Calibrating low-default portfolios, using the cumulative accuracy profile

M. J. van der Burgt
ABN AMRO/Group Risk Management/Models and Tools, PO Box 283 (HQ8040), 1000 EA Amsterdam, The Netherlands; email: marco.van.der.burgt@nl.abnamro.com

In the new Basel II Accord, banks are allowed to develop their own credit rating models. However, the lack of sufficient (default) data for backtesting rating models for "low-default portfolios" is a main concern for the financial industry and regulators. These low-default portfolios are characterized by the lack of sufficient data. In this article we present a method of calibrating low-default portfolios, based on modeling the observed power curve and deriving the calibration from this curve. The curve is determined by a concavity parameter, which can easily be related to the accuracy ratio (AR). The method is demonstrated for sovereign ratings.

1 INTRODUCTION

In June 2004, the Basel Committee on Banking Supervision launched a revised version of the "International Convergence of Capital Measurement and Capital Standards", hereafter denoted as the Basel II Accord (BCBS (2004)). The most important consequence of this accord is that banks are allowed to develop their own counterparty rating models. This is called the Internal Ratings Based (IRB) approach. One of the IRB requirements for the use of internal ratings is that "probability of default (PD) estimates must be grounded in historical experience and empirical evidence, and not based purely on subjective or judgemental considerations" (BCBS (2004, paragraph 449)). In addition, "banks must regularly compare realised default rates with estimated PDs for each grade and be able to demonstrate that the realised default rates are within the expected range for that grade" (BCBS (2004, paragraph 501)). This raises concerns in the financial industry from the regulatory perspective for so-called low-default portfolios (LDPs). Although there is no general definition of LDPs (Benjamin *et al* (2006)), they are defined in this paper as portfolios with limited default experience from which to obtain robust default probabilities (PDs) for Basel II or internal risk management purposes (Wilde and Jackson (2006)). Examples of LDPs are portfolios with exposures to banks, insurance companies, sovereigns, highly-rated corporate obligors and most forms of specialized lending such as project finance.

From a regulatory perspective, a concern has risen that credit risk of LDPs might be underestimated because of data scarcity (Benjamin *et al* (2006)). When PD estimates are based on simple historical averages or just judgemental considerations alone, capital requirements for that portfolio might be underestimated.

Benjamin *et al* show using simulation that portfolios may have a PD in the order of percentages, but these portfolios still have a quite large probability of yielding no defaults in a given year (Benjamin *et al* (2006)). Since the default distributions are highly skewed, the default rates in any given year might lie below the average. The most important concern of participants in the financial industry is that the lack of sufficient statistical data might lead to difficulty in backtesting, ie, testing whether the PDs agree with the observed default rates within a specified confidence interval. When no proper backtesting of risk parameters such as the PDs is feasible, LDPs will be excluded from IRB treatment (BCBS (2005)).

In response to the concerns of both regulators and the financial industry, several methodologies towards LDPs are discussed in the literature. Schuermann and Hanson (2004) propose a methodology to estimate PDs using migration matrices. PDs for high-quality ratings are estimated using borrower migrations to lower grades and eventually default. These migrations are calculated using the duration approach, in which a migration intensity is calculated from all rating changes over the course of the year and then transforming this intensity into a migration probability.

Pluto and Tasche (2005, 2006) propose to estimate PDs using upper confidence bounds while guaranteeing an ordering of PDs that respects the differences in credit quality as indicated by the rating grades. They call this method the *most prudent estimation*. Their papers use the method of most prudent estimation, first under the assumption of independent defaults and then under an assumption of asset correlation, using different confidence levels. The choice of confidence levels is still under discussion, but the authors suggest that confidence levels of less than 95% appear intuitively appropriate. A similar method is described by Forrest (2005), but this method is based on the likelihood approach by working in multiple dimensions. Each dimension corresponds to a rating grade and each point represents a possible choice of grade-level PDs. In this multidimensional space, he identifies a subset of points with a high level of occurrence, conditional on the observed data.

Benjamin *et al* (2006) propose an approach towards LDPs, in which the regulator has a role to play. In their proposal, the regulator provides an objective criterion for LDPs and publishes a look-up table, from which a look-up PD is derived and compared with the weighted average PD of the financial institution's portfolio. Based on this comparison, the financial institution adjusts its PD until the weighted average PD is equal to or above the look-up PD. In a recent publication, Wilde and Jackson (2006) show that PD estimates can be calculated analytically by calibrating the CreditRisk+ model to a Merton model of default behavior.

In this paper we introduce a method for calibrating LDPs. The paper introduces the cumulative accuracy profile (CAP), also known as the power curve or Lorentz curve, and a mathematical function for modeling the CAP. The mathematical function is used in the next section, where the essentials of the method are described. The key parameter in this methodology is the concavity, which defines the shape of the CAP curve. Using the functional form of the CAP and the concavity, a calibration can be calculated by taking the derivative of the closed-form equation for the CAP. The method is tested, using artificial portfolios, and demonstrated for sovereign ratings. In the demonstration, the error in the calibration method is estimated using different scenarios.

2 MODELING THE POWER CURVE

The method is based on the fact that assessing the discriminative power of a credit rating model is easier than calibrating a credit rating model, but calibration can be derived from the discriminative power (Falkenstein *et al* (2000)), ie, the ability to distinguish between defaults and non-defaults. The discriminative power is often measured with a CAP curve, also known as the power curve. Since the power curve is extensively discussed elsewhere (Engelmann *et al* (2003); Keenan and Sobehart (2000)), this concept is only briefly described here.

