Rating model calibration: a modern application

Stefano Bonini
University of Rome, Tor Vergata
bonini.stefano@gmail.com

Giuliana Caivano
University of Rome, Tor Vergata
giuliana.caivano@gmail.com

Abstract: An extensive academic and practitioner’s literature exists on rating models development with well-structured statistical methods, however these models rank the counterparties but not the Probability of Default, then they have to be calibrated according to the economic scenario. During the last years the effect of not well calibrated models has been observed on Credit Lending: usually, in fact, they show a high level of pro-cyclicality that let them lose market credibility and banking usability. The aim of this paper is to present a modern structured calibration approach that takes into consideration specific economic factors. The calibration approach has been finally applied on real data of a Corporate portfolio of a top tier European Bank (under Single Supervisory Mechanism) and a new calibration test, adjusted by the economic cycle, has been performed.

Keywords: Rating Models, Credit Risk Modeling, Economic Cycle, Model Calibration, Quantitative Finance, Binomial Test

JEL Reference: C13, C51, C52, G01, G24, G32

1 Corresponding author
1. Introduction

Today, even because of the financial crises, banks need more and more reliable and usable risk management tools. Moreover within Basel2 and Basel3 Accord the estimation of Probability of Default (PD) is a key step in switching towards Internal Ratings Based Approach. PD models play a key role for an efficient allocation of capital, pricing, credit and client sanctioning, credit monitoring, and finally regulatory compliance. Researcher and practitioners are engaged in development more and more sophisticated rating models that have to show the true picture of the portfolio in present as well as future scenarios [1].

A typical feature of PD models across countries is that they are often based on individual characteristics of clients or they use some information related to the specific credit products, but no information are commonly used for taking into account macroeconomic variables. The only way to align rating models with the economic scenario (as in [2] and [7]) is to apply a sort of “addendum” to the model itself: that is what is commonly known as calibration. When speaking about calibration, it is quite common among banks to refer and distinguish between Point-in-Time (PIT) and Through-the-Cycle (TTC) PDs. While the PIT PDs reflect the probability of going into default given the current point of economic cycle, the TTC PDs reflect a forecast aligned with the average long run historical default rate, neutralizing the economic cycle effect. In capital requirements computation the use of models that reflect the average long run historical default rate is required [6] in order to stabilize capital ratios and optimize capital management policies: in positive economic conditions a run up of capital should kept for negative economic phases. A better stability of capital ratios makes also possible to stem the pro-cyclical effects of financial system and to obtain a better and more efficient management of capital itself.

In this context the use of models that reflect the average long run historical default rate let stabilize capital ratios and optimize capital management policies. Given the relevance of final PDs in reflecting the economic conditions this paper proposes a modern structured calibration approach based on average long run historical default rate (so called Central Tendency, CT) linked with the
economic scenario. The added value of this work in the literature is related to the relevance itself of the topic in the last year: given the current economic downturn Banks has a crucial necessity to have the possibility to properly forecast the portfolios PDs in order to ensure stable capital ratios by neutralizing the effect of default rate volatilities during different macroeconomic scenarios [3]. The advantage of the proposed methodology is that it is both statistical robust and applicable to any Banks portfolios. According to that here an application on real data of a Corporate portfolio of a top tier European Bank (under Single Supervisory Mechanism) has been performed and its goodness has been tested by defining an economics cycle adjusted binomial calibration test.
2. Literature Review

The literature on calibration of probability of default is not so extensive and it is for the most part concentrated starting from 2005 and related to Large Corporate portfolios, typically characterized by a low number of defaults thus different from the big part of portfolios managed by Commercial Banks.

The main findings of the authors that previously studied the calibration topic can be summarized as follows.

In 2005 Pluto and Tasche [8] suggest a methodology for estimating probabilities of default based on upper confidence intervals by use of the most prudent estimation principle. This means to estimate the PDs by upper confidence bounds ensuring at the same time a PD ordering reflecting the differences in credit quality embedded in rating grades. The methodology is easy to apply under the assumption of independent default events.

