Applied Consensus Information and Consensus Rating
A Simulation Study on Rating Aggregation
von
Christoph Lehmann und Daniel Tillich
Applied Consensus Information and Consensus Rating
A Simulation Study on Rating Aggregation

Christoph Lehmann and Daniel Tillich

December 17, 2015

The aggregation of different single ratings to a so called ‘consensus rating’ to get a higher precision of a debtor’s default probability is an idea that is hardly discussed in the literature. Grün, Hofmarcher, Hornik, Leitner & Pichler (2013) came up with a method for rating aggregation, whereby the term ‘consensus rating’ was introduced. For sharpening the whole issue of rating aggregation from a theoretical perspective, Lehmann & Tillich (2014) developed a framework, wherein the terms ‘consensus rating’ and ‘consensus information’ are clearly defined.

The following article tries to connect the two aforementioned contributions and applies the theoretical framework of Lehmann & Tillich (2014) in connection with some practical ideas of Grün et al. (2013). In contrast to Grün et al. (2013), a simulation approach is chosen in order to have a clear benchmark for assessing the rating aggregation outcomes. Thereby, the following questions should be clarified: Does rating aggregation really lead to a higher precision of the estimated default probabilities? Is there a preferable aggregation method? Does the consensus rating, as defined by Lehmann & Tillich (2014), outperforms other aggregation methods? The simulation results show that rating aggregation is an extremely questionable issue.

1 Introduction

The problem of rating aggregation seems not to be discussed so much in the literature in the last years. One of the first articles in this field is Grün et al. (2013), where an approach is developed that combines different ratings into a so called ‘consensus rating’. In contrast, Lehmann & Tillich (2014) discussed the issue from a slightly more basic perspective, what a ‘consensus rating’ is and in which cases this concept makes sense. In Grün et al. (2013) the suggested model for rating aggregation is applied to a real data set. Thereby, the main problem is, that the true default probabilities are unknown and the benchmark chosen is only data driven. The approach of this contribution is a simulation, whereby an artificial world with known default probabilities—based on a logit model—is constructed, i.e. a fixed benchmark is used in contrast to Grün et al. (2013). As already mentioned in Lehmann & Tillich (2014), there are many aspects leading to different estimates of the same default probability, e.g.
different models, estimation methods, data sets and time horizons. Furthermore, the concept of consensus rating seems to be a theoretical issue, because a rating agency strives to get more precise information than its competitors and especially interchange and recombination of information is not desired. Using a simulation—as done in this paper—contains the possibility to control these manyfold aspects. As a consequence, it becomes possible to gain insights on the performance of different aggregation approaches.

This article is organized as follows. In Section 2, some notation and theoretical background is introduced in brief. Section 3 contains the settings of the simulation and the methods of rating aggregation. Section 4 presents the results. Finally, the article ends with some conclusions in Section 5.

2 Information set and consensus information

For the following notation and assumptions cf. Lehmann & Tillich (2014, Section 1 and 2). The credit default of debtor \(i\) is modeled by a random variable \(Y_i\), \(i = 1, \ldots, n\). It takes the value 1 in the case of default of debtor \(i\) and 0 otherwise. Thus, \(P(Y_i = 1)\) is the unconditional default probability of debtor \(i\). In order to estimate individual default probabilities, typically several rating characteristics are taken into account. These rating characteristics are modeled by a subject specific real random vector \(X_i = (X_{i1}, X_{i2}, \ldots, X_{iK})\) with realization \(x_i = (x_{i1}, x_{i2}, \ldots, x_{iK})\) \(\in \mathbb{R}^K\), \(i = 1, \ldots, n\). Then, the probability of interest in a rating process is the conditional default probability \(P(Y_i = 1 | X_i = x_i)\).

