Downturn Loss Given Default: mixture distribution estimation

Raffaella Calabrese[a]

[a] *Essex Business School*
University of Essex
Wivenhoe Park
Colchester CO4 3SQ

Abstract

The internal estimates of Loss Given Default (LGD) must reflect economic downturn conditions, thus estimating the "downturn LGD", as the new Basel Capital Accord Basel II establishes. We suggest a methodology to estimate the downturn LGD distribution to overcome the arbitrariness of the methods suggested by Basel II. We assume that LGD is a mixture of an expansion and recession distribution. In this work, we propose an accurate parametric model for LGD and we estimate its parameters by the EM algorithm. Finally, we apply the proposed model to empirical data on Italian bank loans.

---

*Corresponding author. Tel.: +44 (0) 1206 874569; Fax: +44 (0) 1206 873429.
Email address: `rcalab@essex.ac.uk` (Raffaella Calabrese)

1. Introduction

Loss Given Default (LGD) is the loss incurred by a financial institution when an obligor defaults on a loan, given as the fraction of Exposure At Default (EAD) unpaid after some period of time. In the Basel II framework (Basel Committee on Banking Supervision (BCBS), 2004, paragraph 286-317), banks adopting the advanced Internal Rating Based (IRB) approach are allowed to use their own estimates of LGDs that have to reflect economic downturn conditions. Hence, the "downturn LGD" is the maximum of the long-run default-weighted average LGD and the stressed LGD.

It requires the banks to identify the appropriate downturn conditions and incorporate them so as to produce LGD parameters for the bank's exposures, which are consistent with the identified downturn conditions. The main reason for this requirement is that the Vasicek model (Vasicek, 2002) used in Basel II does not have systematic correlation between Probability of Default (PD) and LGD and, to compensate for this deficiency, downturn LGD are required to be used as input to the model.

Although the downturn LGD is a key variable for banking practice, such a pivotal topic is relatively unexplored in the literature. The main aim of this paper is to propose a methodology to estimate the downturn LGD distribution. To achieve this aim, we consider the 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 distributions, each of these distributions is given by the mixture of a Bernoulli random variable and a beta random variable, as Calabrese (2012) suggested. On the one hand, the Bernoulli random variable allows to reproduce the high

2

concentration of data 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[1] to the modelling of LGDs (Bruche and González-Aguado, 2008; Gupton et al., 1997; Gupton and Stein, 2002). To estimate the parameters of the downturn LGD distribution, we apply the EM algorithm (Dempster et al., 1977). To obtain a finite beta density function, Calabrese and Zenga (2010)'s parametrization is used. With this method banks do not need to identify arbitrarily downturn conditions and, unlike the factor method, data on the default risk and default correlation are not required. Finally, we apply this proposal to a comprehensive Bank of Italy data set (Bank of Italy, 2001) of 149,378 Italian bank loans and we compare it with some methods used in the literature to estimate the downturn LGD.

The present paper is organized as follows. The next section analyses the available literature on downturn LGD. Section 3 describes some approaches to estimate downturn LGD. The following section presents the proposed approach to estimate the downturn LGD distribution. Section 5 describes the dataset of the Bank of Italy and shows the estimation results by applying the proposed model to these data. Finally, the last section is devoted to conclusions.

2. Literature review

An extensive literature suggested a link between LGD and the economic cycle (e.g. Bellotti and Crook, 2012; Calabrese, 2012). The systematic

---

[1]Since LGD lies in the interval [0,1], the beta distribution is a suitable parametric model for LGDs since it has support [0,1] and, in spite of requiring only two parameters, is quite flexible.

correlation between PD and LGD is not taken into account in many models. In the standard rating-based credit risk model developed by Gupton et al. (1997), it is assumed that recoveries on defaulted exposures are random outcomes, independent of the default event. A similar independence assumption is made in the model of Jarrow et al. (1997) and in the Vasicek model (Vasicek, 2002) used in the Basel II Accord. However, if realizations of recoveries are low exactly at times when many firms default, the assumption that recoveries are independent of default rates or constant would result in an underestimation of credit risk. To compensate for this deficiency, downturn LGD estimates are required to be used as an input to the model.

In the Basel II Accord (BCBS, 2005c), two approaches are presented to estimate downturn LGD. One approach would be to apply a mapping function similar to that used for the PDs that would extrapolate downturn LGDs from bank-reported average LGDs. Alternatively, banks could be asked to provide downturn LGD figures based on their internal assessments LGDs during adverse conditions. Provided that data is available, the latter approach is the easiest to implement, so Basel III (BCBS, 2011 paragraph 20) considers only this method to compute downturn LGD.

The drawback is that LGD data is generally sparse and there is very limited industry experience with regard to LGD estimates. Downturn LGD estimation based on historical data is currently not possible for many banks because of the short time periods available or for the lack of an economic downturn during the available period. The first approach of Basel II is an appropriate solution when historical data is not available.

Following the first approach of Basel II, Miu and Ozdemir (2006) suggest that the original LGD assessment by banks, without considering PD and LGD correlation, can be appropriately adjusted by incorporating a certain

degree of conservatism in cyclical LGD estimates within a point-in-time modelling framework. They use Monte Carlo to tabulate the relationship between long-run and downturn LGD. Barco (2007) extends their work to develop an analytical relationship between long-run and downturn LGD.

