Further Investigation of Parametric Loss Given Default Modeling

Phillip Li
Min Qi
Xiaofei Zhang
Xinlei Zhao

Office of the Comptroller of the Currency

Economics Working Paper 2014-2

July 2014

Keywords: loss given default, Tobit regression, smearing estimator, Monte Carlo estimator, transformation regressions, retransformation, inverse Gaussian regression, beta transformation, censored gamma regression, two-tiered gamma regression, inflated beta regression, two-step regression, fractional response regression.

JEL classifications: G21, G28.

All four authors of this paper are with the Office of the Comptroller of the Currency. Phillip Li is a financial economist, Min Qi is a Deputy Director, and Xinlei Zhao is a lead modeling expert in the Credit Risk Analysis Division. Xiaofei Zhang is a senior financial economist in the Market Risk Analysis Division. To comment, please contact Xinlei Zhao at Office of the Comptroller of the Currency, 400 7th St. SW, Mail Stop 6E-3, Washington, DC 20219, or call (202) 649-5544; or e-mail Xinlei.Zhao@occ.treas.gov.

The views expressed in this paper are those of the authors alone and do not necessarily reflect those of the Office of the Comptroller of the Currency or the U.S. Department of the Treasury. The authors would like to thank Jessica Scully for editorial assistance. The authors take responsibility for any errors.

Further Investigation of Parametric Loss Given Default Modeling

Phillip Li
Min Qi
Xiaofei Zhang
Xinlei Zhao

July 2014

Abstract: We conduct a comprehensive study of some new or recently developed parametric methods to estimate loss given default using a common data set. We first propose to use a smearing estimator, a Monte Carlo estimator, and a global adjustment to refine transformation regressions that address loss given default boundary values. Although these refinements only marginally improve model performance, the smearing and Monte Carlo estimators help reduce the sensitivity of transformation regressions to the adjustment factor. We then implement five parametric models (two-step, inflated beta, Tobit, censored gamma, and two-tier gamma regressions) that are not thoroughly studied in the literature but are all designed to fit the unusual bounded bimodal distribution of loss given default. We find that complex parametric models do not necessarily outperform simpler ones, and the non-parametric models may be less computationally burdensome. Our findings suggest that complicated parametric models may not be necessary when estimating loss given default.

Economics Working Paper 2014-2

1. Introduction

Probability of default (PD) and loss given default (LGD) are the two key determinants of the premium of risky bonds, credit default swap spreads, and credit risks of loans and other credit exposures. They are also among the key parameters in the Basel internal ratings-based framework for banks' minimum regulatory capital requirements.[1] Thus, a good understanding of PD and LGD is crucial for fixed-income investors, rating agencies, bankers, bank regulators, and academics. Between the two parameters, LGD is relatively more understudied partly because of the lack of data and risk drivers for it, although LGD research has been growing in recent years.

Besides data limitations and the lack of risk drivers, another challenge in modeling LGD is that the LGD values have an unusual distribution. LGD values are often bounded between 0 and 1 (including observations of exactly 0 or 1), and the distribution tends to be bimodal with modes close to the boundary values. These distributional characteristics make standard statistical models, such as the linear regression model estimated with ordinary least squares (OLS), theoretically inappropriate for LGD modeling.

The importance of accounting for the unusual distribution of LGD is widely acknowledged in the literature,[2] and researchers have attempted to use various statistical methods to address the aforementioned challenges. In general, the semi-parametric and non-parametric methods are found to outperform parametric methods (see Bastos [2010], Loterman et al. [2012], Qi and Zhao [2011], Altman and Kalotay [2014], Hartmann-Wendels, Miller, and Tows [2014], and Tobback et al. [2014]). The papers comparing various parametric methods in the literature, however, are far from exhaustive and do not compare some of the newer parametric models that might be more suitable for fitting the unusual LGD distribution (e.g., the inflated beta distribution [Ospina and Ferrari (2010a, b)] and the gamma regressions [Sigrist and Stahel (2011)]). How these

[1] The Basel II risk parameters are PD, LGD, and exposure at default. Effective maturity is also needed for corporate, sovereign, and bank exposures.

[2] See, for example, Hu and Perraudin (2002), Siddiqi and Zhang (2004), Gupton and Stein (2005), Dermine and Neto de Carvalho (2006), Bastos (2010), Hamerle et al. (2011), Hlawatsch and Ostrowski (2011), and Bellotti and Crook (2012).

sophisticated parametric models perform relative to the simpler parametric models or the non-parametric models that may be less computationally burdensome is not clear from the literature.[3]

We have two main aims in this paper. First, we propose some refinements to the transformation regression methodology that has been used extensively in the literature to explore whether the performance of the current transformation regression methods can be improved. In the literature, an unmentioned criticism of the current transformation regression methods is that the LGD predictions can result in biased estimates due to the inherent nonlinearities in the transformations functions used. To remedy this issue, we propose a smearing estimator based on Duan (1983) and a Monte Carlo (MC) estimator to correct for these biases. Furthermore, we introduce another methodology we call the "global adjustment approach." Transformation regressions typically first apply adjustment factors to LGD values of 0 and 1. Qi and Zhao (2011) show, however, that a small adjustment factor leads to poor model performance. On the other hand, a larger adjustment factor cannot preserve the rank ordering of the raw LGD values, which could potentially affect statistical inference and predictive performance. The global adjustment approach we propose here applies an adjustment factor to all the LGD observations (and not just the boundary values) which retains the rank ordering in LGD values.

Second, we investigate the performance of five recent parametric methods that are designed specifically to fit the unusual distribution of LGD. These include the two-step regression, inflated beta regression (Ospina and Ferrari [2010a, b]), Tobit regression, censored gamma regression (Sigrist and Stahel [2011]), and two-tiered gamma regression (Sigrist and Stahel [2011]) models. These models share a similar structure in that they explicitly model the probability of LGD being 0, 1, or a value in between, but they differ in distributional assumptions. Our primary interest is in whether these recent parametric methods can outperform simpler parametric methods, including transformation regressions, standard linear regression, and fractional response regression (FRR) from Papke and Wooldridge (1996).

[3] A recent study by Yashkir and Yashkir (2013) compares some of the new parametric LGDs models (e.g., inflated beta and censored gamma) and finds much similarity in the goodness of fit among these new parametric models. Yashkir and Yashkir (2013), however, compare only a few models, and it is not clear how their models compare with other simpler parametric models or non-parametric models. Furthermore, their set of explanatory variables does not include the seniority index, the most important determinant of LGD shown in Qi and Zhao (2013).

We use the same data set and explanatory variables as in Qi and Zhao (2011) so that more general conclusions about model performance can be drawn by comparing the models studied in this paper with those investigated by Qi and Zhao (2011). In general, we find that in terms of model fit, all the methods investigated in this paper perform similarly, with in-sample R-squared ranging from 0.449 to 0.458 and slightly worse out-of-sample R-squared ranging from 0.444 to 0.452.

A few additional observations can be made based on our extensive empirical analysis. Regarding our first aim, the three proposed refinements to the transformation regressions can help improve model performance. Although the improvement is only marginal, the smearing and MC estimators can substantially reduce the sensitivity to the value of the adjustment factor in transformation regressions. Although the global adjustment reduces the sensitivity, the transformation regressions are still sensitive to the value of the adjustment factor.

Regarding our second aim, we compare model complexity and computational burden across alternative models and find that simpler parametric models do not necessarily underperform the more complex ones in predictive accuracy and ability to model the bimodal LGD distribution. Although all the methods perform quite similarly, the two-step approach has the best in- and out-of-sample performance, followed by the two-tiered gamma regression. The inflated beta regression performs very closely to the two-tiered gamma regression in sample and slightly outperforms all the transformation regressions (including the refined ones) except for the smearing estimator out of sample. The censored gamma and Tobit regressions perform similarly, with the worst performance among all the methods investigated here. The predictive accuracies of the censored gamma and Tobit models are almost identical, despite the high complexity and computational burden of the censored gamma regression. Estimation of the two-tiered gamma model is challenging because of the complicated likelihood function that is sensitive to the choice of optimization algorithm and the starting values. The two-tiered gamma model does not perform better than the much simpler and easier two-step regression model based on our sample and model setup. Overall, all methods investigated in this paper outperform the linear regression but underperform the FRR and the nonparametric methods investigated in Qi and Zhao (2011).

The findings and conclusions of our study are based on one data set. The relative performance of various models is likely to change if they are applied to different LGD data sets with different sample sizes, distributions, and risk drivers. Thus, it is important for modelers and researchers to be aware of the wide range of possible LGD models and methods, and to choose the one that is appropriate for their particular data set, balancing performance, complexity and computational burden via model validation and benchmarking.

The rest of this paper proceeds as follows. In the section 2, we describe the various models and methods investigated in this study. Section 3 provides details on empirical results and model comparison. Section 4 concludes the paper.

2. Methodology Description

This section discusses alternative methods we use in this study to estimate LGD. In the following subsections, LGD stands for the raw observed values of LGD, and L stands for the LGD values after applying adjustment factors (more details in subsequent sections). All of the models, with the exception of the two-step approach, are estimated by maximum likelihood. We provide the density functions for the data, which can easily be used to form the log likelihood functions. The mean LGD predictions are obtained by plugging in the maximum likelihood estimates into the population mean functions.

2.1 Transformation Regressions

The general idea of transformation regressions is to first convert the LGD observations from [0, 1] to (0, 1) with an adjustment factor, transform these adjusted values into the real line with a transformation function, and then fit linear regressions on the transformed values. In the current literature, the fitted values are then retransformed into LGD predictions by applying the inverse of the transformation function to them. This approach is used in Siddiqi and Zhang (2004), Gupton and Stein (2005), Hamerle et al. (2011), Qi and Zhao (2011), and Hlawatsch and Ostrowski (2011).