In the style of Lehmann & Tillich (2014), it is assumed that all rating agencies use the same vector \(X_i\). This assumption is needed in the sequel for reasonable set operations. In this framework, the differences do not lie in the rating characteristics \(X_i\) themselves, but in the information about them. The situation, where all the values of debtor \(i\) are known, is called complete information. The corresponding information set is \(I_i = \{x_i\} = \{x_{i1}\} \times \{x_{i2}\} \times \ldots \times \{x_{iK}\}\). It is a singleton. Typically, rating agencies don’t have complete information, but incomplete information. They know and use only subvectors of \(x_i\). The subvector belonging to rating agency \(r = 1, \ldots, R\) is denoted by \(x_{ri}\). Its corresponding information set \(I_{ri}\) differs from the complete information set \(I_i\) in that the unknown values are replaced by the real numbers. It is assumed that a rating agency either knows the exact value of a rating characteristic or nothing about it. It follows \(I_i \subseteq I_{ri}\).

Example 1 Rating agency \(r = 1\) has information only about the rating characteristics 1, 3, 5 and 7, i.e. \(x_{11} = (x_{i1}, x_{i3}, x_{i5}, x_{i7})\). Assuming vector \(x_i\) consists of \(K = 8\) components, the resulting information set is \(I_{1i} = \{x_{i1}\} \times \mathbb{R} \times \{x_{i3}\} \times \mathbb{R} \times \{x_{i5}\} \times \mathbb{R} \times \{x_{i7}\} \times \mathbb{R}\).

Because of their different information sets, rating agencies estimate different conditional probabilities \(P(Y_i = 1 | X_i \in I_{ri})\). In order to estimate the same probability and for getting close to the complete information as much as possible, the information sets of the rating agencies should be merged in this framework. This leads to a consensus information. Since

---

1 In the following, an institution that assigns ratings is called ‘rating agency’. This also could be a bank which evaluates the debtors.

2 This example corresponds to the information set of rating agency 1 in the simulation scenario B2 below (cf. Section 3, Table 3).
there is no contradictory information, the consensus information set of the i-th debtor $I_i$ can be defined as (cf. Definition 1 in Lehmann & Tillich (2014))

$$I_i \triangleq \bigcap_{r=1}^{R} I_{ri}.$$  \hspace{1cm} (1)

Here, it holds $I_i \subseteq I_i \subseteq I_{ri}$. Based on the consensus information set $I_i$, the subvector $x_i$ can be constructed. It contains all values of the rating characteristics that are known to at least one rating agency.

**Example 2** Let

$$I_{1i} = \{x_{i1}\} \times \mathbb{R} \times \{x_{i3}\} \times \mathbb{R} \times \{x_{i5}\} \times \mathbb{R} \times \{x_{i7}\} \times \mathbb{R}$$

as in Example 2 and

$$I_{2i} = \{x_{i1}\} \times \{x_{i2}\} \times \mathbb{R} \times \{x_{i5}\} \times \mathbb{R} \times \{x_{i8}\}.$$  

Then the consensus information set corresponding to the information sets of the two rating agencies is

$$I_i = I_{1i} \cap I_{2i} = \{x_{i1}\} \times \{x_{i2}\} \times \{x_{i3}\} \times \mathbb{R} \times \{x_{i5}\} \times \{x_{i7}\} \times \{x_{i8}\}.$$  

The resulting subvector of the rating characteristics is $x_i = (x_{i1}, x_{i2}, x_{i3}, x_{i5}, x_{i7}, x_{i8}).$

### 3 Simulation

Every information set leads to a conditional default probability, e.g. $P(Y_i = 1|X_i \in I_{ri}) \in \]0,1[$. A rating agency calculates an estimate for this unknown default probability. Rating aggregation typically means to aggregate such different estimates. To get a consensus rating in the sense of estimating the same target, rating agencies have to use the same information set, e.g. the consensus information set. In this section, some different aggregation variations as well as the concept of consensus rating are simulated and compared. Having in mind the theoretical considerations from above, it should be clarified, whether a real consensus rating performs better than other forms of rating aggregation.

