UCD GEARY INSTITUTE
DISCUSSION PAPER SERIES

Modelling Downturn Loss Given Default

Raffaella Calabrese
University of Milano-Bicocca

Geary WP2012/16
November 2012

Modelling Downturn Loss Given Default

Raffaella Calabrese[a]

[a] *University of Milano-Bicocca, Milan, Italy*

Abstract

Basel II requires that the internal estimates of Loss Given Default (LGD) reflect economic downturn conditions, thus modelling the "downturn LGD". In this work we suggest a methodology to estimate the downturn LGD distribution. Under the assumption that LGD is a mixture of an expansion and a recession distribution, an accurate parametric model for LGD is proposed and its parameters are estimated by the EM algorithm. Finally, we apply the proposed model to empirical data on Italian bank loans.

Keywords: Downturn LGD; Mixed random variable; Mixture; Beta density

1. Introduction

The Basel II Accord (Basel Committee on Banking Supervision (BCBS), 2004, paragraph 286-317) considers the "Loss Given Default"(LGD) as the loss quota in the case of the borrower's default. In this framework, banks adopting the advanced Internal-Rating-Based (IRB) approach are allowed to use their own estimates of LGDs. Basel II requires that the internal estimates reflect economic downturn conditions wherever necessary to capture risk accurately (BCBS, 2004, paragraph 468). Moreover, banks must

*Corresponding author. Tel.: +39 02 64483193; fax: +39 02 64483105.
Email address: raffaella.calabrese1@unimib.it (Raffaella Calabrese)

account for the possibility that the LGD may exceed the weighted average value when credit losses are higher than average, thus modelling the so-called "downturn LGD". In the assessment of capital adequacy the downturn LGD is also useful for stress testing purposes (BCBS, 2004, paragraph 434). Although the need to estimate the downturn LGD is clearly framed (BCBS, 2005), Basel II does not provide a specific approach that banks must use in calculating this variable.

The main aim of this paper is to propose an accurate model for LGD that allows to estimate the distribution function of the downturn LGD. We consider a dynamic behaviour of LGD over the economic cycle characterized by two regimes: expansion and recession. We assume that the LGD is a mixture of an expansion and a recession distribution. Under the assumption introduced by Calabrese (2012) in the regression framework, the expansion and the recession distributions of the LGD are assumed to be two mixed random variables, each of them is given by the mixture of a Bernoulli random variable and a beta random variable. On the one hand, the Bernoulli random variable allows to reproduce the high concentration of data on LGDs at total recovery and total loss (Calabrese and Zenga, 2010; Renault and Scaillet, 2004; Schuermann, 2003). On the other hand, the beta distribution is well suited to the modelling of LGDs (Gupton et al., 1997; Gupton and Stein, 2002), as it has support [0,1] and, in spite of requiring only two parameters, is quite flexible.

To estimate the parameters of the LGD distribution, we apply the EM algorithm. To obtain a finite beta density function, Calabrese and Zenga (2010)'s parametrization is used. The suggested methodology allows to estimate the distribution of the downturn LGD and finally is applied to a comprehensive survey on loan recovery process of Italian banks.

The present paper is organized as follows. The next section presents the approach, here proposed, to estimate the downturn LGD distribution. Section 3 describes the dataset of the Bank of Italy and shows the estimation results by applying the proposed model to these data.

2. A model for downturn LGD

Two different distributions of LGDs are considered to hold over expansion and recession periods, so LGDs can be seen as realization of these two distributions. By considering two regimes of the economic cycle, expansion and recession, LGDs are drawn from a mixture of an expansion (E) and a recession (R) distributions

$$F_{LGD}(lgd) = \pi F_{LGD/E}(lgd) + (1-\pi)F_{LGD/R}(lgd), \tag{1}$$

where $F_{LGD/S}(lgd)$ is the cumulative distribution function of the LGD over a given period conditional on the state S $(S = E, R)$ of the economic cycle and π is the probability of the expansion regime (Filardo, 1994).

