Internal LGD Estimation in Practice

Peter Glößner, Achim Steinbauer, Vesselka Ivanova
d-fine 28 King Street, London EC2V 8EH, Tel (020) 7776 1000, www.d-fine.co.uk

1 Introduction

Driven by a competitive market and motivated by the new Basel Capital Accord (Basel II), banks have put a lot of effort into development and improvement of their methods to assess the creditworthiness of their obligors and to deduce the probability of default (PD). However, not only the probability of default but also the economic loss in the case of default have to be estimated to quantify credit risk and to calculate the Basel II capital requirements under the advanced approach.

Basel II (2004) decomposes the estimated economic loss in the event of default into two factors: the exposure at default (EAD) and the economic loss relative to this exposure. The latter ratio is called *loss given default* (LGD). From the Basel point of view, both values have to be estimated individually for each claim treated under the advanced IRB approach. Observe that this estimation takes place prior to a possible default. LGD and EAD may therefore be interpreted as estimated expectations of random variables.

In the following we focus on the estimation of LGD. While the notion of LGD is easy to understand, there are numerous difficulties that make LGD estimation a demanding task in practice.

- In many asset classes, the economic loss of a defaulted claim is generally not realized immediately just after default. It rather may take years until it turns out whether the transaction re-performs or else realizes an economic loss (as made manifest by writing it off completely).
- The data base is much smaller than the one used for estimating PD: for PD estimation, the "denominator of the relative frequencies" is the whole of all claims within the respective asset class, while when estimating LGD, just the (hopefully) small number of defaults makes up the totality on which to work.

The most widely used approach relies on the segmentation of observed defaults from the past years, followed by the calculation of average loss rates within each segment. This average loss rate is used as LGD-estimate for all current performing claims associated with the respective segment.

While convincingly simple in theory, this approach becomes impractical if the data basis of worked out defaulted claims is small, since the segmentation employed imposes further splits of already scarce data. This issue is aggravated by the need to isolate data from economic downturn conditions to calculate a *downturn LGD* compliant with §468 of the Basel II framework (cf. 2005).

In the following we introduce a method that largely avoids these difficulties. It is based on a LGD-score that represents a relative order of expected loss severity and a calibration process that takes into account all of the internal loss history and reduces further splitting of this data basis.

2 LGD Score

While LGD itself is an estimate of the expected loss rate in case of default, the LGD score proposed here only aims to reflect the relative **order** of expected loss rates. All of the relevant information, that would otherwise be used for a segmentation of claims, condenses in this LGD score. The gap to absolute LGD estimates is closed by the calibration procedure (described in Section 3) that takes into account the internal loss history.

Just like PD, the estimated LGD is associated with a fixed time horizon, i.e. it refers to the expectation value of **all** future economic losses of the respective claim **given** a default occurs within the **following** year (while the losses may substantiate much after). This is important since for a given claim, LGD score and LGD estimate my depend on attributes that change over time–think for example of expected recoveries on real estate collateral or unsecured claims.

For retail and commercial loans, expected loss rates mainly depend on collateralization and expected recovery on uncollateralized exposure. A simple score function would therefore be of the form

$$\text{score} = 1 - \left(1 - \frac{c \cdot (1-h)}{\chi}\right) \cdot l_{uc}, \tag{1}$$

where χ is the estimated exposure at default, c is the market value of collateral, h is an appropriate haircut on the collateral value ($h \in [0,\ 1]$), and l_{uc} is the estimated loss rate on uncollateralized exposure ($l_{uc} \in [0,\ 1]$). In general, h and l_{uc} will be modeled as functions themselves. The haircut will primarily depend on the type of collateral and (secondarily and depending on the type of collateral) on the state of the economy. The loss rate on uncollateralized exposure is influenced by the properties of the obligor (ability to service debt from cash flows and assets other than collateral, legal form, jurisdiction, region, industry sector etc.), the seniority of the claim and the state of the economy (recoveries on uncollateralized claims can be significantly lower under economic downturn conditions). See Julian Franks et al. (2004) for empirical results and Gupton and Stein (2005) for details on a specific LGD scoring method.