The basic idea for the simulation is as following: At first, an artificial world is constructed by a complete specified logit model where all real default probabilities are known. In the second step, defaults are simulated based on these known probabilities. On this data set, the default probabilities are estimated, given different subsets of the information set that is used at the beginning. Fourth and finally, the estimated probabilities are aggregated in different ways. Thus, a comparison between the different estimated probabilities and the „true“ probabilities is possible. The second to fourth step are replicated in a Monte-Carlo-simulation. As a basis of the simulation the logit function is needed, that is defined as

$$G(z) \triangleq \frac{1}{1 + \exp(-z)} \quad \text{for all } z \in \mathbb{R}.$$  

Its inverse is

$$G^{-1}(\pi) = \ln \left( \frac{\pi}{1 - \pi} \right) \quad \text{for all } 0 < \pi < 1.$$  

In the following, all steps are described in detail.
Step 1: At first, the rating characteristics and the corresponding default probability for all of the \( n \) debtors are needed. To this, \( n \) realizations \( x_i = (x_{i1},\ldots,x_{iK})' \) of random vectors \( X_i = (X_{i1},\ldots,X_{iK})' \) i.i.d. \( F_X \) are generated with an arbitrary distribution \( F_X \). Thus, \( I_i = \{x_i\} \). Next, the true default probability for debtor \( i \) is calculated by\
\[
\pi(I_i) \overset{\text{def}}{=} P(Y_i = 1|X_i \in I_i) \overset{\text{def}}{=} G(\beta_0 + \beta' x_i), \quad i = 1,\ldots,n, \tag{2}
\]
where \( \beta_0 \in \mathbb{R} \) and \( \beta = (\beta_1,\ldots,\beta_K)' \in \mathbb{R}^K \) denote the known parameters.

Step 2: Based on the true default probability \( \pi(I_i) \) one realization \( y_i \) of the default variable \( Y_i \) is simulated for every debtor \( i = 1,\ldots,n \).

Step 3: a) Based on the default data from Step 2 the parameters \( \beta_0,\beta_1,\ldots,\beta_K \) of model (2) are estimated. With the estimates \( \hat{\beta}_0 \) and \( \hat{\beta} = (\hat{\beta}_1,\ldots,\hat{\beta}_K)' \), the default probability \( \pi(I_i) \) is estimated by\
\[
\hat{\pi}(I_i) \overset{\text{def}}{=} G(\hat{\beta}_0 + \hat{\beta}' x_i).
\]
b) Additionally, some modifications of model (2) are estimated. The modifications refer to the consideration of different rating agency-specific information sets \( I_{ri} \). In detail, the rating agencies know and/or use only subvectors of \( x_i \). These subvectors \( x_{ri} \) have dimension \( 0 < K_r \leq K \). From these subvectors and the default data, a reduced number of parameters is estimated, namely \( \beta_{r0} \) and \( \beta_r = (\hat{\beta}_{r1},\ldots,\hat{\beta}_{rK_r})' \). With the estimates \( \hat{\beta}_{r0} \) and \( \hat{\beta}_r = (\hat{\beta}_{r1},\ldots,\hat{\beta}_{rK_r})' \), the agency-specific default probabilities\
\[
\pi(I_{ri}) = P(Y_i = 1|x_i \in I_{ri}), \quad i = 1,\ldots,n, \quad r = 1,\ldots,R, \tag{3}
\]
are estimated by\
\[
\hat{\pi}(I_{ri}) \overset{\text{def}}{=} G(\hat{\beta}_{r0} + \hat{\beta}_r' x_{ri}). \tag{4}
\]
c) Last, the consensus information set \( I^\cap_i \) is used for estimation. From the corresponding subvectors \( x^\cap_i \) and the default data the default probabilities \( \pi(I^\cap_i) \) are estimated by \( \hat{\pi}(I^\cap_i) \) analogously to Step 3b).

All estimations are done by ML-estimation.