Since the incidence of LGDs equal to 0 or 1 is high (Renault and Scaillet, 2004; Schuermann, 2003), to supply accurate estimations for the extreme values, Calabrese (2012) propose to consider LGD as a mixed random variable, obtained as the mixture of a Bernoulli random variable and the beta random variable B. This means that the distribution function of LGD $F_{LGD/S}$ conditional on the state S of the economic cycle is defined as

$$F_{LGD/S}(lgd) = \begin{cases} p_0^s & lgd = 0 \\ p_0^s + [1 - p_0^s - p_1^s]F_{B/S}(lgd) & lgd \in (0,1) \\ 1 & lgd = 1 \end{cases} \tag{2}$$

where $F_{B/S}$ denotes the distribution function of the beta random variable B conditional on the state S of the economic cycle and $p_j^s = P\{LGD = j/S\}$ is the conditional probability that the LGD is equal to j with $j = 0, 1$ given the state S of the economic cycle.

The probability density function of the beta random variable B given S of parameters $\dfrac{\theta^s}{\sigma^s} + 1$ and $\dfrac{1-\theta^s}{\sigma^s} + 1$ is

$$g_{B/S}(lgd; \theta^s, \sigma^s) = \frac{\Gamma\left(\frac{\theta^s}{\sigma^s} + 1\right)\Gamma\left(\frac{1-\theta^s}{\sigma^s} + 1\right) lgd^{\frac{\theta^s}{\sigma^s}} (1 - lgd)^{\frac{1-\theta^s}{\sigma^s}}}{\Gamma\left(\frac{\theta^s}{\sigma^s} + 1, \frac{1-\theta^s}{\sigma^s} + 1\right)}, \tag{3}$$

where $0 < lgd < 1$, $0 < \theta^s < 1$ is the mode of the beta density function and $\sigma^s > 0$ is the dispersion parameter. The parametrization of the beta density function $g_{B/S}(lgd; \theta^s, \sigma^s)$ is applied by Calabrese and Zenga (2010) in the nonparametric framework. The main advantage of this parametrization is that the beta density function is finite for all values of θ^s and σ^s since the beta parameters $\frac{\theta^s}{\sigma^s} + 1$ and $\frac{1-\theta^s}{\sigma^s} + 1$ are higher than one.

The beta random variable B given S of parameters $\frac{\theta^s}{\sigma^s} + 1$ and $\frac{1-\theta^s}{\sigma^s} + 1$ has an expected value

$$E(B/S) = \frac{\theta^s + \sigma^s}{2\sigma^s + 1}$$

and variance

$$V(B/S) = \frac{\sigma^s[\theta^s - (\theta^s)^2 + \sigma^s + (\sigma^s)^2]}{(1 + 2\sigma^s)^2(1 + 3\sigma^s)}. \tag{4}$$

To understand the influence of the parameter σ^s, we apply Maclaurin series to equation (4) obtaining

$$V(B/S) = \sigma^s\theta^s(1 - \theta^s) + O[(\sigma^s)^2],$$

this means that the variance $V(B/S)$ increases by increasing σ^s and maintaining constant θ^s.

Assuming the states S of the economic cycle are unobserved, the probability π and the parameters $(\boldsymbol{p}_0, \boldsymbol{p}_1, \boldsymbol{\theta}, \boldsymbol{\sigma})$ need to be estimated, where $\boldsymbol{p}_0 = [p_0^e, p_0^r]'$, $\boldsymbol{p}_1 = [p_1^e, p_1^r]'$ $\boldsymbol{\theta} = [\theta^e, \theta^r]'$ and $\boldsymbol{\sigma} = [\sigma^e, \sigma^r]'$.

2.1. Calibrating the downturn LGD distribution

We use the Expectation-Maximization (EM) algorithm (Dempster et al., 1977) to estimate the parameters π and $(\boldsymbol{p}_0, \boldsymbol{p}_1, \boldsymbol{\theta}, \boldsymbol{\sigma})$.