The CAP curve is constructed by sorting the debtors in order from bad ratings to good ratings, ie, by decreasing credit risk. Plotting the cumulative percentage of defaults as a function of the cumulative percentage of debtors leads to one of the curves in Figure 1. The curve, as represented by the dashed line, is observed for a rating system with considerable discriminative power. The curve shows that, for example, 80% of the defaults occur in 20% of the most risky debtors. When the rating system has no discriminative power and randomly assigns ratings to obligors, the cumulative percentage of defaults increases proportionally with the cumulative percentage of debtors, as demonstrated by the gray curve in Figure 1. On the other extreme, when the model works perfectly, all defaults are observed in the worst risk class and the curve corresponds to a perfect model.

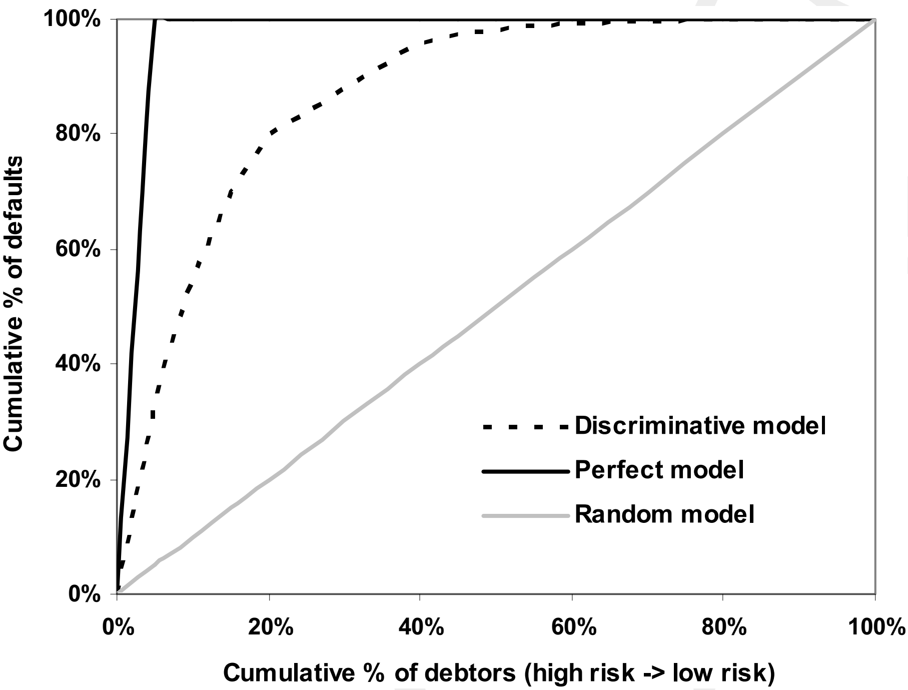
The concave shape of the CAP curve can be easily modeled using a mathematical function. This function gives the cumulative percentage of defaults (hereafter denoted as y) as a function of the cumulative percentage of debtors (hereafter denoted as x). The function should obey the following requirements.

- When $x = 0$, $y = 0$: this is clear from Figure 1, which shows that the CAP starts from the origin.
- When $x = 1$, $y = 1$: when the cumulative number of debtors is 100%, the cumulative percentage of defaults is also 100%.
- The derivative of the CAP can be used as calibration (Falkenstein *et al* (2000)). Assuming that the PD increases exponentially with the rating class, the derivative of the function must be exponential.

The last requirement is based on a crucial assumption, ie, the PD varies exponentially as a function of the rating class. In order to test the validity of this assumption, PDs are calculated for each S&P rating class as one-year default rates from S&P data[1] and averaged over the period from 1981 until 2005. The long-term average is used, because the Basel II Accord requires that "PD estimates must be a long-run average of one-year default rates for borrowers in the grade" (BCBS

[1] Data is retrieved from Standard & Poor's CreditPro® v7.02 (http://creditpro.sandp.com).

FIGURE 1 Cumulative accuracy profile (power curve) for a perfect model (—), a discriminatory model (- - -) and a model with no discriminative power at all (—).



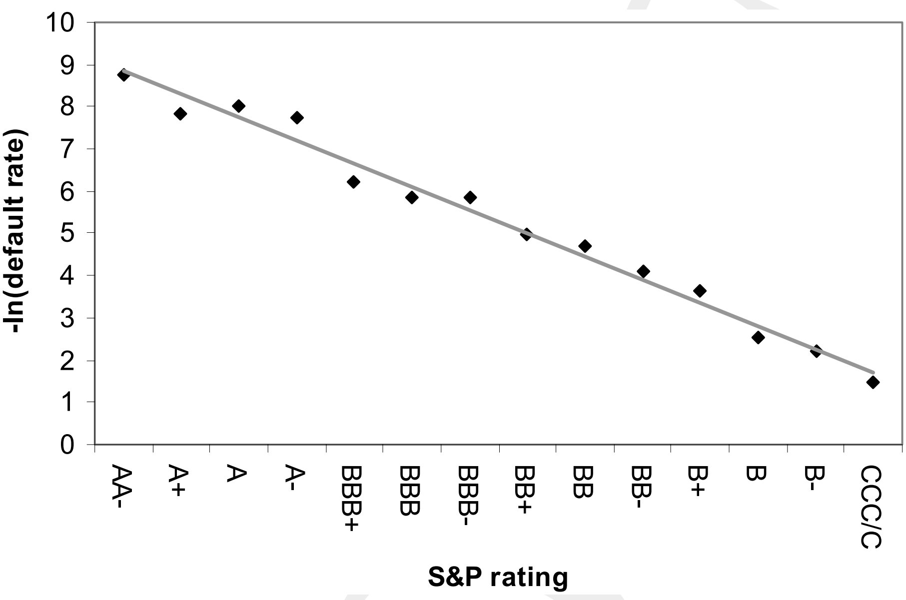
(2004, paragraph 447)). Figure 2 shows the logarithm of the average one-year default rate versus the S&P rating class. The logarithmic of the default rate is linear as a function of the rating scale and supports the assumption that the default rate is an exponential function of the rating grades.