Step 4: For all debtors \( i = 1,\ldots,n \), aggregated (compromise) default probabilities are derived from the rating agencies’ estimates \( \hat{\pi}(I_{ri}), \quad r = 1,\ldots,R \), in (4). This is the investor’s or external perspective, where no information about the rating characteristics is used.

a) An aggregated default probability is calculated as an arithmetic mean:
\[
\bar{\pi}_i \overset{\text{def}}{=} \frac{1}{R} \sum_{r=1}^{R} \hat{\pi}(I_{ri}), \quad i = 1,\ldots,n.
\]

b) Another form of aggregation is done by using the geometric mean:
\[
\bar{\pi}_{G_{ii}} \overset{\text{def}}{=} \sqrt[\prod_{r=1}^{R} \hat{\pi}(I_{ri})}, \quad i = 1,\ldots,n.
\]
c) Taking into account the benchmark idea of Grün et al. (2013, p.82), a third aggregation method based on the so called Z-scores is used. Formally, the Z-scores are simply the estimated linear predictors of the logit model in (4), i.e. 
\[ z_{ri} \overset{\text{def}}{=} G^{-1}(\hat{\pi}(I_{ri})) = \hat{\beta}_r + \hat{\beta}'_r \mathbf{x}_{ri}. \]
Calculating their arithmetic mean 
\[ \bar{z}_i \overset{\text{def}}{=} \frac{1}{R} \sum_{r=1}^{R} z_{ri}, \quad i = 1, \ldots, n, \]
finally leads to the corresponding aggregated default probability \( \bar{\pi}_S_i \overset{\text{def}}{=} G(\bar{z}_i) \).

The Steps 2 to 4 are replicated \( m \) times, i.e. \( m \) defaults are simulated for each individual (based on the default probability from Step 1). In contrast to Grün et al. (2013), there are no time dynamics considered, it is a simulation only for one period of time.

There are mainly three advantages of the simulation framework above.

1. In the artificial world a real consensus information in the sense of (1) is possible and therewith a real consensus rating.
2. The true default probability as a fixed benchmark is known.
3. The aggregated ratings are based on the same model, estimation method, data set and time horizon. Only the information sets differ.

Simulations as described above are performed with settings as following:

- Number of rating agencies: \( R = 3 \),
- Number of debtors: \( n = 4000 \),
- Number of MC-replications: \( m = 100 \).

All simulations were implemented with GAUSS Vers. 15.0.7 (seed for random numbers: 1664525) using the MAXLIK-package Vers. 5.0.9. The graphics were generated with R Vers. 3.1.2 (R Core Team (2014)) and the package ggplot2 Vers. 1.0.1.

Two basic scenarios, A and B, of simulations are investigated. Their specifications are given in columns 2 and 3 of Table 2 and Table 3. Both scenarios contain the same types of distributions for the regressor variables, namely Lognormal, Poisson and Bernoulli distribution. The Lognormal distribution producing only positive real numbers could indicate income for example. The Poisson distribution produces non-negative integers and could stand for the number of loans or the number of credit cards a person already has. Finally the Bernoulli distribution with values zero and one indicates some dichotomous characteristic like sex. The coefficients \( \beta_0, \beta_1, \ldots, \beta_K \) are set in this way to get two very different scenarios for the generated default probabilities. Scenario A contains much higher default probabilities than scenario B. Thus scenario B forms the more practical scenario, especially in the case of credit defaults. Nonetheless, in the face of a crisis scenario A seems not to be so absurd. Histograms of the true default probabilities \( \pi(I_i) \) for A and B as calculated in Step 1 are illustrated in Figure 1. Additionally, Table 1 provides some descriptive statistics.

For each Scenario A and B of default probabilities, four subscenarios are considered. The subscenarios 1–4 differ in the information sets of the rating agencies and therefore in the resulting consensus information set. The rating characteristics used for estimating can be
Figure 1: Histograms of the true default probabilities (log scale) for simulation scenarios A and B

Table 1: Descriptive statistics of true default probabilities for scenarios A and B
read from the column „Choice matrix“ of Tables 2 and 3 as follows: Every column of the choice matrix stands for one of the rating characteristics that are used, i.e. every column symbolizes one regressor variable from the logit model in 3. Thereby, the first column represents the intercept (this is the coefficient \( \beta_0 \)) and the second column up to the last one represent the rating characteristics \( x_{i1}, \ldots, x_{iK} \) (with the coefficients \( \beta_1, \ldots, \beta_K \)). Every row of the matrix stands for one of the model modifications in 2 and 3. More precisely, the first row represents the „true-world“ model, the second to the fourth represent the \( R = 3 \) different rating agencies and the fifth row contains the case of the consensus information (cf. Step 3c)). Thereby, the choice matrix only contains ones and zeros, that indicate whether the \( k \)-th regressor variable is used (one) or not (zero) within the considered model modification. Within the framework from Section 2 for every regressor variable not used by rating agency \( r \) it holds \( X_{ik} \in \mathbb{R} \). Therefore, the regressor variable \( X_{ik} \) can be omitted in the estimation. In the case of used regressor variables it holds \( X_{ik} = x_{ik} \), i.e. the rating agency knows the correct value of debtor \( i \).

