolio can be significantly underestimated. This has a critical consequence on risk management practice. In the Basel II Accord this issue is addressd by the notion of "downturn LGD".

The insight of LGD being subject to systematic risk dates back to Frye (2000), in which the LGD is modeled by a normal distribution. An obvious problem with this model is that LGD is unbounded in $\mathbb{R}$ and can thus be negative. To ensure the nonnegativity of LGD, Pykhtin (2003) employs a truncated log-normal distribution for the LGD. Andersen & Sidenius (2004) propose the use of a probit transform of the LGD such that the transformed LGD is normally distributed. The probit transformation guarantees that the LGD stays in the interval $[0, 1]$. In a similar manner Düllmann & Trapp (2004), Rösch & Scheule (2005) employ a logit transform of the LGD. Rather different from the above approaches, Giese (2006), Bruche & González-Aguado (2008) extend the static Beta distribution assumption in CreditMetrics and KMV Portfolio Manager by modeling the LGD by a mixture of Beta distributions that depend on the systematic risk.

In this article we propose a *Generalized Beta Regression* (GBR) *framework* to model the Loss-Given-Default. This framework generalizes the Beta Regression model proposed by Ferrari & Cribari-Neto (2004) and is very similar to a class of models derived from Generalized Linear Models (GLM). Our models are called Generalized Beta Regression Models since the LGD is always assumed to be (conditionally) Beta distributed. The models by Giese (2006), Bruche & González-Aguado (2008) can be regarded as special examples in our GBR framework. The quantities in our models have simple economic interpretation. Compared to the transformed LGD models, GBR models accommodate better skewness and possible heteroscedastic errors. Inference of models in this framework can be unified. Moreover, the GBR framework allows both the normal approximation and the saddlepoint approximation to efficiently calculate the portfolio loss distribution.

The rest of the article is organized as follows. In section 2 we introduce the Vasicek's Gaussian one-factor model as the default model and give a brief summary of existing random LGD models. Section 3 elaborates on the GBR framework including the basic Beta regression model and two extensions. In Section 4 we discuss methods for parameter estimation and provide a calibration example.

Section 5 explains techniques for efficient loss distribution approximation in the GBR framework.

2 Credit portfolio loss

Consider a credit portfolio consisting of n obligors, each with exposure at default (EAD) w_i and probability of default (PD) p_i. Obligor i is subject to default after a fixed time horizon and the default can be modeled as a Bernoulli random variable D_i such that

$$D_i = \begin{cases} 1 & \text{with probability } p_i, \\ 0 & \text{with probability } 1 - p_i. \end{cases}$$

Let Loss-Given-Default (LGD), the proportion of the exposure that will be lost if a default occurs, be denoted by Λ, then the loss incurred due to the default of obligor i is given by $L_i = w_i \Lambda_i D_i$. It follows that the portfolio loss is given by

$$L = \sum_{i=1}^{n} L_i = \sum_{i=1}^{n} w_i \Lambda_i D_i.$$

To evaluate the distribution of L, a key issue is essentially to model the various dependence effects, including the dependence between defaults, the dependence between between LGDs and the dependence between PD and LGD. A convenient approach is to utilize a latent factor model and introduce systematic risk in both PD and LGD.

2.1 Default model

We consider the Vasicek (2002) Gaussian one-factor model as our default model. Based on Merton's firm value model, the Vasicek model evaluates the default of an obligor in terms of the evolution of its asset value. For obligor i, default occurs when the standardized asset log-return X_i, is less than some pre-specified threshold γ_i, where X_i is normally distributed and $\mathbb{P}(X_i < \gamma_i) = p_i$. X_i is decomposed into a systematic part Y, representing the state of the economy, and an idiosyncratic part Z_i, such that

$$X_i = \sqrt{\rho} Y + \sqrt{1-\rho} Z_i, \tag{1}$$

where Y and all Z_i are i.i.d standard normal random variables and ρ is the common pairwise correlation. It is now easily deduced that X_i and X_j are conditionally independent given the realization of Y. This implies that L_i and L_j are also conditionally independent given Y.

Denote by $p_i(y) = \mathbb{P}[L_i = 1 | Y = y]$, i.e., the probability of default conditional on the common factor $Y = y$. Then

$$p_i(y) = \mathbb{P}[L_i = 1 | Y = y] = \mathbb{P}[X_i < \gamma_i | Y = y] = \Phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho} y}{\sqrt{1-\rho}}\right), \tag{2}$$

where Φ is the cdf of the standard normal distribution.

2.2 LGD models

A variety of models in which LGD is subject to systematic risk can be found in the literature. Within a one-factor framework, Frye (2000) proposed a model in which the LGD is normally distributed and influenced by the same systematic factor Y that drives the PD, so that

$$\Lambda = \mu + \sigma\xi,$$
$$\xi = \sqrt{\tilde{\rho}}Y + \sqrt{1-\tilde{\rho}}\epsilon,$$

where ξ and ϵ are both standard normally distributed. In this way both the dependence between LGDs and the dependence between PD and LGD are modeled simultaneously. The parameters μ and σ can be understood to be the expected LGD and the LGD volatility, respectively. Unfortunately, the LGD is unbounded in $\mathbb{R}$ and can thus be negative. To ensure the nonnegativity of LGD, Pykhtin (2003) employs a log-normal distribution for the LGD,

$$\Lambda = \left(1 - e^{\mu+\sigma\xi}\right)^{+},$$

Other extensions include Andersen & Sidenius (2004), choosing a probit transformation

$$\Lambda = \Phi(\mu + \sigma\xi),$$

where Φ is again the cdf of the standard normal distribution and Düllmann & Trapp (2004), Rösch & Scheule (2005) that employ a logit transformation

$$\Lambda = \frac{1}{1+\exp(\mu+\sigma\xi)}.$$

All three transformations for the LGD above guarantee that the LGD lies in the interval $[0, 1]$. However the parameters μ and σ no longer have a convenient economic interpretation as in Frye's model.

A rather different approach from the above extends the static Beta distribution assumption as it is present in CreditMetrics and KMV Portfolio Manager. Giese (2006), Bruche & González-Aguado (2008) model the LGD by a mixture of Beta distributions

$$\Lambda \sim Beta(\alpha, \beta)$$

where both α and β are functions of common factor Y.