Based on the functional requirements, the following function is introduced for modeling the concave shape of the CAP:

$$y(x) = \frac{1 - \mathrm{e}^{-kx}}{1 - \mathrm{e}^{-k}} \tag{1}$$

where parameter k is called the concavity. The concavity can be interpreted as a measure of discriminative power, as demonstrated by Figure 1. When $k \to \infty$, (1) gives $y \approx 1$ and the CAP corresponds to a rating system, which perfectly discriminates between defaults and non-defaults. In this case, the area (A_{perfect}) under the CAP curve approaches 1. When $k \to 0$, (1) gives $y \approx x$ and the CAP corresponds to a random model with no discriminative power. In this case the area (A_{random}) is equal to 0.5. In practice, k will be between 0 and ∞. The CAP function has only a concave shape when $k > 0$, ie, $y(x) > x$ on the interval [0, 1]. The CAP function is convex when k is negative. This is demonstrated in Figure 3.

FIGURE 2 Logarithmic plot of the average one-year default rate as a function of the S&P rating classes.



The one-year default rate is an average default rate, observed over the period from 1981 to 2005. The solid line represents a linear fit. (Source: S&P CreditPro® v7.02; http://creditpro.sandp.com.)

A convex CAP occurs when most of the defaults occur in the best rating classes and hardly any defaults in the worst rating classes. Therefore, a negative k value corresponds to a rating system that assigns ratings in reverse order of default risk. This is called adverse selection,[2] which was recently observed in the credit crisis of 2007: defaults of subprime mortgage loans had a negative impact on hedge funds, investment banks and conduits (for an overview of the current crisis, see Specht *et al* (2007)) by means of US mortgage securitized pools. Most of these securitized pools were rated AA, whereas the underlying risk was high. External rating agencies have admitted this by downgrading over 1,000 mortgage-backed securities. From (1), it can be concluded that the concavity k might be an indicator of the soundness of a rating system:

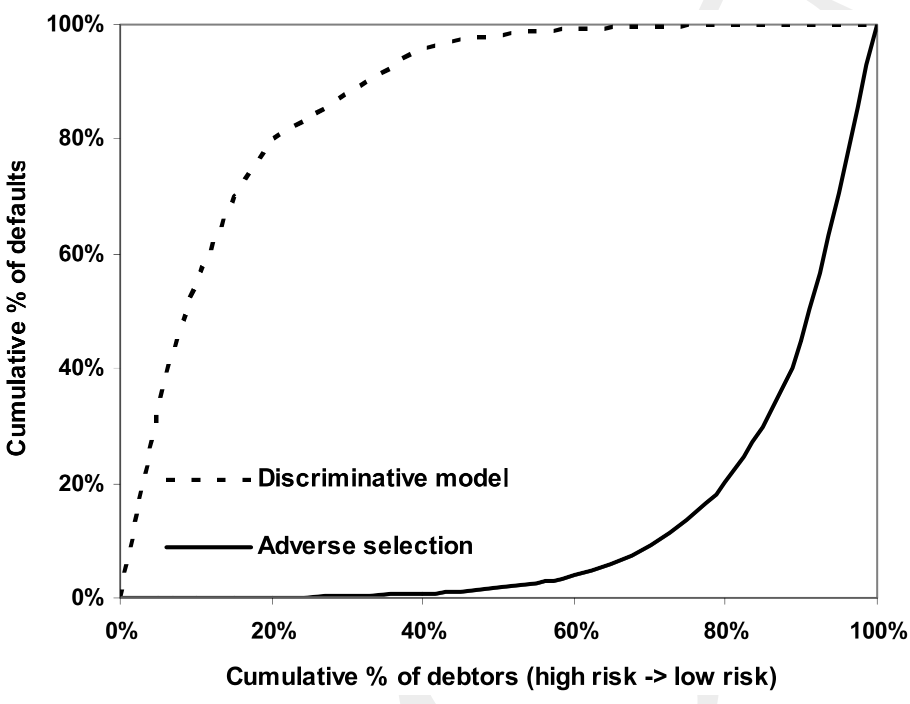
1. when k is high and positive, the rating system has good discriminatory power;
2. when $k = 0$, the rating system has no discriminatory power; and
3. when k is negative, the rating system is subject to adverse selection.

Although the concavity might be interpreted as a measure of discriminatory power, in best-practice credit risk management it is common to use the accuracy

[2]Sometimes this is also referred to as anti-selection or negative selection.

Marked Proof Ref: vanderBurgt1(4)/36858ae February 27, 2008

FIGURE 3 Cumulative accuracy profile (power curve) for a discriminatory model (- - -) and a model that is based on adverse selection (—).



The curve for the model with adverse selection is convex, which corresponds to a negative concavity parameter k.

ratio (AR), which is also derived from the CAP curve. The relation between the AR and the area (A) under the CAP curve is calculated as (Engelmann *et al* (2003); Keenan and Sobehart (2000); Tasche (2005)):

$$\mathrm{AR} = \frac{A - A_{\mathrm{random}}}{A_{\mathrm{perfect}} - A_{\mathrm{random}}} \approx 2A - 1 \tag{2}$$

The AR measure varies between 0 for a random rating system and approaches 1 for an extremely predictive rating system. Since both AR and k measure discriminatory power, it is not surprising that a relation exists between both quantities. In Appendix A, a relation between AR and k is derived.

