Smoothing Algorithms by Constrained Maximum Likelihood

Yang, Bill Huajian

June 2017

SMOOTHING ALGORITHMS BY CONSTRAINED MAXIMUM LIKELIHOOD*
-Methodologies and implementations for CCAR stress testing and IFRS9 ECL estimation

Bill Huajian Yang

Abstract
In the process of loan pricing, stress testing, capital allocation, modeling of PD term structure, and IFRS9 expected credit loss estimation, it is widely expected that higher risk grades carry higher default risks, and that an entity is more likely to migrate to a closer non-default rating than a farther away non-default rating. In practice, sample estimates for rating level default rate or rating migration probability do not always respect this monotonicity rule, and hence the need for smoothing approaches. Regression and interpolation techniques are widely used for this purpose. A common issue with these approaches is that the risk scale for the estimates is not fully justified, leading to a possible bias in credit loss estimates. In this paper, we propose smoothing algorithms for rating level PD and rating migration probability. The smoothed estimates obtained by these approaches are optimal in the sense of constrained maximum likelihood, with a fair risk scale determined by constrained maximum likelihood, leading to more robust credit loss estimation. The proposed algorithms can be easily implemented by a modeller using, for example, the SAS procedure PROC NLMIXED. The approaches proposed in this paper will provide an effective and useful smoothing tool for practitioners in the field of risk modeling.

Keywords: Credit loss estimation**,** risk scale, constrained maximum likelihood, PD term structure, rating migration probability

1. Introduction

Given a risk-rated portfolio with k ratings $\{R_i \mid 1 \le i \le k\}$, we assume that rating R_1 is the best quality rating and R_k is the worst rating, i.e., the default rating. It is widely expected that higher risk ratings carry higher default risk, and that an entity is more likely to be downgraded (resp. upgraded) to a closer non-default rating than a farther away non-default rating. The following constraints are therefore required:

$$0 \le p_1 \le p_2 \le \ldots \le p_{k-1} \le 1 \tag{1.1}$$

$$p_{i\,i+1} \ge p_{i\,i+2} \ge \ldots \ge p_{ik-1} \tag{1.2}$$

$$p_{i\,1} \le p_{i\,2} \le \ldots \le p_{i\,i-1} \tag{1.3}$$

where $p_i, 1 \le i \le k-1$, denotes the probability of default for rating R_i, and $p_{i\,j}, 1 \le i, j \le k-1$, is the migration probability from a non-default initial rating R_i to a non-default rating R_j.

Estimates that satisfy the above monotonicity constraints are called smoothed estimates. Smoothed estimates are widely expected for rating level PD and rating migration probability in process of loan pricing, capital allocation, CCAR stress testing ([2]), modeling of probability of default (PD) term structure, and IFRS9 expected credit loss (ECL) estimation ([1]).

In practice, sample estimates for rating level PD and rating migration probability do not always respect these monotonicity rules. This calls for smoothing approaches. Regression and interpolation methods have been widely used for this purpose. A common issue with these approaches is that the risk scale for the estimates is not fully justified, leading to a possible bias estimate for the credit loss.

In this paper, we propose smoothing algorithms based on constrained maximum likelihood (CML). These CML smoothed estimates are optimal in the sense of constrained maximum likelihood, with a fair risk scale determined by constrained maximum likelihood, leading to a fair and more justified loss estimation. As shown by the empirical examples for rating level PD in section 2.3, the CML approach is more robust, compared to the logistic model and the log-linear model, with quality being measured based on the resulting likelihood ratio, the predicted portfolio level PD, and the impacted ECL.

Bill Huajian Yang, Royal Bank of Canada, 155 Wellington Street West, Toronto, Canada, M5V 3H6. The views expressed in this article are not necessarily those of Royal Bank of Canada or any of its affiliates. Please direct any comments to bill.yang@rbc.com

The paper is organized as follows: In section 2, we propose smoothing algorithms for smoothed rating level PD, for the cases with and with no default correlation. A smoothing algorithm for multinomial probability is proposed in section 3. Empirical examples are given accordingly in sections 2 and 3, where in section 2, we benchmark the CML approach for rating level PD with logistic model proposed by Tasche ([5]) and log-linear model proposed by Burgt ([6]).