Before we describe our refinements, we describe transformation regressions more formally. Let $L_i \in (0,1)$ denote the i-th LGD observation after the adjustment factors have been applied. Let Z_i denote a transformed value of L_i, where $Z_i = h(L_i; a)$, or $L_i = h^{-1}(Z_i; a)$. The function h and its inverse h^{-1} are assumed to be nonlinear, monotonic, and continuously differentiable. We refer to h as the *transformation* and h^{-1} as the *retransformation*. The vector a consists of known constants (i.e., the predetermined parameters in the transformation/retransformation functions). The codomain of Z_i is chosen to be the entire real line, in which case, it is reasonable to use linear regression models for Z_i, $Z_i = x_i\beta + e_i$. The usual OLS estimates of the regression coefficients $\hat{\beta}$ and the variance of the error term $\hat{s}^2$, as well as the prediction for the transformed scale $\hat{Z}_i = x_i\hat{\beta}$, are unbiased and also consistent if the design matrix is asymptotically non-degenerate. We refer to this as the "transformation regression."

As in Qi and Zhao (2011), we use two particular transformation functions: an inverse standard Gaussian cumulative distribution function (CDF) and a combination of inverse standard Gaussian and beta CDFs, which leads to the inverse Gaussian regression model (IGR) and inverse Gaussian regression with beta transform model (IGR-BT). For IGR, the vector a is equal to (0, 1), representing a mean of 0 and a variance of 1 for the standard Gaussian distribution; similarly, for IGR-BT, the vector a consists of the same mean and variance, but also the two beta distribution parameters calibrated to the LGD data.

2.1.1 Refinements to Transformation Regressions

The transformation regressions are simple, straightforward, and easy to implement; however, the optimal predictions on the untransformed scale are generally not equal to the inversions of the optimal predictions on the transformed scale. It seems natural to obtain $\hat{L}_i$, the predictor for L_i, by inverting $\hat{Z}_i = x_i\hat{\beta}$ to produce the retransformed predictor $\hat{L}_i = h^{-1}\left(\hat{Z}_i;a\right) = h^{-1}\left(x_i\hat{\beta};a\right)$, which we call the naïve estimator in the rest of this paper. This is the approach taken in the current LGD literature. The naïve estimator $\hat{L}_i = h^{-1}\left(x_i\hat{\beta};a\right)$, however, is neither unbiased nor consistent for $E(L_i)$ unless the transformation is linear. Obviously, the transformation functions in the LGD studies are nonlinear (e.g., the inverse Gaussian CDF in IGR). The literature (e.g., Duan [1983]) widely recognizes that, as long as the transformation is not linear, even if the true parameters are known, $h^{-1}(x_i\beta)$ is not the correct "estimate" of $E(L_i)$:

$$E\left(L_i\right) = E\left(h^{-1}(x_i\beta + e_i;a)\right) \neq h^{-1}\left(x_i\beta;a\right) \ . \tag{1}$$

The main difficulty in obtaining the optimal predictions lies in finding the mean of $L_i = h^{-1}\left(x_i\beta + e_i;a\right)$. Note that the distribution for L_i is easy to obtain (e.g., by using the Jacobian change of variables theorem). Its mean and other population quantities of interest, however, do not generally have closed form solutions. We propose two ways of obtaining the optimal predictions $E(L_i|x_i)$ in this subsection: a smearing estimator and an MC estimator.

2.1.1.1 A Smearing Estimator

Duan (1983) proposes a non-parametric smearing estimate for the mean $E\left(L_i \mid x_i\right) = \int h^{-1}\left(x_i\beta + e_i;a\right) f_{e_i}\left(t\right) dt$. Its intuition can be understood in three steps. First, the empirical CDF of the estimated residuals is computed as

$$\hat{F}_N\left(r\right) = \frac{1}{N}\sum_{j=1}^{N} I\left(\hat{e}_j \leq r\right) \tag{2}$$

where $\hat{e}_j = Z_j - x_j\hat{\beta}$, N is the number of observations, and $I(A)$ denotes the indicator function of the event "A". Second, using the empirical CDF, an estimate of the mean is expressed as

$$\bar{E}\left(L_i \mid x_i\right) = \frac{1}{N}\sum_{j=1}^{N} h^{-1}\left(x_i\beta + \hat{e}_j;a\right) \tag{3}$$

Because β is unknown, the third step is to plug in the OLS estimator and obtain

$$\hat{E}\left(L_i \mid x_i\right) = \frac{1}{N}\sum_{j=1}^{N} h^{-1}\left(x_i\hat{\beta} + \hat{e}_j; a\right) \tag{4}$$

which is referred to as Duan's smearing estimator. This is a simple quantity to compute in practice. One basically computes the *N* OLS residuals, plugs the residuals and OLS estimate of β into (4), and then takes the sample average to produce the estimate.

Rigorous proofs for the consistency of (4) are in Duan (1983). Note that this is a non-parametric estimate as the normality of e_j is not used. This can be viewed as inexpensive insurance against possible departures from normality.

2.1.1.2 An MC Estimator

MC methods can also be used to estimate the conditional mean. To understand our MC estimator, first note that if G independent draws of e_i can be obtained from f_{e_i}, then the sample average of

$$\frac{1}{G}\sum_{g=1}^{G} h^{-1}\left(x_i\beta + e_i^{(g)}; a\right) \tag{5}$$

converges to the conditional mean from the law of large numbers. Because β and σ^2 are unknown, we can plug in the OLS estimators into (5) and form the MC estimator

$$\tilde{E}\left(L_i \mid x_i\right) = \frac{1}{G}\sum_{g=1}^{G} h^{-1}\left(x_i\hat{\beta} + \hat{e}_i^{(g)}; a\right) \tag{6}$$

This quantity converges to the desired quantity for a large G by application of the continuous mapping theorem and law of large numbers.

Succinctly, for each observation i, the MC algorithm is as follows:

1. Use $\hat{s}^2$ from the OLS estimation, draw G values of the disturbance term $\hat{e}_i$ from $\mathrm{N}\left(0, \hat{s}^2\right)$ and denote them as $\hat{e}_i^{(1)},\ \hat{e}_i^{(2)}, \ldots, \hat{e}_i^{(G)}$.
2. Use $\hat{\beta}$ from the OLS estimation, obtain the G values of $h^{-1}\left(x_i\hat{\beta} + \hat{e}_i^{(1)}\right), h^{-1}\left(x_i\hat{\beta} + \hat{e}_i^{(2)}\right), \ldots, h^{-1}\left(x_i\hat{\beta} + \hat{e}_i^{(G)}\right)$.
3. Compute $\tilde{E}\left(L_i | x_i\right)$, which equals the sample average of the G values from the previous step.

Note that this approach is different from the smearing estimator as the MC method uses the normality assumption.

2.1.2 Transformation Regressions With Global Adjustment

Usually the small adjustment factor is applied only to the boundary LGD values of 0 or 1 prior to fitting the transformation regressions. This adjustment approach can create some inconsistency between adjusted values and unadjusted values and may result in LGD values that do not rank order, particularly if a large adjustment factor ε is used. Qi and Zhao (2011) find the transformation regression results are very sensitive to the magnitude of ε, and it is not clear how much of the sensitivity might be attributed to the adjustment factor that applies only to LGD values of 0 and 1. We aim to shed light on this question by investigating an alternative adjustment method in this paper. Specifically, we propose to adjust all LGD observations from [0, 1] to (*b*, 1-*b*) through

$$L = b + (1 - 2b) \times LGD \tag{7}$$

where b is a predetermined adjustment factor. These adjusted LGDs are transformed with the function h and used in the transformation regressions. The fitted values from the regressions, $\hat{L}$, are retransformed to the scale (0, 1), and the retransformed values are then converted back to [0, 1] by applying the following reverse adjustment:

$$L\hat{G}D = \frac{\left(\hat{\text{L}} - b\right)}{\left(1-2b\right)} \qquad (8)$$

We investigate various values of b in section 3. We call this approach "global adjustment" as the adjustment factor b is applied to all LGD observations. We call the typical adjustment approach in the literature (e.g., Qi and Zhao [2011] and Altman and Kalotay [2014]) "local adjustment" because the adjustment factor ε is applied only to the LGD values of 0 or 1. The LGD estimates produced from the reverse adjustment in equation 8 above can be less than 0 if $\hat{L} < b$, or greater than 1 if $\hat{L} > 1 - b$. The LGD estimates can be floored at 0 and capped at 1 after the reverse adjustment if desired.

2.2 Models to Account for the Unusual LGD Distribution

We discuss five methods that specifically account for the unusual bounded and bimodal distribution of LGD.

2.2.1 Two-Step Approach

This approach allows for the possibility that the processes governing whether the LGD equals 0 or 1, or any value in between, may be different. This approach is similar to the two-step estimation in Gurtler and Hibbeln (2013). We estimate LGDs in two steps. In step 1, we run an ordered logistic regression on the probability of LGD falling into one of three categories:
0, (0, 1), or 1

$$P_i = \begin{cases} P_0^i = Logistic\left(\gamma_0 - x_i\beta\right) & if\ LGD = 0 \\ P_{0,1}^i = Logistic\left(\gamma_1 - x_i\beta\right) - Logistic\left(\gamma_0 - x_i\beta\right) & if\ LGD \in \left(0,1\right) \\ P_1^i = 1 - Logistic\left(\gamma_1 - x_i\beta\right) & if\ LGD = 1 \end{cases} \qquad (9)$$

where Logistic() denotes the logistic function, and γ_0 and γ_1 are cut-point parameters to be estimated. This first step is used to account for the mass concentrated at 0 or 1.

In step 2, we run OLS using all the LGD observations within the range (0, 1) on the explanatory variables, and we call the predicted LGD from the second regression $\hat{\mu}^i = x_i\hat{\beta}$ for observations in (0, 1). We then predict the *i*th LGD as

$$\hat{\mathrm{E}}\left(\mathrm{LGD_i}\right) = \hat{\mu}^i \times \left(1 - \hat{P}_0^i - \hat{P}_1^i\right) + \hat{P}_1^i \tag{10}$$

Note that the predicted LGD generated from equation (10) is a weighted average of the model outputs from steps 1 and 2. It is not mathematically bounded between 0 and 1.