3 DESCRIPTION OF THE METHOD

The method of calibration is performed in a few steps. The first step is to construct the CAP from the observations. The fitting of the CAP to the function in (1) is applied. This fitting procedure is executed by minimizing root mean square (RMS) error[3] E:

$$E = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\{y_i - \frac{1-\exp(-kx_i)}{1-\exp(-k)}\right\}^2} \tag{3}$$

which results in the value of the concavity parameter k. Here N is the number of rating classes and x_i and y_i are the observed cumulative percentages of debtors and defaults respectively in each rating class i. In (3), it is implicitly assumed that the residuals $y_i - (1-\exp(-kx_i))/(1-\exp(-k))$ are normally distributed, which will be tested in Sections 4 and 5 by measuring the skewness and kurtosis of the residuals. The skewness and kurtosis are used in the Jarque–Bera (JB) test for normality in residuals (Bera and Jarque (1980)):

$$\mathrm{JB} = \frac{N}{6}\left[S^2 + \left(\frac{K-3}{4}\right)^2\right] \tag{4}$$

where N is the number of observations, S is the skewness and K is the kurtosis. This test statistic is χ^2-distributed with two degrees of freedom and results in a p-value. The H_0 hypothesis that the residuals are normally distributed is rejected when the p-value is smaller than 5%.

A much simpler method than optimizing (3) is making use of the relation between the area under the CAP curve (A) and the concavity k. In Appendix A we show that:

$$k \approx \frac{1}{1-A} \tag{5}$$

when A is larger than 0.8.

After deriving the concavity, the PDs can be derived from the CAP curve, using the following equation of Falkenstein *et al* (2000):

$$\mathrm{PD}(R) = \langle D\rangle \frac{\mathrm{d}y}{\mathrm{d}x} \tag{6}$$

where $\langle D\rangle$ is the average observed default rate, ie, the total number of defaults divided by the total number of obligors for the whole portfolio. Combining Equation (1) with Equation (6) gives the following equation for calibration:

$$\mathrm{PD}(R) = \frac{k\langle D\rangle}{1-\mathrm{e}^{-k}}\exp\{-kx_R\} \tag{7}$$

Equation (7) is the derivative of the CAP function in (1). The symbol x_R represents the cumulative percentage of counterparties in rating class R. The value of x_R is

[3]The minimum of the RMS in (3) is calculated numerically by applying a Newton–Raphson procedure, ie, a new k_1 is calculated from the initial value $k_0 = 1$, k_2 is calculated from k_1, etc, using the following iteration ($i = 1, 2, \ldots$):

$$k_{i+1} = k_i - \left[\frac{\partial E/\partial k}{\partial^2 E/\partial k^2}\right]_{k=ki}$$

In this paper, 20 iterations are used.

calculated as the midpoint between the cumulative percentage of counterparties in rating class R and $R-1$:

$$x_R = \frac{z_N + z_{N-1} + \cdots + z_{R-1} + (z_R/2)}{z} \tag{8}$$

where z is the total number of counterparties and z_i is the number of counterparties in rating class i. In (8), rating N represents the most risky rating class.

The described method adds value when the number of observations in a specific rating class is low. Traditional methods calculate the default rate or odds ratio per rating class or per score bucket. These methods rely on the number of defaults and number of counterparties, which are observed in a rating class. When the number of counterparties in a rating class is low, the default rate or odds ratio of that rating class cannot be properly estimated owing to the low number of observations. Section 5 of this paper gives an example for a sovereign portfolio, in which the number of counterparties is smaller than 10 in each rating bucket. In these cases, a reliable estimation of the PD by calculating the default rate as the number of defaults divided by the number of counterparties cannot be obtained. An advantage of the new method is that it is not only data from a specific rating class that is used for calculating the default rate, but implicitly all the available data by fitting the CAP to (1) and deriving the default rate from this function. As such, the method is based on economic use of the data, which is available in a limited way.

In light of the credit crisis of 2007, a key question is how adverse selection has an effect on the outcomes of the method in this paper. Adverse selection is caused by information asymmetry: in the crisis, securitization of subprime mortgage loans transferred risk from banks, which have incentives to monitor their borrowers, to other institutions such as hedge funds, which do not have such an incentive, see for example Rajan (2005). Banks need to disclose their risk under Basel II and are highly regulated, whereas hedge funds, by definition, are not. Owing to this transfer, information on the underlying portfolios was sparse. As mentioned before, the method as described in this paper might add value in these cases, which are characterized by the limited availability of data. However, the method fails when the rating does not properly reflect the true risk of the rated entity, just like any other calibration method.

4 DEMONSTRATION FOR ARTIFICIAL PORTFOLIOS

Before application of the method on actual data, the method is first tested on artificial portfolios, which are constructed from counterparties with S&P ratings. For every rating class, it is assumed that real PDs are known and denoted as PD_{real}. The PD_{real} are calculated as long-term averages of one-year default rates from the S&P data. These averages are also used to demonstrate the exponential behavior of default rates in Figure 2. Using the number of counterparties and defaults in each rating class, the total default rate of the whole portfolio and the concavity is calculated. Both quantities are used to calculate the PDs, using (7).

These estimated PDs are referred to as the $\mathrm{PD}_{\mathrm{est}}$. The method tests whether the $\mathrm{PD}_{\mathrm{est}}$ falls within a 95% confidence level around the $\mathrm{PD}_{\mathrm{real}}$. This confidence interval is calculated as Tasche (2005):

$$\left[\mathrm{PD}_{\mathrm{real}} + N^{-1}[(1+\alpha)/2] \cdot \sqrt{\frac{\mathrm{PD}_{\mathrm{real}} \cdot (1 - \mathrm{PD}_{\mathrm{real}})}{N}},\right.$$
$$\left. \mathrm{PD}_{\mathrm{real}} - N^{-1}[(1+\alpha)/2] \cdot \sqrt{\frac{\mathrm{PD}_{\mathrm{real}} \cdot (1 - \mathrm{PD}_{\mathrm{real}})}{N}}\right] \quad (9)$$

where α is the confidence level, which is chosen as 95%, and $N^{-1}[\cdots]$ is the inverse of the cumulative normal distribution. This test is frequently referred to as the binomial or Wald test (Schuermann and Hanson (2004)).