Along this direction, we here propose the Generalized Beta Regression (GBR) framework for random LGD. The GBR framework includes Giese (2006), Bruche & González-Aguado (2008) as special examples. The class of models is very flexible and the quantities in our models have an easy economic interpretation. Compared to the transformed LGD models, the GBR models accommodate better skewness and possible heteroscedastic errors.

3 Generalized Beta Regression Models

3.1 Parameterization of a beta distribution

Recall that the probability density function of a beta distribution with parameters $\alpha > 0$, $\beta > 0$ reads

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1},$$

where $B(\cdot,\cdot)$ denotes the beta function and $\Gamma(\cdot)$ the Gamma function.

The Beta distribution is well-known to be very flexible, modeling quantities constrained in the interval $[0,1]$. Depending on the choice of parameters, the probability density function can be unimodal, U-shaped, J-shaped or uniform. The expectation and variance of a beta distributed variable X are given by

$$\mu = E(X) = \frac{\alpha}{\alpha+\beta}, \tag{3}$$

$$\sigma^2 = Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{\mu(1-\mu)}{\alpha+\beta+1}. \tag{4}$$

Let $\varphi = \alpha + \beta$, then φ can be regarded as a *dispersion parameter* in the sense that, for a given μ, the variance is determined by the size of φ.

The parameters α and β can be formulated in terms of the mean and dispersion in the following way

$$\alpha = \mu\varphi, \quad \beta = (1-\mu)\varphi. \tag{5}$$

Apparently, a beta distribution can also be uniquely determined by its mean and dispersion.

3.2 Beta Regression Model

The Generalized Beta Regression framework is characterized by the following elements,

1. the LGD is assumed to be beta distributed, conditional on some covariates,
2. the beta distribution is parameterized by its mean and dispersion, rather than its natural parameters (α, β).

This framework generalizes the Beta Regression model proposed by Ferrari & Cribari-Neto (2004) for modeling rates and proportions. The models from the GBR framework are similar to a class of models derived from Generalized Linear Models (GLM).

The Generalized Linear Models have been developed since the seminal paper Nelder & Wedderburn (1972) as an extension to the classical linear regression models. In a GLM, the response variable X is in the exponential family. Its density can be represented in the form

$$f(x;\theta,\varphi) = e^{a(\varphi)[x\theta - b(\theta)] + c(\varphi,x)}, \tag{6}$$

For a comprehensive exposition of GLM we refer to McCullagh & Nelder (1989).

We start the explanation of the GBR framework with the basic Beta Regression model proposed in Ferrari & Cribari-Neto (2004). This basic approach only models the mean μ and treats the dispersion parameter φ as a nuisance parameter. With some abuse of language we also call the model for the mean, μ, a GLM (although the probability density function of the beta distribution cannot be written in the form (6) and therefore it does not fit in the framework of GLM). The mean model has the following two components:

- a *linear predictor* η

  $$\eta = \alpha\zeta \tag{7}$$

  where ζ is a vector of explanatory variables and α is a vector of the corresponding regression coefficients. As convention the first element of ζ is set to be 1, so that the first element of α is an intercept term.

- a monotonic, differentiable *link function* g

  $$g(\mu) = \eta. \tag{8}$$

In particular, the inverse of the link function, $g^{-1}(\cdot)$, should form a mapping from $\mathbb{R}$ to $[0,1]$, which is exactly the range of μ. This can be achieved by a variety of link functions, such as the logit link

$$\mu = \frac{e^{\eta}}{1+e^{\eta}},\ \eta = \log\left(\frac{\mu}{1-\mu}\right), \tag{9}$$

or the probit link

$$\mu = \Phi(\eta),\ \eta = \Phi^{-1}(\mu). \tag{10}$$

Both the logit and probit link functions have a symmetric form about $\mu = 1/2$. If however it is believed that symmetric links are not justified, asymmetric link functions like the scaled probit link and the complementary log-log link can be used instead.

The most parsimonious model for LGD subject to systematic risk is a one-factor model with $\zeta = [1, Y]^T$, where Y is the common factor that also drives the default process. An example of such a model is given in Giese (2005), where the mean is modeled by

$$\mu = 1 - a_0\left(1 - p_i(Y)^{a_1}\right)^{a_2} \tag{11}$$

and φ is considered a nuisance parameter. According to Schuermann (2004), other factors that have a significant effect on LGD include seniority, collateral and type of industry.

3.3 Extensions

The basic beta regression model above can be readily extended in various ways. One extension is to model the mean and dispersion jointly, rather than treating the dispersion parameter φ as a nuisance parameter which is either fixed or known. This is in the same spirit as the Joint Generalized Linear Model (JGLM) from the GLM framework, see e.g., Nelder & Lee (1991), Lee & Nelder (1998).

The dispersion φ can be modeled by a separate GLM,

$$h(\varphi) = b\zeta,$$

where h is also a link function. A simple way to ensure $\varphi > 0$ is to use a log link so that

$$\varphi = e^{b\zeta}. \tag{12}$$

A model of this type, but using a different version of the dispersion parameter, has been suggested in Bruche & González-Aguado (2008), where the Beta distribution is parameterized by its two parameters α and β.

Since both α and β are positive, Bruche & González-Aguado (2008) suggest to employ the following log-linear model for the two parameters,

$$\alpha = e^{c\zeta},\ \beta = e^{d\zeta}, \tag{13}$$

where, as usual, ζ is a vector of covariates and c and d are vector coefficients. However, α and β are both shape parameters, and an economic interpretation of such a model is very difficult. In this regard, we note that by substituting (13) into (3) we obtain

$$\mu = \frac{\alpha}{\alpha+\beta} = \frac{e^{(c-d)\zeta}}{1+e^{(c-d)\zeta}},$$

which is simply a logit model with vector coefficient $c - d$. The variance is then given by

$$\sigma^2 = \frac{\mu(1-\mu)}{\alpha+\beta+1} = \frac{\mu^2(1-\mu)}{\alpha+\mu},$$

so that the following dispersion parameter is adopted,

$$\varphi = \alpha = e^{c\zeta}.$$

A second extension is that the mean parameter μ can be modeled by a Generalized Linear Mixed Model (GLMM). GLMM extends GLM by adding normally distributed random effects in the linear predictor η. The simplest mixed model is the *random intercept model*

$$g(\mu) = \eta = a\zeta + \nu, \tag{14}$$

where, in addition to the fixed effect $a\zeta$, η also has a single component of random effect ν that follows a univariate normal distribution $N(0, \sigma_\nu^2)$. In our setting

ν can be thought of as a latent factor for the LGD independent of the fixed effects.