4 Results

In the following, comparisons are made between the results of the different approaches from above. At first, the plausibility of the estimated values is assessed by error rates. The error rates describe the portion of debtors, for which the true default probability does not lie between the minimum and the maximum of the estimated default probabilities over all MC-replications. For formal notation let \( |A| \) denote the number of elements of an arbitrary set \( A \). Then the error rates are defined as follows, \( r = 1, 2, 3 \):

\[
e_r \overset{\text{def}}{=} 1 - \frac{|A_r|}{n}, \quad e_r' \overset{\text{def}}{=} 1 - \frac{|A_r'|}{n}, \quad e_{\cap} \overset{\text{def}}{=} 1 - \frac{|A_{\cap}|}{n}, \quad e_{\bar{G}} \overset{\text{def}}{=} 1 - \frac{|A_{\bar{G}}|}{n},
\]

with

\[
A \overset{\text{def}}{=} \left\{ i : \min_{j=1}^m \hat{\pi}^{(j)}(I_i) \leq \pi(I_i) \leq \max_{j=1}^m \hat{\pi}^{(j)}(I_i) \right\}, \quad A_r \overset{\text{def}}{=} \left\{ i : \min_{j=1}^m \hat{\pi}^{(j)}(I_{ir}) \leq \pi(I_i) \leq \max_{j=1}^m \hat{\pi}^{(j)}(I_{ir}) \right\}, \quad A_{\cap} \overset{\text{def}}{=} \left\{ i : \min_{j=1}^m \hat{\pi}^{(j)}(I_i^\cap) \leq \pi(I_i) \leq \max_{j=1}^m \hat{\pi}^{(j)}(I_i^\cap) \right\}, \quad A_{\bar{G}} \overset{\text{def}}{=} \left\{ i : \min_{j=1}^m \hat{\pi}_{G1}^{(j)} \leq \pi(I_i) \leq \max_{j=1}^m \hat{\pi}_{G1}^{(j)} \right\}, \quad A_{\bar{S}} \overset{\text{def}}{=} \left\{ i : \min_{j=1}^m \hat{\pi}_{S1}^{(j)} \leq \pi(I_i) \leq \max_{j=1}^m \hat{\pi}_{S1}^{(j)} \right\},
\]

where the term \( (j) \) indicates the corresponding value of the \( j \)-th MC-replication. E.g. \( \hat{\pi}^{(j)}(I_i^\cap) \) denotes the estimated default probability of debtor \( i \) in MC-replication \( j \) using the consensus information set \( I_i^\cap \).

In the following some important results/remarks on Tables 2 and 3 are reported.
Table 2: Simulation scenarios A1 to A4 with $r = 1, 2, 3, i = 1, \ldots, 4000, j = 1, \ldots, 100$ and corresponding error rates