2.2.2 Inflated Beta Regression

Ospina and Ferrari (2010a) propose inflated beta distributions that are mixtures between a beta distribution and a Bernoulli distribution degenerated at 0, 1, or both 0 and 1. Ospina and Ferrari (2010b) then further develop inflated beta regressions by assuming the response distribution to follow the inflated beta and by incorporating explanatory variables into the mean function. Ospina and Ferrari (2010a) propose that the probability function for the *i*th observation is

$$P_i\left(LGD; P_0^i, P_1^i, \mu^i, \phi\right) = \begin{cases} P_0^i & if\ LGD = 0 \\ \left(1 - P_0^i - P_1^i\right) f\left(LGD; \mu^i, \phi\right) & if\ LGD \in (0,1) \\ P_1^i & if\ LGD = 1 \end{cases} \tag{11}$$

where $0 < \mu^i < 1$, $\phi > 0$, and $f(.)$ is a beta probability density function (PDF), i.e.,

$$f\left(LGD; \mu^i, \phi\right) = \frac{\Gamma(\phi)}{\Gamma\left(\mu^i\phi\right)\Gamma\left(\left(1-\mu^i\right)\phi\right)} LGD^{\mu^i\phi - 1}\left(1 - LGD\right)^{\left(1-\mu^i\right)\phi - 1} \tag{12}$$

Note that μ^i is the mean of the beta distribution, and ϕ is interpreted as a dispersion parameter. The mean function is $E\left(LGD_i\right) = P^i{}_1 + \mu^i\left(1 - P_0{}^i - P_1^i\right)$. The connection between explanatory variables x_i and the expected LGD is through the three equations as follows:

$$P_0^i = e^{x_i\alpha} / (1 + e^{x_i\alpha} + e^{x_i\beta}) \quad (13)$$

$$P_1^i = e^{x_i\beta} / (1 + e^{x_i\alpha} + e^{x_i\beta}) \quad (14)$$

$$\mu^i = e^{x_i\gamma} / (1 + e^{x_i\gamma}) \quad (15)$$

where the parameters α, β, γ are model coefficients. These coefficients along with ϕ are estimated by maximizing the log likelihood function. For details on the inflated beta regression in general, see Ospina and Ferrari (2010b), Pereira and Cribari-Neto (2010), and Yashkir and Yashkir (2013).[4]

Note that the two-step approach and the inflated beta regression are quite similar. They differ in that the parameters of the inflated beta model are estimated from a parametric model, while the parameters from the two-step approach are estimated in two separate steps. The two-step method might perform better than the inflated beta regression due to its flexibility in predicting the observations in (0, 1).[5] On the other hand, because we assume a parametric model, equation (11) guarantees that the predicted LGDs are within the [0, 1] boundary, while such an outcome is not ensured in equation (10).

2.2.3 Tobit Regression

Tobit regression is often used to describe the relationship between a random variable that is censored and some explanatory variables. In our case, the basic assumption in this modeling approach is that our dependent variable LGD is censored to the closed interval [0, 1]. Observed LGD is a censored version of the latent variable L^*, where L^* may be less than 0 or greater than 1

[4] Our parameterizations of the probabilities in (15) and (16) are slightly different from the ones in Yashkir and Yashkir (2013). Our parameterizations ensure that each probability is positive and that the mixture weights in (13) sum to 1, while the parameterizations in Yashkir and Yashkir (2013) do not guarantee that $P_0^i + P_1^i < 1$, resulting in mixture weights in (13) that may be negative for $L \in (0, 1)$.

[5] In the two-step approach, $\hat{\mu}^i$ is estimated freely without considering the ordered logit in the second step while $\hat{\mu}^i$ in the beta regression is estimated from the likelihood derived from the beta distribution. Given that the data generating process is unknown, the latter case might be too restrictive in its form and the first approach is more flexible.

for various reasons. The original data from Moody's Ultimate Recovery Database include some observations with negative LGDs, and we floor those LGDs, which leads to censoring from below at 0. L^* can also be greater than 1 if the lender extends more loans to the obligor post default, which leads to censoring from above at 1. The Tobit model can be estimated by standard statistical software. Mathematically, the Tobit model for LGD is

$$P_i\left(LGD;\theta_i\right)=\begin{cases}P\left[LGD=0\right]=\Phi\left(-\theta_i/\sigma\right)\\ P\left[LGD\in\left(l,l+dl\right)\right]=\varphi\left((l-\theta_i)/\sigma\right)/\sigma\,dl,\ if\ 0<l<1\\ P\left[LGD=1\right]=1-\Phi\left((1-\theta_i)/\sigma\right)\end{cases}\qquad(16)$$

where $\varphi(.)$ and $\Phi(.)$ are the PDF and CDF of a standard normal random variable, respectively, and $\theta_i = x_i\beta$. See Amemiya (1984) for an expression of the mean function and associated details.

2.2.4 Censored Gamma Regression

Sigrist and Stahel (2011) introduce gamma regression models to estimate LGD. The probability function for the ith observation is

$$P_i\left(LGD;\xi,\alpha,\theta_i\right)=\begin{cases}P\left[LGD=0\right]=G\left(\xi,\alpha,\theta_i\right)\\ P\left[LGD\in\left(l,l+dl\right)\right]=g\left(l+\xi,\alpha,\theta_i\right)dl,\ if\ 0<l<1\\ P\left[LGD=1\right]=1-G\left(1+\xi,\alpha,\theta_i\right)\end{cases}\qquad(17)$$

where $g(u;\alpha,\theta_i)=\frac{1}{\theta_i^{\alpha}\Gamma(\alpha)}u^{\alpha-1}e^{-u/\theta_i}$ and $G(u;\alpha,\theta_i)=\int_0^u g(x;\alpha,\theta_i)\,dx$ are the PDF and CDF for a gamma random variable, respectively. Also, $\alpha>0$, $\xi>0$, and $\theta_i>0$. Note that Sigrist and Stahel (2011) define the underlying latent variable to follow a gamma distribution shifted by $-\xi$. The use of a gamma distribution with a shifted origin, instead of a standard gamma distribution, is motived by the fact that the lower censoring occurs at zero.

The connection between explanatory variables x_i and the expected LGD for the ith observation is through the linear equations as follows:

$$\left.\begin{aligned} \log \alpha &= \alpha^*, \\ \log \xi &= \xi^*, \\ \log \theta_i &= x_i \beta \end{aligned}\right\} \quad (18)$$

where β is the vector of model coefficients. These coefficients and the parameters α^* and ξ^* are estimated by maximizing the log likelihood function. The resulting estimates are then used to obtain LGD predictions: $E(LGD_i) = \alpha\theta_i[G(1+\xi, \alpha+1, \theta_i) - G(\xi, \alpha+1, \theta_i)] + (1+\xi)\big(1 - G(1+\xi, \alpha, \theta_i)\big) - \xi(1 - G(\xi, \alpha, \theta_i))$. For more detail on the censored gamma regression, refer to Sigrist and Stahel (2011).

The censored gamma regression model is quite similar to a Tobit model. The only difference is that the underlying latent variable in the censored gamma model has a shifted gamma distribution, while the Tobit model assumes a normal distribution for the latent variable. It is not trivial to maximize the likelihood function of the censored gamma regression model analytically or numerically, while Tobit models can be fairly easily estimated in most statistical software.

2.2.5 Two-Tiered Gamma Regression

Sigrist and Stahel (2011) extend the censored gamma model into the two-tiered gamma model. This extension allows for two underlying latent variables, one that governs the probability of LGD being 0 and another for LGD being in (0, 1). The extension is useful in that it allows each latent variable to have its own set of explanatory variables and parameters.

More specifically, the two-tiered gamma regression assumes that there are two latent variables: the first latent variable, L_1^*, which follows a shifted gamma distribution with density function $g\left(L_1^* + \xi, \alpha, \tilde{\theta}_i\right)$, and the second variable, L_2^*, which is a shifted gamma distribution lower truncated at zero with the density function $g\left(L_2^* + \xi, \alpha, \theta_i\right)$. These two latent variables are then related to L through

$$L = \begin{cases} 0 & if \quad L_1^* < 0 \\ L_2^* & if \quad 0 < L_1^* \ and \ L_2^* < 1 \\ 1 & if \quad 0 < L_1^* \ and \ 1 \le L_2^* \end{cases} \tag{19}$$

The distribution of LGD can be characterized as follows:

$$P_i\left(LGD;\xi,\alpha,\tilde{\theta}_i,\theta_i\right) = \begin{cases} P\left[LGD=0\right] = G\left(\xi,\alpha,\tilde{\theta}_i\right) \\ P\left[LGD \in \left(l,l+dl\right)\right] = g\left(l+\xi,\alpha,\theta_i\right)\dfrac{1-G\left(\xi,\alpha,\tilde{\theta}_i\right)}{1-G\left(\xi,\alpha,\theta_i\right)} dl, if \ 0<l<1 \\ P\left[LGD=1\right] = 1-G\left(1+\xi,\alpha,\theta_i\right)\dfrac{1-G\left(\xi,\alpha,\tilde{\theta}_i\right)}{1-G\left(\xi,\alpha,\theta_i\right)} \end{cases} \tag{20}$$

The connection between the explanatory variables x_i and the expected LGD is through the linear equations as follows:

$$\left.\begin{aligned} \log\alpha &= \alpha^* \\ \log\xi &= \xi^* \end{aligned}\right\} \tag{21}$$

$$\log\tilde{\theta}_i = x_i\beta \tag{22}$$

$$\log\theta_i = x_i\gamma \tag{23}$$

where β,γ are vectors of model coefficients. These coefficients and the parameters α^* and ξ^* are estimated by maximizing the log likelihood. The mean LGD is calculated as

$$E(LGD) = \Pr(LGD=1) + \Pr\big(LGD \in (0,1)\big)\, E(LGD|LGD \in (0,1))$$

where

$$\Pr(LGD=1) = (1-G(1+\xi,\alpha,\theta))\frac{1-G(\xi,\alpha,\tilde{\theta})}{1-G(\xi,\alpha,\theta)}$$

$$\Pr(LGD=0) = G(\xi,\alpha,\tilde{\theta})$$

$$\Pr\big(LGD \in (0,1)\big) = 1-\Pr(LGD=0) - \Pr(LGD=1)$$

$$E\big(LGD \big| LGD \in (0,1)\big) = \frac{\alpha\theta\big(G(1+\xi,\alpha+1,\theta) - G(\xi,\alpha+1,\theta)\big) - \xi\big(G(1+\xi,\alpha,\theta) - G(\xi,\alpha,\theta)\big)}{G(1+\xi,\alpha,\theta) - G(\xi,\alpha,\theta)}$$

As this expectation is not provided in Sigrist and Stahel (2011), we provide the derivation in the appendix. For more information on the two-tiered gamma regression, refer to Sigrist and Stahel (2011).