Table 1 presents the test results for a homogeneous portfolio, ie, a portfolio with the same number of counterparties in each rating class. The table shows that the $\mathrm{PD}_{\mathrm{est}}$ all fall within the 95% confidence levels around $\mathrm{PD}_{\mathrm{real}}$ and all estimated PDs pass the binomial test. The distribution of residuals in (3) exhibits a skewness of 0.32 and a kurtosis of 4.35. The positive kurtosis indicates that the distribution of errors is peaked. However, the JB test of normality gives a p-value of 44%, so the null hypothesis that the residuals are normally distributed is accepted (Bera and Jarque (1980)).

The results in Table 1 are based on a homogeneous portfolio with the same number of counterparties in each rating class. However, these types of portfolios are hardly encountered in practice. In Table 2 a similar test is performed, but here it is assumed that most counterparties are in the moderate rating classes, whereas a low number of counterparties exists in the highest and lowest rating class. In this case, the binomial test shows that the $\mathrm{PD}_{\mathrm{est}}$ all fall within the 95% confidence levels around $\mathrm{PD}_{\mathrm{real}}$. The distribution of errors shows a skewness and kurtosis of 1.28 and 3.88, respectively. The JB normality test results in a p-value of 6% and therefore the hypothesis of normality is accepted.

5 DEMONSTRATION FOR A LOW-DEFAULT PORTFOLIO: SOVEREIGNS

The method as described in the Section 3 is demonstrated for a low-default portfolio with exposures to 86 sovereigns. Their Foreign Currency (FC) ratings of March 2004 and March 2005 are collected from S&P. Defaults are based on migrations in the year after March 2004. Only the governments of Grenada and the Dominican Republic migrated to a default in the period between March 2004 and March 2005, no defaults are observed for the other 84 countries.

Table 3 presents the number of sovereigns and the number of defaults per rating class, sorted by decreasing credit risk. Cumulative percentages of sovereigns (X) and defaults (Y) are calculated and used to construct the CAP in Figure 4. The black curve gives the observed CAP and the gray curve results from the fitted CAP function. The concavity is found to be 8.03, at which the RMS error has a minimum value of 0.15. The distribution of residuals in (3) shows a skewness of

TABLE 1 Demonstration of the new method for an artificial portfolio assuming that PD_{real} is known.

Rating	PD_{real}	Number of counterparties	Number of defaults	PD_{est}	High	Low	Binomial test
CCC/C	22.92%	100	22	21.98%	31.16%	14.68%	TRUE
B−	10.83%	100	10	10.19%	16.92%	4.74%	TRUE
B	7.97%	100	7	4.73%	13.28%	2.66%	TRUE
B+	2.59%	100	2	2.19%	5.70%	0.00%	TRUE
BB−	1.64%	100	1	1.02%	4.14%	0.00%	TRUE
BB	0.90%	100	0	0.47%	2.75%	0.00%	TRUE
BB+	0.70%	100	0	0.22%	2.34%	0.00%	TRUE
BBB−	0.28%	100	0	0.10%	1.32%	0.00%	TRUE
BBB	0.29%	100	0	0.05%	1.35%	0.00%	TRUE
BBB+	0.20%	100	0	0.02%	1.08%	0.00%	TRUE
A−	0.04%	100	0	0.01%	0.45%	0.00%	TRUE
A	0.03%	100	0	0.00%	0.39%	0.00%	TRUE
A+	0.04%	100	0	0.00%	0.43%	0.00%	TRUE
AA−	0.03%	100	0	0.00%	0.37%	0.00%	TRUE
AA	0.03%	100	0	0.00%	0.37%	0.00%	TRUE
AA+	0.03%	100	0	0.00%	0.37%	0.00%	TRUE
AAA	0.03%	100	0	0.00%	0.37%	0.00%	TRUE

The probability of default as calculated by the method PD_{est} is compared with the PD_{real} using the binomial test. The results of the binomial test are given in the last column. The portfolio default rate is 2.47% and the concavity is 13.06.

0.69 and a kurtosis of 5.57. The JB test on normality results in a p-value of 3%. Therefore, the hypothesis that the residuals are normally distributed is rejected. The minimum of the RMS error only gives a relative measure about the quality of the fit. Therefore, the area under the CAP is compared with the area under the fitted CAP function. The area under the CAP is 0.89 whereas the area that is based on integrating (1) over the interval [0, 1] is equal to 0.88. Since both values are close to each other, it is concluded that the fit of the CAP is quite accurate.

The calculated concavity $k = 8.03$ and the average default rate of 2.33%, which results from observing two defaults of 86 sovereigns, is used to calculate the PD curve with use of (7). In Figure 5, this PD curve is represented by the solid line. The PD curve is also presented in the last column of Table 3. Figure 5 and Table 3 show that the PD curve is a smooth function of the rating classes. When the default rate was calculated in the traditional way, by dividing the number of defaults by the number of counterparties, there was a default rate of 100% in rating class CC, a default rate of 25% in rating class BB− and the default rate in all other rating classes would be 0%. This would certainly not reflect the default risk per rating classes and these default rates are not reliable owing to the low observation of data.

The method might be very sensitive to the rating classes, in which the defaults are observed. For example, the concavity would change when the default is not observed in the CC class rather than in the adjacent rating class CCC+.

TABLE 2 Demonstration of the new method for an artificial portfolio assuming that PD_{real} is known.