The details of the score function (attributes and functional form of h and l_{uc}) will be based on statistical analysis as well as expert judgment. Although the score described above apparently allows for the interpretation as an recovery estimate itself, it will only be used as means of relative ordering of expected loss rates. This way, any distortion of the LGD-score can be compensated in the calibration process. The latter ensures that the estimated average loss rate for all claims with equal score is close to the average observed loss rate for similar claims in the past.

3 Calibration

Calibration of a LGD score is the process of assigning absolute LGD estimates to score values based on the observed loss history. In this section we propose a calibration procedure that draws its stability form a separation into four main building blocks: The LGD score, the distribution of exposures across this score, the calibration of LGD in units of the average portfolio loss rate, and the estimation of the average portfolio loss rate itself. This separation largely avoids the otherwise necessary segmentation of default data into small homogeneous subsets.

3.1 Distribution of Exposure and Loss Across Scores

Here and in the following we will refer to the set of all claims defaulted in a certain year in the past as *(default) cohort*. The empirical distribution of the exposure at default across the values of the LGD-score can easily be calculated for each cohort. In general, these distributions will differ between cohorts.

Suppose that m claims of a certain cohort have so far been resolved, and let $\hat{\chi}_i$, $\hat{\ell}_i$ and $\hat{s}_i$ denote the observed exposures, losses and LGD scores, respectively. Assume that the LGD-score (after suitable normalization) takes integer values between $s_{\min}$ and $s_{\max}$ and let $X(s)$ and $L(s)$ denote the sum of exposures and losses for all defaulted claims up to LGD-score s, i.e.

$$X(s) := \sum_{\{i|\hat{s}_i \le s\}} \hat{\chi}_i \quad \text{and} \quad L(s) := \sum_{\{i|\hat{s}_i \le s\}} \hat{\ell}_i.$$

Then, $\Phi_x(s) := X(s)/X(s_{\max})$ is the distribution of exposure and $\Phi_l(s) := L(s)/L(s_{\max})$ the distribution of loss across scores.

Since Φ_x and Φ_l are calculated from the **resolved** (i.e. *decided*) claims of the respective cohort, both quantities may evolve over time until **all** members of the cohort have been resolved. However, in many practical cases this time dependence vanishes for the calibration mapping calculated from the (time dependent) distributions Φ_x and Φ_l as discussed in the following section. For ease of notation we therefore omit the explicit indication of time dependence.

3.2 Calibration Mapping

Let $LR(s)$ denote the observed loss rate for claims with score s, i.e. for $\Phi_x(s) \neq \Phi_x(s-1)$,

$$LR(s) := \frac{\Phi_l(s) - \Phi_l(s-1)}{\Phi_x(s) - \Phi_x(s-1)} \cdot \frac{L(s_{\max})}{X(s_{\max})}. \tag{2}$$

If the points

$$(\Phi_x(s),\ \Phi_l(s)),\ \ s = s_{\min}, \ldots, s_{\max} \tag{3}$$

are approximated by a differentiable function

$$F : [0,\ 1] \to [0,\ 1],\ \ z \mapsto F(z) \text{ with } F(0) = 0 \text{ and } F(1) = 1,$$

then $F'(\Phi_x(s))$ can be used as an approximation of the left hand side difference quotient

$$\frac{\Phi_l(s) - \Phi_l(s-1)}{\Phi_x(s) - \Phi_x(s-1)}$$

in equation (2) and $LR(s)$ can be approximated by

$$LR(s) = F'(\Phi_x(s)) \cdot \frac{L(s_{\max})}{X(s_{\max})}. \tag{4}$$

The function F (and consequently its derivative F') turns out to be a very robust object in many practical cases. In particular, F does in these cases not depend on the specific cohort the observations $\hat{\chi}_i$ and $\hat{\ell}_i$ come from. In this sense, the functional form of F is stable over time. All of the time dependence is moved to the score s and the average portfolio loss rate $L(s_{\max})/X(s_{\max})$.

This draws the attention to the adequate estimation of the average portfolio loss rates.

3.3 Estimating the Average Portfolio Loss Rate

The loss rate $L(s_{\max})/X(s_{\max})$ in equation (2) is the exposure weighted average loss rate in the data basis. If F is independent of the cohort used as data basis, data from different cohorts can be aggregated in the estimation of F, but cohorts should still enter separately into the calculation of the average portfolio loss rate.