As the two-tier gamma regression involves a mixture of two shifted gamma distributions, maximizing its log likelihood function is quite challenging.

3. Summary of Empirical Results

To facilitate model performance comparison, we use the same data set as in Qi and Zhao (2011) with the same explanatory variables. This data set is based on Moody's Ultimate Recovery Database, which covers U.S. corporate default events with over $50 million in debt at the time of default. There are a total of 3,751 observations from 1985 to 2008. Refer to Qi and Zhao (2011) for a more detailed description of the data construction and summary statistics. It is worth noting that 30 percent of the observations in the sample have LGD values equal to 0, and 6 percent of the observations have LGD values equal to 1.

We describe in this section the estimation results from different modeling methods for LGD using the same set of explanatory variables in all models. In all the models, we use subordinated bonds as the base instrument and “most assets” as the base collateral type. Also, to be comparable with Qi and Zhao (2011), we present in-sample and out-of-sample (i.e., 10-fold cross validation) results for each method. As a benchmark, Qi and Zhao (2011) report the in-sample R-squared for the linear regression and the FRR as 0.448 and 0.463, respectively, and the out-of-sample R-squared as 0.443 and 0.457, respectively.

3.1 Refined Transformation Regressions

3.1.1 Improved Retransformation Methods—the Smearing and MC Estimators

We follow Qi and Zhao (2011) and adjust the boundary LGD values by ε before transforming the adjusted LGDs to the real line and then applying the retransformation methods discussed in section 2.1.1. The same local adjustment factor values of ε as in table 4 of Qi and Zhao (2011) are investigated. Table 1 reports the R-squared and sum of squared errors (SSE) from the naïve transformation method investigated in Qi and Zhao (2011), (i.e., $\hat{L}_i = h^{-1}\left(x_i\hat{\beta};a\right)$), and from the smearing estimator and the MC estimator. Panel A shows the results for IGR and panel B shows those for IGR-BT.[6,7] The in-sample results are shown in panels A1 and B1 and the out-of-sample 10-fold cross-validation results are displayed in panels A2 and B2. The bolded rows represent the results corresponding to the optimal cutoffs for ε, where optimal is defined as the cutoff that leads to the highest R-squared values.

These panels show that there is little difference in the results between the IGR and IGR-BT, a finding similar to that in Qi and Zhao (2011). The performance of the two retransformation estimators is much less sensitive to the choice of ε than the naïve retransformation. The advantage of the retransformation estimators is the most obvious for small values of ε, and the discrepancies disappear at ε values beyond 0.01. This is because the transformations (e.g., inverse Gaussian) are sensitive or very nonlinear at values close to 0 or 1 (i.e., small ε), so properly accounting for nonlinearities with the smearing or the MC estimators yields more accurate predictions. On the other hand, the transformations are close to linear for values away from the boundaries (i.e., larger ε), and little difference exists between the naïve and the smearing (or the MC) estimators.

---

[6] In IGR-BT, the two beta parameters are chosen so that the mean and variance of the beta distribution match the sample mean and variance of the original LGD data. After calibration, we use a beta (0.3104, 0.3751) distribution, which implies a mean and variance of 0.453 and 0.147, respectively.

[7] We have also investigated the inverse non-standard Gaussian, inverse non-standard Gaussian with beta transformation, and the logit transform regressions. The results are qualitatively similar and thus are not reported to save space.

In addition, the two retransformation estimators perform very similarly across every scenario, suggesting that the normality assumption in the MC estimator is not overly restrictive. The optimal ε under the naïve approach is 0.05, but is 0.01 under both the smearing and MC estimators. Furthermore, the optimal values under the two retransformation estimators are slightly higher than those under the naïve approach. These methods perform better than the linear regression, but they still underperform the FRR. These conclusions hold for both in-sample and out-of-sample predictions.

In summary, relative to the naïve estimator, the retransformation estimators are not as sensitive to different values of ε and result in more stable R-squared and SSE values. The retransformation estimators are thus especially useful if the optimal ε value is not stable across different subsets of the estimation sample. Therefore, although the retransformation estimators show only marginal improvement over the naïve approach, they are helpful when a modeler chooses to use these transformation methods but is unsure about the optimal adjustment factor.

3.1.2 Transformation Regressions With Global Adjustment

Table 2 reports the results from transformation regressions with global adjustment using a different adjustment factor b that ranges from 1e-11 to 0.45.[8] Results using IGR are reported in panel A, and those using IGR-BT are shown in panel B. As in Qi and Zhao (2011), there is little difference in the results between the IGR and IGR-BT at the optimal value of $b = 0.1$.

The model fit improves dramatically as b increases from 1e-11 to 0.0001. As b increases further, the performance continues to improve but at a declining rate until it reaches peak performance at around $b = 0.1$. These results hold both in sample and out of sample in the 10-fold cross-validation. Note that the optimal value of b under the global adjustment approach is

[8] This adjustment method is undefined for $b = 0.5$.

much larger than the optimal local adjustment factor ε of 0.05 as reported in Qi and Zhao (2011).

Similar to the results in table 4 of Qi and Zhao (2011), the global adjustment method also performs poorly at very small values of b —it dramatically underperforms the linear regression at b = 0.001 and b = 0.005. Its performance catches up with that of the linear regression at b = 0.05. At the optimal point of b = 0.1, the global adjustment approach marginally outperforms the peak performance of the IGR and IGR-BT under the local adjustment approach. Furthermore, these methods outperform the linear regression but underperform the FRR. Therefore, the transformation regressions are generally very sensitive to the adjustment factor, regardless of the local or the global adjustment method, and even at the optimum, they do not lead to superior performance.

Note that the results in table 2 do not floor or cap the predicted LGDs. Applying the floor and the cap does not dramatically change the conclusions, except that the optimal point for b goes up to 0.2, and the in-sample and the 10-fold cross-validation R-squared are at 0.455 and 0.450, respectively, slightly higher than those reported in table 2 but still lower than those for the FRR. These results are not reported to save space.

3.2 Models to Account for the Unusual LGD Distribution

3.2.1 Two-Step Approach

Table 3 reports the ordered logistic regression results from step 1 and the OLS results from step 2. On one hand, the coefficient estimates from the ordered logit model are largely intuitive. Term loans, loans secured by inventory, accounts receivables, cash, and exposures with guarantees and to the utility industry are more likely to fall into the group with an LGD of 0. Unsecured loans, loans secured by capital stock, and third liens are more likely to fall into the group with an LGD of 1. Most macroeconomic and industry condition variables have statistically significant coefficients as expected. The signs of the coefficients from the ordered logit in step 1 are largely consistent with those from the OLS in table 3 of Qi and Zhao (2011), with some differences in significance levels. This is not surprising, given the main difference between the two regressions is that the OLS models continuous LGD values, whereas the ordered logit models discrete LGD groups.

On the other hand, there is much difference in the coefficient estimates from steps 1 and 2. For example, some seniority dummy variables have a change in statistical significance: the term "loan dummy variable" loses its significance, while the senior secure dummy gains statistical significance. Many of the collateral types also show changes in significance levels. For instance, equipment gains statistical significance, while inventory, accounts receivable, and cash lose statistical significance. These results are interesting, suggesting that some explanatory variables are important in explaining only the probability of LGD falling into the groups of 0, (0, 1), and 1, but not the LGD variations within (0, 1), and vice versa.[9] Because of this flexibility, the two-step regression approach might outperform the OLS.

Table 3 indeed shows that the two-step approach has higher R-squared and lower SSE values than the OLS, both in sample and in the 10-fold cross validation. The two-step approach, however, still slightly underperforms the FRR.

Despite that the LGD prediction from the two-step approach is not bounded between 0 and 1 in theory, we find that only one of the predictions in our empirical exercise falls outside the boundary [0, 1].[10] The two-step approach can be easily estimated using any standard statistical software, and we also find it intuitive.

[9] For example, 50 percent of term loans have zero losses, and over 85 percent of debt is collateralized by inventory, accounts receivable, and cash experience zero losses. Therefore, these collateral types are important when predicting LGD=0, but may not be particularly important when predicting LGDs in (0, 1].

[10] We did not apply a floor for this method in this paper. Alternatively, a transformation regression instead of OLS can be applied in the second step to ensure that all LGD predictions from the two-step approach are bounded between 0 and 1. The results are not expected to differ much given only one observation falls outside the [0, 1] boundary.

3.2.2 Inflated Beta Regression

Results from the inflated beta regression are reported in table 4. Similar to the two-step results, much difference exists between the three sets of coefficient estimates. For example, the capital stock dummy and the equipment dummy are associated with lower probability of LGD being 0 in equation (13) and also lower probability of LGD being 1 in equation (14). By contrast, the inventory, receivables, and cash dummy are connected with higher probability of LGD being 0 in equation (13) but also higher probability of LGD being 1 in equation (14). Such results suggest that the process governing the three equations in the inflated beta regression has some differences. As a result, the ordered logit model in the first step of the two-step approach might be too simplified. On the other hand, however, the variables that Qi and Zhao (2013) show to be the most critical factors in determining LGD, such as the seniority index and the industry distance-to-default, have consistent positive or negative correlation with LGD in all three equations. This suggests that the differences in the three equations in the inflated beta regression might be minor or caused by noise, and an ordered logit model might be sufficient.

Table 4 shows that the inflated beta regression does not outperform the two-step approach, both in sample and out of sample based on the 10-fold cross-validation. Such a finding suggests that the differences reflected in equations (13), (14), and (15) in table 4 are not of first order importance, and resorting to a more complicated model like the inflated beta regression might not be necessary.

3.2.3 Tobit Regression

Table 5 reports the coefficient estimates from the Tobit regression, with censoring at 0 and 1. Some differences exist between this table and the OLS results in table 3 of Qi and Zhao (2011). For instance, the loan dummy parameter loses its statistical significance. In addition, several variables show major changes in the magnitude of the coefficient estimates. The coefficient for inventory, accounts receivable, and cash increases by more than 200 percent. These results show that accounting for censoring at both ends can change the relationship between LGD and the explanatory variables quite dramatically. Although censoring at 0 and 1 are accounted for, the Tobit model outperforms the OLS only trivially, and it still underperforms the FRR.