Moreover, Sabato and Schmid (2008) suggest a simple mapping function to estimate downturn LGD. They investigate the relationship between LGD and the credit cycle over the period from 2002 to 2007 using data covering a set of retail loans. The linear mapping function proposed by the Board of Governors of the Federal Reserve System (2006) can be considered as a particular case of Sabato and Schmid's proposal, as explained in Section 3.

Both the approaches presented by Basel II show the pivotal drawback of how economic downturn or mapping function of average LGD should be defined and identified. This arbitrariness leads to very different approaches being implemented across banks and countries, and significant effects on the level of capital requirements. As many authors have shown (e.g., Saurina and Trucharte, 2004; Altman and Sabato, 2005), Basel II Advanced-IRB capital requirements are highly sensible to LGD values in particular for retail asset classes. Hence, there is the necessity of suggesting a method to estimate downturn LGD to overcome the arbitrariness of the approaches suggested by Basel II.

3. Downturn LGD estimation models

Following the first approach of Basel II (BCBS, 2005c), Sabato and Schmid (2006) suggest the following linear mapping function to estimate downturn LGD on unsecured positions

$$DLGD - \mu^{LGD} = LGDSF(1 - \mu^{LGD}) \tag{1}$$

where μ^{LGD} is the long-term average [2] LGD, DLGD is the expected Downturn LGD and LGDSF is the LGD Stressing Factor given by

$$LGDSF = \frac{\text{stressed LGD} - \text{LGD}}{\text{LGD}}. \qquad (2)$$

Sabato and Schmid (2006) suggest to compute the stressed LGD as a function (not specified) of the stressed PD, given by the average PD plus the standard deviation of the observed default rates. The mapping function (1) implies that debts with relatively low historical LGD rates (e.g. senior bank loans) should have relatively large adjustments to their long-term average LGD rates, while debts with high historical LGD rates (e.g. subordinated bonds) should have relatively small adjustments. In other words, the difference between downturn LGD and the long-term average LGD varies inversely with the long-term average LGD.

In recognition that banks may be unable to estimate the LGD stressing factor, the Board of Governors of the Federal Reserve System (2006) proposes the following particular case of the equation (1) with $LGDSF = 0.08$

$$DLGD = .08 + .92\mu^{LGD}. \qquad (3)$$

where μ^{LGD} equals the long-term average LGD and DLGD is the expected Downturn LGD. It is worth noting that the magnitudes of the proposed adjustments to LGD are relatively modest, with a maximum adjustment of only eight percentage points. The Federal Reserve has offered no justification for the linear mapping function (3) except perhaps its intuitive appeal that debts with the lowest historical average LGD rates receive the largest upward downturn LGD adjustments.

---

[2]μ^{LGD} is known as *long-run default-weighted average loss rate given default* in Basel II (BCBS, 2005).

The methodology suggested by Miu and Ozdemir (2006) and then generalized by Barco (2007) also belongs to the first approach of Basel II. To analyse Barco's model, we start with the well-known Merton framework where the log return of the obligor i's asset value is given by

$$A_i = \rho_i^{PD} X_m + \sqrt{1 - (\rho_i^{PD})^2} X_{is}$$

where ρ_i^{PD} is known as correlation of asset returns. The independent standard normal random variables X_m and X_{is} are the systematic factor and obligor-specific idiosyncratic factor, respectively. Barco (2007) models the value of the asset of the creditor with a lognormal distribution. This is achieved by first establishing the following relationship for the standardised asset return

$$R_i = \rho_i^R X_m + \sqrt{1 - (\rho_i^R)^2} Y_m$$

where Y_m is an independent standard normal random variable representing the residual systematic asset return not explained by X_m. The parameter ρ_i^R correlates the assets of the obligor to the systemic factor X_m. Assuming obligor exposure is one unit with mean default rate denoted by μ_i^{PD}, its loss random variable is defined by

$$L_i = 1_{\{A_i < \Phi^{-1}(\mu_i^{PD})\}} L_i^R$$

where $1_{\{\}}(\cdot)$ is the indicator function and $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution.

Assuming a fully granular portfolio, L_i is contained in a homogeneous portfolio and continue to suppose that its exposure is equal to one unit, so Barco (2007) defines

$$L^\infty = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} L_i.$$

In a model with two systematic factors, from Proposition 3 in Gordy (2003) and the strong law of large numbers Barco (2007) obtains

$$q_\alpha(L^\infty) = q_\alpha \left[\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} L_i \right] = q_\alpha \left[E(L_i | X_m, Y_m) \right]$$

where q_α is the α quantile.

An important result obtained by Barco (2007) is the analytical expression of the downturn LGD

$$DLGD = \frac{q_\alpha \left[E(L_i | X_m, Y_m) \right]}{\Phi \left(\frac{\Phi^{-1}(\mu_i^{PD}) + \rho_i^{PD} \Phi^{-1}(\alpha)}{\sqrt{1 - (\rho_i^{PD})^2}} \right)} \tag{4}$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution.

The main drawback of the first approach of Basel II is that the mapping function between long-run and downturn LGD can change from model to model. This difference can have a significant effect on capital requirements. For example, Barco (2007), Miu and Ozedemir (2006) using the normal distribution underestimate tail contributions.