2. Smoothing Rating Level Probability of Default

2.1. The proposed smoothing algorithm for rating level PD assuming no default correlation

Cross-section or within section default correlation may arise, due to some commonly shared risk factors. In which case, we assume that the sample is at a point-in-time, given the commonly shared risk factors; and that defaults occur independently given the commonly shared risk factors.

Let d_i and $(n_i - d_i)$ be respectively the observed default and non-default frequencies for a non-default risk rating R_i. Let p_i denote the probability of default for an entity with a non-default initial rating R_i. With no default correlation, we can assume that the default frequency follows a binomial distribution. Then the sample log likelihood is given by:

$$LL = \sum_{i=1}^{k-1} [(n_i - d_i)\log(1 - p_i) + d_i \log(p_i)] \tag{2.1}$$

up to a summand given by the logarithms of the related binomial coefficients, which are independent of $\{p_i\}$. By taking the derivative of (2.1) with respect to p_i and setting it to zero, we have:

$$-(n_i - d_i)/(1 - p_i) + d_i / p_i = 0$$
$$d_i(1 - p_i) = (n_i - d_i) p_i$$
$$\Rightarrow p_i = d_i / n_i$$

Therefore, the unconstrained maximum likelihood estimate for p_i is just the sample default rate d_i / n_i.

We propose the following smoothing algorithm for the case when no default correlation is assumed.

Algorithm 2.1. (Smoothing rating level PD assuming no default correlation)

(a) Parameterize the probability of default for a non-default rating R_i by:

$$p_i = \exp(b_1 + b_2 + ... + b_{k-i}) \tag{2.2}$$

where

$$b_{k-1} \le -\varepsilon_1,\ b_{k-2} \le -\varepsilon_2, ..., b_2 \le -\varepsilon_{k-2},\ b_1 \le 0 \tag{2.3}$$

for given constants $\varepsilon_i \ge 0,\ 1 \le i \le k-2$.

(b) Maximize, under constraint (2.3), the log likelihood (2.1) for parameters $\{b_1, b_2, ..., b_{k-1}\}$. Derive the smoothed estimates using (2.2). □

By (2.2) and (2.3), we have:

$$p_{k-1} = \exp(b_1) \le \exp(0) = 1,\quad p_i / p_{i-1} = \exp(-b_{k-i+1}) \ge \exp(\varepsilon_{i-1}) \ge 1$$
$$\Rightarrow 0 \le p_1 \le p_2 \le ... \le p_{k-1} \le 1$$

Thus monotonicity (1.1) is satisfied. When $\varepsilon_1 = \varepsilon_2 = ... = \varepsilon_{k-2} = \varepsilon \ge 0$, let $\rho = \exp(\varepsilon)$. Then ρ is the maximum lower bound for all the ratios $\{p_i / p_{i-1}\}$ of the smoothed estimates $\{p_i\}$.

2.2. The proposed smoothing algorithms for rating level PD assuming default correlation

Default correlation can be modelled by the asymptotic single risk factor (ASRF) model using asset correlation. Under the ASRF model framework, the risk for an entity is governed by a latent random variable z , called the firm's normalized asset value, which splits into two parts as ([3]):

$$z = s\sqrt{\rho} + \varepsilon\sqrt{1-\rho}\,,\ 0 < \rho < 1,\ s \sim N(0,1),\ \varepsilon \sim N(0,1) \qquad (2.4)$$

where s denotes the common systematic risk and ε is the idiosyncratic risk independent of s. The quantity ρ is called the asset correlation. It is assumed that there exist threshold values (i.e., the default points) $\{b_i\}$ such that an entity with an initial risk rating R_i will default when z falls below the threshold value b_i. The long-run PD for rating R_i is then given by $p_i = \Phi(b_i)$, where Φ denotes the standard normal cumulative distribution function.