3.2.4 Censored Gamma Regression

Results from the censored gamma regression are presented in table 6. Note that the variables with statistically significant signs are almost the same as the Tobit regression, and the model fit, both in sample and in the 10-fold cross-validation, is nearly identical in tables 5 and 6. This might indicate that the estimated shifted gamma distribution from our sample resembles a normal distribution; hence the censored gamma and Tobit models would behave similarly. Given the similarity in performance between the Tobit and censored gamma models, resorting to the more complicated model incorporating the gamma distribution for LGD modeling might not be necessary.

3.2.5 Two-Tiered Gamma Regression

The two-tiered gamma regression results are presented in table 7. It is clear that there are differences in the parameter estimates between the two latent variables. Sigrist and Stahel (2011) show that this model provides better model fit than the censored gamma regression, which is confirmed here. The improvement, however, is quite marginal, and this model still underperforms the two-step approach and the FRR. Such results again raise doubts on the value added by choosing more complicated LGD models over simpler ones.

3.3 Distribution of the Actual and Predicted LGDs

We plot in figure 1 the histogram of the actual and predicted LGDs from the various models investigated in this study. For the transformation regressions, since the results from IGR-BT and IGR are very close, we show only the histograms for IGR. In addition, the histograms shown in figure 1 are based on the optimal adjustment factor used in the transformation regressions (i.e.,

ε = 0.01 for the smearing and the MC estimators, 0.05 for the naïve estimator, and b = 0.1 for the global adjustment approach).

It is clear from figure 1 that, although all methods yield some degree of bimodality, the predictive distributions differ from the distribution of the actual LGDs. In particular, the predicted LGDs are much more concentrated in the interval (0, 1) than at the peaks of 0 and 1 compared with the actual LGD values and the proportion of LGD predictions falling in the [0.9, 1] bucket is particularly low. Such a pattern holds even for the models that are particularly designed to handle the unusual LGD distribution. This finding is consistent with that of Qi and Zhao (2011), again indicating the difficulty to adequately account for the bounded bimodal distributions.

In addition, although the predicted LGDs from the two-step approach can be outside the range [0, 1] in theory, the top panel of figure 1 shows that only one of the fitted LGDs from this method actually falls outside the [0, 1] boundary. This finding suggests that the theoretical concern might not be a significant problem in practice. Figure 1 suggests that adding the floor of 0 and cap of 1 in the global adjustment approach could better capture the bimodal distributions; however, the predictive distribution still falls short of the actual degree of the peaks.

3.4 Model Performance, Complexity, and Computational Burden

We summarize model performance, complexity, and computational burden of all the parametric LGD models discussed in this study in table 8. The models are sorted and ranked based on the in-sample SSE. To show the model stability, we also report model ranking based on the 10-fold cross-validation SSE. We assess model complexity and computational burden using the high, medium, and low ratings.

Several observations can be made from table 8. First, all models we investigate in this study perform similarly within a very narrow range: less than two percentage points difference exist between the best and the worst performing models. For only the models considered in this paper, the in-sample R-squared and SSE range from 0.449 to 0.458 and 298.673 to 303.748, respectively, and the out-of-sample R-squared and SSE are slightly worse, ranging from 0.444 to 0.452 and 301.369 to 305.922, respectively. They all perform better than the OLS but worse than the FRR as reported in Qi and Zhao (2011).

Second, across all these parametric models, complex or computationally intense models do not necessarily perform better than their simpler or less computationally intense counterparts. For example, the top two models are rated either low or medium in complexity and computation burden.

Third, despite the desirable statistical properties of the new or recently proposed parametric LGD models investigated in this study, none of these models perform better than the nonparametric models investigated in Qi and Zhao (2011). Thus, the dominance of non-parametric LGD models, such as the decision tree and the neural network, remains unchallenged. It is also worth noting that some of these parametric models are more complex and computationally burdensome than the non-parametric models.

These observations suggest that within the parametric models family, the more complex and computationally burdensome parametric LGD models might not have much value added in terms of model fit. Our analysis indicates that the best options are the fractional response regression and the two-step approach, both of which have the best model performance but are relatively simple and easy to implement without much computational burden. If model complexity and computational burden are not constraints, then nonparametric models, such as the decision tree and the neural network, might be better choices than their complex and burdensome parametric counterparts, such as the censored gamma and the IGR-BT with the MC estimator.

4. Summary and Conclusions

We conduct a comprehensive study of some new or recently developed parametric models for LGD, a semi-continuous random variable that lies in the interval of [0, 1] and that often follows a bimodal distribution. The first group of models consists of three methods that we propose to refine the transformation regressions. These methods include a smearing estimator and an MC estimator for retransforming the transformation regression outputs to LGD predictions, and a global adjustment method for handling the boundary LGD values. The second group of models consists of five regression models suitable for the unusual distribution of LGD: a two-step approach, inflated beta, Tobit, censored gamma, and two-tier gamma.

We find that the performance of the transformation regression with global adjustment is still sensitive to the adjustment factor. In addition, the smearing estimator and the MC estimator can help reduce the sensitivity of the transformation regression to the adjustment factor, and thus can be very useful if one would like to use the transformation regression but is not sure what adjustment factor to use. Even with these refinements and the optimal adjustment factor, however, the transformation regressions still do not drastically outperform the OLS, and they all underperform the FRR investigated in Qi and Zhao (2011).

Among the second group, five regression models designed to fit the unusual distribution of LGD, the two-step approach is similar to the inflated beta regression but is simpler and theoretically less sound, while the censored gamma regression is similar to Tobit regression but is more complicated and theoretically more appealing. We find, however, that the two-step approach slightly outperforms all methods investigated in this paper in our sample, and the performance of the censored gamma regression is essentially the same as Tobit regression. The two-tiered gamma regression is the most complicated and computationally challenging, but it does not outperform the simpler two-step approach. Our findings suggest that complicated parametric models may not be necessary when estimating LGD.

Despite the special design of each model to fit the unusual LGD distribution and the complexity in some of the models, all models investigated in this study tend to generate LGD predictions that are more concentrated between the two boundary values, and thus cannot reproduce the bimodal shape of the observed LGD distribution. This might be because the models investigated in this and other studies thus far produce mean LGD predictions, and it is difficult for these models to produce mean predictions that are far off in the tails. It may be fruitful for future research to investigate prediction methods based on other quantiles or alternatives to mean prediction. The findings and conclusions of our study are based on one data set. The relative performance of various models is likely to change with different LGD data sets that have different sample sizes, distributions, and risk drivers; we intend to explore these scenarios in future research.

5. References

Altman, E., and V.M. Kishore. 1996. “Almost Everything You Wanted to Know About Recoveries on Defaulted Bonds.” *Financial Analysts Journal* 52(6): 57–64.

Altman, E., and E.A. Kalotay. 2014. “Ultimate Recovery Mixtures.” *Journal of Banking and Finance* 40: 116–129.

Amemiya, Takeshi. 1984. "Tobit Models: A Survey." *Journal of Econometrics* 24(1): 3–61.

Bastos, J.A. 2010. “Forecasting Bank Loans Loss-Given-Default.” *Journal of Banking and Finance* 34: 2510–2517.

Bellotti, T., and J. Crook. 2012. “Loss Given Default Models Incorporating Macroeconomic Variables for Credit Cards.” *International Journal of Forecasting* 28(1): 171–182.

Dermine, J., and C. Neto de Carvalho. 2006. “Bank Loan Losses-Given Default: A Case Study.” *Journal of Banking and Finance* 30: 1219–1243.

Duan, N. 1983. “Smearing Estimate: A Nonparametric Retransformation Method.” *Journal of the American Statistical Association* 78 (383): 605–610.

Gupton, G. M., and R. M. Stein. 2005. LossCalc V2: Dynamic Prediction of LGD Modeling Methodology, Moody’s KMV.

Gurtler, M., and M. Hibbeln. 2013. “Improvements in Loss Given Default Forecasts for Bank Loans.” *Journal of Banking and Finance* 37: 2354–2366.

Hartmann-Wendels, T., P. Miller, and E. Tows. 2014. “Loss Given Default for Leasing: Parametric and Nonparametric Estimations.” *Journal of Banking and Finance* 40: 364–375.

Hamerle, A., M. Knapp, and N. Wildenauer. 2011. “Modeling Loss Given Default: A Point in Time-Approach,” in: *The Basel II Risk Parameters, Springer*, pp. 137–150.

Hlawatsch, S., and S. Ostrowski. 2011. “Simulation and Estimation of Loss Given Default.” *The Journal of Credit Risk* 7(3): 39–73.

Hu, Y., and W. Perraudin. 2002. “The Dependence of Recovery Rates and Defaults.” Working paper, Birkbeck College.

Loterman, G., I. Brown, D. Martens, C. Mues, and B. Baesens. 2012. "Benchmarking Regression Algorithms for Loss Given Default Modeling." *International Journal of Forecasting* 28: 161–170.

Ospina, R., and S. L. P. Ferrati. 2010a. "Inflated Beta Distributions." *Statistical Papers* 51: 111–126.

Ospina, R., and S. L. P. Ferrati. 2010b. "Inflated Beta Regression Models." Working paper, Universidade Federal de Pernambuco and Universidade de Sao Paulo.

Papke, L. E., and J. M. Wooldridge. 1996. "Econometric Methods for Fractional Response Variables With an Application to 401(k) Plan Participation Rates." *Journal of Applied Econometrics* 11: 619–632.

Pereira, T.L, and F. Cribari-Neto. 2010. "A Test for Correct Model Specification in Inflated Beta Regressions." Working paper, Institute de Matematica, Estatistica e Computaceo Cientifica Universidade Estadual de Campinas.

Qi, M., and X. Zhao. 2011. "A Comparison of Methods to Model Loss Given Default." *Journal of Banking and Finance* 35: 2842–2855.

Qi, M., and X. Zhao. 2013. "Debt Structure, Market Value of Firm and Recovery Rate." *Journal of Credit Risk* 9(1): 3–37.

Siddiqi, N., and M. Zhang. 2004. "A General Methodology for Modeling Loss Given Default." *RMA Journal* 86(8): 92–95.

Sigrist, Fabio, and Werner A. Stahel, 2011. "Using the Censored Gamma Distribution for Modeling Fractional Response Variables With an Application to Loss Given Default." ASTIN Bulletin 41: 673-710doi: 10.2143/AST.41.2.2136992.