Such a GLMM, along with the probit link (10), is employed to model the mean LGD in Hillebrand (2006). Other applications of GLMM for portfolio credit default and migration risk can be found in McNeil & Wendin (2006, 2007).

Note that the two extensions above can be readily combined to form a new model that jointly models the mean and dispersion by means of GLMMs, i.e., fixed and random effects can be included in the modeling of both mean and dispersion. Further extensions are possible, e.g., replacing the linear predictor by a Generalized Additive Model (GAM), see Hastie & Tibshirani (1990), or adding multi-level random effects in the GLMM.

4 Estimation

In this section, we discuss the parameter estimation in the GBR framework by

1. least squares,
2. maximum likelihood estimation (MLE).

The former requires only the knowledge of yearly mean LGD and LGD volatility and can be used as the first approximation to the MLE.

Suppose we have a time series of LGD data for $T \in \mathbb{N}$ years. Let K_t be the number of defaulted obligors in year t and $\lambda_{t,k}$ be the observed LGD for defaulted obligor k, $t = 1, \dots, T$, $k = 1, \dots, K_t$. Each year, a realization of the common factor Y_t can be inferred from the default model and historical default data. The value of $Y_t, t = 1, \dots, T$ should be considered a known fixed effect in the LGD model.

From now on we call the three models in the GBR framework GBR-GLM, GBR-JGLM and GBR-GLMM, respectively. The parameters to be estimated are: $\{a, \varphi\}$ in GBR-GLM, $\{a, b\}$ in GBR-JGLM or $\{a, \varphi, \sigma_\nu\}$ in GBR-GLMM, where a represent the vector coefficients in the linear predictor (7), b the vector coefficients in the linear predictor (12), φ is the dispersion parameter and σ_ν^2 is the variance of the random effect ν in (14).

4.1 Least squares

The method of least squares we propose here only requires the knowledge of the yearly mean LGD and LGD volatility for parameter estimation. The estimates of the yearly mean LGD and LGD volatility for $t = 1, \dots, T$ can be obtained by matching the first and second moments of the LGD realizations $\lambda_{t,k}$ such that

$$m_t = \frac{1}{K_t} \sum_{k=1}^{K_t} \lambda_{t,k}, \quad \sigma_t^2 = \frac{1}{K_t} \sum_{k=1}^{K_t} \lambda_{t,k}^2 - m_t^2.$$

Estimation of a and μ

The estimate for parameter a can be obtained by employing a linear regression of the transformed mean LGD $g(m_t)$ on Y_t and other covariates,

$$g(m_t) = \hat{a}\zeta_t + \nu_t, \tag{15}$$

where ν_t is the residual term. In the GBR-GLM and GBR-JGLM

$$\hat{\mu}_t = g^{-1}(\hat{a}\zeta_t). \tag{16}$$

And in the GBR-GLMM ν_t is taken to be the realized random effect in year t so that

$$\hat{\mu}_t = g^{-1}(\hat{a}\zeta_t + \nu_t). \tag{17}$$

Estimation of b or φ

The estimation of the parameters b or φ takes the prediction of $\hat{\mu}_t$, produced by (16) or (17), as an input. From (4), we obtain

$$\varphi_t = \frac{\hat{\mu}_t(1 - \hat{\mu}_t)}{\sigma_t^2} - 1.$$

In both the GBR-GLM and GBR-GLMM the dispersion parameter φ is treated as a nuisance parameter. Its method-of-moments estimator is simply

$$\hat{\varphi} = \frac{1}{T}\sum_{t=1}^{T}\varphi_t.$$

In the GBR-JGLM, the coefficient b can be calculated by linear regression of the transformed dispersion $h(\varphi_t)$ on covariate vector ζ such that

$$h(\varphi_t) = \hat{b}\zeta_t + \epsilon_t.$$

Estimation of σ_ν in GBR-GLMM

The moment based estimate for σ_ν^2 is given by

$$\hat{\sigma}_\nu^2 = \frac{1}{T}\sum_{t=1}^{T}\nu_t^2,$$

where ν_t is the residual term in (15).

4.2 Maximum likelihood estimation

Parameter estimation by the method of maximum likelihood is also straightforward in the GBR framework. In the models without random effects, i.e., GBR-GLM and GBR-JGLM, the log-likelihood function to be maximized reads

$$\ell(\mu,\varphi) = \sum_{t=1}^{T}\sum_{k=1}^{K_t}\{(\mu\varphi-1)\log(\lambda_{t,k}) + [(1-\mu)\varphi-1]\log(1-\lambda_{t,k}) + \\ + \log\Gamma(\varphi) - \log\Gamma(\mu\varphi) - \log\Gamma[(1-\mu)\varphi]\} \quad (18)$$

The score function, the gradient of the log-likelihood function and the Fisher information matrix, i.e., the variance of the score, can be formulated explicitly in terms of polygamma functions. They are given in Appendix A. Asymptotic standard errors of the maximum likelihood estimates of the parameters can be computed from the Fisher information matrix.

Since the corresponding estimating equations do not admit a closed form solution, numerical maximization of the log-likelihood is necessary. Estimates by the method of least squares may be used as the initial approximations to the solutions of the likelihood equations.

We remark that the maximum likelihood estimation in Ferrari & Cribari-Neto's Beta Regression Model is already implemented in the statistical computing software R (www.r-project.org) in package 'betareg' so that it can be used immediately.

Marginal likelihood in GBR-GLMM

With the presence of random effects, the samples are no longer independent. In the random intercept model (14), the LGD's in year t are only independent conditional on the random effect ν_t. Since we are only interested in inference of the variance of the random component ν, but not in its realizations, the random effect needs to be integrated out. Therefore we maximize the marginal log-likelihood,

$$\ell_m(a,\varphi,\sigma_\nu) = \sum_{t=1}^{T}\log\left(\int\prod_{k=1}^{K_t} L(a,\varphi,\zeta_t,\nu_t;\lambda_{t,k})p_{\sigma_\nu}(\nu_t)d\nu_t\right)$$

where $p_{\sigma_\nu}(\cdot)$ is the pdf of a normal distribution with mean 0 and variance σ_ν^2, and $L(\ ;\lambda_{t,k})$ is the likelihood of $\{LGD = \lambda_{t,k}\}$ given ν_t. The integral can be efficiently evaluated by Gaussian quadrature. Alternatively, the marginal likelihood can be approximated analytically by the use of the Laplace approximation to the integral, such as the penalized quasi-likelihood (PQL) estimation (Breslow & Clayton 1993) and the h-likelihood (Lee & Nelder 2001), thus avoiding numerical integration.