In the mixture framework, the observed data $\boldsymbol{lgd} = (lgd_1, lgd_2, ..., lgd_n)$ is completed with a component-label vector $\boldsymbol{Z} = (Z_1, Z_2, ..., Z_n)$ whose elements are assumed to be independent and are defined as

$$z_{is} = \begin{cases} 1 & \text{if } lgd_i \text{ comes from the state } s \\ 0 & \text{otherwise.} \end{cases}$$

Under the assumption that LGD_i are independent and identically distributed random variables with cumulative distribution function (1), the complete-data log-likelihood function is

$$\begin{aligned} l_c(\pi, \boldsymbol{p}_0, \boldsymbol{p}_1, \boldsymbol{\theta}, \boldsymbol{\sigma}) &= \ln \pi \sum_{i=1}^{n} z_{ie} + \ln(1-\pi) \sum_{i=1}^{n} z_{ir} \\ &+ \sum_{i=1}^{n} \sum_{S=e,r} z_{is} \ln f_{LGD/S}\left(lgd_i; p_0^s, p_1^s, \theta^s, \sigma^s\right). \end{aligned} \tag{5}$$

The EM algorithm iteratively maximizes the log-likelihood function (5). Each iteration of the EM algorithm includes the E-step, which computes the conditional expectation of the complete-data log-likelihood (5) given the observed data $\boldsymbol{lgd}$, and the M-step, which obtains the maximum of the complete-data log-likelihood function (5).

To compute the inital value $\pi^{(0)}$, analogously to Ji et al. (2005) we assign the smallest 50% of lgd_i to the expansion component and the others to the recession component. Maximizing (5), we obtain the initial values $(\boldsymbol{p}_0^{(0)}, \boldsymbol{p}_1^{(0)}, \boldsymbol{\theta}^{(0)}, \boldsymbol{\sigma}^{(0)})$. The algorithm follows the sequence:

(1) On the $(k+1)$-th iteration, the **E-step** requires the calculation of the conditional expectation of the complete-data log-likelihood (5). Since $\boldsymbol{Z}$ is non-observed data, $z_{is}^{(k)}$ is replaced by the conditional expectation of Z_{is} given the observed data $\boldsymbol{lgd}$ and we consider the parameter estimates $(\pi^{(k)}, \boldsymbol{p}_0^{(k)}, \boldsymbol{p}_1^{(k)}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\sigma}^{(k)})$ from the k-th iteration of the M-step

$$z_{is}^{(k+1)} = \frac{P^{(k)}(S) f_{LGD/S}\left(lgd_i; p_0^{(k)s}, p_1^{(k)s}, \theta^{(k)s}, \sigma^{(k)s}\right)}{f_{LGD}\left(lgd_i; \boldsymbol{p}_0^{(k)}, \boldsymbol{p}_1^{(k)}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\sigma}^{(k)}\right)}$$

where $P^{(k)}(S) = \pi^{(k)}$ for $S = R$, $P^{(k)}(S) = 1 - \pi^{(k)}$ otherwise.

(2) On the $(k+1)$-th iteration, the **M-step** requires the maximization of the complete-data log-likelihood function (5) with respect to π, $\boldsymbol{p}_0$, $\boldsymbol{p}_1$, $\boldsymbol{\theta}$, $\boldsymbol{\sigma}$ replacing z_{is} by z_{is}^{k+1}. Firstly, the updated estimates $\pi^{(k+1)}$, $\boldsymbol{p}_0^{(k+1)}$, $\boldsymbol{p}_1^{(k+1)}$ are obtained as

$$\begin{aligned} \pi^{(k+1)} &= \frac{\sum_{i=1}^{n} z_{ie}^{(k+1)}}{n} \\ p_j^{(k+1)s} &= \frac{\sharp\{LGD_i = j/S\} P^{(k+1)}(S)}{n} \end{aligned}$$

where $j = 0, 1$.

The updated estimates $\boldsymbol{\theta}^{(k+1)}$ and $\boldsymbol{\sigma}^{(k+1)}$ are the solution of the following system

$$\begin{cases} \dfrac{\partial l_c\left(\pi, \boldsymbol{p}_0, \boldsymbol{p}_1, \boldsymbol{\theta}, \boldsymbol{\sigma}\right)}{\partial \boldsymbol{\theta}} = \mathbf{0} \\[2ex] \dfrac{\partial l_c\left(\pi, \boldsymbol{p}_0, \boldsymbol{p}_1, \boldsymbol{\theta}, \boldsymbol{\sigma}\right)}{\partial \boldsymbol{\sigma}} = \mathbf{0} \end{cases}$$

Since the updated estimates $\boldsymbol{\theta}^{(k+1)}$ and $\boldsymbol{\sigma}^{(k+1)}$ do not have a closed-form, they are obtained by using a nonlinear optimization algorithm.