Tobback, E., D. Martens, T.V. Gestel, and B. Baesens. 2014. "Forecasting Loss Given Default Models: Impact of Account Characteristics and Macroeconomic State." *Journal of the Operational Research Society* 65: 376–392.

Yashkir, O., and Y. Yashkir. 2013. "Loss Given Default Modeling: A Comparative Analysis." *Journal of Risk Model Validation* 7: 25–59.

Table 1. Transformation Regression With Improved Retransformation

Panel A: Inverse Gaussian Regression (IGR)						
Panel A1: In-Sample						
	Naïve		**Smearing estimator**		**MC estimator**	
ε	R^2	**SSE**	R^2	**SSE**	R^2	**SSE**
1.00E-11	0.171	456.934	0.404	328.520	0.404	328.819
0.000	0.327	370.845	0.438	309.608	0.437	310.475
0.001	0.357	354.239	0.444	306.401	0.442	307.479
0.001	0.372	346.286	0.447	305.006	0.445	305.929
0.005	0.408	326.231	0.452	302.225	0.451	302.875
0.010	0.424	317.389	**0.453**	**301.662**	**0.451**	**302.397**
0.050	**0.452**	**302.294**	0.445	306.200	0.443	307.052
0.080	0.450	302.926	0.435	311.665	0.432	313.343
0.100	0.446	305.145	0.427	315.749	0.424	317.804
0.200	0.406	327.276	0.383	340.393	0.374	344.979
0.300	0.345	361.317	0.325	371.988	0.315	377.766
0.500	0.165	460.335	0.167	459.456	0.158	464.117

Table 1. (Continued)

Panel A2: 10-Fold Cross-Validation												
	Naïve				Smearing estimator				MC estimator			
ε	**R^2**	**(Std)**	**SSE**	**(Std)**	**R^2**	**(Std)**	**SSE**	**(Std)**	**R^2**	**(Std)**	**SSE**	**(Std)**
1.00E-11	0.165	(0.089)	459.261	(13.968)	0.401	(0.058)	329.763	(9.514)	0.400	(0.053)	330.118	(8.486)
0.000	0.322	(0.078)	373.340	(12.383)	0.435	(0.053)	311.080	(8.546)	0.431	(0.051)	312.962	(8.145)
0.001	0.352	(0.075)	356.761	(11.866)	0.441	(0.052)	307.909	(8.279)	0.437	(0.050)	309.859	(7.908)
0.001	0.366	(0.073)	348.819	(11.588)	0.443	(0.051)	306.528	(8.135)	0.441	(0.049)	307.729	(7.770)
0.005	0.403	(0.068)	328.785	(10.750)	0.448	(0.048)	303.784	(7.698)	0.446	(0.047)	304.823	(7.430)
0.010	0.419	(0.065)	319.946	(10.274)	**0.449**	**(0.047)**	**303.237**	**(7.445)**	**0.447**	**(0.045)**	**304.507**	**(7.178)**
0.050	**0.446**	**(0.056)**	**304.818**	**(8.722)**	0.441	(0.042)	307.809	(6.596)	0.439	(0.041)	309.033	(6.491)
0.080	0.445	(0.052)	305.418	(8.087)	0.431	(0.040)	313.284	(6.238)	0.426	(0.039)	315.781	(6.104)
0.100	0.441	(0.050)	307.615	(7.744)	0.424	(0.039)	317.372	(6.043)	0.419	(0.039)	320.098	(6.045)
0.200	0.401	(0.043)	329.623	(6.458)	0.379	(0.034)	342.026	(5.340)	0.370	(0.033)	346.617	(5.120)
0.300	0.340	(0.038)	363.515	(5.531)	0.322	(0.030)	373.614	(4.973)	0.311	(0.030)	379.489	(4.535)
0.500	0.161	(0.033)	462.141	(4.545)	0.163	(0.028)	461.039	(5.509)	0.156	(0.027)	464.668	(3.963)

Table 1. (Continued)

Panel B: Inverse Gaussian Regression With Beta Transformation (IGR-BT)						
Panel B1: In-Sample						
	Naïve		**Smearing estimator**		**MC estimator**	
ε	**R^2**	**SSE**	**R^2**	**SSE**	**R^2**	**SSE**
1.00E-11	0.149	469.081	0.401	330.392	0.403	329.321
0.000	0.313	378.836	0.437	310.536	0.435	311.481
0.001	0.346	360.802	0.443	307.015	0.442	307.670
0.001	0.361	352.098	0.446	305.464	0.445	305.881
0.005	0.401	330.004	0.452	302.311	0.450	303.184
0.010	0.419	320.220	**0.453**	**301.614**	**0.451**	**302.605**
0.050	**0.450**	**303.391**	0.445	306.083	0.442	307.662
0.080	0.449	303.840	0.435	311.625	0.431	313.815
0.100	0.445	306.050	0.427	315.752	0.422	318.873
0.200	0.404	328.404	0.383	340.387	0.372	346.089
0.300	0.342	362.530	0.326	371.538	0.312	379.313
0.500	0.164	460.968	0.171	457.057	0.161	462.631

Table 1. (Continued)

Panel B2: 10-Fold Cross-Validation												
	Naïve				Smearing estimator				MC estimator			
ε	**R^2**	**(Std)**	**SSE**	**(Std)**	**R^2**	**(Std)**	**SSE**	**(Std)**	**R^2**	**(Std)**	**SSE**	**(Std)**
1.00E-11	0.143	(0.090)	471.382	(14.093)	0.398	(0.059)	331.636	(9.695)	0.399	(0.053)	330.833	(8.444)
0.000	0.307	(0.080)	381.317	(12.651)	0.433	(0.054)	312.003	(8.701)	0.432	(0.052)	312.857	(8.235)
0.001	0.340	(0.077)	363.311	(12.128)	0.440	(0.052)	308.519	(8.416)	0.437	(0.051)	309.800	(8.191)
0.001	0.356	(0.075)	354.619	(11.842)	0.442	(0.052)	306.984	(8.262)	0.440	(0.049)	308.523	(7.836)
0.005	0.396	(0.070)	332.550	(10.972)	0.448	(0.049)	303.870	(7.795)	0.445	(0.048)	305.247	(7.696)
0.010	0.413	(0.067)	322.772	(10.475)	**0.449**	**(0.047)**	**303.187**	**(7.527)**	**0.447**	**(0.047)**	**304.536**	**(7.485)**
0.050	**0.444**	**(0.057)**	**305.922**	**(8.865)**	0.441	(0.042)	307.686	(6.646)	0.438	(0.042)	309.523	(6.566)
0.080	0.443	(0.053)	306.345	(8.216)	0.431	(0.040)	313.236	(6.284)	0.425	(0.040)	316.287	(6.233)
0.100	0.439	(0.051)	308.536	(7.867)	0.424	(0.039)	317.367	(6.090)	0.418	(0.038)	320.646	(5.935)
0.200	0.399	(0.043)	330.779	(6.573)	0.379	(0.034)	342.013	(5.403)	0.369	(0.034)	347.629	(5.133)
0.300	0.337	(0.039)	364.766	(5.650)	0.322	(0.030)	373.161	(5.053)	0.309	(0.030)	380.632	(4.521)
0.500	0.159	(0.034)	462.823	(4.655)	0.167	(0.028)	458.644	(5.575)	0.156	(0.027)	464.989	(4.008)

Table 2. Transformation Regressions With Global Adjustment

Panel A: Inverse Gaussian Regression (IGR)						
	In-sample		10-fold cross-validation			
b	R^2	SSE	R^2	(Std)	SSE	(Std)
1.00E-11	0.171	456.934	0.165	(0.089)	459.421	(13.958)
0.0001	0.328	370.498	0.322	(0.078)	373.126	(12.359)
0.0005	0.358	353.887	0.352	(0.075)	356.532	(11.833)
0.001	0.372	346.017	0.366	(0.073)	348.668	(11.549)
0.005	0.407	326.722	0.401	(0.068)	329.379	(10.702)
0.01	0.422	318.585	0.416	(0.065)	321.239	(10.242)
0.05	0.449	303.946	0.443	(0.057)	306.570	(8.950)
0.08	0.452	301.831	0.447	(0.054)	304.441	(8.532)
0.1	**0.453**	**301.298**	**0.448**	**(0.053)**	**303.900**	**(8.332)**
0.2	0.453	301.641	0.447	(0.049)	304.229	(7.720)
0.3	0.450	302.971	0.445	(0.047)	305.563	(7.393)
0.45	0.448	304.232	0.443	(0.046)	306.834	(7.172)
Panel B: Inverse Gaussian Regression With Beta Transformation (IGR-BT)						
	In-sample		**10-fold cross-validation**			
b	R^2	**SSE**	R^2	**(Std)**	**SSE**	**(Std)**
1.00E-11	0.111	490.334	0.105	(0.091)	492.748	(14.174)
0.0001	0.310	380.364	0.304	(0.080)	382.983	(12.714)
0.0005	0.346	360.552	0.340	(0.077)	363.189	(12.128)
0.001	0.363	351.298	0.357	(0.075)	353.942	(11.810)
0.005	0.403	329.021	0.397	(0.069)	331.674	(10.867)
0.01	0.420	319.845	0.414	(0.066)	322.496	(10.361)
0.05	0.449	303.923	0.443	(0.057)	306.545	(8.970)
0.08	0.453	301.769	0.447	(0.054)	304.378	(8.533)
0.1	**0.454**	**301.256**	**0.448**	**(0.053)**	**303.857**	**(8.326)**
0.2	0.453	301.702	0.447	(0.049)	304.289	(7.706)
0.3	0.450	303.032	0.445	(0.047)	305.624	(7.384)
0.45	0.448	304.237	0.443	(0.046)	306.840	(7.172)