Finally we note that the likelihood ratio test based on large sample inference can be employed for model selection. Information criteria such as Akaike's information criterion (AIC) or the Bayesian information criterion (BIC) can also be used.

4.3 A simulation study

As we presently have no access to historical LGD data, we show in this section how the models in the GBR framework can be calibrated to fit some simulated data. As a result the focus here is not to identify possible covariates that influence LGD but only to show that parameter estimation and model selection in the GBR framework can be easily dealt with.

The simulated LGD observations are based on data from Bruche & González-Aguado (2008), which give the annual default frequency, number of defaults, mean LGD and LGD volatility for the years 1982 - 2005. For completeness the data are reproduced in Appendix B. For each year, a realization of the LGD is simulated for each defaulted obligor from a beta distribution matching the empirical mean and variance. This gives in total $1,123$ LGD observations in $T = 24$ years.

First, we fit the Vasicek default model. We assume that all obligors in the portfolio have the same probability of default p and asset correlation ρ. Denote by p_t the annual default frequency. We take the MLE's for ρ and p according to Düllmann & Trapp (2004),

$$\rho = \frac{Var\left[\Phi^{-1}(p_t)\right]}{1 + Var\left[\Phi^{-1}(p_t)\right]}, \quad p = \Phi\left(\frac{\sum_{t=1}^{T}\Phi^{-1}(p_t)}{T\sqrt{1 + Var\left[\Phi^{-1}(p_t)\right]}}\right),$$

where $Var[\delta] = \frac{1}{T}\sum_{t=1}^{T}\delta_t^2 - \left(\frac{1}{T}\sum_{t=1}^{T}\delta_t\right)^2$. This yields

$$\rho = 0.0569, \quad p = 0.0153. \tag{19}$$

The common factor Y_t for year t can be estimated as follows,

$$Y_t = \frac{\Phi^{-1}(p) - \sqrt{1-\rho}\,\Phi^{-1}(p_t)}{\sqrt{\rho}}.$$

As for the LGD model, we only consider one covariate, which is the common factor Y in the default model. The mean LGD is fitted using a logit link

$$\mu = \frac{e^{a_1 + a_2 Y}}{1 + e^{a_1 + a_2 Y}}$$

in the GBR-GLM and GBR-JGLM and

$$\mu = \frac{e^{a_1 + a_2 Y + \nu}}{1 + e^{a_1 + a_2 Y + \nu}}$$

in the GBR-GLMM. In the GBR-JGLM model, the dispersion parameter is modeled to be

$$\varphi = e^{b_1 + b_2 Y}.$$

The estimates given by the method of least squares are presented in Table 1. Note that the coefficient a_2 is negative, indicating negative relationship between Y and mean LGD, just as expected. These estimates are used as the first approximation to the MLE.

	GBR-GLM	GBR-JGLM	GBR-GLMM
a_1	0.3718	0.3718	0.3718
a_2	-0.3054	-0.3054	-0.3054
φ	4.1914	-	4.0907
b_1	-	1.3505	-
b_2	-	-0.0033	-
σ_ν	-	-	0.2686

Table 1: Estimates given by the method of least squares for different models.

The maximum likelihood estimates for the various parameters are given in Table 2. For the GBR-GLM model, we also report in parenthesis the asymptotic standard errors of the estimates. The Wald test confirms that both a_1 and a_2 are statistically significant (both p-values < 0.0001).

The log-likelihood ratio statistics of GBR-JGLM and GBR-GLMM to GBR-GLM are $-402.74 - (-403.34) = 0.6$ and $-402.74 - (-468.78) = 66.04$, respectively. They correspond to p-value 0.44 and < 0.0001 for the chi-square distribution with one degree of freedom. It is clear that GBR-GLMM provides a significant improvement over the basic GBR-GLM, whereas GBR-JGLM fails to do so. AIC and BIC lead to the same conclusion (see Table 2). We remark that this however does not suggest that GBR-JGLM should be abandoned in general since the idea of jointly modeling mean and dispersion may be meaningful if we include other covariates.

	GBR-GLM	GBR-JGLM	GBR-GLMM
a_1	0.3459 (0.0359)	0.3471	0.3319
a_2	-0.3213 (0.0298)	-0.3246	-0.3307
φ	3.0276 (0.1149)	-	3.3240
b_1	-	1.0879	-
b_2	-	-0.0306	-
σ_ν	-	-	0.2943
-2ℓ	-402.74	-403.34	-468.78
AIC	-398.74	-395.34	-460.78
BIC	-381.67	-375.25	-440.69

Table 2: Maximum Likelihood Estimates of various models.

It is also interesting to see how much the choice of a LGD model can influence the VaR at the portfolio level. We consider a portfolio of 100 obligors with uniform PD p and correlation ρ as in (19) and exposures as follows

$$w_i = \begin{cases} 1, & k = 1, \ldots, 20 \\ 4, & k = 21, \ldots, 40 \\ 9, & k = 41, \ldots, 60 \\ 16, & k = 61, \ldots, 80 \\ 25, & k = 81, \ldots, 100. \end{cases}$$

We compare three models for the LGD, (i) the GBR-GLM, (ii) the GBR-GLMM and (iii) the constant LGD model. For the GBR-GLM and GBR-GLMM, the LGD parameters are taken from Table 2. In the constant LGD model, we take for all obligors $\Lambda = 0.58$, matching the expected LGD $E_Y[\mu(Y)]$ in the GBR-GLM model, where $E_Y(\cdot)$ denotes the expectation obtained by integrating over Y.

The portfolio loss distributions plotted in Figure 1(a) are based on Monte Carlo simulation with two hundred thousand scenarios. On the one hand the curves of GBR-GLMM and GBR-GLM are almost identical, with GBR-GLMM producing a slightly heavier tail. On the other hand the loss distribution under the constant LGD model deviates substantially from the other two models with random LGD.

We then look at the portfolio VaR at three particular confidence levels 99%, 99.9%, 99.99%, which are illustrated in Figure 1(b). Compared to the constant LGD model, the GBR-GLM (GBR-GLMM) increases the VaR at the three levels by a factor of 1.26 (1.26), 1.32 (1.36) and 1.36 (1.41), respectively. It is apparent that ignoring the systematic risk in the LGD significantly underestimates risk. Moreover, the further in the tail, the higher the degree of underestimation.