4. A new modelling approach for downturn LGD

We consider a dynamic behaviour of LGD over the economic cycle characterized by two distributions of LGDs held over expansion and recession periods. 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

$$F_{LGD}(y) = \pi F_{LGD/E}(y) + (1 - \pi) F_{LGD/R}(y) \quad 0 \le y \le 1 \tag{5}$$

where $F_{LGD/S}(y)$ 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). We assume that the parameter π is fixed over a given period of time.

In order to fit the model (5) to LGDs data, we first specify the cumulative distribution functions $F_{LGD/E}(\cdot)$ and $F_{LGD/R}(\cdot)$, which later we estimate jointly with the parameter π. Since the incidence of LGDs equal to 0 or 1 is high (Calabrese and Zenga, 2010; Renault and Scaillet, 2004; Schuermann, 2003), to supply accurate estimations for the extreme values, Calabrese (2012) proposes 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 $F_{LGD/S}(\cdot)$ conditional on the state S of the economic cycle is defined as

$$F_{LGD/S}(y) = \begin{cases} p_0^s & y = 0 \\ p_0^s + [1 - p_0^s - p_1^s]F_{B/S}(y) & y \in (0,1) \\ 1 & y = 1 \end{cases} \quad (6)$$

where $F_{B/S}(\cdot)$ 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.

Consistently with the hypothesis that LGD is a mixed random variable, we define the cumulative distribution estimator of LGD as

$$\hat{F}_{LGD}(y; \pi, \boldsymbol{\theta}, \boldsymbol{\sigma}) = \begin{cases} \dfrac{n_0}{n} & y = 0 \\ \dfrac{n_0}{n} + \left(1 - \dfrac{n_0}{n} - \dfrac{n_1}{n}\right)\hat{F}_M(y; \pi, \boldsymbol{\theta}, \boldsymbol{\sigma}) & y \in (0,1) \\ 1 & y = 1. \end{cases} \quad (7)$$

9

where $n_0 = \sharp\{LGD_i = 0\}$ and $n_1 = \sharp\{LGD_i = 1\}$ indicate the frequency of the LGDs equal to 0 and 1, respectively, and n is the number of LGDs in the portfolio.

The cumulative distribution estimator of the mixture M of two beta random variables is given by

$$\hat{F}_M(y; \boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{\sigma}) = \frac{1}{n - n_0 - n_1} \int_0^y \sum_{i=1}^{n-n_0-n_1} [\pi f_e(lgd_i; \theta_e, \sigma_e) + (1-\pi) f_r(lgd_i; \theta_r, \sigma_r)] d\, lgd_i$$

(8)

where π is the probability of the expansion regime. Moreover, $f_s(lgd; \theta_s, \sigma_s)$ is the probability density function of the beta random variable B_s (with $s = r, e$) of parameters $\frac{\theta_s}{\sigma_s} + 1$ and $\frac{1-\theta_s}{\sigma_s} + 1$

$$f_s(lgd; \theta_s, \sigma_s) = \frac{\Gamma\left(\frac{1}{\sigma_s} + 2\right) lgd^{\frac{\theta_s}{\sigma_s}} (1 - lgd)^{\frac{1-\theta_s}{\sigma_s}}}{\Gamma\left(\frac{\theta_s}{\sigma_s} + 1\right) \Gamma\left(\frac{1-\theta_s}{\sigma_s} + 1\right)},$$

(9)

where θ_s is the mode of the s-th beta density function and $\sigma_s > 0$ is the dispersion parameter. The parametrization of the beta density function $f_s(lgd; \theta_s, \sigma_s)$ is applied by Calabrese and Zenga (2010) in the nonparametric estimation. The main advantage of this parametrization is that every beta density function of the mixture M 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_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

$$V(B_s) = \frac{\sigma_s(\theta_s - \theta_s^2 + \sigma_s + \sigma_s^2)}{(1 + 2\sigma_s)^2(1 + 3\sigma_s)}$$

(10)

for $S = r, e$. To understand the influence of the parameter σ_s, in Figure 1 we plot the variance $V(B_s)$ as a function of σ_s by fixing θ_s. From this plot

10

we can deduce that the variance of B_s increases by increasing σ_s and with constant θ_s. An interesting result is that the limit of the variance $V(B_s)$, as σ_s tends to infinity, is 0.08, the variance of a uniform random variable with support the unit interval [0,1]. This outcome is due to the convergence of the beta density function (9) to a uniform density function as σ_s diverges to infinity.

Figure 1 around here

When σ_s goes to zero, we can apply Maclaurin series to equation (10) obtaining

$$V(B_s) = \sigma_s\theta_s(1 - \theta_s) + O(\sigma_s^2),$$

this means that by fixing θ_j the variance $V(B_s)$ tends to zero as σ_s tends to zero too, as Figure 1 shows. From this approximation we can also deduce that by fixing σ_s near to zero, the variance $V(B_s)$ becomes a parabola with maximum in $\theta_s = 0.5$.

In this paper we propose a method to estimate the unknown parameters in the mixture of two beta density functions. We use the Expectation-Maximization (EM) algorithm (Dempster et al., 1977) to estimate the parameters π and $\boldsymbol{\theta}, \boldsymbol{\sigma}$ in equation (8). For notation simplicity, we consider the parameter vector $\boldsymbol{\pi} = [\pi, 1-\pi]'$, so we aim at estimating the parameters vector $(\boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{\sigma})$.