Let $p_i\,(s)$ denote the probability of default for an entity with an initial risk rating R_i given the systematic risk s. It is shown ([8]) that

$$p_i\,(s) = \Phi(b_i\sqrt{1+r^2} - rs) \qquad (2.5)$$

where

$$r = \sqrt{\rho}\,/\sqrt{1-\rho}$$

Let $n_i(t)$ and $d_i\,(t)$ denote respectively the number of entities and the number of defaults at time t for $t = t_1, t_2, ..., t_q$. Given the latent factor s , we propose the following smoothing algorithm for rating level correlated long-run PDs by using (2.5).

Algorithm 2.2. (Smoothing rating level correlated long-run PDs given the latent systematic risk factor)

(a) Parameterize $p_i(s)$ for a non-default rating R_i by (2.5) with

$$b_i = (c_1 + c_2 + ... + c_{k-i}) \qquad (2.6)$$

where, for a given constants $\varepsilon \ge 0$, the following constraints are satisfied:

$$c_{k-1} \le -\varepsilon,\ c_{k-2} \le -\varepsilon, ..., c_2 \le -\varepsilon,\ c_1 \le 0 \qquad (2.7)$$

(b) Estimate parameters $\{c_1, c_2, ..., c_{k-1}\}$ by maximizing, under constraint (2.7), the log likelihood below:

$$LL = \sum_{h=1}^{q}\ \sum_{i=1}^{k-1}[(n_i(t_h) - d_i(t_h))\log(1 - p_i(s) + d_i(t_h)\log(p_i(s))] \qquad (2.8)$$

Set $p_i = \Phi(b_i)$. Then monotonicity (1.1) for $\{p_i\}$, i.e., the rating level long-run PDs, follows from constraints (2.6) and (2.7). □

Optimization with a random effect can be implemented by using, for example, SAS PROC NLMIXED ([4]).

When some key risk factors $x = (x_1, x_2, ..., x_m)$, common to all ratings, are observed, we assume the following decomposition for the systematic risk factor s :

$$s = -\lambda ci(x) - e\sqrt{1-\lambda^2}, \quad e \sim N(0,1), \ \ 0 < \lambda < 1$$

where $ci(x) = [a_1 x_1 + a_2 x_2 + ... + a_m x_m - u] / v$ is a linear combination of variables $x_1, x_2, ..., x_m$ with u and v being the mean and standard deviation of $a_1 x_1 + a_2 x_2 + ... + a_m x_m$.

Let $p_i(x)$ denote the probability of default given the scenario x. Assume that $ci(x)$ is standard normal independent of e. Then we have ([8, Theorem 2.2])

$$p_i(x) = \Phi[b_i \sqrt{1+\tilde{r}^2} + \tilde{r}\, ci(x)] \tag{2.9}$$

for some $\tilde{r}$.

Let $ci(x(t))$ denote the value of $ci(x)$ at time t for $t = t_1, t_2, ..., t_q$. Given the common index $ci(x)$, we propose the following smoothing algorithm for rating level correlated long-run PDs and rating level point-in-time PDs by using (2.9).

Algorithm 2.3. (Smoothing rating level correlated PDs given the common index $ci(x)$)

(c) Parameterize $p_i(x(t))$ for a non-default rating R_i by (2.6) with

$$b_i = (c_1 + c_2 + ... + c_{k-i}) \tag{2.10}$$

where, for a given constants $\varepsilon \geq 0$, the following constraints are satisfied:

$$c_{k-1} \leq -\varepsilon,\ c_{k-2} \leq -\varepsilon, ..., \ c_2 \leq -\varepsilon,\ c_1 \leq 0 \tag{2.11}$$