Rating	PD_{real}	Number of counterparties	Number of defaults	PD_{est}	High	Low	Binomial test
CCC/C	22.92%	50	11	15.32%	34.57%	11.27%	TRUE
B−	10.83%	75	8	11.65%	17.86%	3.79%	TRUE
B	7.97%	100	7	7.94%	13.28%	2.66%	TRUE
B+	2.59%	150	3	4.59%	5.13%	0.05%	TRUE
BB−	1.64%	225	3	2.02%	3.31%	0.00%	TRUE
BB	0.90%	300	2	0.64%	1.97%	0.00%	TRUE
BB+	0.70%	400	2	0.14%	1.52%	0.00%	TRUE
BBB−	0.28%	500	1	0.02%	0.75%	0.00%	TRUE
BBB	0.29%	550	1	0.00%	0.74%	0.00%	TRUE
BBB+	0.20%	500	1	0.00%	0.60%	0.00%	TRUE
A−	0.04%	400	0	0.00%	0.25%	0.00%	TRUE
A	0.03%	250	0	0.00%	0.26%	0.00%	TRUE
A+	0.04%	225	0	0.00%	0.30%	0.00%	TRUE
AA−	0.03%	150	0	0.00%	0.31%	0.00%	TRUE
AA	0.03%	100	0	0.00%	0.37%	0.00%	TRUE
AA+	0.03%	75	0	0.00%	0.42%	0.00%	TRUE
AAA	0.03%	50	0	0.00%	0.51%	0.00%	TRUE

The probability of default as calculated by the method PD_{est} is compared with PD_{real} using the binomial test. The results of the binomial test are given in the last column. The portfolio default rate is 0.95% and the concavity is 17.97.

FIGURE 4 Observed CAP and CAP, modeled using (1): the observed CAP is based on the defaults and non-defaults of sovereigns.

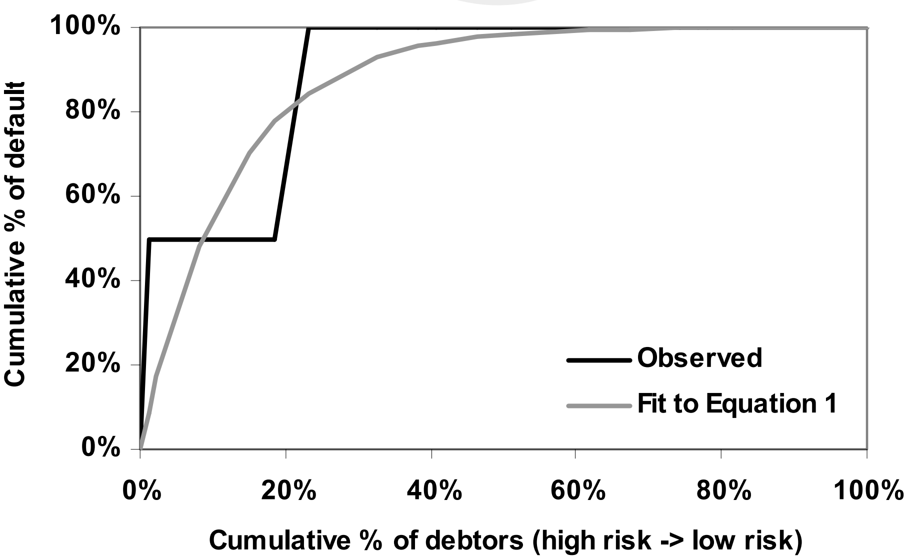


TABLE 3 Data used for demonstrating the method of calibrating LDP.

Rating	Sovereigns	Defaults	X	Y	PD curve
			0%	0%	
CC	1	1	1%	50%	17.83%
CCC+	1	0	2%	50%	16.24%
B−	5	0	8%	50%	12.27%
B	6	0	15%	50%	7.34%
B+	3	0	19%	50%	4.82%
BB−	4	1	23%	100%	3.48%
BB	8	0	33%	100%	1.99%
BB+	5	0	38%	100%	1.08%
BBB−	2	0	41%	100%	0.78%
BBB	5	0	47%	100%	0.56%
BBB+	4	0	51%	100%	0.37%
A−	9	0	62%	100%	0.20%
A	5	0	67%	100%	0.10%
A+	6	0	74%	100%	0.06%
AA−	2	0	77%	100%	0.04%
AA	1	0	78%	100%	0.04%
AA+	3	0	81%	100%	0.03%
AAA	16	0	100%	100%	0.01%
Total	86	2			

The S&P FC ratings of 86 sovereigns are collected from March 2004. The defaults are based on migrations from March 2004 to March 2005. Only two sovereigns migrated to default between March 2004 and March 2005: the Dominican Republic and Grenada. The last column gives the PD, which is calculated by the method as described in Section 3.

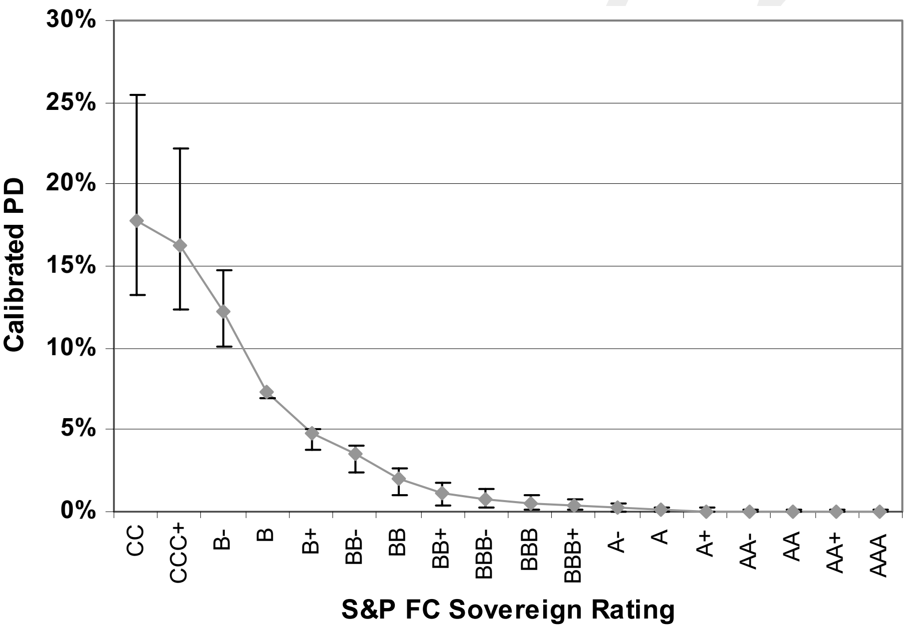
In order to assess the accuracy of the method, the concavity k is calculated for several scenarios, in which defaults are shifted from the rating class in which they are originally observed towards adjacent rating classes. The concavities for all of these scenarios are presented in Table 4. The table shows that the average concavity is equal to 8.22 and the standard deviation in the concavity is 2.28, respectively. Based on the standard deviation and assuming a normal distribution of the concavity, a 95% confidence interval around 8.22 can be defined[4] as 8.22 ± 4.47. The average value agrees with the originally calculated value of 8.03 for the observed data within the confidence interval.