In the mixture framework, the observed data $\boldsymbol{lgd} = (lgd_1, lgd_2, ..., lgd_{n-n_0-n_1})$ are completed with a component-label vector $\boldsymbol{Z} = (Z_1, Z_2, ..., Z_{n-n_0-n_1})$

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 (5), the complete-data log-likelihood function is

$$
\begin{aligned}
l_c\left(\pi, \boldsymbol{p}_0, \boldsymbol{p}_1, \boldsymbol{\theta}, \boldsymbol{\sigma}\right) &= \ln \pi \sum_{i=1}^{n-n_0-n_1} z_{ie} + \ln(1-\pi) \sum_{i=1}^{n-n_0-n_1} z_{ir} \qquad (11) \\
&+ \sum_{i=1}^{n-n_0-n_1} \sum_{S=e,r} z_{is} \ln f_{LGD/S}\left(lgd_i; \theta^s, \sigma^s\right).
\end{aligned}
$$

The EM algorithm (Dempster et al., 1977; McLachlan and Krishnan, 1997) maximizes the log-likelihood function (11). Each iteration of the EM algorithm includes the "E-step", which computes the conditional expectation of the complete-data log-likelihood (11) given the observed data $\boldsymbol{lgd}$, and the "M-step", which obtains the maximum of the complete-data log-likelihood function (11).

To compute the initial value $\boldsymbol{\pi}^{(0)}$, analogously to Ji et al. (2005) we assign the smallest 50% of lgd_i belonging to the interval (0,1) to the expansion component, and the highest 50% of lgd_i to the recession component. Maximizing (11), we obtain the initial values $(\boldsymbol{\pi}^{(0)}, \boldsymbol{\theta}^{(0)}, \boldsymbol{\sigma}^{(0)})$.

Afterwards, the algorithm[3] 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 (11). Since $\boldsymbol{Z}$ is non-observed data, $z_{is}^{(k)}$ is replaced by the conditional expectation

---

[3]The code in R language is available on request.

of Z_{is} given the observed data **lgd** and we consider the parameter estimates $(\boldsymbol{\pi}^{(k)}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\sigma}^{(k)})$ from the k-th iteration of the M-step

$$
\begin{aligned}
z_{is}^{(k+1)} &= E\left(Z_{is} \,\middle|\, lgd_i, \boldsymbol{\pi}^{(k)}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\sigma}^{(k)}\right) \\
&= P\left(Z_{is} = 1 \,\middle|\, lgd_i, \boldsymbol{\pi}^{(k)}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\sigma}^{(k)}\right) \\
&= \frac{\pi_s^{(k)} f_s\left(lgd_i; \theta_s^{(k)}, \sigma_s^{(k)}\right)}{\displaystyle\sum_{s=e,r} \pi_s^{(k)} f_s\left(lgd_i; \theta_s^{(k)}, \sigma_s^{(k)}\right)}
\end{aligned}
$$

with $i = 1, 2, ..., n - n_0 - n_1$ and $S = r, e$.

(2) On the $(k+1)$-th iteration, the **M-step** requires the maximization of the complete-data log-likelihood function (11) with respect to $\boldsymbol{\pi}$, $\boldsymbol{\theta}$, $\boldsymbol{\sigma}$ replacing z_{is} by $z_{is}^{(k+1)}$. Firstly, the elements of the updated estimates $\boldsymbol{\theta}^{(k+1)}$ are obtained as

$$
\widehat{\pi}_s^{(k+1)} = \frac{\displaystyle\sum_{i=1}^{n-n_0-n_1} z_{is}^{(k+1)}}{n - n_0 - n_1},
$$

with $S = r, e$.

To estimate $\boldsymbol{\theta}^{(k+1)}$ and $\boldsymbol{\sigma}^{(k+1)}$, we compute the score functions for $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}$ (see Appendix). The updated estimates $\boldsymbol{\theta}^{(k+1)}$ and $\boldsymbol{\sigma}^{(k+1)}$ are the solution of the following system

$$
\begin{cases}
\dfrac{\partial l_c\left(\boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{\sigma}\right)}{\partial \boldsymbol{\theta}} = \mathbf{0} \\[3mm]
\dfrac{\partial l_c\left(\boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{\sigma}\right)}{\partial \boldsymbol{\sigma}} = \mathbf{0}
\end{cases}
$$

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

13

The E-step and the M-step are alternatively repeated until the difference between two consecutive values of the complete-data log-likelihood (11) is negligible.

5. Empirical evidence

5.1. Data set

A comprehensive survey on the loan recovery process of Italian banks was conducted in the years 2000-2001 by the Bank of Italy. Its purpose was to gather information on the main characteristics of the Italian recovery process.

The survey is divided into three main sections. The first focuses on organizational aspects, the second regards credit recovery procedures and the last concerns analytical data on individual debts. About 250 banks were surveyed by means of a questionnaire. The database is composed of 149,378 defaulted borrowers and, since they covered nearly 90% of total domestic loans, this is the most recent and the most important survey on the Italian recovery process.