In 2008, Kiefer [5] proposes the use of Bayesian approach to final PD estimation (through calibration process). He formalizes a framework defined by the following steps: first of all, the construction of a prior distribution based on an expert rank of the default probabilities could be observed. It determines actually four parameters necessaries to identify a Beta-distribution. Next, it computes a posterior distribution by applying Bayes Theorem. The estimation of the probability of default is simply the expected value of the posterior distribution.

Iqbal and Ali [4] in 2012 propose an actuarial methodology of “convolution” calculating both Bayesian and Real Probability on each rating class under different default scenarios and generate an implied distribution of each of them by using convolution techniques.

In 2013 Tasche [9] and [10] tries to bypass some discussions arose from his previous work, concerning which confidence level can be used for the PD estimation. The Bayesian estimator for the PD based on the uninformed, uniform prior distribution is an obvious alternative that avoids the choice of a confidence level. In this paper, Tasche tries to demonstrate that in the case of
independent default events the upper confidence bounds can be represented as quantiles of a Bayesian posterior distribution based on a prior that is slightly more conservative than the uninformed prior. The comparison leads the author to suggest a constrained version of the uninformed (neutral) Bayesian estimator as an alternative to the upper confidence bound estimators.

As underlined by literature review, studies on PD calibration stress the use of Bayesian techniques for identifying the interval confidence that makes possible the estimation of final PDs. These studies are for the most part related to low default portfolios and assess the goodness of calibration process through traditional tests like binomial test. This paper contributes to the existing literature with a new calibration framework applied on a Portfolio of Medium size Corporate Clients and assessed by an economic cycle adjusted binomial test that incorporates the effect of macroeconomic cycle adding an asset correlation factor in calculation of confidence interval.
3. Calibration Methodological Framework

Main goal of this section is to provide a description of the theoretical framework proposed by this paper. The calibration philosophy at the basis of this work follows a “Through-The Cycle” (TTC) approach, aimed to neutralize the effects of volatility deriving from changes of macroeconomic scenarios in order to estimate more stable PDs.

The framework proposed can be summarized in this three following steps:

1. Central Tendency definition;
2. Calibration function estimation;
3. Calibration function fitting to application portfolio.

3.1. Central Tendency Definition

An important constraint to be considered during the calibration of PDs is that the average PDs of the application portfolio (that is usually different from the one used for PD estimation because of the data treatment) must be aligned to a long run average default rate, named Central Tendency. When the above mentioned constraint is not satisfied, the risk is to overestimate or – worst case – systematically underestimate the real risk of portfolio, with wrong credit sanctioning or non-optimal pricing definition. That is the reasons why the Central Tendency choice is the crucial phase of the calibration process.

Before defining the long run average default rate, it is then important to establish which is the best “long run” period along which computing the Central Tendency.

In order to gain an overall vision of the last macroeconomic trend, an analysis of annual Italian GDP\(^2\) has been performed on the last 10 years:

\(^2\) Source: ISTAT
The trend of Italian annual GDP during the last 10 years shows a certain degree of stability between 2002 and 2007, followed by a negative peak between 2008 and 2009, a new relative positive economic condition in the following biennium (2010 – 2011) and a new negative peak after 2011. These evidences support the choice of the last five years (2008 – 2012) as long run period for computing the average default rate as Central Tendency and neutralizing the cyclical effects: this period covers a whole “negative – positive – negative” sequence of cycles.

3.2. Calibration function estimation

The calibration function is the intermediate step of the calibration process required for combining the scores, of the rating model, with the Central Tendency in order to define a relation between statistical scores and PDs. It is the calibration function. Commonly rating models for retail and corporate counterparties are based on logistic models and the result is a vector of scores that provide the level of risk of each counterparties related with the others. This vector then has to be calibrated in order to obtain Probability of Defaults that could reflect the long run average risk profile embedded in Central Tendency value.
From a theoretical point of view, the starting point is the basic concept of a logistic regression, which means modeling the relation between score and log-odds (that is the linear). In order to identify this relation, the following equation derived from a logistic distribution can be used:

$$\log(\text{odds}) = \alpha + \beta \times \text{score}$$  \[1\]

where:

$$\text{odds} = \frac{DR}{1 - DR}$$  \[2\]

In order to estimate consistent parameters, it is important to ensure that an adequate number of defaults are used for each score: thus the fitting curve can be defined using an ordinal transformation of the score distribution instead of the continuous one, thus the fitting curve \[1\] becomes:

$$\log(\text{odds}) = \alpha + \beta \times \text{score}_{\text{mean},j}$$  \[3\]

where:

$$\text{score}_{\text{mean}} = \text{average score of } j-\text{th bucket}$$  \[4\]

The buckets are defined considering the ventiles of scores and a monotonic default rate has to be ensured.