(d) Estimate parameters $\{c_1, c_2, ..., c_{k-1}\}$ by maximizing, under constraint (2.11), log likelihood below:

$$LL = \sum_{h=1}^{q} \sum_{i=1}^{k-1} [(n_i(t_h) - d_i(t_h)) \log(1 - p_i(x(t_h)) + d_i(t_h) \log(p_i(x(t_h)))] \tag{2.12}$$

Set $p_i = \Phi(b_i)$. Then monotonicity (1.1) for $\{p_i\}$, i.e., the rating level long-run PDs, and for $\{p_i(x(t_h))\}$ at time $t = t_h$, follows from constraints (2.10) and (2.11). □

2.3. Empirical examples: smoothing of rating level PDs

A. Example 1: Smoothing rating level long-run PDs assuming no default correlation

Table 1 shows the record count and default rate (DF Rate) for a sample created synthetically with 6 non-default risk ratings (RR):

Table 1. Sample count by rating

RR	1	2	3	4	5	6	Portfolio Level
DF	1	11	22	124	62	170	391
Count	5529	11566	29765	52875	4846	4318	108899
DF Rate	0.0173%	0.0993%	0.0739%	0.2352%	1.2833%	3.9442%	0.3594%

Algorithm 2.1 will be benchmarked by the following methods:

LGL1- With this approach, the PD for rating R_i is estimated by $p_i = \exp(a+bx)$, where x denotes the index for rating R_i, i.e., $x = i$ for rating R_i. Parameters a and b are estimated by a linear regression of the form below, using logarithm of the sample default rate for a rating:

$$\log(r_i) = a + bx + e,\ e \sim N(0, \sigma^2)$$

A common issue with this approach is the unjustified uniform risk scale b (in the log-space) for all ratings. Besides, this approach in general causes the portfolio level PD to be underestimated, due to the convexity of the exponential function (the 2nd derivative of the function $\exp(-)$ is positive):

$$E(y|x) = E(\exp(a+bx+e)|x) = \exp(a+bx+\sigma^2/2) > \exp(a+bx)$$

LGL2 – Like method LGL1, rating level PD is estimated by $p_i = \exp(a+bx)$. However, parameters a and b are estimated by maximizing the log likelihood given in (2.1). With this approach, the bias for portfolio PD can generally be avoided, though the issue with the unjustified uniform risk scale remains.

EXP-CDF – The method proposed by Burgt ([6]). With this approach, the rating level PD is estimated by $p_i = \exp(a+bx)$, where x denotes, for rating R_i, the adjusted sample cumulative distribution:

$$x(i) = (n_1 + n_2 + \ldots + n_{i-1} + n_i/2)/(n_1 + n_2 + \ldots + n_{k-1}) \qquad (2.13)$$

Instead of estimating parameters via cap ratio ([6]), we estimate parameters by maximizing the log likelihood given in (2.1).

LGST-INVCDF – The method proposed by Tasche ([5]). With this approach, the rating level PD is estimated by $p_i = 1/(1+\exp(a+b\Phi^{-1}(x)))$, where x is as in (2.13). Parameters are estimated by maximizing the log likelihood given in (2.1).

Estimation quality is measured by the following:

P-Value – The p-value calculated from the likelihood ratio chi-squared test with degree freedom equal to the number of restrictions. Higher p-value indicates a better model.

ECL Ratio – The ratio of expected credit loss based on the smoothed rating level PDs in relative to that based on the realized rating level PDs, given the EAD and LGD parameters for each rating. A significantly lower ECL ratio value indicates a possible underestimation of the credit loss.

PD Ratio – The ratio of the portfolio level PD aggregated from the smoothed rating level PDs in relative to the portfolio level PD aggregated from the realized rating level PDs. A value significantly lower than 100% for the PD ratio indicates a possible underestimation for the PD at portfolio level.

Table 2 shows results for Algorithm 2.1 (labelled as CML) when $\varepsilon_1 = \varepsilon_2 = \ldots = \varepsilon_{k-2} = 0$, and the benchmarks, where smoothed rating level PDs are listed in columns 2-7.