It is important to specify that data refer to individual loans which are privately held, and not listed on the market. In particular, loans are towards Italian resident debtors non-performing on the 31/12/1998 and entirely written off within the end of 1999.

The definition of default chosen in the survey is narrower than the one proposed by the BCBS (2004, paragraph 452). The difference is the inclusion of transitory non-performing debts. Finally, the Bank of Italy established the discount rates for each quarter from 1975 to 1999 by relying on the interest rates on short-term loans.

We highlight that to constrain LGD within the interval [0,1], we apply the expression proposed by Calabrese and Zenga (2010)

$$LGD = \frac{LOSS}{TE} = \frac{LOSS}{EAD + I + L}.$$

where $LOSS$ is the loss amount and TE denotes the Total Exposure, given by the sum of the Exposure At Default (EAD), the legal costs (L) and the interest on delayed payment (I).

Table 1 around here

Table 1 reports the main characteristics of the Bank of Italy's data. Our results are consistent with the values suggested by the Basel II Accord (BCBS, 2004 paragraph 287-8) under the foundation IRB approach: LGD is fixed at 45% for all senior unsecured debts and at 75% for all subordinated exposures. Moreover, Schuermann (2003) obtain similar values for positions with different seniority levels and instrument types. For the dispersion, the coefficient of variation (55.17% of the average LGD) is similar to the results obtained by Schuermann (2003) (53.63% of the average recovery rate) and by Renault and Scaillet (2004) (60.31% of the average recovery rate) on corporate bonds.

Figure 2 around here

Figure 2 shows the LGD distribution. 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%). Many authors obtain

a bimodal LGD distribution for different countries, for example, Araten et al. (2004) and Asarnow and Edwards (1995) for the US, Hurt and Felso-valyi (1998) for Latin America, Franks et al. (2004) for France, Grunert and Weber for Germany (2005) and Dermine and Neto de Carvalho (2006) for Portugal. For instance, in Friedman and Sandow's (2003) study about 10% and 20% of the observations showed recovery rates respectively equal to zero and one.

5.2. Estimation results and comparative analysis

Table 2 around here

In order to estimate the downturn LGD, we apply the methodology proposed in this work and the four methods analysed in Section 2 and 3 to the Bank of Italy's database. The results[4] are reported in Table 2. Even if the database concerns loans, our proposal is also applicable to bonds.

The main advantage of the methodology here proposed is that it allows to estimate the downturn LGD distribution, shown in Figure 3, without the arbitrary definition of an economic downturn period or a mapping function of long-run LGD. Moreover, the high concentration of data at total recovery and total loss (Calabrese and Zenga, 2010; Renault and Scaillet, 2004; Schuermann, 2003) can be replicated by the model suggested in this work.

Figure 3 around here

The approach proposed in this paper is compared with the four models analysed in Section 2 and 3. For the model suggested by Sabato and Schmid

---

[4]To obtain these results we apply the algorithm in Appendix.

(2008) (*SS model*) we need to estimate the LGD stressing factor defined in equation (2). Following the authors' suggestion, the LGD stressing factor is given by the sum of the mean ($\mu^{PD} = 0.0063$) and the standard deviation ($\sigma^{PD} = 0.0055$) of the default rates. These results are obtained from data of the International Monetary Fund by assigning to each loan the annual default rate evaluated at the time of default. Furthermore, for both the SS model and the one proposed by the Federal Reserve System (*FED model*), we consider the average LGD (0.6154) from Table 1.

For the factor model suggested by Barco (2007) we consider $\alpha = 0.999$. We specify that from equation (4) we should estimate μ_i^{PD} and ρ_i^{PD} for each loan. Since these data are not available in the Bank of Italy's survey, we assign to each loan the annual PD evaluated at default time and supplied by the International Monetary Fund. This means that we assign the same PD to loans defaulted the same year. Moreover, we compute the correlation coefficient ρ^{PD} of the assigned PDs and, for lack of data, it is assumed constant for all observations. Finally, compliant to Basel II adverse conditions are identified as a negative GDP growth rate, occurred only in 1993. This means that in the second approach of Basel II (*AC model*) the downturn LGD is given by the mean of 12,992 LGDs defaulted in 1993.

Furthermore, we apply the methodology suggested in this paper and the obtained estimates are reported in Table 2. If we focus our attention only on the recession distribution, this represents the downturn LGD distribution and it is shown in Figure 3. We compute the average of this distribution and we report it in Table 3. The downturn LGDs computed by the methods outlined in Section 2 and 3 are also reported in Table 3.

To understand how we should interpret the results in Table 3, we underline that for a bank the underestimation of the downturn LGD is more

17

risky than the overestimation. The first important result of this comparison is that the approach suggested by Basel II and III of using estimates during adverse conditions strongly underestimates the downturn LGD. We can deduce it because the estimate of the downturn LGD 0.6370 is very close to the long-run default-weighted average loss rate $\mu^{LGD} = 0.6154$. This result could be due to the long-term influence of the GDP growth rate on LGDs, coherent with Calabrese (2012) and Bellotti and Crook (2009). In this way, we would show the arbitrariness of identifying adverse conditions.