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Rating characteristics</th>
<th>$(\beta_0, \beta_1, \ldots, \beta_K)'$</th>
<th>Choice matrix</th>
<th>Error rates</th>
<th>Error rate $\bar{e}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td></td>
<td>$\begin{bmatrix} 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 \ 1 &amp; 1 &amp; 1 &amp; 1 &amp; 0 &amp; 0 &amp; 0 &amp; 0 \ 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 0 &amp; 1 &amp; 0 &amp; 0 \ 1 &amp; 1 &amp; 1 &amp; 1 &amp; 0 &amp; 0 &amp; 1 &amp; 0 &amp; 0 \ 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 0 &amp; 0 &amp; 0 \ \end{bmatrix}$</td>
<td>$e = 0.00$</td>
<td>$e_1 = 0.89$ $\bar{e} = 0.89$ $e_2 = 0.85$ $\bar{e}_G = 0.90$ $e_3 = 0.86$ $\bar{e}<em>S = 0.89$ $e</em>{\cap} = 0.47$</td>
<td></td>
</tr>
<tr>
<td>A2</td>
<td>$X_{i1} \sim LN(0.25, 0.35)$, $X_{i2}, X_{i3}, X_{i4} \sim Poi(0.4)$, $X_{i5}, X_{i6}, X_{i7}, X_{i8} \sim Ber(0.5)$; All regressor variables are mutually independent.</td>
<td>$\begin{bmatrix} 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 \ 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 0 &amp; 0 &amp; 0 &amp; 1 &amp; 1 \ 1 &amp; 0 &amp; 1 &amp; 1 &amp; 0 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 \ 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 0 &amp; 0 &amp; 1 &amp; 1 &amp; 0 \ 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 0 &amp; 1 &amp; 1 &amp; 1 \ \end{bmatrix}$</td>
<td>$e = 0.00$</td>
<td>$e_1 = 0.89$ $\bar{e} = 0.83$ $e_2 = 0.62$ $\bar{e}_G = 0.81$ $e_3 = 0.86$ $\bar{e}<em>S = 0.83$ $e</em>{\cap} = 0.00$</td>
<td></td>
</tr>
<tr>
<td>A3</td>
<td>$X_{i2}, X_{i3}, X_{i4} \sim Poi(0.4)$, $X_{i5}, X_{i6}, X_{i7}, X_{i8} \sim Ber(0.5)$; All regressor variables are mutually independent.</td>
<td>$\begin{bmatrix} 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 \ 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 \ 1 &amp; 0 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 \ 1 &amp; 1 &amp; 1 &amp; 0 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 \ 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 \ \end{bmatrix}$</td>
<td>$e = 0.00$</td>
<td>$e_1 = 0.72$ $\bar{e} = 0.61$ $e_2 = 0.37$ $\bar{e}_G = 0.58$ $e_3 = 0.76$ $\bar{e}<em>S = 0.58$ $e</em>{\cap} = 0.00$</td>
<td></td>
</tr>
<tr>
<td>A4</td>
<td>$X_{i2}, X_{i3}, X_{i4} \sim Poi(0.4)$, $X_{i5}, X_{i6}, X_{i7}, X_{i8} \sim Ber(0.5)$; All regressor variables are mutually independent.</td>
<td>$\begin{bmatrix} 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 \ 1 &amp; 1 &amp; 0 &amp; 1 &amp; 0 &amp; 0 &amp; 0 &amp; 1 &amp; 1 &amp; 1 \ 1 &amp; 1 &amp; 0 &amp; 1 &amp; 0 &amp; 1 &amp; 0 &amp; 0 &amp; 1 &amp; 1 \ 1 &amp; 0 &amp; 1 &amp; 1 &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 \ 1 &amp; 1 &amp; 1 &amp; 1 &amp; 0 &amp; 1 &amp; 0 &amp; 0 &amp; 1 &amp; 1 \ \end{bmatrix}$</td>
<td>$e = 0.00$</td>
<td>$e_1 = 0.90$ $\bar{e} = 0.93$ $e_2 = 0.90$ $\bar{e}_G = 0.92$ $e_3 = 0.93$ $\bar{e}<em>S = 0.93$ $e</em>{\cap} = 0.85$</td>
<td></td>
</tr>
<tr>
<td>Scenario</td>
<td>Rating characteristics</td>
<td>$(\beta_0, \beta_1, \ldots, \beta_K)'$</td>
<td>Choice matrix</td>
<td>Error rates</td>
<td>Error rate $\bar{e}$</td>
</tr>
<tr>
<td>----------</td>
<td>------------------------</td>
<td>------------------------------------</td>
<td>---------------</td>
<td>-------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>B1</td>
<td></td>