Using (5) and the area under the CAP, which is found to be 0.89, the concavity is calculated as 9.01. The concavity, which is calculated by minimizing the RMS error, is equal to 8.03. From the difference between these results, it is concluded that the approximation in (5) provides a proper estimation of the concavity.

PD curves are also calculated using the minimum concavity of 5.87 and the maximum concavity of 11.70. Figure 5 shows the corresponding curves as

[4]This confidence interval is calculated as $\mu \pm z\sigma$, in which μ is the average value 8.22, z is the 95% percentile (1.96) and σ is the standard deviation (2.28).

FIGURE 5 Calibration of the S&P FC sovereign ratings, derived from the modeled CAP curve.



The error bars present the PDs, calculated at the minimum ($k = 5.87$) and maximum concavity ($k = 11.70$).

TABLE 4 Different scenarios that are used to assess the accuracy of the concavity.

Scenario	Concavity
One default in CC, 1 default in BB	6.15
One default in CC, 1 default in BB−	8.03
One default in CC, 1 default in B+	11.70
One default in CCC+, 1 default in BB	5.87
One default in CCC+, 1 default in BB−	7.47
One default in CCC+, 1 default in B+	10.10
Standard deviation	2.28
95% Confidence level	4.47
Average	8.22

The concavity varies between 5.87 and 11.70 in all scenarios. In addition, the average and standard deviation are also shown.

error bars on the curve for a concavity of 8.03. A general observation from Figure 5 is that the PD curve becomes flatter when the concavity is lower. This is explained by the fact that a low concavity corresponds to a rating system with low discriminatory power, ie, the likelihood of default in every rating class is about the same and therefore the PD curve has a flat shape. Figure 5 also shows that the most drastic effects of changing the concavity are observed in the worst rating classes CC and CCC+.

6 CONCLUSION

In this paper we have presented a method for the calibration of rating systems that are applied to LDPs. The method is based on fitting the CAP to a concave function. Using the derivative of the concave function and the average default rate, a calibration can be performed. The method adds value to the existing methods such as calculating default rates and odd ratios per rating bucket when the number of observations is low. A key parameter in the method is the concavity. When the concavity is positive, the rating system has high ability to discriminate between defaults and non-defaults. When the concavity is zero, the rating system has no discriminatory power. A negative concavity means that the rating system is subject to adverse selection.

The method is demonstrated for a LDP of sovereigns, but can be applied to any portfolio when defaults are observed. Traditionally, default rates are calculated as the total number of defaults divided by the total number of obligors for each rating class. Although this approach is simple and straightforward, it has certain drawbacks: the estimation of the default rate depends on the number of observations in a specific rating class. When the number of obligors in a rating bucket is small, the default rate in that rating bucket cannot be properly estimated. In the method in this paper, all observations in all rating buckets are included in the concavity parameter, from which the default rates are derived.

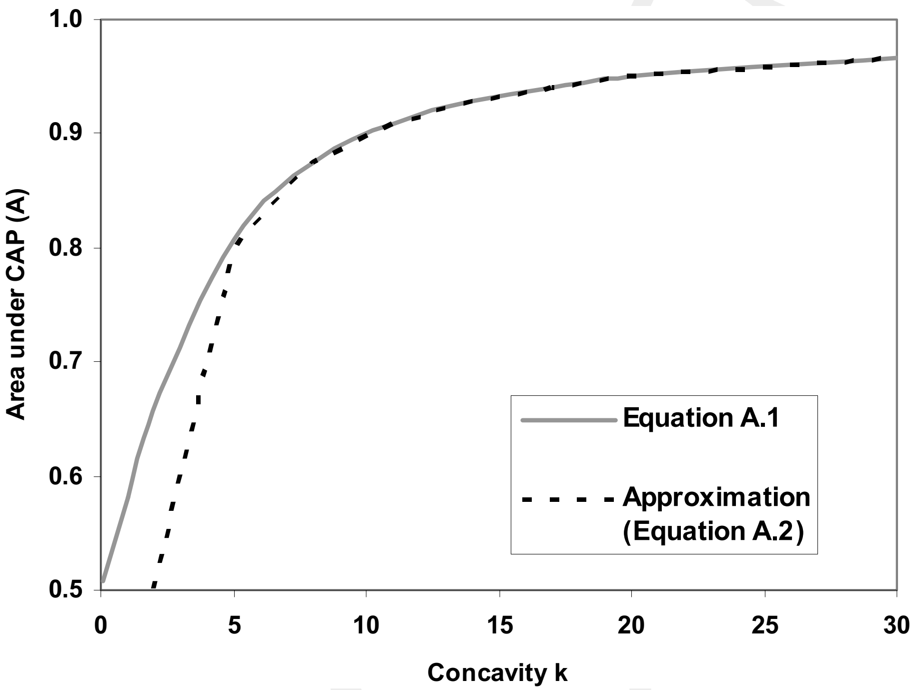
Since the method is based on modeling the CAP, the method does not work when no defaults are observed at all. In this case, other solutions should be selected before the method can be applied. In this case, the principle of the most prudent estimator of Pluto and Tasche (2005, 2006) seems appropriate. Other solutions for backtesting LDPs are related to data enhancement. Several approaches in this direction are suggested by the Basel Committee Accord Implementation Group's Validation Subgroup, such as pooling of data with other banks, combining portfolios with similar risk characteristics, using the lowest non-default rating as a proxy for default and combining rating categories (BCBS (2005)).