### 3.3. Final PDs calibration

The application of \(\alpha\) and \(\beta\) as defined in \[3\] produce an average PD aligned with the observed default rate of the historical sample on which the statistical scores have been estimated. In order to ensure that the statistical PDs deriving from logistic model are aligned to the long run default rate (Central Tendency), an optimization problem must be solved subject to the following constraint:

$$\sum_{i=1}^{n} \frac{PD_i}{n} = CT$$  \[5\]

where:

\(n = \text{total number of observations;}\)
$PD_i =$ final PD at client level;

$CT =$ Central Tendency.

It is possible to identify the shift to be applied to $\alpha$, with the same $\beta$, when the following equation is satisfied:

$$F(\text{shift}) = \min_{\text{shift}} F(\text{shift})$$  \hspace{1cm} \text{[6]}$$

where:

$$F(\text{shift}) = \text{abs} \left( E \left[ \frac{1}{1 + e^{-\text{score} \beta - \alpha \text{shift}}} \right] - CT \right)$$ \hspace{1cm} \text{[7]}$$

3.4. Binomial test

Banks usually do not use individual PDs in the credit process but the mapping on a rating scale, that make possible an easier comparison and ranking of clients inside banking portfolio. Thus, the final results of calibration must be tested on rating classes, in order to verify if the final PDs are higher or lower than the real default rate of portfolio.

In order to assess the goodness of calibration process over the real portfolio, the use of an economic cycle adjusted binomial test is here defined and applied.

The traditional binomial test is based on the strong assumption of independence between default events:

$$P^* = \phi^{-1}(\alpha) \sqrt{\frac{\pi_r (1 - \pi_r)}{n} + \pi_r}$$ \hspace{1cm} \text{[8]}$$

Where:

$\alpha =$ confidence interval chosen for test;

$\pi_r =$ average calibrated PD for each r-rating class

$n =$ number of observations.
On the other hand the Basel2/3 Accord recognizes that the assumption under the traditional binomial test is unrealistic, because the correlation between economic cycles and defaults implies a correlation between defaults. According to that the use of an adjusted binomial test is proposed, where the confidence interval is extended in order to take into account the correlation effect:

$$Q = \phi\left(\frac{\sqrt{\rho} \phi^{-1}(\alpha) + \phi^{-1}(PD_k)}{\sqrt{1-\rho}}\right)$$

[9]

where:

$$\rho = \text{asset correlation used as proxy of defaults correlation, using Basel2 formulae of R (asset correlation for capital requirements calculation)}^3.$$
4. Calibration Results

The following section provides the main results deriving from the application of the proposed framework to Top tier European Bank portfolio.

4.1. Sample description and Central Tendency calculation

The calibration framework has been applied to a Corporate portfolio of 219.318 counterparties starting from the statistical PDs defined as logistic transformation of scores estimated on a sample of 188.228 observations, and reflecting an average default rate of 2.44%.