<td>$\begin{bmatrix} 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 \end{bmatrix}$</td>
<td>$e = 0.000$</td>
<td>$\bar{e} = 0.183$</td>
<td>$\bar{e}_G = 0.000$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\begin{bmatrix} 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 0 &amp; 1 &amp; 0 \end{bmatrix}$</td>
<td>$e_1 = 0.004$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\begin{bmatrix} 1 &amp; 0 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 \end{bmatrix}$</td>
<td>$e_2 = 0.045$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\begin{bmatrix} 1 &amp; 1 &amp; 1 &amp; 0 &amp; 1 &amp; 1 &amp; 1 &amp; 1 \end{bmatrix}$</td>
<td>$e_3 = 0.254$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\begin{bmatrix} 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 &amp; 1 \end{bmatrix}$</td>
<td>$e_\cap = 0.000$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B2</td>
<td>$X_{i1}, X_{i2} \overset{i.i.d.}{\sim} LN(0.25, 0.75)$, $X_{i3}, X_{i4} \overset{i.i.d.}{\sim} Pois(0.8)$, $X_{i5}, X_{i6}, X_{i7}, X_{i8} \overset{i.i.d.}{\sim} Ber(0.1)$; All regressor variables are mutually independent.</td>
<td>$\begin{bmatrix} -2 \end{bmatrix}$</td>
<td>$e = 0.000$</td>
<td>$\bar{e} = 0.672$</td>
<td>$\bar{e}_G = 0.082$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\begin{bmatrix} -1 \end{bmatrix}$</td>
<td>$e_1 = 0.270$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\begin{bmatrix} -2 \end{bmatrix}$</td>
<td>$e_2 = 0.475$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\begin{bmatrix} -3 \end{bmatrix}$</td>
<td>$e_3 = 0.459$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\begin{bmatrix} -1 \end{bmatrix}$</td>
<td>$e_\cap = 0.000$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B3</td>
<td></td>
<td>$\begin{bmatrix} 2 \end{bmatrix}$</td>
<td>$e = 0.000$</td>
<td>$\bar{e} = 0.763$</td>
<td>$\bar{e}_G = 0.195$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\begin{bmatrix} 2 \end{bmatrix}$</td>
<td>$e_1 = 0.505$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\begin{bmatrix} 1 \end{bmatrix}$</td>
<td>$e_2 = 0.316$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\begin{bmatrix} 1 \end{bmatrix}$</td>
<td>$e_3 = 0.589$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\begin{bmatrix} 1 \end{bmatrix}$</td>
<td>$e_\cap = 0.000$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B4</td>
<td></td>
<td>$\begin{bmatrix} 1 \end{bmatrix}$</td>
<td>$e = 0.000$</td>
<td>$\bar{e} = 0.354$</td>
<td>$\bar{e}_G = 0.285$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\begin{bmatrix} 1 \end{bmatrix}$</td>
<td>$e_1 = 0.281$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\begin{bmatrix} 1 \end{bmatrix}$</td>
<td>$e_2 = 0.257$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\begin{bmatrix} 1 \end{bmatrix}$</td>
<td>$e_3 = 0.224$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\begin{bmatrix} 1 \end{bmatrix}$</td>
<td>$e_\cap = 0.084$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Simulation scenarios B1 to B4 with $r = 1, 2, 3$, $i = 1, \ldots, 4000$, $j = 1, \ldots, 100$ and corresponding error rates
1. Basically, all results of aggregated ratings are quite unsatisfactory, referred to the obtained error rates. Error rates about 80% are quite high and such outcomes seem not to be appropriate for practical use. Additionally, take into account, that an error rate as defined above constitutes a quite weak measure of quality.