5 Loss distribution approximations

An important advantage of adopting the Generalized Beta Regression framework for random LGD is that it allows both the normal approximation and the saddlepoint approximation to efficiently calculate the portfolio loss distribution, thus avoiding the need for time-consuming simulation. For simplicity, we derive the formulas only for the basic GBR-GLM with a single covariate Y, or equivalently, a single-factor model. Generalization to more complex models is rather straightforward.

5.1 Normal approximation

First of all, in the case of a large homogeneous portfolio, the expected loss from obligor i conditional on Y reads

$$E[L_i(Y)] = w_i E[D_i(Y)] E[\Lambda_i(Y)] = w_i p_i(Y) \mu_i(Y). \tag{20}$$

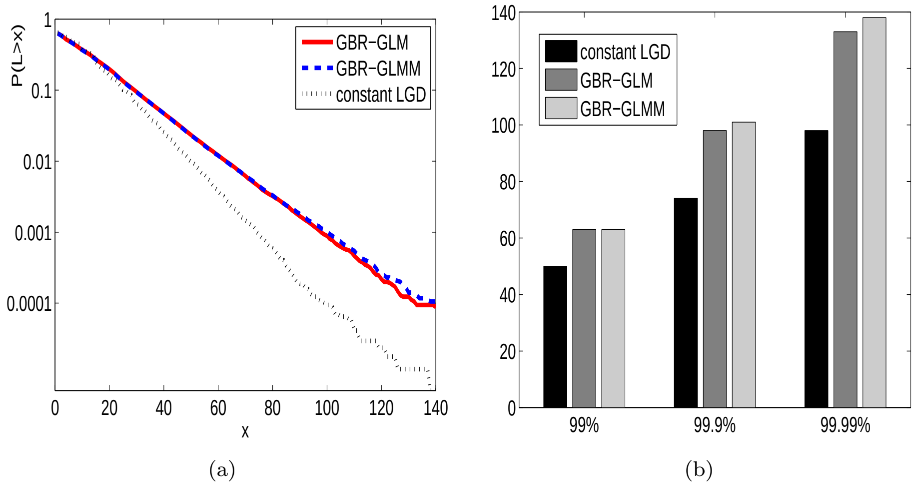


Figure 1: (a) The portfolio loss distributions and (b) the portfolio VaR at three confidence levels under the three LGD models. The results are based on Monte Carlo (MC) simulation of two hundred thousand scenarios. For GBR-GLM and GBR-GLMM, the LGD parameters are taken from Table 2. In the constant LGD model $\Lambda = 0.58$ for all obligors.

A version of the large homogeneous portfolio approximation similar to that in the Vasicek model can also be obtained for random LGD:

$$\frac{L(Y)}{\sum_{i=1}^{n} w_i} \to \frac{\sum_{i=1}^{n} w_i p_i(Y)\mu_i(Y)}{\sum_{i=1}^{n} w_i} \quad a.s.$$

When the portfolio is however not sufficiently large or not very homogeneous, unsystematic risk arises. The normal approximation improves on the large homogeneous portfolio approximation by taking into account the variability of portfolio loss L conditional on the common factor Y. The conditional portfolio loss $L(Y)$ can be approximated by a normally distributed random variable with mean $M(Y)$ and variance $V^2(Y)$ such that

$$M(Y) = \sum_{i=1}^{n} w_i p_i(Y)\mu_i(Y),$$

$$V^2(Y) = \sum_{i=1}^{n} E[L_i^2(Y)] - \sum_{i=1}^{n} E[L_i(Y)]^2,$$

where

$$\begin{aligned} E[L_i^2(Y)] &= w_i^2 E[D_i(Y)]E[\Lambda_i^2(Y)] = w_i^2 p_i(Y)E[\Lambda_i^2(Y)] \\ &= w_i^2 p_i(Y)[\mu_i^2(Y) + Var(\Lambda|Y)] \\ &= w_i^2 p_i(Y)[\mu_i^2(Y) + \mu_i(Y)(1-\mu_i(Y))/(1+\varphi_i)]. \end{aligned}$$

The conditional tail probability is $P(L > x|Y) = \Phi\left(\frac{M(Y)-x}{V(Y)}\right)$ and it follows that the unconditional tail probability reads

$$P(L > x) = \int \Phi\left(\frac{M(Y)-x}{V(Y)}\right) dP(Y) = E_Y\left[\Phi\left(\frac{M(Y)-x}{V(Y)}\right)\right]. \tag{21}$$

5.2 Saddlepoint approximation

The saddlepoint approximation has been presented by Huang, Oosterlee & van der Weide (2007) as an efficient tool to estimate the portfolio credit loss distribution in the Vasicek model. More importantly, it handles well exposure concentration when the portfolio is dominated by a few loans significantly larger than the rest. In Huang, Oosterlee & van der Weide (2007) the LGD was however assumed to be constant.

The use of the saddlepoint approximation only requires the existence of the moment generating function (MGF), which makes the beta distribution assumption for LGD in our framework very attractive. Recall that the MGF of a beta distributed random variable with parameters (α, β) is a confluent hypergeometric function as follows,

$$\begin{aligned}
MGF(t) &= \int e^{tx}\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}dx = \sum_{k=0}^{+\infty}\int \frac{t^k x^k}{k!}\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}dx \\
&= \sum_{k=0}^{+\infty}\frac{t^k}{k!}\frac{B(\alpha+k,\beta)}{B(\alpha,\beta)}\int \frac{x^{\alpha+k-1}(1-x)^{\beta-1}}{B(\alpha+k,\beta)}dx \\
&= \sum_{k=0}^{+\infty}\frac{t^k}{k!}\frac{\Gamma(\alpha+k)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha+\beta+k)} \\
&= {}_1F_1(\alpha, \alpha+\beta; t).
\end{aligned}$$

By basic differentiation, we obtain the following first and second derivatives of the MGF

$$\begin{aligned}
MGF'(t) &= \sum_{k=1}^{+\infty}\frac{t^{k-1}}{(k-1)!}\frac{\Gamma(\alpha+k)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha+\beta+k)} = \sum_{k=0}^{+\infty}\frac{t^k}{k!}\frac{\Gamma(\alpha+k+1)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha+\beta+k+1)} \\
&= \sum_{k=0}^{+\infty}\frac{t^k}{k!}\frac{\Gamma(\alpha+k+1)\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+1)\Gamma(\alpha+\beta+k+1)}\frac{\alpha}{\alpha+\beta} \\
&= {}_1F_1(\alpha+1, \alpha+\beta+1; t)\frac{\alpha}{\alpha+\beta}, \\
MGF''(t) &= {}_1F_1(\alpha+2, \alpha+\beta+2; t)\frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}.
\end{aligned}$$