APPENDIX A RELATION BETWEEN THE CONCAVITY AND THE AREA UNDER THE CAP CURVE

In this appendix, a simple relation is derived between the AR and the concavity k. First, the area under the CAP curve is calculated by integrating (1) over the

Marked Proof Ref: vanderBurgt1(4)/36858ae February 27, 2008

FIGURE A.1 Comparing (A.1), which relates k to A (area under CAP), with the approximation in (A.2): the approximation can be used when $A > 0.8$.



interval [0, 1]:

$$A = \int_0^1 \left\{ \frac{1 - e^{-kx}}{1 - e^{-k}} \right\} dx = \frac{1}{1 - e^{-k}} - \frac{1}{k} \tag{A.1}$$

The area A approaches 1 when $k \to \infty$. This is the case when the rating system has perfect discriminatory power. When $k \to 0$ the area A approaches 0.5, which corresponds to a rating system with no discriminatory power. As e^{-k} tends to go to zero very fast, (A.1) can be approximated by the following expression:

$$A \approx 1 - \frac{1}{k} \quad \Rightarrow \quad k \approx \frac{1}{1 - A} \tag{A.2}$$

Figure A.1 compares (A.1) with (A.2) and shows that (A.2) can be used as a good approximation when $A > 0.8$. For rating systems, which exhibit good discriminatory power, (A.2) provides a simple method for calculating the concavity k, which is used in the calculation of the PD for each rating class.

Although the area under the CAP curve (A) can be interpreted as a measure of discriminatory power, it is widely accepted to use the AR as a measure of discriminatory power (Engelmann *et al* (2003); Keenan and Sobehart (2000); Tasche (2005)). A relation between k and the AR measure is obtained by combining (2) and (A.1):

$$\mathrm{AR} = 2A - 1 = 2\left[\frac{1}{1-\mathrm{e}^{-k}} - \frac{1}{k}\right] - 1 \tag{A.3}$$

Using the approximation in (A.2), a simple relation between the AR and the concavity k can also be derived:

$$\mathrm{AR} \approx 2A - 1 = 1 - \frac{2}{k} \quad \Rightarrow \quad k \approx \frac{2}{1-\mathrm{AR}} \tag{A.4}$$

Although (A.4) gives a simple relation between the AR and the concavity, it should be noted that this equation is based on two approximations. The first approximation is that the relation between the AR and the area under the CAP curve (A) is given in (2). In this equation, it is assumed that $A_{\mathrm{perfect}} \approx 1$. For a perfect discriminatory model, all defaults are observed in the worst rating class and therefore A_{perfect} will be slightly less than 1 (for an extensive description see Engelmann *et al* (2003); Keenan and Sobehart (2000); Tasche (2005)). The second approximation is based on (A.2) and assumes that the area $A \geq 0.8$, which corresponds to $\mathrm{AR} \geq 0.6$.

Combining the approximation in (A.4) and (7) results in a relation between the PD and the AR:

$$\mathrm{PD}(R) = \frac{2\langle D\rangle \exp\{-2x_R/(1-\mathrm{AR})\}}{(1-\exp\{-2/(1-\mathrm{AR})\})(1-\mathrm{AR})} \tag{A.5}$$

Equation (A.5) shows that the PD curve rises steeply for high-risk rating classes when the rating system has a high AR and, therefore, a high discriminatory power.

REFERENCES

Basel Committee on Banking Supervision (BCBS) (2004). International Convergence of Capital Measurement and Capital Standards, A Revised Framework. Bank for International Settlements, Basel, July.

Basel Committee on Banking Supervision (BCBS) (2005). Validation of low-default portfolios in the Basel II Framework. Basel Committee Newsletter No. 6, Bank for International Settlements, September.

Benjamin, N., Cathcart, A., and Ryan, K. (2006). Low-default portfolios: a proposal for conservative estimation of default probabilities. Financial Services Authority.

Bera, A. K., and Jarque, C. M. (1980). Efficient test for normality, homoskedasticity and serial independence of regression residuals. *Economic Letters* **6**(3), 255–259.

Engelmann, B., Hayden, E., and Tasche, D. (2003). Measuring the discriminative power of rating systems. Discussion Paper No. 01/2003, Series 2: Banking and Financial Supervision, Deutsche Bundesbank.

Falkenstein, E., Boral, A., and Carty, L. (2000). RiskCalc for private companies: Moody's default model: rating methodology. Moody's Investor Service.

Forrest, A. (2005). Likelihood approaches to low default portfolios. Joint Industry Working Group Discussion Paper, Credit Research Center (CRC), University of Edinburgh.

Keenan, S., and Sobehart, J. (2000). A credit risk catwalk. *Risk* (July), 84–88.

Pluto, K., and Tasche, D. (2005). Thinking positively. *Risk* **18**(8), 72–78.

Pluto, K., and Tasche, D. (2006). Estimating probabilities of default for low default portfolios. *The Basel II Risk Parameters*, Engelmann, B. and Rauhmeier, R. (eds). Springer, Berlin, pp. 79–103.

Ranjan (2005). Has financial development made the world easier? *Proceedings Jackson Hole Symposium*, Federal Reserve Bank of Kansas City, September.

Schuermann, T., and Hanson, S. (2004). Estimating probabilities of default. Staff Report No. 190, Federal Reserve Bank of New York.

Specht, B., Ratul, R., Neve, J. M., and Green, J. D. (2007). What really backs ABCP or should it be who? *Structured Credit Products Global Report*, CitiGroup, p. 5.

Tasche, D., (2005). Studies on validation of internal rating systems. Working Paper No 14, Basel Committee on Banking Supervision, p. 28.

Wilde, T., and Jackson, L. (2006). Low-default portfolios without simulation. *Risk* (August), 60–63.