Since the workout period for defaulted claims can average at years, the calculation of the average loss rate from a specific cohort will always be restricted to the *decided* fraction of claims. To quantify the uncertainty introduced by still undecided claims, it can be useful to plot the cumulative exposure versus the cumulative loss of **decided claims** for each cohort from the year of default until today. The resulting curves can be extrapolated to give an indication of the cumulative loss expected for **all** claims in this cohort. Natural bounds for this quantity are given by the assumption of loss rate either *zero* or *one* for still undecided claims. For the interpretation of the plot it is advantageous to express all amounts as fractions of the overall exposure of the cohort. Figure 1 gives an illustrative example.

4 Example

In the following we illustrate the calibration procedure described in Section 3 on basis of real data from a retail portfolio. Our sample data set consists of 806 retail claims, defaulted in 1999 and 2000 with an overall exposure at default of EUR 184.6mn. During the observation period until end 2003, a total of 312 claims has *reperformed*, 126 have been finally *charged off* with economic loss, and 368 claims have remained *undecided*. The average length of the workout period is estimated to be five years.

4.1 Distribution of LGD-Score, Exposure and Economic Loss

Prior to default, an LGD-score between 0 and 100 had been assigned to each claim, based on equation (1). This score is not supposed to be optimal, but serves for illustration purposes. Figure 2 shows the distribution of exposures and losses across the values of the LGD-score for all claims from the 1999- and 2000-cohort that where *decided* at the end of 2003.

Figure 3 shows the distributions $\Phi_x(s)$ and $\Phi_l(s)$ of exposure and economic loss across the values of the LGD-Score for the *decided* claims from the 1999- and 2000-cohort.