In our setting, the obligors are independent conditional on the common factor Y and (α, β) conditional on Y can be determined by (5). The conditional MGF and the cumulant generating function (CGF), denoted by κ, of the portfolio loss are then given by

$$MGF(t,Y) = \prod_{i=1}^{n} \left[1 - p_i + p_i \, {}_1F_1(\alpha, \alpha+\beta; w_i t)\right],$$

$$\kappa(t,Y) = \log(MGF(t,Y)) = \sum_{i=1}^{n} \log\left[1 - p_i + p_i \, {}_1F_1(\alpha, \alpha+\beta; w_i t)\right].$$

For simplicity of notation, we have suppressed the explicit dependence of p_i and (α, β) on the common factor Y.

The derivatives of the conditional CGF up to second order are

$$\kappa'(t,Y) = \sum_{i=1}^{n} \frac{w_i p_i \, {}_1F_1(\alpha+1, \alpha+\beta+1; w_i t)}{1 - p_i + p_i \, {}_1F_1(\alpha, \alpha+\beta; w_i t)} \frac{\alpha}{\alpha+\beta},$$

$$\kappa''(t,Y) = \sum_{i=1}^{n} \left\{ \frac{w_i^2 p_i \alpha(\alpha+1) \, {}_1F_1(\alpha+2, \alpha+\beta+2; w_i t)}{(\alpha+\beta)(\alpha+\beta+1)[1 - p_i + p_i \, {}_1F_1(\alpha, \alpha+\beta; w_i t)]} - \frac{w_i^2 p_i^2 \alpha^2 \, {}_1F_1(\alpha+1, \alpha+\beta+1; w_i t)^2}{(\alpha+\beta)^2 [1 - p_i + p_i \, {}_1F_1(\alpha, \alpha+\beta; w_i t)]^2} \right\}.$$

After finding the saddlepoint $\tilde{t}$ that solves $\kappa'(\tilde{t}, Y) = x$ for the loss level x, the tail probability conditional on Y can be approximated by the Lugannani & Rice (1980) formula

$$\mathbb{P}(L > x|Y) = 1 - \Phi(z_l) + \phi(z_l) \left(\frac{1}{z_w} - \frac{1}{z_l} \right), \tag{22}$$

where $z_w = \tilde{t}\sqrt{\kappa''(\tilde{t}, Y)}$, $z_l = sgn(\tilde{t})\sqrt{2[x\tilde{t} - \kappa(\tilde{t}, Y)]}$ and ϕ is the pdf of the standard normal distribution.

Integrating over Y gives the unconditional tail probability $\mathbb{P}(L > x)$, from which the portfolio Value at Risk (VaR) can be derived. Formulas for the calculation of other risk measures like VaR contribution, Expected shortfall (ES) and ES contribution can be found in Huang, Oosterlee & van der Weide (2007).

5.3 Numerical results

We now illustrate the performance of the normal and saddlepoint approximations in loss distribution calculation. We take a homogeneous portfolio with $n = 100$ obligors, each with

$$w = 1, \; p = 0.005, \; \rho = 0.18,$$

The parameters in the LGD are

$$a = [0.37, -0.32], \; \varphi = 3.16,$$

with a logit link for mean LGD. This leads to the following specification of (conditional) mean LGD

$$\mu = \frac{1}{1 + e^{-0.37+0.32Y}}.$$

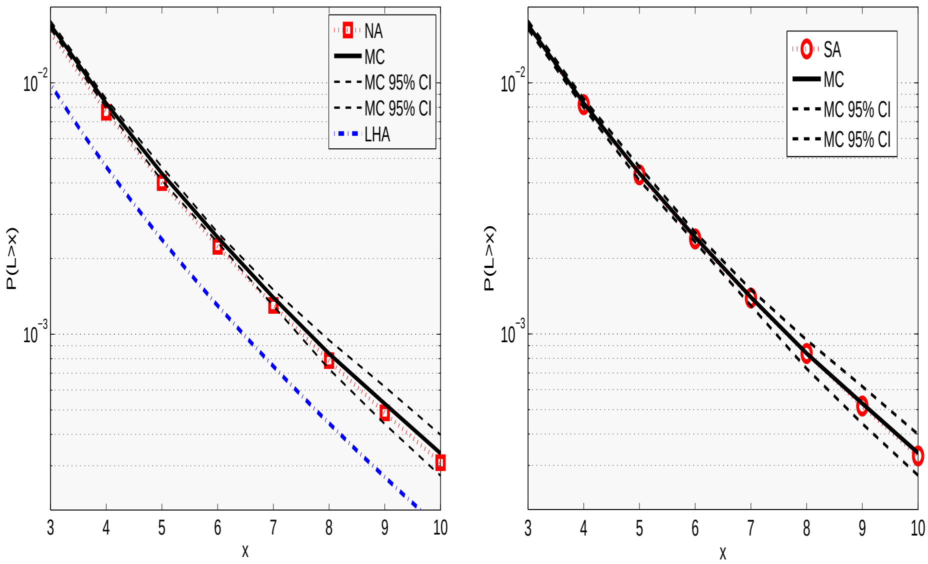
We compare the loss distributions obtained from various approximation methods to the results from a Monte Carlo (MC) simulation. Our benchmark is the sample mean and the accompanying 95% confidence intervals obtained by 10 subsamples of Monte Carlo simulation with 20 thousand replications each. The performance of the approximations is demonstrated in Figure 2(a)-(b).

The large homogeneous approximation (LHA) results deviate considerably from our benchmark. This is not surprising as the size of the portfolio is rather small. The normal approximation (NA) provides a significant improvement over the LHA and underestimates risk only slightly. Some of its tail probability estimates however fall out of the 95% confidence interval. By comparison, the saddlepoint approximation (SA) is able to give all tail probability estimates within the 95% confidence interval. The loss distribution given by the saddlepoint approximation is indistinguishable from the benchmark. A remark is that the calculation of the loss distribution in MATLAB costs roughly 4 seconds for the normal approximation and 4 minutes for the saddlepoint approximation on a Pentium 4 2.8 GHz desktop.