Table 2. Smoothed results by Algorithm 2.1 and benchmarks

								Portfolio Level	
Method	P1	P2	P3	P4	P5	P6	P-Value	ECL Ratio	PD Ratio
CML	0.0173%	0.0810%	0.0810%	0.2352%	1.2833%	3.9442%	95.92%	99.91%	100.00%
LGL1	0.0165%	0.0416%	0.1053%	0.2663%	0.6732%	1.7022%	0.00%	46.09%	72.57%
LGL2	0.0032%	0.1468%	0.2901%	0.4333%	0.5763%	0.7191%	0.00%	27.58%	100.07%
EXP-CDF	0.0061%	0.0086%	0.0294%	0.3431%	1.9081%	2.5057%	0.00%	72.92%	100.21%
LGST-INVCDF	0.0104%	0.0188%	0.0585%	0.2795%	1.5457%	3.4388%	0.00%	90.46%	100.00%

Results show, the Algorithm 2.1 outperforms significantly the other benchmarks by p-value, impacted ECL, and aggregated portfolio level PD. The first log-linear model (LGL1) underestimates the portfolio level PD significantly. All log linear models LGL1, LGL2, and EXP-CDF underestimate the ECL significantly.

Table 3 illustrates the strictly monotonic smoothed rating level PDs by Algorithm 2.1 when $\varepsilon_1 = \varepsilon_2 = ... = \varepsilon_{k-2} = \varepsilon > 0.$ While p-value deteriorates quickly as ε increases from 0 to 1, the impacted ECL does not change that much, however.

Table 3. Strictly monotonic smoothed rating level PDs

								Portfolio Level	
Epsilon	P1	P2	P3	P4	P5	P6	P-Value	ECL Ratio	PD Ratio
0	0.0173%	0.0810%	0.0810%	0.2352%	1.2833%	3.9442%	95.92%	99.91%	100.00%
0.1	0.0173%	0.0753%	0.0832%	0.2352%	1.2833%	3.9442%	89.06%	99.88%	100.00%
0.5	0.0173%	0.0552%	0.0910%	0.2352%	1.2833%	3.9442%	36.63%	99.79%	100.00%
1	0.0120%	0.0327%	0.0890%	0.2419%	1.2833%	3.9442%	2.54%	99.63%	100.00%

B. Example 2: Smoothing rating level long-run PDs in presence of default correlation

The sample created synthetically contains the quarterly default count by rating for a portfolio with 6 non-default ratings between 2005Q1 and 2014Q4. Point-in-time default rate (rating level or portfolio level) is calculated for each quarter and then averaged over the sample window by dividing the number of quarters (44) to get the estimate for the long-run average realized PD (labelled as AVG PD). Sample distribution (labelled as Overall Distribution) by rating is calculated by combining all 44 quarters. Table 4 displays sample statistics (with heavy size concentration at rating R_4):

Table 4. Long-run default rate by rating calculated from the sample

RR	1	2	3	4	5	6	Portfolio
Long-Run AVG PD	0.0215%	0.1027%	0.0764%	0.2731%	1.1986%	3.8563%	0.3818%
Overall Distribution	5.07%	10.61%	27.47%	48.32%	4.52%	4.01%	100.00%

Table 5 shows the smoothed correlated rating level long-run PD for all 6 non-default ratings by using Algorithm 2.2.

Estimation quality is measured by the following:

AIC – Akaike information criterion. Lower AIC indicates a better model.

PD Ratio - The ratio of the long-run average predicted portfolio level PD (labelled AVG PD) relative to the long-run average realized portfolio level PD. A value significantly lower than 100% for this ratio indicates a possible underestimation for the PD at portfolio level.

The first row in Table 5 shows results for the case when no default correlation is assumed (labelled as "no correl") and ε (labelled as Epsilon) is chosen to be 0, while for the 2nd row, results are for the case when default correlation is assumed (labelled as "w correl") and ε is equal to 0.