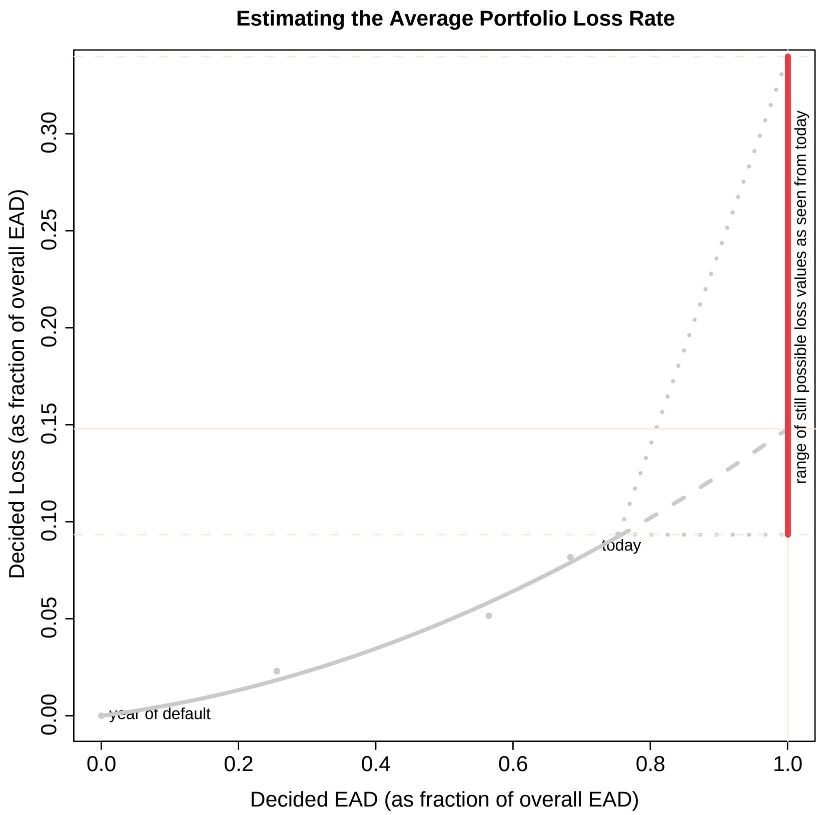

Figure 1: Estimating the average portfolio loss rate

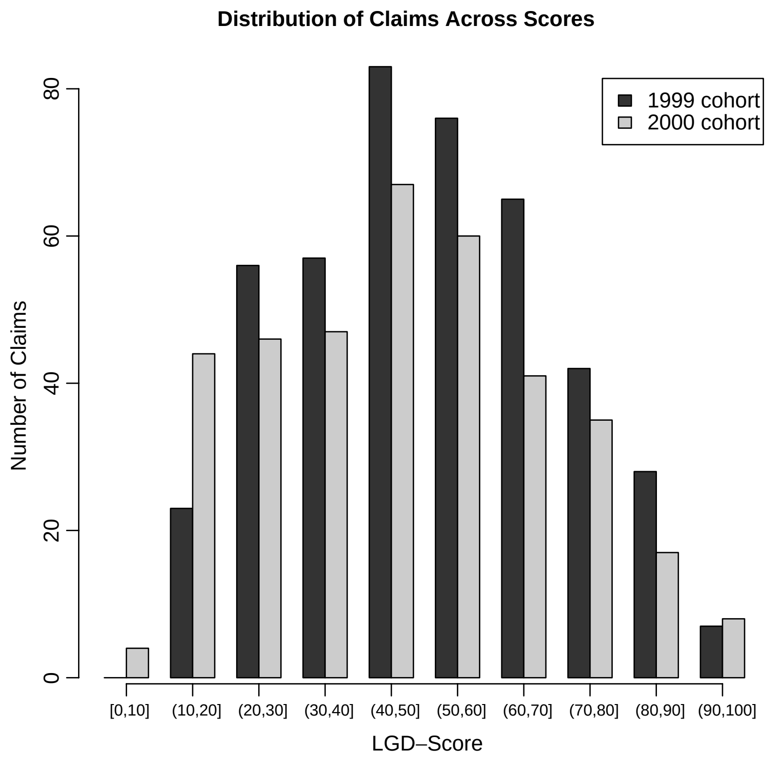

Figure 2: Distribution of defaulted claims across LGD-Scores

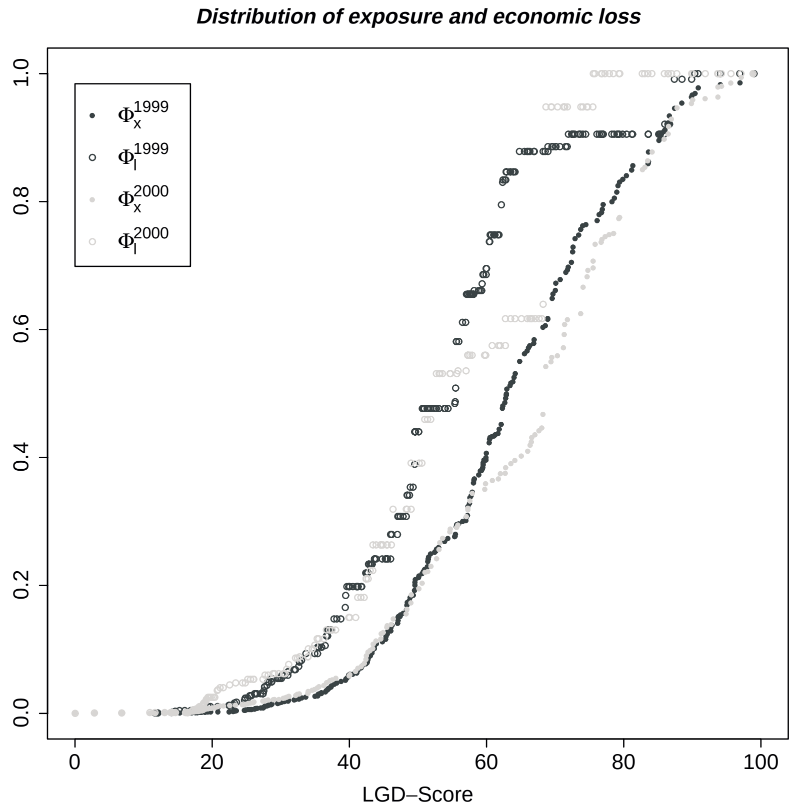

Figure 3: Distribution of exposure and economic loss across LGD-Scores

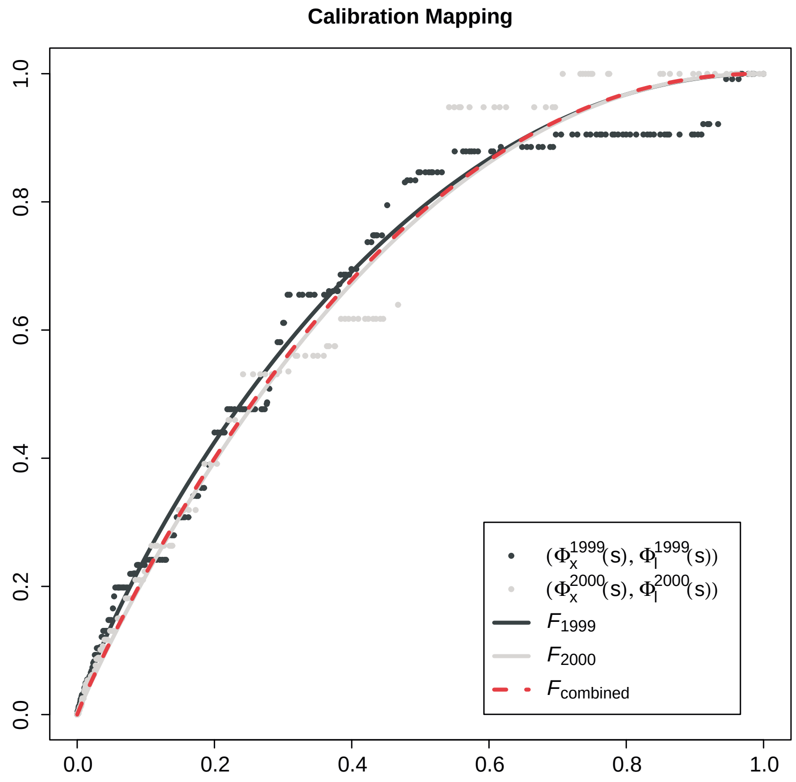

Figure 4: Calibration mapping

4.2 Calibration Mapping

The calibration mapping as described in Section 3.2 is based on a differentiable approximation of the set $(\Phi_x(s), \Phi_l(s))$ with $s = 0, \ldots, 100$. Figure 4 shows these sets for the 1999- and 2000-cohort as well as the respective approximations F_{1999} and F_{2000}. These approximations are chosen as Beta-distribution with suitable parameters[1] and turn out to be nearly identical for both cohorts. This fact confirms the stability assumption from Section 3.2 and justifies to merge both cohorts for the determination of a joint approximation $F_{combined}$. The possibility to merge cohorts in this step is the great advantage of the proposed method. Instead of subsequent splitting of rare data into homogeneous pools, the calculatory object used here allows to use all available data at once.

In order to deduce a score-specific LGD estimate, the derivative f of F has to be evaluated at $\Phi_x(s)$. Figure 5 shows the results for $F_{combined}$ in conjunction with Φ_x^{1999} and Φ_x^{2000}. While the cumulative distributions $\Phi_x(s)$ for the cohort years 1999 and 2000 turn out to be very similar in our case, such an effect is generally not to be expected. The distribution of the defaulted claims subportfolio over the score axis can shift over time, or at least show random fluctuations. The question arises, which distribution to take for the quantity $F'(\Phi_x(s))$ in our formula (4) for $LR(s)$? The choice should take into