Finally we calculate the VaR for the portfolio considered in §4.3 with LGD modeled by the GBR-GLM. The results are given in Table 3. The MC results are based on two hundred thousand simulated scenarios and can be regarded as our benchmark. In this example the saddlepoint approximation is again very accurate. The normal approximation is however rather unsatisfactory: at all three levels relative errors are around 8%. This is certainly due to the existence of exposure concentration as the variation in the exposures is not negligible. For more details on how robust the normal approximation and saddlepoint approximation are in terms of handling exposure concentration, we refer to Huang, Oosterlee & Mesters (2007).

	$\mathrm{VaR}_{99\%}$	$\mathrm{VaR}_{99.9\%}$	$\mathrm{VaR}_{99.99\%}$
MC	63	98	133
NA	58	90	123
SA	63	97	133

Table 3: Approximations to the portfolio VaR at three confidence levels. The LGD model adopted here is GBR-GLM. The MC results are based on two hundred thousand simulated scenarios and can be regarded as our benchmark.



(a)

(b)

Figure 2: The loss distribution obtained from (a) the large homogeneous approximation (LHA), the normal approximation (NA) and (b) the saddlepoint Approximation (SA) compared to results based on Monte Carlo (MC) simulation of two hundred thousand scenarios. The MC 95% confidence interval (CI) are based on the standard deviation calculated using 10 simulated sub-samples of 20 thousand scenarios each.

6 Conclusion

In this paper we have proposed the Generalized Beta Regression framework for modeling systematic risk in Loss-Given-Default (LGD) in the context of credit portfolio losses. The GBR framework provides great flexibility in random LGD modeling. The quantities in the GBR models have simple economic interpretation. We have shown that parameter estimation and model selection are straightforward in this framework. Moreover, it has been demonstrated that the portfolio loss distribution can be efficiently evaluated by both the normal approximation and the saddlepoint approximation.

References

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

Andersen, L. & Sidenius, J. (2004), 'Extensions to the Gaussian copula: Random recovery and random factor loadings', *Journal of Credit Risk* **1**(1), 29–70.

Breslow, N. & Clayton, D. (1993), 'Approximate inference in generalized linear mixed models', *Journal of the American Statistical Association* **88**, 9–25.

Bruche, M. & González-Aguado, C. (2008), Recovery rates, default probabilities and the credit cycle, Discussion papers, CEMFI. Available at SSRN: http://ssrn.com/abstract=934348.

Düllmann, K. & Trapp, M. (2004), Systematic risk in recovery rates - an empirical analysis of US corporate credit exposures, Deutsche Bundesbank Discussion Paper Series 2: Banking and Financial Supervision, No 02/2004.

Ferrari, S. & Cribari-Neto, F. (2004), 'Beta regression for modeling rates and proportions', *Journal of Applied Statistics* **31**, 799–815.

Frye, J. (2000), 'Depressing recoveries', *RISK* (November), 108–111.

Giese, G. (2005), 'The impact of PD/LGD correlations on credit risk capital', *RISK* (April), 79–84.

Giese, G. (2006), 'A saddle for complex credit portfolio models', *RISK* (July), 84–89.

Gupton, G. & Stein, R. (2002), *LossCalc: Model for predicting loss given default (LGD)*, Moody's Investors Service.

Gupton, G., Finger, C. & Bhatia, M. (1997), *CreditMetrics – Technical document*, J.P. Morgan, New York.

Hastie, T. J. & Tibshirani, R. J. (1990), *Generalized Additive Models*, New York: Chapman & Hall.

Hillebrand, M. (2006), 'Modelling and estimating dependent loss given default', *RISK* (September), 120–125.

Hu, Y. T. & Perraudin, W. (2002), The dependence of recovery rates and defaults, Working paper, Birkbeck College.

Huang, X., Oosterlee, C. W. & Mesters, M. (2007), 'Computation of var and var contribution in the Vasicek portfolio credit loss model: A comparative study', *Journal of Credit Risk* **3**(3), 75–96.

Huang, X., Oosterlee, C. W. & van der Weide, J. A. M. (2007), 'Higher order saddlepoint approximations in the Vasicek portfolio credit loss model', *Journal of Computational Finance* **11**(1), 93–113.

Lee, Y. & Nelder, J. A. (1998), 'Generalized linear models for the analysis of quality-improvement experiments', *Canadian Journal of Statistics* **26**(1), 95–105.

Lee, Y. & Nelder, J. A. (2001), 'Hierarchical generalized linear models: a synthesis of generalized linear models, random effect models and structured dispersions', *Biometrika* **88**, 987–1006.

Lugannani, R. & Rice, S. (1980), 'Saddlepoint approximations for the distribution of the sum of independent random variables', *Advances in Applied Probability* **12**, 475–490.

McCullagh, P. & Nelder, J. (1989), *Generalized Linear Models*, London: Chapman & Hall.

McNeil, A. J. & Wendin, J. P. (2006), 'Dependent credit migrations', *Journal of Credit Risk* **2**(3), 87–114.

McNeil, A. J. & Wendin, J. P. (2007), 'Bayesian inference for generalized linear mixed models of portfolio credit risk', *Journal of Empirical Finance* **14**, 131–149.

Nelder, J. A. & Lee, Y. (1991), 'Generalized linear models for the analysis of Taguchi-type experiments', *Applied Stochastic Models and Data Analysis* **7**, 107–120.

Nelder, J. A. & Wedderburn, R. W. M. (1972), 'Generalized linear models', *Journal of the Royal Statistical Society: Series A* **135**(3), 370–384.

Pan, J. & Thompson, R. (2007), 'Quasi-monte Carlo estimation in generalized linear mixed models', *Computational Statistics & Data Analysis* **51**, 5756–5775.

Pykhtin, M. (2003), 'Unexpected recovery risk', *RISK* (August), 74–78.

Rösch, D. & Scheule, H. (2005), 'A multi-factor approach for systematic default and recovery risk', *Journal of Fixed Income* **15**(2), 63–75.

Schuermann, T. (2004), What do we know about loss given default?, *in* D. Shimko, ed., 'Credit Risk Models and Management, Second Edition', Risk Books, London, pp. 249–274.

Vasicek, O. (2002), 'Loan portfolio value', *RISK* (December), 160–162.