Even using a mapping function, as SS and FED models do, the estimate of downturn LGD are 8.08% and 5.60% lower, respectively, than the estimate obtained by our proposal. The FED model shows the clear drawback of using a mapping function with fixed parameters. Even if these parameters are estimated in the SS model, we obtain a downturn LGD lower than that of the FED model. We presume that this result could be due to the lack of an accurate method to estimate the LGD stressing factor LGDSF defined in equation (2). This means that also this method is subject to arbitrariness.

Finally, Barco's model supplies a downturn LGD a bit lower than that of our proposal. This means that Barco's model overcomes the drawback of the underestimation of downturn LGD but it needs some information (i.e. μ_i^{PD} and ρ_i^{PD}) for each loan that are often not available. We highlight that we do not need this information or any kind of additional information to apply our model.

From this empirical analysis we can conclude that, on the one hand, the model proposed in this paper overcomes the drawback of arbitrariness of some methods used in the literature. On the other hand, it is an accurate model that allows to represent the high concentration of data at total loss. In this way it overcomes the drawback of the underestimation of the downturn

LGD, without using any additional information on borrowers.

Different types of portfolios (e.g. for bonds by Renault and Scaillet (2004) and Schuerman (2003)) and different geographic areas (e.g. Araten et al. (2004) for US, Grunert and Weber for Germany (2005)) show a high concentration of data at total loss, we could expect also for these datasets that our proposal outperforms the methods used in the literature. In conclusion, our proposal could supply accurate estimate of the downturn LGD for different types of portfolios with different geographic locations.

Table 2 around here

6. Concluding remarks

This paper aims at proposing an accurate method to estimate the downturn LGD distribution to compensate for the fact that the Vasicek regulatory capital model does not incorporate systematic LGD. To represent the dynamic behaviour of LGD over the economic cycle, this paper proposes that the LGD distribution is given by a mixture of an expansion and a recession distributions. To represent the high concentration of data at total loss and total recovery, we suggest an accurate parametric model for each state of the economic cycle and a method to estimate their parameters.

The main advantages of this proposal are that banks do not need to identify arbitrarily downturn conditions and additional information on the default risk for each borrower. Finally, we apply our proposal to a comprehensive dataset of Italian bank loans. The main result of this empirical analysis is that our model can replicate the high concentration of data at total loss, unlike the other analysed models, so it overcomes the drawback of underestimating the downturn LGD. Our hope is to provide a framework

19

that banks can use based on their internal data to estimate downturn LGD for different portfolios.

References

Abramowitz M., Stegun I. A., 1972. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover, New York.

Altman E. I., Sabato G., 2005. Effects of the new Basel Capital Accord on bank capital requirements for SMEs, Journal of Financial Services Research, 28, 15-42.

Altman E. I., Resti A., Sironi, A., 2005. Loss given default: A review of the literature. In: Altman E. I., Resti A., Sironi A. (Eds.), The Next Challenge in Credit Risk Management. Riskbooks: London.

Araten M., Jacobs Jr. M., Varshney P., 2004. Measuring LGD on commercial loans: An 18-year internal study. Journal of Risk Management Association 4, 96-103.

Asarnow E., Edwards D., 1995. Measuring loss on default bank loans: A 24-year study. Journal of Commercial Lending 77, 11-23.

Bank of Italy, 2001. Principali risultati della rilevazione sull'attivitÃ di recupero dei crediti. Bollettino di Vigilanza 12, December.

Barco M., 2007. Going downturn. Risk Australia, Spring, 39-44.

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.

Basel Committee on Banking Supervision, 2005b. Working Paper No. 14 Studies on the Validation of Internal Rating Systems. Bank for International Settlements: Basel, September.

Basel Committee on Banking Supervision, 2005c. An Explanatory Note on the Basel II IRB Risk Weight Functions. Bank for International Settlements: Basel, September.

Basel Committee on Banking Supervision, 2011. Basel III: A global regulatory framework for more resilient banks and banking systems. Bank for International Settlements: Basel, June.

Bellotti T., Crook J. (2009) Loss Given Default models for UK retail credit cards. Credit Research Centre, working paper.

Board of Governors of the Federal Reserve Bank, 2006. Basel II Capital Accord-Notice of Proposed Rulemaking. Federak Reserve Bank, September.

Bruche M., González-Aguado C., 2008. Recovery rates, default probabilities and the credit cycle. Working paper, CEMFI.

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

Chalupka R. and Kopecsni K. (2009) Modelling Bank Loan LGD of Corporate and SME Segments: A Case Study. Czech journal of Economics and Finance, 59(4), 360-382.

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.

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

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

Franks J., de Servigny A. and Davydenko, S. (2004) A comparative analysis of the recovery process and recovery rates for private companies in the U.K., France and Germany, Standard & Poors Risk Solutions Report, 1-93.

Friedman C., Sandow S., 2003. Ultimate recoveries. Risk 16, 69-73.

Gaspar R. M., Slinko I., 2005. Correlation Between Intensity and Recovery in Credit Risk Models. Working Paper Series in Economics and Finance 614, Stockholm School of Economics.

Gordy M. B., 2003. A risk-factor model foundation for ratings-based bank capital rules, Journal of Financial Intermediation, 12(3), 199-232.

Grunert J., Weber M., 2009. Recovery rate of commercial lending: Empirical evidence for German companies. Journal of Banking and Finance 33, 505-513.

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.