The E-step and the M-step are alternatively repeated until the difference between two consecutive values of the complete-data log-likelihood (5) is negligible.

3. Empirical analysis

The Bank of Italy conducts a comprehensive survey on the loan recovery process of Italian banks in the years 2000-2001. By means of a questionnaire, about 250 banks are surveyed. Since they cover nearly 90% of total domestic assets of 1999, the sample is representative of the Italian recovery process. The database comprises 149,378 defaulted borrowers. We highlight that the data concern individual loans which are privately held and not listed on the market. In particular, loans are towards Italian resident defaulted borrowers on the 31/12/1998 and written off by the end of 1999.

The definition of default the Bank of Italy chooses in its survey is tighter than the one BCBS (2004, paragraph 452) proposes. The difference is given by the inclusion of transitory non-performing debts.

Figure 1 around here

To compute the LGD, we apply the expression proposed by Calabrese and Zenga (2010). Figure 1 shows the Bank of Italy's data. Firstly, the mode of the LGD distribution is the extreme value 1, with 23% of the observations. Besides, LGD equal to 0 exhibits also a high percentage (7.78%). The average LGD is 0.7158, the median value is 0.7777 and the standard deviation is 0.3400.

	$\hat{\pi}$	$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\sigma}$
Expansion	0.7337	0.0571	0.1687	0.3925	0.5968
Recession	0.2663	0.0207	0.0612	0.9171	0.1014

Table 1: Estimates from the Bank of Italy's data.

We apply the methodology proposed in this work to the Bank of Italy's database and the results[1] are reported in Table 1. The main advantages of the methodology here proposed are that it allows to estimate the downturn LGD distribution, shown in Figure 2, and to replicate the multimodality of LGDs on (0,1) (Renault and Scaillet, 2004; Schuermann, 2003).

Figure 2 around here

References

Basel Committee on Banking Supervision, 2004. International convergence of capital measurement and capital standards: A revised framework. Bank for International Settlements: Basel, June.

Basel Committee on Banking Supervision, 2005. Guidance on paragraph 468 of the framework document. Bank for International Settlements: Basel, July.

Calabrese R., Zenga M., 2010. Bank loan recovery rates: Measuring and

[1]We obtain these results by R-program and the algorithm is available upon request to the authors.

nonparametric density estimation. Journal of Banking and Finance 34(5), 903-911.

Calabrese R., 2012. Predicting bank loan recovery rates with mixed continuous-discrete model. Applied Stochastic Models in Business and Industry. To appear

Committee on the Global Financial System, 2005. The role of ratings in structured finance: Issues and implications. Bank for International Settlements: Basel.

Dempster A.P., Laird N.M., Rubin D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society B 39(1), 1-38.

Filardo A. J., 1994. Business-cycles phases and their transitional dynamics. Journal of Business and Economic Statistics 12 (3), 299-308.

Gupton G. M., Finger C. C., Bhatia M., 1997. CreditMetrics. Technical document, J. P. Morgan.

Gupton G. M., Stein R. M., 2002. Losscalc^{TM}: Model for predicting Loss Given Default (LGD). Moody's Investors Service.

Ji, Y. and Wu, C. and Liu, P. and Wang, J. and Coombes, K.R., 2005. Applications of beta-mixture models in bioinformatics. Bioinformatics, 21(9), 2118-2122.

Renault O., Scaillet O., 2004. On the way to recovery: A nonparametric bias free estimation of recovery rate densities. Journal of Banking and Finance 28, 2915-2931.

Schuermann T., 2003. What do we know about loss given default? Recovery risk. Working paper, Federal Reserve Bank of New York.