Table 5. Smoothed correlated long-run rating level PDs

								Portfolio Long-Run PD	
Epsilon	P1	P2	P3	P4	P5	P6	AIC	AVG PD	PD Ratio
0 (no correl)	0.0179%	0.0836%	0.0836%	0.2371%	1.3076%	4.0372%	694.02	0.3710%	97.17%
0 (w correl)	0.0183%	0.0828%	0.0828%	0.2545%	1.1951%	3.9340%	594.62	0.3843%	100.66%
0.1 (w correl)	0.0183%	0.0483%	0.0966%	0.2541%	1.1942%	3.9318%	600.79	0.3842%	100.64%
0 .2 (w correl)	0.0035%	0.0176%	0.0754%	0.2775%	1.1859%	3.9237%	617.96	0.3842%	100.64%
0.3 (w correl)	0.0010%	0.0086%	0.0560%	0.2905%	1.1961%	3.9342%	637.25	0.3845%	100.71%

Results in the 1st row show, the estimated long-run portfolio level PD for the case assuming no default correlation is lower than the case when default correlation is assumed (2nd row), which suggests we may have underestimated the long-run rating level PD when assuming no default correlation. The high AIC value in the first row implies that the assumption of no default correlation may not be appropriate.

It is worth mentioning that, when applying Algorithm 2.2 to the sample used in Example 1, assuming no default correlation, we got exactly the same estimates as in Example 1.

3. Smoothing Algorithms for Multinomial Probability

3.1. Unconstrained maximum likelihood estimates for multinomial probability

For n independent trials, where each trial results in exactly one of h fixed outcomes, the probability of observing frequencies $\{n_i\}$, with frequency n_i for the i^{th} ordinal outcome, is:

$$\frac{n!}{n_1!n_2!\ldots n_h!} x_1^{n_1} x_2^{n_2} \ldots x_h^{n_h} \tag{3.1}$$

where $x_i > 0$ is the probability of observing the i^{th} ordinal outcome in a single trial, and

$$n = n_1 + n_2 + \ldots + n_h, \quad x_1 + x_2 + \ldots + x_h = 1$$

The natural logarithm of the likelihood is:

$$LL = n_1 \log x_1 + n_2 \log x_2 + \ldots + n_h \log x_h \tag{3.2}$$

up to a constant given by the logarithm of some multinomial coefficient independent of parameters $\{x_1, x_2, \ldots, x_h\}$. By using the relation $x_h = 1 - x_1 - x_2 - \ldots - x_{h-1}$ and setting to zero the derivative of (3.2) with respect to x_i, $1 \le i \le h-1$, we have:

$$n_i / x_i - n_h / (1 - x_1 - x_2 - \ldots - x_{h-1}) = 0$$
$$\Rightarrow n_i / x_i = n_h / x_h$$

Since this holds for each i and for the fixed h, we conclude that the vector $(x_1, x_2, \ldots, x_h)$ is in proportion with $(n_1, n_2, \ldots, n_h)$. Thus, the maximum likelihood estimate for x_i is the sample estimate:

$$x_i = n_i / (n_1 + n_2 + \ldots + n_h) = n_i / n \tag{3.3}$$

3.2. The proposed smoothing algorithm for multinomial probability

We propose the following smoothing algorithm for multinomial probability under the constraint below:

$$0 \le x_1 \le x_2 \le \ldots \le x_h \le 1 \tag{3.4}$$

Algorithm 3.1. (Smoothing multinomial probability)

(a) Parameterize the multinomial probability by:

$$x_i = \exp(b_1 + b_2 + \ldots + b_{h+1-i}) / \Delta \tag{3.5}$$
$$\Delta = \exp(b_1) + \exp(b_1 + b_2) + \ldots + \exp(b_1 + b_2 + \ldots + b_h)$$

(b) Maximize (3.2), with x_i being given by (3.5), for parameters $b_1, b_2, \ldots, b_h$ subject to:

$$b_h \le -\varepsilon_1,\ b_{h-1} \le -\varepsilon_2,\ \ldots,\ b_2 \le -\varepsilon_{h-1},\ b_1 \le 0 \tag{3.6}$$

for $\varepsilon_i \geq 0,\ 1 \leq i \leq h-1.$ Derive the CML smoothed estimates by using (3.5). Then monotonicity (3.4) for the estimates follows from (3.5) and (3.6).