Table 3. Two-Step Regression

	Step 1 (Ordered logit)		Step 2 (OLS)	
Explanatory variable	**Coefficient**	**(SE)**	**Coefficient**	**(SE)**
Seniority index	4.747	(0.257)***	0.347	(0.032)***
Revolvers	-0.872	(0.248)***	-0.052	(0.031)*
Term loans	-0.524	(0.253)**	-0.026	(0.032)
Senior secured bonds	0.190	(0.272)	-0.068	(0.034)**
Senior unsecured bonds	-0.744	(0.169)***	-0.042	(0.019)**
Senior subordinated bonds	-0.165	(0.182)	0.022	(0.021)
Junior bonds	-0.188	(0.314)	0.044	(0.041)
Capital stock	0.414	(0.188)**	-0.028	(0.026)
Equipment	-0.049	(0.266)	0.115	(0.031)***
Guarantees	-2.376	(1.079)**	-0.427	(0.247)*
Intellectual	1.225	(1.034)	0.282	(0.125)**
Inter-company debt	1.385	(2.827)	0.363	(0.248)
Inventory, receivables, and cash	-1.462	(0.276)***	-0.107	(0.065)
Other	-0.831	(0.410)**	0.165	(0.077)**
Unsecured	1.038	(0.209)***	0.064	(0.027)**
Second lien	-0.138	(0.234)	0.101	(0.031)***
Third lien	1.919	(0.488)***	0.067	(0.061)
Industry distance-to-default	-4.479	(0.981)***	-0.940	(0.130)***
Aggregate default rate	0.981	(0.604)	0.296	(0.080)***
Trailing 12-month market return	-0.060	(0.029)**	-0.015	(0.004)***
Three-month T-bill rate	0.216	(0.227)	0.206	(0.031)***
Utility dummy	-1.769	(0.190)***	0.004	(0.028)
Intercept			0.409	(0.050)***
γ_0	0.446	(0.373)***		
γ_1	5.704	(0.387)***		
Observations	3,751		2,380	
R^2	0.458			
SSE	298.673			
10-fold cross-validation				
R^2 (Std)	0.452	(0.047)		
SSE (Std)	301.369	(7.504)		

*, **, *** indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

Table 4. Inflated Beta Regression

	α: Equation (13)		β: Equation (14)		γ:Equation (15)	
Explanatory variable	**Coefficient**	**(SE)**	**Coefficient**	**(SE)**	**Coefficient**	**(SE)**
Seniority index	-4.385	(0.202)***	4.749	(0.403)***	1.663	(0.128)***
Revolvers	0.277	(0.144)**	-0.571	(0.308)**	-0.205	(0.102)**
Term loans	-0.026	(0.142)	0.098	(0.265)	-0.094	(0.106)
Senior secured bonds	-0.991	(0.162)***	-5.309	(0.719)***	-0.201	(0.113)**
Senior unsecured bonds	-0.175	(0.154)	-1.082	(0.188)***	-0.112	(0.071)*
Senior subordinated bonds	-1.138	(0.119)***	-0.638	(0.192)***	0.149	(0.080)**
Junior bonds	-0.155	(0.294)	-0.443	(0.276)*	0.228	(0.168)*
Capital stock	-0.467	(0.192)***	-2.762	(0.712)***	-0.070	(0.107)
Equipment	-0.129	(0.206)	-1.297	(0.165)***	0.237	(0.121)**
Guarantees	2.215	(0.286)***	-0.204	(0.086)***	-1.343	(0.323)***
Intellectual	-1.743	(0.513)***	0.045	(0.014)***	0.916	(0.129)***
Inter-company debt	-1.008	(0.491)**	0.091	(0.010)***	1.197	(0.253)***
Inventory, receivables, and cash	1.745	(0.221)***	3.115	(0.566)***	-0.224	(0.145)*
Other	0.923	(0.241)***	-0.889	(0.381)***	0.505	(0.073)***
Unsecured	-0.796	(0.132)***	0.984	(0.345)***	0.293	(0.101)***
Second lien	0.139	(0.217)	0.066	(0.458)	0.381	(0.126)***
Third lien	-0.902	(0.175)***	1.933	(0.384)***	0.307	(0.188)*
Industry distance-to-default	4.100	(0.544)***	-4.841	(1.845)***	-3.348	(0.257)***
Aggregate default rate	-2.785	(0.400)***	-4.136	(0.614)***	1.263	(0.208)***
Trailing 12-month market return	0.008	(0.029)	-0.185	(0.045)***	-0.070	(0.016)***
Three-month T-bill rate	-0.478	(0.211)**	-1.002	(0.188)***	0.981	(0.120)***
Utility dummy	1.776	(0.169)***	-0.988	(0.294)***	-0.101	(0.103)
Intercept	1.453	(0.211)***	-3.824	(0.519)***	-0.680	(0.148)***
Phi (φ)	2.574	(0.026)				
Observations	3,751					
R^2	0.455					
SSE	300.571					
10-fold cross-validation						
R^2 (Std)	0.448	(0.048)				
SSE (Std)	303.600	(7.499)				

*, **, *** indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

Table 5. Tobit Regression

	Censoring at 0 and 1	
Explanatory variable	**Coefficient**	**(SE)**
Seniority index	0.924	(0.044)***
Revolvers	-0.149	(0.043)***
Term loans	-0.047	(0.044)
Senior secured bonds	0.057	(0.047)
Senior unsecured bonds	-0.086	(0.027)***
Senior subordinated bonds	0.029	(0.030)
Junior bonds	0.009	(0.058)
Capital stock	0.055	(0.036)
Equipment	0.109	(0.045)**
Guarantees	-0.693	(0.227)***
Intellectual	0.374	(0.185)**
Inter-company debt	0.526	(0.401)
Inventory, receivables, and cash	-0.330	(0.053)***
Other	-0.128	(0.086)
Unsecured	0.224	(0.038)***
Second lien	0.066	(0.042)
Third lien	0.315	(0.084)***
Industry distance-to-default	-1.357	(0.176)***
Aggregate default rate	0.393	(0.108)***
Trailing 12-month market return	-0.018	(0.005)***
Three-month T-bill rate	0.176	(0.042)***
Utility dummy	-0.307	(0.035)***
Intercept	-0.056	(0.067)
Observations	3,751	
R^2	0.450	
SSE	303.282	
10-fold cross-validation		
R^2 (Std)	0.445	(0.047)
SSE (Std)	305.571	(7.545)

*, **, *** indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

Table 6. Censored Gamma Regression

Explanatory variable	**Coefficient**	**(SE)**
Seniority index	0.035	(0.005)***
Revolvers	-0.006	(0.003)**
Term loans	-0.002	(0.003)
Senior secured bonds	0.002	(0.003)
Senior unsecured bonds	-0.003	(0.001)**
Senior subordinated bonds	0.001	(0.002)
Junior bonds	0.000	(0.003)
Capital stock	0.002	(0.002)
Equipment	0.004	(0.002)**
Guarantees	-0.026	(0.012)**
Intellectual	0.014	(0.011)
Inter-company debt	0.020	(0.017)
Inventory, receivables, and cash	-0.012	(0.004)***
Other	-0.005	(0.005)
Unsecured	0.008	(0.002)***
Second lien	0.002	(0.002)
Third lien	0.012	(0.005)***
Industry distance-to-default	-0.051	(0.012)***
Aggregate default rate	0.015	(0.007)**
Trailing 12-month market return	-0.001	(0.000)**
Three-month T-bill rate	0.007	(0.002)***
Utility dummy	-0.012	(0.002)***
Intercept	-5.124	(0.126)***
Alpha (α)	4350.793	(1085.157)***
Xi (ξ)	25.952	(3.262)***
Observations	3,751	
R^2	0.449	
SSE	303.748	
10-fold cross-validation		
R^2 (Std)	0.445	(0.047)
SSE (Std)	305.636	(7.925)