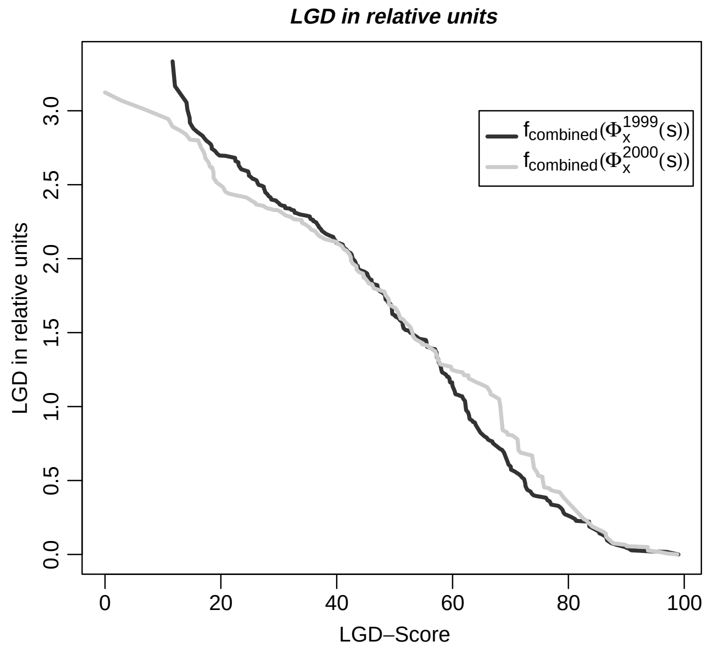

Figure 5: LGD in units of the average portfolio loss rate

account that the target function $LR(s)$ is meant as a predictor for the future and at the same time has to be a conservative estimator. Possible ways towards finding a suitable $\Phi_x(s)$ are

- Analyze the subportfolio distributions over the score s for the last cohorts. If there is evidence for stability (as in our case) or a trend can be discerned, this should be used.
- Instead of a *backward looking* choice for Φ_x, it might be adequate for the prediction of LGD to use the *expected* distribution of defaulted exposures across LGD scores, i.e the distribution of $PD \cdot EAD$ for the current (performing) portfolio.

In either case, the result is a mapping of the LGD score to an LGD estimate in units of the average portfolio loss rate.

According to equation (4), it is now this average portfolio loss rate $L(s_{\max})/X(s_{\max})$ that has to be supplied to complete the function LR. This quantity, though, is likely to be volatile in time and has to be estimated from the available loss history in accordance with the Basel II requirement of a *downturn LGD* (cf. 2005). From the decided claims of the 1999- and 2000-cohorts we calculate the average portfolio loss rates $\overline{LR}_{1999} = 12.4\%$ and $\overline{LR}_{2000} = 16.3\%$, respectively. Figure 6 shows the corresponding LGD-estimates as functions of the score s.