□

In the case when $\varepsilon_1 = \varepsilon_2 = ... = \varepsilon_{h-1} = \varepsilon \geq 0,$ let $\rho = \exp(\varepsilon).$ Then ρ is the maximum lower bound for all the ratios $\{x_i / x_{i-1}\}$.

3.3. An empirical example: Smoothing transition probability matrix

Rating migration matrix models ([3], [7]) are widely used for IFRS9 expected credit loss estimation and CCAR stress testing. Given a non-default risk rating R_i, let n_{ij} be the observed long-run transition frequency from R_i to R_j at the end of the horizon, and $n_i = n_{i1} + n_{i2} + ... + n_{ik}$. Let p_{ij} be the long-run transition probability from R_i to R_j. By (3.3), the maximum likelihood estimate for p_{ij} observing the long-run transition frequencies $\{n_{ij}\}$ for a fixed i is:

$$p_{ij} = n_{ij} / n_i \tag{3.7}$$

It is widely expected that higher risk grades carry greater default risk, and that an entity is more likely to be downgraded (resp. upgraded) to a closer non-default rating than a farther away non-default rating. The following constraints are hence required:

$$p_{i\,i+1} \geq p_{i\,i+2} \geq ... \geq p_{ik-1} \tag{3.8}$$

$$p_{i1} \leq p_{i2} \leq ... \leq p_{i\,i-1} \tag{3.9}$$

$$p_{1k} \leq p_{2k} \leq ... \leq p_{k-1\,k} \tag{3.10}$$

The constraint (3.10) is for rating level probability of default, which has been discussed in section 2.

Smoothing the long-run migration matrix involves the following steps:

(a) Rescale migration probabilities $\{p_{i1}, p_{i2}, ..., p_{i\,i-1}\}$ in (3.9) to make them sum to 1, then find the CML smoothed estimates by using Algorithm 3.1, and rescale these CML estimates back to have the same summed value for $\{p_{i1}, p_{i2}, ..., p_{i\,i-1}\}$ as that before smoothing. Do the same for (3.8).

(b) Find the CML smoothed estimates by using Algorithm 2.1 for rating level default rate. Keep these CML default rate estimates unchanged, while rescaling for each non-default rating R_i the non-default migration probabilities $\{p_{i1}, p_{i2}, ..., p_{i\,k-1}\}$ so that the entire row $\{p_{i1}, p_{i2}, ..., p_{ik}\}$ sums to 1.

Table 6 below shows empirical results using Algorithms 2.1 and 3.1 for smoothing the long-run migration matrix, where for Algorithm 3.1 all ε_i are set to zero.

The sample used here is created synthetically. It consists of historical quarterly rating transition frequency for a commercial portfolio from 2005Q1 to 2015Q4. There are 7 risk ratings, with R_1 as best quality rating and R_7 being default rating.

The left-hand-side of the table shows sample estimates for long-run transition probabilities before smoothing, while the right-hand-side shows CML smoothed estimates. There are three rows as highlighted in pink in the left-hand-side of the table, where sample estimates violate (3.8) or (3.9) (but (3.10) is satisfied). Rating level sample default rates (the column labelled as "p.7") does not require smoothing.

As shown in the table, the CML smoothed estimates are the simple average of the relevant non-monotonic sample estimates (For the structure of CML smoothed estimates for multinomial probabilities, we show theoretically in a separate paper that the CML smoothed estimate for an ordinal class is either the sample estimate or the simple average of the sample estimates for some consecutive ordinal classes including the named class).