*, **, *** indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

Table 7. Two-Tiered Gamma Regression

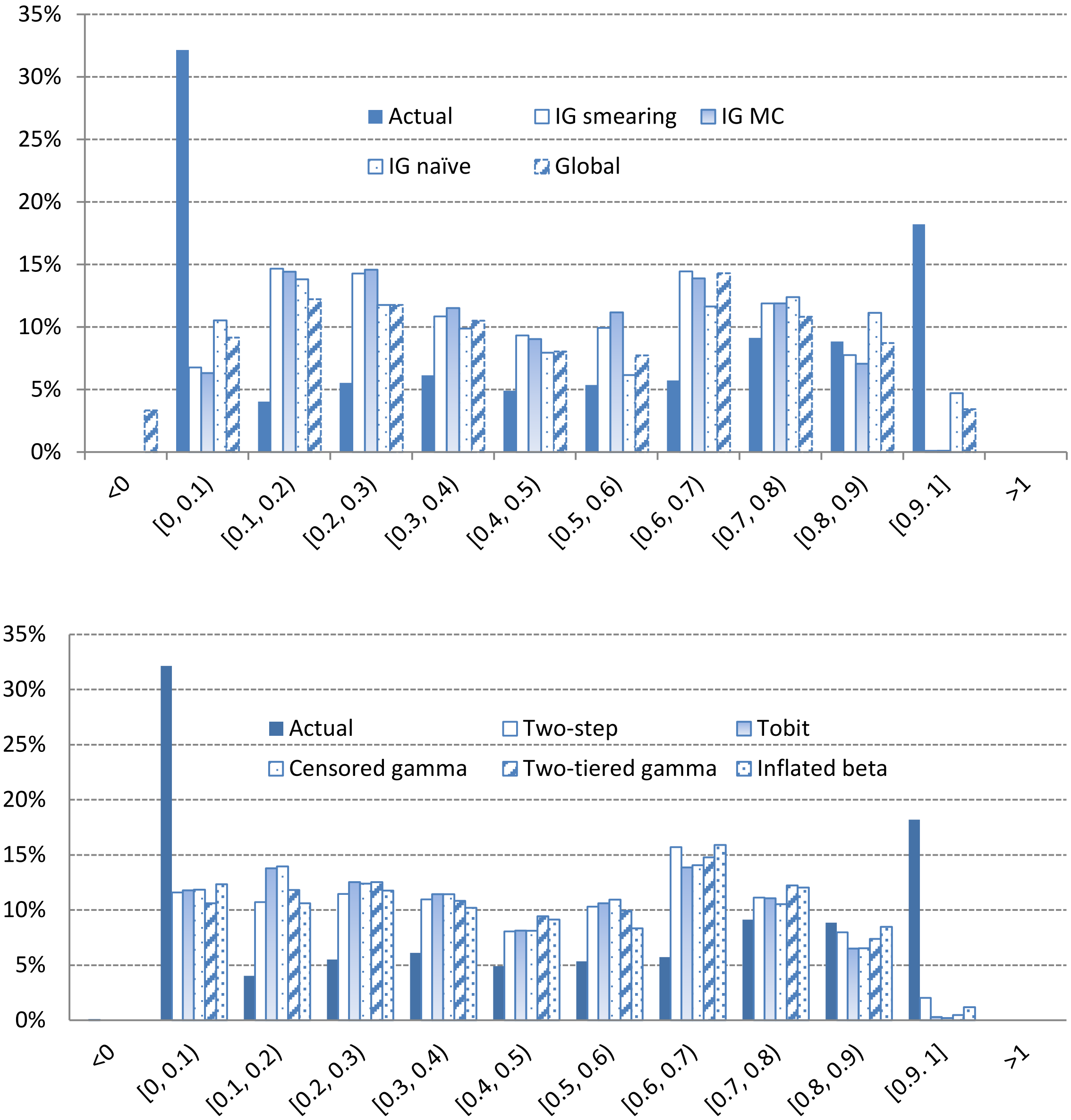
	β: Equation (22)		γ: Equation (23)	
Explanatory variable	**Coefficient**	**(SE)**	**Coefficient**	**(SE)**
Seniority index	0.065	(0.010)***	0.048	(0.008)***
Revolvers	-0.005	(0.007)	-0.008	(0.006)*
Term loans	0.000	(0.008)	-0.004	(0.006)
Senior secured bonds	0.015	(0.008)**	-0.012	(0.006)**
Senior unsecured bonds	0.002	(0.005)	-0.010	(0.004)***
Senior subordinated bonds	0.014	(0.007)**	-0.003	(0.004)
Junior bonds	0.003	(0.014)	0.000	(0.007)
Capital stock	0.007	(0.005)*	-0.002	(0.005)
Equipment	0.004	(0.008)	0.011	(0.006)**
Guarantees	-1.330	(0.023)***	-0.056	(0.042)*
Intellectual	0.024	(0.025)	0.030	(0.020)*
Inter-company debt	0.767	(0.010)***	0.040	(0.035)
Inventory, receivables, and cash	-0.022	(0.006)***	0.008	(0.010)
Other	-0.015	(0.010)*	0.020	(0.012)**
Unsecured	0.013	(0.006)**	0.009	(0.006)**
Second lien	-0.001	(0.006)	0.010	(0.006)*
Third lien	0.016	(0.016)	0.017	(0.011)*
Industry distance-to-default	-0.063	(0.022)***	-0.104	(0.021)***
Aggregate default rate	0.041	(0.016)***	0.013	(0.013)
Trailing 12-month market return	0.000	(0.001)	-0.002	(0.001)***
Three-month T-bill rate	0.007	(0.006)	0.016	(0.005)***
Utility dummy	-0.027	(0.005)***	-0.003	(0.004)
Intercept	-4.991	(0.104)***	-4.929	(0.111)***
Alpha (α)	1462.067	(325.831)***		
Xi (ξ)	10.147	(1.205)***		
Observations	3,751			
R^2	0.455			
SSE	300.311			
10-fold cross-validation				
R^2 (Std)	0.450	(0.045)		
SSE (Std)	302.820	(7.172)		

*, **, *** indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

Table 8. Summary of Model Performance, Complexity, and Computational Burden of Alternative Models

	In-sample			10-fold cross-validation			Complexity	Computational burden
	R^2	SSE	Rank	R^2	SSE	Rank		
FRR	0.463	296.120	1	0.457	298.600	1	Medium	Low
Two-step	0.458	298.673	2	0.452	301.369	2	Low	Low
Two-tiered gamma	0.455	300.311	3	0.450	302.820	3	High	High
Inflated beta	0.455	300.571	4	0.448	303.600	6	High	High
IGR-BT global	0.454	301.256	5	0.448	303.857	7	High	Medium
IGR global	0.453	301.298	6	0.448	303.900	8	Low	Low
IGR-BT smearing	0.453	301.614	7	0.449	303.187	4	Medium	Medium
IGR smearing	0.453	301.662	8	0.449	303.237	5	Medium	Medium
IGR naïve	0.452	302.294	9	0.446	304.818	11	Low	Low
IGR MC	0.451	302.397	10	0.447	304.507	9	Medium	High
IGR-BT MC	0.451	302.605	11	0.447	304.536	10	High	High
Tobit	0.450	303.280	12	0.445	305.571	12	Low	Low
IGR-BT naïve	0.450	303.391	13	0.444	305.922	14	Medium	Medium
Censored gamma	0.449	303.748	14	0.445	305.636	13	High	High
OLS	0.448	304.320	15	0.443	306.920	15	Low	Low

Figure 1. Distribution of the Actual and Fitted LGDs



Appendix

This appendix derives $E(L)$ under the two-tiered gamma model specification. From the definition of expectations, we have that

$$\begin{aligned} E(L) &= Pr(L=0)E(L|L=0) + Pr(L=1)E(L|L=1) + Pr(L\in(0,1))E(L|L\in(0,1)) \\ &= Pr(L=1) + Pr(L\in(0,1))E(L|L\in(0,1)) \\ &= Pr(L=1) + (1-Pr(L=0)-Pr(L=1))E(L_2^*|L_1^*>0, L_2^*\in(0,1)) \\ &= Pr(L=1) + (1-Pr(L=0)-Pr(L=1))E(L_2^*|L_2^*\in(0,1)) \end{aligned}$$

The expression in the third line follows from the setup of the model, and the fourth line follows from the independence of L_1^* and L_2^*. The probabilities $Pr(L=0)$ and $Pr(L=1)$ are given in Sigrist and Stahel (2011) and in the main body of this paper, so the only term that needs to be derived is $E(L_2^*|L_2^*\in(0,1))$.

In the model formulation, L_2^* is a shifted gamma truncated at 0, and in order to calculate the aforementioned expectation, we need to truncate L_2^* again to $(0,1)$ and find its expectation. We will proceed by deriving the distribution of the shifted gamma truncated at 0, followed by the distribution of the previous distribution truncated again to $(0,1)$, and then the expectation of the resulting distribution.

Generally, the PDF of a shifted gamma is $f_Y(y) = g_{\alpha,\theta}(y+\xi)$ for $y\in(-\xi,\infty)$, where $g_{\alpha,\theta}(\cdot)$ denotes the PDF of a typical gamma distribution with parameters α and θ. The CDF is $F_Y(y) = G_{\alpha,\theta}(y+\xi)$, where $G_{\alpha,\theta}(\cdot)$ is the CDF of a typical gamma distribution with parameters α and θ. These are derived in Sigrist and Stahel (2011). Using these quantities, the PDF for the truncated version of this distribution to $(0,\infty)$ is by definition

$$\begin{aligned} f_{Y|Y\in(0,\infty)}(y) &= \frac{f_Y(y)}{F_Y(\infty)-F_Y(0)} \\ &= \frac{g_{\alpha,\theta}(y+\xi)}{1-G_{\alpha,\theta}(\xi)} \end{aligned}$$

for $y \in (0, \infty)$. The CDF is thus

$$
\begin{aligned}
F_{Y|Y\in(0,\infty)}(y) &= \int_0^y \frac{g_{\alpha,\theta}(t+\xi)}{1-G_{\alpha,\theta}(\xi)}dt \\
&= \int_\xi^{y+\xi} \frac{g_{\alpha,\theta}(u)}{1-G_{\alpha,\theta}(\xi)}du \\
&= \frac{G_{\alpha,\theta}(y+\xi)-G_{\alpha,\theta}(\xi)}{1-G_{\alpha,\theta}(\xi)}
\end{aligned}
$$

where the second line uses the variable substitution $u = t + \xi$.

Now we truncate the previous distribution to $(0, 1)$. The PDF is

$$
\begin{aligned}
f_{Y|Y\in(0,1)}(y) &= \frac{f_{Y|Y\in(0,\infty)}(y)}{F_{Y|Y\in(0,\infty)}(1)-F_{Y|Y\in(0,\infty)}(0)} \\
&= \frac{\frac{g_{\alpha,\theta}(y+\xi)}{1-G_{\alpha,\theta}(\xi)}}{\frac{G_{\alpha,\theta}(1+\xi)-G_{\alpha,\theta}(\xi)}{1-G_{\alpha,\theta}(\xi)}-\frac{G_{\alpha,\theta}(0+\xi)-G_{\alpha,\theta}(\xi)}{1-G_{\alpha,\theta}(\xi)}} \\
&= \frac{g_{\alpha,\theta}(y+\xi)}{G_{\alpha,\theta}(1+\xi)-G_{\alpha,\theta}(\xi)}
\end{aligned}
$$

for $y \in (0, 1)$. The last step is to take the expectation:

$$
\begin{aligned}
E_{Y|Y\in(0,1)}(Y) &= \int_0^1 y f_{Y|Y\in(0,1)}(y)dy \\
&= \int_0^1 y\frac{g_{\alpha,\theta}(y+\xi)}{G_{\alpha,\theta}(1+\xi)-G_{\alpha,\theta}(\xi)}dy \\
&= \frac{\int_\xi^{1+\xi}(u-\xi)g_{\alpha,\theta}(u)du}{G_{\alpha,\theta}(1+\xi)-G_{\alpha,\theta}(\xi)} \\
&= \frac{\left[\int_\xi^{1+\xi} u g_{\alpha,\theta}(u)du - \int_\xi^{1+\xi} \xi g_{\alpha,\theta}(u)du\right]}{G_{\alpha,\theta}(1+\xi)-G_{\alpha,\theta}(\xi)} \\
&= \frac{\alpha\theta(G_{\alpha+1,\theta}(1+\xi)-G_{\alpha+1,\theta}(\xi))-\xi(G_{\alpha,\theta}(1+\xi)-G_{\alpha,\theta}(\xi))}{G_{\alpha,\theta}(1+\xi)-G_{\alpha,\theta}(\xi)}
\end{aligned}
$$

where the third line uses the substitution $u = y + \xi$, the first integral in the numerator of the fourth line uses equation 47 from Sigrist and Stahel (2013), and the second integral on the same line is equal to the difference in CDFs evaluated at the limits of integration once ξ is factored out. Putting these quantities together, we obtain the expectation of interest.