Hurt L., Felsovalyi A., 1998. Measuring loss on Latin American defaulted bank loans, a 27-year study of 27 countries. The Journal of Lending & Credit Risk Management. (October), 41-46.

Jarrow R., Lando D., Turnbull S., 1997. A Markov model for the term structure of credit risk spreads. Review of Financial Studies, 481-523.

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

McLachlan G., Krishnan T., 2007. The EM algorithm and extensions. John Wiley & Sons, New York.

Miu P., Ozdemir B., 2006. Basel requirements of downturn LGD: modeling and estimating probability of default and loss given default correlations. Journal of Credit Risk 2(2), 43-68.

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.

Sabato G., Schmid M., 2008. Estimating conservative loss given default. Swiss Institute of Banking and Finance, University of St. Gallen working paper.

Saurina J., Trucharte C., 2007. An Assessment of Basel II Procyclicality in Mortgage Portfolios. Journal of Financial Services Research, 32, 81-101.

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

Vasicek, O., 2002. Loan portfolio value. Risk (December), 160-162.

7. Appendix

In this appendix, we obtain the score functions for $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}$. The notation used here is defined in Section 4. The derivatives of the complete-data log-likelihood function (11) with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}$ are

$$\frac{\partial l_c\left(\pi, \boldsymbol{p}_0, \boldsymbol{p}_1, \boldsymbol{\theta}, \boldsymbol{\sigma}\right)}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n} \sum_{S=e,r} z_{is}^{(k+1)} \frac{\partial \ln g_s\left(lgd_i; \theta_s, \sigma_s\right)}{\partial \boldsymbol{\theta}}$$

$$\frac{\partial l_c\left(\boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{\sigma}\right)}{\partial \boldsymbol{\sigma}} = \sum_{i=1}^{n} \sum_{S=e,r} z_{is}^{(k+1)} \frac{\partial \ln g_s\left(lgd_i; \theta_s, \sigma_s\right)}{\partial \boldsymbol{\sigma}}.$$

To compute the score functions, we consider the following results

$$\frac{\partial \ln g_s\left(lgd_i; \theta_s, \sigma_s\right)}{\partial \theta_s} = \frac{1}{\theta_s} \left\{ \left[\psi\left(\frac{1-\theta_s}{\sigma_s}+1\right) - \psi\left(\frac{\theta_s}{\sigma_s}+1\right) \right] + \ln\left(\frac{lgd_i}{1-lgd_i}\right) \right\}$$

and

$$\frac{\partial \ln g_s\left(lgd_i; \theta_s, \sigma_s\right)}{\partial \sigma_s} = \frac{1}{\sigma_s^2} \left\{ \left[\theta_s \psi\left(\frac{\theta_s}{\sigma_s}+1\right) + (1-\theta_s)\psi\left(\frac{1-\theta_s}{\sigma_s}+1\right) \right] + \right.$$
$$\left. -\psi\left(\frac{1}{\sigma_s}+2\right) \right\} - \ln \frac{(lgd_i)^{\theta_s}}{(1-lgd_i)^{1-\theta_s}}$$

where $\psi(\cdot)$ is called digamma function (Abramowitz and Stegun, 1972, p. 380) $\psi(z) = \dfrac{\partial log\Gamma(z)}{\partial z} = \dfrac{\partial \Gamma(z)}{\Gamma(z)}$ for $z > 0$.

	Loss	Total Exposure	LGD
Mean	0.1319	0.2275	0.6154
Median	0.0067	0.0133	0.6667
Standard deviation	4.4815	14.0504	0.3395
Minimum	0	$\simeq 0$	0
Maximum	1212.3954	5135.7263	1
I quartile	0.0020	0.0045	0.3333
II quartile	0.0067	0.0133	0.6667
III quartile	0.0233	0.0429	0.9545

Table 1: Some characteristics of the Bank of Italy's data set. The amounts "Loss" and "Total Exposure" are expressed in Thousands of Euros.

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

Table 2: Estimates obtained by the mixture model from the Bank of Italy's data.

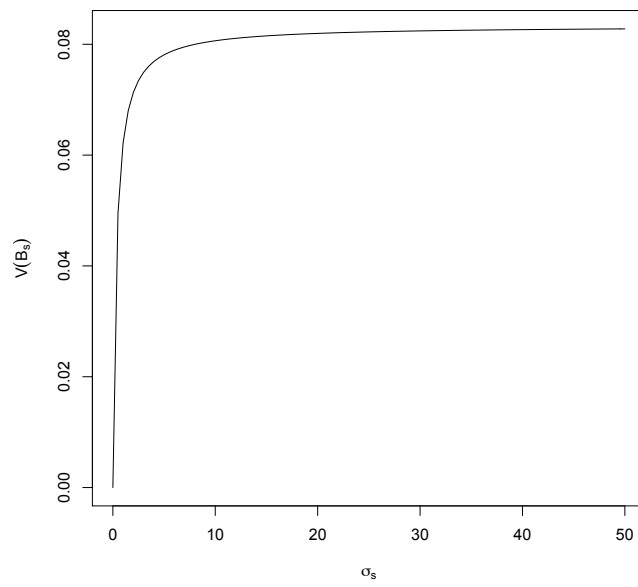|	Mixture model	Barco's model	FED model	SS model	AC model
Downturn LGD	0.7823	0.7673	0.7385	0.7191	0.6370