A Score function and Fisher information matrix

In this appendix we give details about the score function and the Fisher information matrix for the parameters appearing in the GBR-GLM and GBR-JGLM models. The score function may help to accelerate the convergence in the MLE procedure and the Fisher information matrix leads to the asymptotic standard errors of the maximum likelihood estimates of the parameters in the models. In the GBR-GLMM the corresponding formulas get more complicated and lengthy and therefore they are omitted here. We refer the interested reader to Pan & Thompson (2007) for an example.

The score function, i.e., the partial derivative of the log-likelihood function with respect to parameters (μ, φ), reads

$$\frac{\partial \ell}{\partial \mu} = \varphi \left\{ \log\left(\frac{\lambda}{1-\lambda}\right) - \Psi(\mu\varphi) + \Psi[(1-\mu)\varphi] \right\}, \tag{23}$$

$$\frac{\partial \ell}{\partial \varphi} = \mu \log \lambda + (1-\mu)\log(1-\lambda) + \Psi(\varphi) - \mu\Psi(\mu\varphi) - (1-\mu)\Psi[(1-\mu)\varphi], \tag{24}$$

where λ is a realization of the LGD and $\Psi(\cdot)$ is the digamma function.

The second order partial derivatives of the log-likelihood function with respect to parameters (μ, φ) are

$$\frac{\partial^2 \ell}{\partial \mu^2} = -\varphi^2\{\Psi'(\mu\varphi) + \Psi'[(1-\mu)\varphi]\}, \tag{25}$$

$$\frac{\partial^2 \ell}{\partial \varphi^2} = \Psi'(\varphi) - \mu^2\Psi'(\mu\varphi) - (1-\mu)^2\Psi'[(1-\mu)\varphi], \tag{26}$$

$$\frac{\partial^2 \ell}{\partial \mu \partial \varphi} = \frac{1}{\varphi}\frac{\partial \ell}{\partial \mu} - \varphi\{\mu\Psi'(\mu\varphi) - (1-\mu)\Psi'[(1-\mu)\varphi]\}. \tag{27}$$

where $\Psi'(\cdot)$ is the trigamma function.

In the GBR-GLM, the parameters to be estimated are a and φ. The score function for φ is given by (24); the score function with respect to a_i, the i-th element of a, is given by

$$\frac{\partial \ell}{\partial a_i} = \frac{\partial \ell}{\partial \mu}\frac{\partial \mu}{\partial a_i} = \varphi \left\{ \log\left(\frac{\lambda}{1-\lambda}\right) - \Psi(\mu\varphi) + \Psi[(1-\mu)\varphi] \right\} \frac{\zeta_i}{g'(\mu)}. \tag{28}$$

The Fisher information matrix is the negative of the expectation of the second derivative of the log-likelihood with respect to the parameters. The entries in the Fisher information matrix are

$$-E\left(\frac{\partial^2 \ell}{\partial \varphi^2}\right) = -\frac{\partial^2 \ell}{\partial \varphi^2}, \tag{29}$$

$$-E\left(\frac{\partial^2 \ell}{\partial a_i \partial a_j}\right) = -\frac{\partial^2 \ell}{\partial \mu^2}\frac{\zeta_i\zeta_j}{(g'(\mu))^2}, \tag{30}$$

$$-E\left(\frac{\partial^2 \ell}{\partial a_i \partial \varphi}\right) = \varphi\{\mu\Psi'(\mu\varphi) - (1-\mu)\Psi'[(1-\mu)\varphi]\}\frac{\zeta_i}{g'(\mu)}. \tag{31}$$

In the GBR-JGLM, the parameters to be estimated are a and b. The score function for the coefficient a is given by (28) and that for b_i, the i-th element of b, is as follows

$$\begin{aligned}\frac{\partial \ell}{\partial b_i} = \frac{\partial \ell}{\partial \varphi}\frac{\partial \varphi}{\partial b_i} = &\{\mu \log \lambda + (1-\mu)\log(1-\lambda) + \\ &+\Psi(\varphi) - \mu\Psi(\mu\varphi) - (1-\mu)\Psi[(1-\mu)\varphi]\}\frac{\zeta_i}{h'(\varphi)}.\end{aligned} \tag{32}$$

The Fisher information matrix contains $-E\left(\frac{\partial^2 \ell}{\partial a_i \partial a_j}\right)$ given by (30) and

$$-E\left(\frac{\partial^2 \ell}{\partial b_i \partial b_j}\right) = -\frac{\partial^2 \ell}{\partial \varphi^2} \frac{\zeta_i \zeta_j}{(h'(\varphi))^2}, \tag{33}$$

$$-E\left(\frac{\partial^2 \ell}{\partial a_i \partial b_j}\right) = \varphi\{\mu \Psi'(\mu\varphi) - (1-\mu)\Psi'[(1-\mu)\varphi]\} \frac{\zeta_i \zeta_j}{g'(\mu) h'(\varphi)}. \tag{34}$$

B LGD Statistics by Year

In this section we present a table of the LGD statistics by year, from 1982 until 2005. This table is taken from Bruche & González-Aguado (2008), where mean recovery rate (RR) is reported instead of LGD. The column of mean LGD here is calculated to be 1 minus RR, i.e., LGD=1-RR.

Year	PD	# of defaults	mean LGD	LGD volatility
1982	1.18%	12	60.49%	14.9%
1983	0.75%	5	51.07%	23.53%
1984	0.9%	11	51.19%	17.38%
1985	1.1%	16	54.59%	21.87%
1986	1.71%	24	63.91%	18.82%
1987	0.94%	20	46.64%	26.94%
1988	1.42%	30	63.43%	17.97%
1989	1.67%	41	56.54%	28.78%
1990	2.71%	76	74.76%	22.28%
1991	3.26%	95	59.95%	26.09%
1992	1.37%	35	45.55%	23.38%
1993	0.55%	21	62.46%	20.11%
1994	0.61%	14	54.46%	20.46%
1995	1.01%	25	57.1%	25.25%
1996	0.49%	19	58.1%	24.68%
1997	0.62%	25	46.54%	25.53%
1998	1.31%	34	58.9%	24.56%
1999	2.15%	102	71.01%	20.4%
2000	2.36%	120	72.49%	23.36%
2001	3.78%	157	76.66%	17.87%
2002	3.6%	112	69.97%	17.18%
2003	1.92%	57	62.67%	23.98%
2004	0.73%	39	52.19%	24.1%
2005	0.55%	33	41.37%	23.46%