Table 6. Long-run transition probability matrices before and after smoothing

Transition probability before smoothing							Transition probability after smoothing						
p.1	p.2	p.3	p.4	p.5	p.6	p.7	p.1	p.2	p.3	p.4	p.5	p.6	p.7
0.97162	0.01835	0.00312	0.00554	0.00104	0.00017	0.00017	0.97162	0.01835	0.00433	0.00433	0.00104	0.00017	0.00017
0.00621	0.94528	0.03071	0.01284	0.00215	0.00257	0.00025	0.00621	0.94528	0.03071	0.01284	0.00236	0.00236	0.00025
0.00071	0.01028	0.93803	0.04089	0.00659	0.00277	0.00074	0.00071	0.01028	0.93803	0.04089	0.00659	0.00277	0.00074
0.00024	0.00069	0.01260	0.96726	0.01261	0.00543	0.00118	0.00024	0.00069	0.01260	0.96726	0.01261	0.00543	0.00118
0.00039	0.00118	0.00790	0.07996	0.82725	0.07048	0.01283	0.00039	0.00118	0.00790	0.07996	0.82725	0.07048	0.01283
0.00022	0.00133	0.00266	0.04498	0.01197	0.89940	0.03944	0.00022	0.00133	0.00266	0.02847	0.02847	0.89940	0.03944

Conclusions. Regression and interpolation approaches are widely used for smoothing rating transition probability and rating level probability of default. A common issue with these methods is that the risk scale for the estimates is not on a strong mathematical footing, leading to possible bias in credit loss estimation. In this paper, we propose smoothing algorithms based on constrained maximum likelihood for rating level probability of default and for rating migration probability. These smoothed estimates are optimal in the sense of constrained maximum likelihood, with a fair risk scale determined by constrained maximum likelihood, leading to a fair and more justified credit loss estimation. These algorithms can be implemented by a modeller using, for example, the SAS PROC NLMIXED.

Acknowledgements: The author would like to thank both referees for suggesting the extended discussion to cover the case when default correlation is assumed, and the likelihood ratio test for the constrained maximum likelihood estimates. Special thanks to Carlos Lopez for his consistent inputs, insights, and supports for this research. Thanks also go to Clovis Sukam and Biao Wu for their critical reading for this manuscript, and Zunwei Du, Wallace Law, Glenn Fei, Kaijie Cui, Jacky Bai, and Guangzhi Zhao for many valuable conversations.

REFERENCES

[1] Ankarath, N., Ghost, T.P., Mehta, K.J., Alkafaji, Y. A. (2010), Understanding IFRS Fundamentals, John Wiley & Sons, Inc.

[2] Board of Governors of the Federal Reserve System (2016). Comprehensive Capital Analysis and Review 2016 Summary and Instructions, January 2016.

[3] Miu, P., Ozdemir, B. (2009). Stress testing probability of default and rating migration rate with respect to Basel II requirements, Journal of Risk Model Validation, Vol. 3 (4) Winter 2009

[4] SAS Institute Inc., Cary, NC. (2009). SAS 9.2 User's Guide, The NLMIXED Procedure.

[5] Tasche, D. (2013). The art of PD curve calibration, Journal of Credit Risk, 9 (4), December 2013, 63-103. DOI: 10.21314/JCR.2013.169

[6] Van der Burgt, M. (2008), Calibrating low-default portfolios, using the cumulative accuracy profile, Journal of Risk Model validation, 1 (4), 17-33.

[7] Yang, B. H., Du, Zunwei (2016). Rating Transition Probability Models and CCAR Stress Testing, Journal of Risk Model Validation 10 (3), 2016, 1-19. DOI: 10.21314/JRMV.2016.155

[8] Yang, B. H., Point-in-time PD term structure models for multi-period scenario loss projections, Journal of Risk Model Validation, Vol 11 (1), Spring 2017. DOI:10.21314/JRMV.2017.164