Table 3: Average downturn LGD

Figure 1: The variance $V(B_S)$ of the beta random variable B_s as a function of the parameter σ_S with $\theta_S = 0.5$.
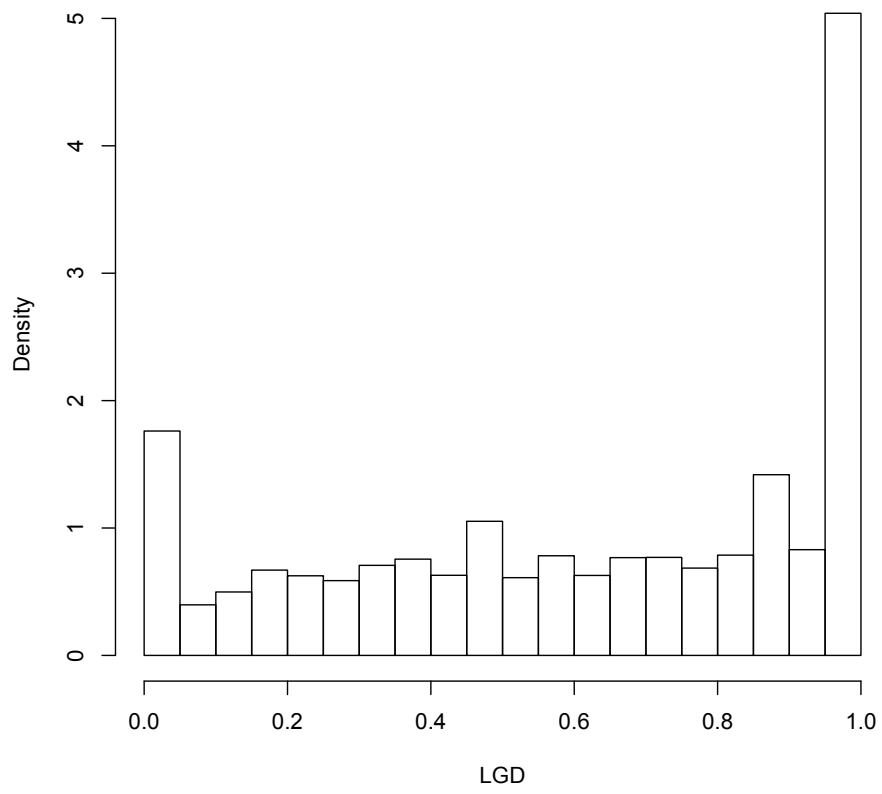
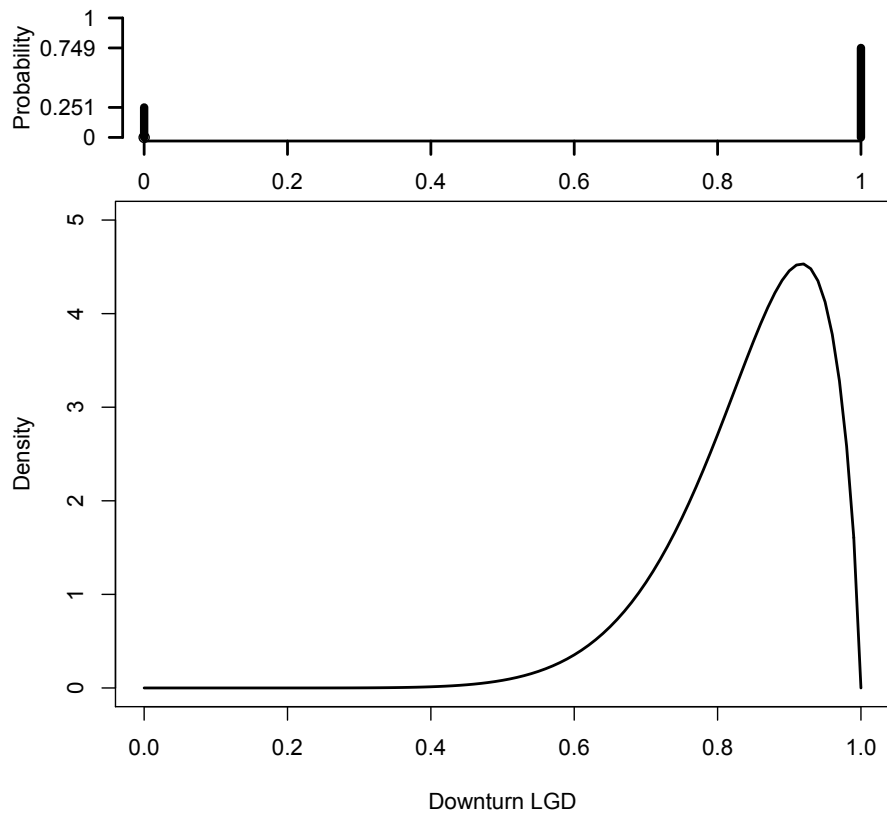Figure 2: The LGD distribution of the Bank of Italy data set.

Figure 3: Estimate of downturn LGD distribution from the Bank of Italy data set. The upper part of the plot is the estimate of the extreme values 0 and 1 (the discrete component). The lower part of the plot is the estimate of the values belonging to the interval (0,1) (the continuous component).