<table>
<thead>
<tr>
<th>Period</th>
<th>Event</th>
<th># obs.</th>
<th>Total obs.</th>
<th>DR</th>
<th># obs.</th>
<th>Total obs.</th>
<th>DR</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>In Bonis</td>
<td>32.144</td>
<td>32.808</td>
<td>2.02%</td>
<td>39.871</td>
<td>40.864</td>
<td>2.43%</td>
</tr>
<tr>
<td></td>
<td>Default</td>
<td>664</td>
<td></td>
<td></td>
<td>993</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td>In Bonis</td>
<td>35.677</td>
<td>36.552</td>
<td>2.39%</td>
<td>41.334</td>
<td>42.623</td>
<td>3.02%</td>
</tr>
<tr>
<td></td>
<td>Default</td>
<td>875</td>
<td></td>
<td></td>
<td>1289</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>In Bonis</td>
<td>37.641</td>
<td>38.464</td>
<td>2.14%</td>
<td>42.214</td>
<td>43.590</td>
<td>3.16%</td>
</tr>
<tr>
<td></td>
<td>Default</td>
<td>823</td>
<td></td>
<td></td>
<td>1376</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>In Bonis</td>
<td>38.222</td>
<td>39.244</td>
<td>2.60%</td>
<td>43.365</td>
<td>44.841</td>
<td>3.29%</td>
</tr>
<tr>
<td></td>
<td>Default</td>
<td>1022</td>
<td></td>
<td></td>
<td>1476</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>In Bonis</td>
<td>39.943</td>
<td>41.160</td>
<td>2.96%</td>
<td>45.711</td>
<td>47.400</td>
<td>3.56%</td>
</tr>
<tr>
<td></td>
<td>Default</td>
<td>1217</td>
<td></td>
<td></td>
<td>1689</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>188.228</td>
<td>2.44%</td>
<td></td>
<td>219.318</td>
<td>3.11%</td>
<td></td>
</tr>
</tbody>
</table>

The calibration sample is different from the estimation sample because of data treatment and default rate definitions across years\(^4\), then different default rate over the time between estimation and calibration sample is observed. According to the choice of time period 2008 – 2012 for calculating Central Tendency, the calibration framework has been applied to counterparties in a non-default

\(^4\) Cfr. Italian rules from Bank of Italy, n. 285 – Title 2 Chapter 1. The technical past-due have been not considered in the estimation phase but included for the calibration because if they move to a non-technical status they must be considered as default
status (good) at the end of January of each year starting from 2008 and observing their performance status during the following 12 months.

Starting from this sample, the Central Tendency has been computed as the average of the 5-years default rate on calibration sample and weighted for the volume of clients in portfolio on each year considered. The final values is equal to 3,11%, a bit higher than the one observed in the estimation sample.

4.2. Calibration function parameters

The best fit between score and log (odds) has been derived on the estimation sample after the score discretization process, as shown in Table 2 below:

Table 2 – Discretization of scores

<table>
<thead>
<tr>
<th>BUCKET</th>
<th># bonis</th>
<th># default</th>
<th># oss.</th>
<th>% oss.</th>
<th>DR</th>
<th>score_mean</th>
<th>score_min</th>
<th>score_max</th>
<th>LOG_ODDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.444</td>
<td>3</td>
<td>2.447</td>
<td>1%</td>
<td>0,12%</td>
<td>-6,6390</td>
<td>-7,9497</td>
<td>-6,0948</td>
<td>-2,910980</td>
</tr>
<tr>
<td>2</td>
<td>3.433</td>
<td>7</td>
<td>3.440</td>
<td>2%</td>
<td>0,20%</td>
<td>-5,8610</td>
<td>-6,0837</td>
<td>-5,5267</td>
<td>-2,690576</td>
</tr>
<tr>
<td>3</td>
<td>4.486</td>
<td>23</td>
<td>4.509</td>
<td>2%</td>
<td>0,51%</td>
<td>-5,4798</td>
<td>-5,5226</td>
<td>-5,3655</td>
<td>-2,290131</td>
</tr>
<tr>
<td>4</td>
<td>6.899</td>
<td>44</td>
<td>6.943</td>
<td>4%</td>
<td>0,63%</td>
<td>-5,1913</td>
<td>-5,3643</td>
<td>-5,0500</td>
<td>-2,195333</td>
</tr>
<tr>
<td>5</td>
<td>8.155</td>
<td>78</td>
<td>8.233</td>
<td>4%</td>
<td>0,95%</td>
<td>-4,5440</td>
<td>-5,0439</td>
<td>-4,4778</td>
<td>-2,019329</td>
</tr>
<tr>
<td>6</td>
<td>10.966</td>
<td>147</td>
<td>11.113</td>
<td>6%</td>
<td>1,32%</td>
<td>-4,4016</td>
<td>-4,4776</td>
<td>-4,3983</td>
<td>-1,872731</td>
</tr>
<tr>
<td>7</td>
<td>12.498</td>
<td>233</td>
<td>12.731</td>
<td>7%</td>
<td>1,83%</td>
<td>-4,1315</td>
<td>-4,3974</td>
<td>-3,9493</td>
<td>-1,729485</td>
</tr>
<tr>
<td>8</td>
<td>14.344</td>
<td>300</td>
<td>14.644</td>
<td>8%</td>
<td>2,05%</td>
<td>-3,5684</td>
<td>-3,9489</td>
<td>-3,4723</td>
<td>-1,679549</td>
</tr>
<tr>
<td>9</td>
<td>17.932</td>
<td>444</td>
<td>18.376</td>
<td>10%</td>
<td>2,42%</td>
<td>-3,2592</td>
<td>-3,4719</td>
<td>-3,1904</td>
<td>-1,606246</td>
</tr>
<tr>
<td>10</td>
<td>19.865</td>
<td>544</td>
<td>20.409</td>
<td>11%</td>
<td>2,67%</td>
<td>-3,1362</td>
<td>-3,1833</td>
<td>-3,1206</td>
<td>-1,562490</td>
</tr>
<tr>
<td>11</td>
<td>22.994</td>
<td>700</td>
<td>23.694</td>
<td>13%</td>
<td>2,95%</td>
<td>-2,7885</td>
<td>-3,1181</td>
<td>-2,3179</td>
<td>-1,516516</td>
</tr>
<tr>
<td>12</td>
<td>26.655</td>
<td>843</td>
<td>27.498</td>
<td>15%</td>
<td>3,07%</td>
<td>-2,1387</td>
<td>-2,3167</td>
<td>-1,8910</td>
<td>-1,499951</td>
</tr>
<tr>
<td>13</td>
<td>32.956</td>
<td>1235</td>
<td>34.191</td>
<td>18%</td>
<td>3,61%</td>
<td>-0,0661</td>
<td>-1,8903</td>
<td>2,6425</td>
<td>-1,426268</td>
</tr>
</tbody>
</table>

| TOTAL   | 183.627 | 4.601       | 188.228 | 2,44% |

Starting from this sample, the following parameters have been estimated in order to establish the relationship between score and log(odds). Their application produces an average PD equal to 2,43%.

Table 3 – Fitting curve parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>-0.97002</td>
</tr>
</tbody>
</table>
\[
\beta = 0.24195
\]

| Average PD | 2.43% |
| CT | 3.11% |

In order to produce final PDs aligned with the long run average default rate (Central Tendency) equal to 3.11%, the solving of an optimization problem on calibration population produce a shift to be applied to \( \alpha \) and equal to -0.365. Here the final formula for calculating PD after calibration process:

\[
PD = \frac{1}{1 + \exp^{\left(-0.24195 \times \text{score} - 0.97002 - 0.365\right)}}
\]

Figure 2 – Calibration results

4.3. Binomial test results

The goodness of calibration has been tested using the adjusted binomial test proposed in formula [9]. In particular, a “two tails” test has been performed considering \( \alpha = 0.05 \).

<table>
<thead>
<tr>
<th>Table 4 - Adjusted binomial test results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating class</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
</tbody>
</table>
The adjusted binomial test shows that on the whole period considered (2008 – 2012) for calibration, the observed default rates are always within the confidence interval.

Figure 2 - Adjusted confidence interval
5. Conclusions

In this paper a new calibration approach has been proposed in order to align Banks internal rating models to the economic cycle by avoiding the volatility and pro-cyclical effect embedded in statistical PDs estimation, based on 1-year default rate observation. In particular, a new approach for the definition of Central Tendency has been proposed by observing the historical default rates and their trend with respect to the macroeconomic cycle on a long run period (5 years). The Central Tendency has been used for fitting the calibration curve of the application portfolio represented by Corporate exposures of a top tier European bank between 2008 and 2012. The goodness of calibration methodology has been assessed performing a new binomial test adjusted for the economic scenario, in order to avoid the underestimation of PDs during the last recent years. The test shows that the calibration methodology correctly assess the client probability of default without suffering of pro-cyclicity effect related to macroeconomic volatility.
References