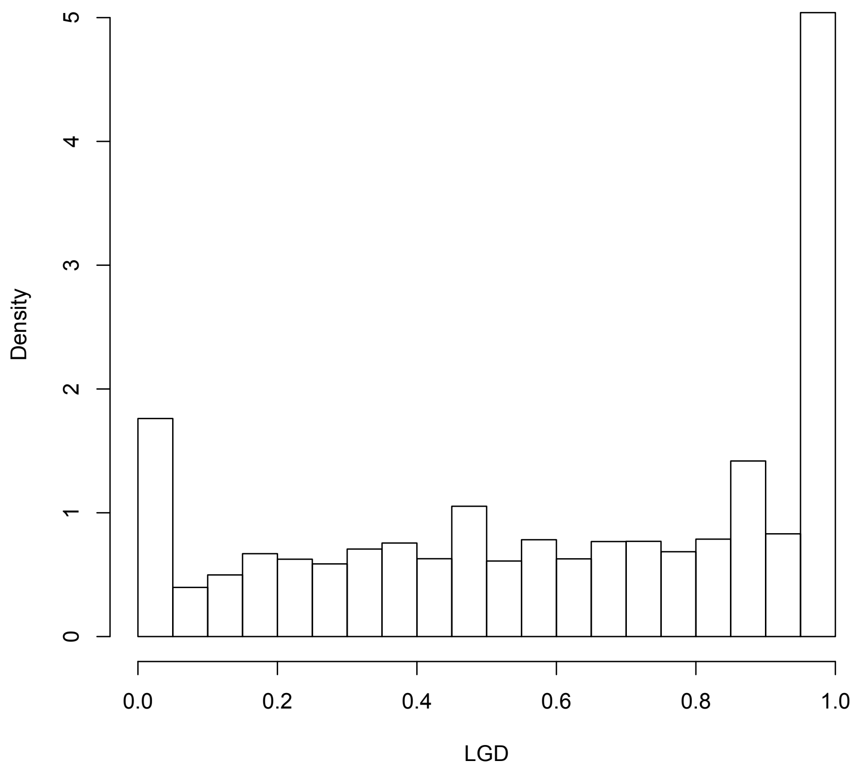
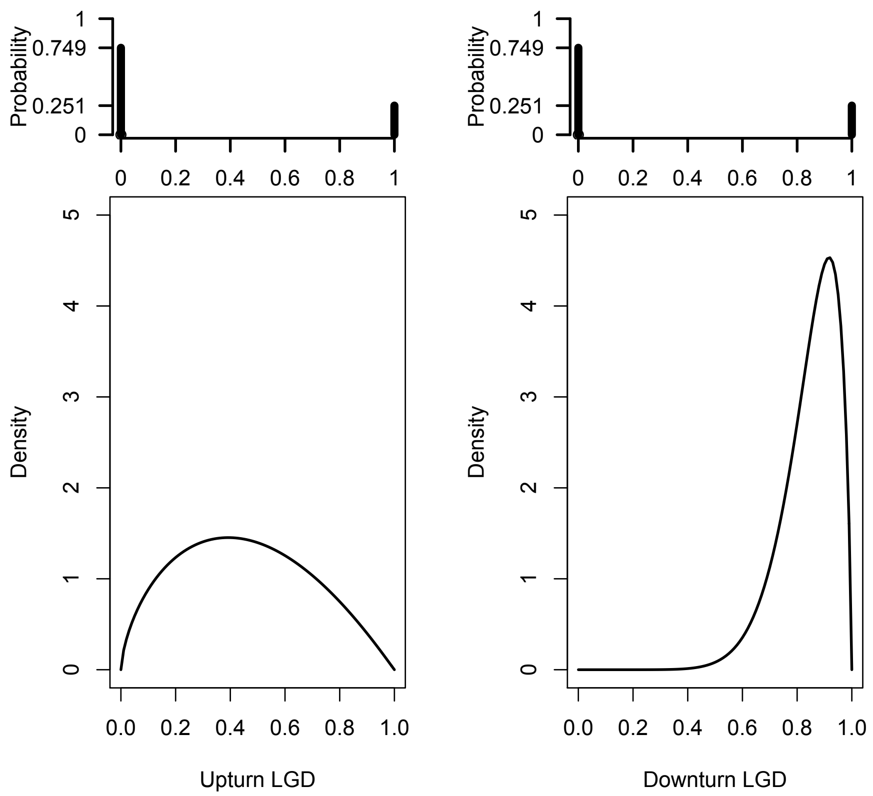


Figure 1: The distribution of the Bank of Italy's LGDs.


Probability
1
0.749
0.251
0
0
0.2
0.4
0.6
0.8
1
Density
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
Upturn LGD
Downturn LGD