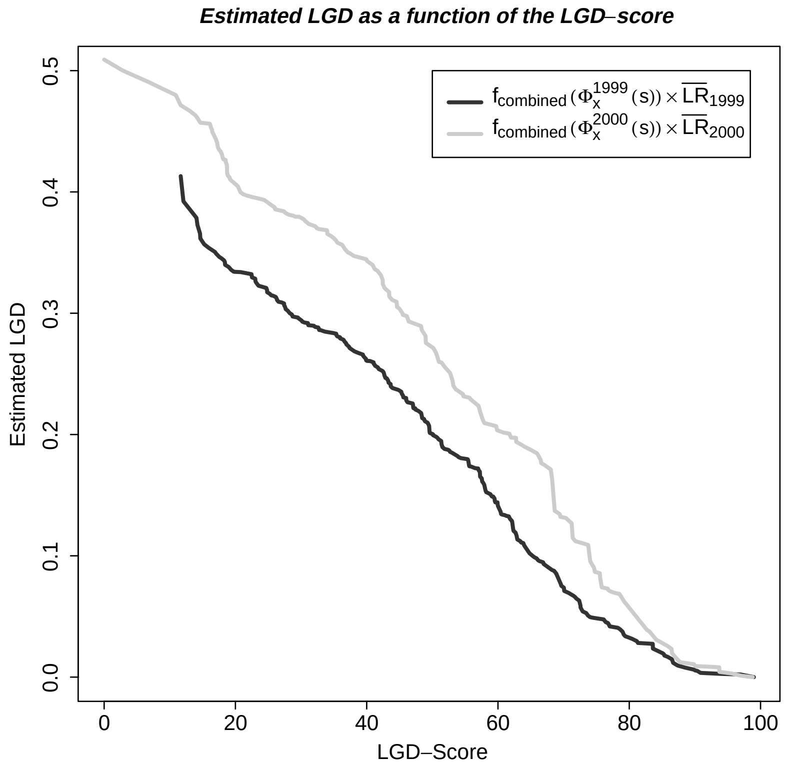

Figure 6: LGD in units of the average portfolio loss rate

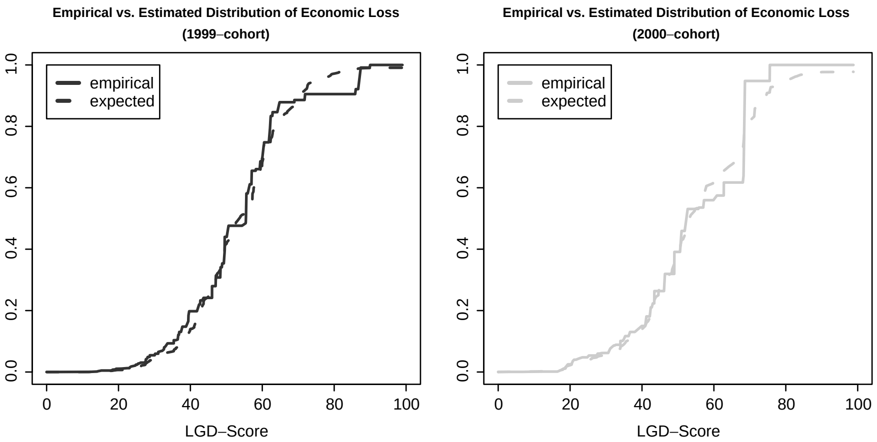

Figure 7: Empirical vs. expected distribution of economic loss

To assess the accordance of the calculated LGD-estimate with the underlying data, we compare the empirical distribution $\Phi_l(s)$ of economic loss (as of end 2003) with the corresponding expected distribution $\tilde{\Phi}_l(s) = \sum_{\{i|s_i \leq s\}} LGD(\hat{s}_i) \cdot \hat{\chi}_i$, where $\hat{s}_i$ and $\hat{\chi}_i$ denote the LGD-score and exposure of the i-th decided claim in the respective cohort and $LGD(.)$ denotes the LGD-estimate from Figure 6. The results displayed in Figure 7 show a good accordance of empirical (observed) and expected (calculated) cumulative distributions of loss across scores.

5 Conclusion

LGD estimation is a demanding task that so far has mainly been tackled by calculating average loss rates on sub-segments of observed losses. Instead of segmentation of observed loss data, we recommend to introduce an LGD score which, in a second step, is calibrated to the internal loss history. The cornerstone of the calibration procedure is a calibration mapping that shows to be a very robust object in practice. This offers the possibility to merge data from different cohorts and to thereby improve accuracy, without losing the cohort-dependent time resolution of the losses' realization process. An example with real world data proves practicability and exactness of the calibration procedure.

■ *Dr Peter Glößner is Manager at d-fine, the consulting firm. His client work at major banks has for the past five years mostly been focused on credit risk and Basel II related topics such as IRB modelling and validation, Basel II risk parameter identification and estimation, and credit portfolio modelling. His co-authors Dr Achim Steinbauer and Dr Vesselka Ivanova are Senior Consultants at d-fine and are advising banks on credit risk-related topics similarly to Peter.*

FOOTNOTE & REFERENCES

1. The choice of a Beta-distribution as means of the parametric fit is not intrinsic, but common due to its flexibility (cf. Gupton and Stein (2005) and Dirk Tasche (2004)).

■ Basel Committee on Banking Supervision. *International Convergence of Capital Measurement and Capital Standards*. Bank for International Settlements, June 2004.
■ Basel Committee on Banking Supervision. *Guidance on Paragraph 468 of the Framework Document*. Bank for International Settlements, July 2005.
■ Dirk Tasche. *The single risk factor approach to capital charges in case of correlated loss given default rates*. February 2004 (working paper).
■ Greg M. Gupton, Roger M. Stein *LossCalc V2: Dynamic Prediction of LGD: Modeling Methodology*. Moody's KMV Company, January 2005.
■ Iain Maclachlan *Choosing the Discount Factor for Estimating Economic LGD*. May 2004 (working paper).
■ Julian Franks, Arnaud de Servigny, and Sergei Davydenko *A Comparative Analysis of the Recovery Process and Recovery Rates for Private Companies in the U.K., France, and Germany*. Risk Solutions, Standard & Poors, May 2004.
■ Michael Araten, Michael Jacobs Jr., and Peeyush Varshney *Measuring LGD on Commercial Loans: An 18-year Internal Study*. RMA Journal, May 2004.
■ Til Schuermann. *What do we know about loss given default?*. Wharton Financial Institutions Center, January 2004 (working paper).
■ Vesselka Ivanova *LGD-Rating for a Portfolio of Retail Loans*. University of Oxford, 2004.