2. Regarding the aggregation method, neither the arithmetic mean, the geometric mean nor the Z-score mean seems to be preferable in scenarios A1 to A4.

   In A4 the smallest amount of information about the rating characteristics is used, especially the consensus information (6 out of 9 regressor variables) is not identical to the full model as in scenarios A2 and A3. This is also the case in scenario A1. But interestingly, in A1 the error rate $e_1$ is essentially lower than in A4. This illustrates the influence of single regressor variables on the linear predictor in the logit model. The influence of single regressor variables on the linear predictor and the outcome also can be seen in A3, where only one variable is missing for every rating agency. These little differences lead to very different error rates, especially referring to $e_2$. These varying outcomes depend strongly on the type of regressor variables as well as their interaction with the linear predictor.

3. The error rates of the B-scenarios are remarkably lower than in the A-scenarios.

   In all B-scenarios (except from $e_3$ in B1), the single ratings perform better than the aggregation with the arithmetic mean.

   Using the geometric mean or the Z-score for aggregation leads to lower error rates than the arithmetic mean, but it does not lead to an improvement against the single ratings in every case.

   Calculating only the error rates, does not take into account any considerations referring to the preciseness of the results. Having small error rates is not enough. The calculated default probability should come along with a small distance to the true default probability. This is assessed by the mean distance between the true probability $\pi(I_i)$ and the corresponding estimated or aggregated probability:

   $s_i = \frac{1}{m} \sum_{j=1}^{m} \left| \hat{\pi}(j)(I_i) - \pi(I_i) \right|, \tag{5}$

   $s_{\cap i} = \frac{1}{m} \sum_{j=1}^{m} \left| \hat{\pi}(j)(I_i^{\cap}) - \pi(I_i) \right|, \tag{6}$

   $\bar{s}_i = \frac{1}{m} \sum_{j=1}^{m} \left| \bar{\pi}_i(j) - \pi(I_i) \right|, \tag{7}$

   $\bar{s}_{Gi} = \frac{1}{m} \sum_{j=1}^{m} \left| \bar{\pi}_{Gi}(j) - \pi(I_i) \right|, \tag{8}$

   $\bar{s}_{Si} = \frac{1}{m} \sum_{j=1}^{m} \left| \bar{\pi}_{Si}(j) - \pi(I_i) \right|, \tag{9}$

   where the term $(j)$ again indicates the $j$-th MC-replication. Tables 4 and 5 contain an overview on descriptive statistics for the mean absolute deviations $\bar{s}_i$ – $\bar{s}_{Si}$ for all simulation scenarios.
| A1 | 1st Quart. | 0.0018 | 0.0063 | 0.0235 | 0.0214 | 0.0295 |
|    | Median     | 0.0088 | 0.0181 | 0.0634 | 0.0615 | 0.0746 |
|    | 3rd Quart. | 0.0199 | 0.0682 | 0.1885 | 0.1634 | 0.1976 |
| A2 | 1st Quart. | 0.0018 | 0.0018 | 0.0158 | 0.0135 | 0.0177 |
|    | Median     | 0.0088 | 0.0088 | 0.0512 | 0.0420 | 0.0537 |
|    | 3rd Quart. | 0.0199 | 0.0199 | 0.1375 | 0.1113 | 0.1383 |
| A3 | 1st Quart. | 0.0018 | 0.0018 | 0.0109 | 0.0064 | 0.0111 |
|    | Median     | 0.0088 | 0.0088 | 0.0297 | 0.0213 | 0.0252 |
|    | 3rd Quart. | 0.0199 | 0.0199 | 0.0737 | 0.0561 | 0.0510 |
| A4 | 1st Quart. | 0.0018 | 0.0175 | 0.0344 | 0.0294 | 0.0387 |
|    | Median     | 0.0088 | 0.0606 | 0.1203 | 0.1097 | 0.1398 |
|    | 3rd Quart. | 0.0199 | 0.1902 | 0.2191 | 0.2011 | 0.2401 |

Table 4: Mean absolute deviations for A-scenarios

| B1 | 1st Quart. | 0.00001 | 0.00001 | 0.00010 | 0.00003 | 0.00005 |
|    | Median     | 0.00015 | 0.00015 | 0.00050 | 0.00024 | 0.00033 |
|    | 3rd Quart. | 0.00113 | 0.00113 | 0.00216 | 0.00131 | 0.00168 |
| B2 | 1st Quart. | 0.00001 | 0.00001 | 0.00080 | 0.00022 | 0.00031 |
|    | Median     | 0.00015 | 0.00015 | 0.00185 | 0.00074 | 0.00112 |
|    | 3rd Quart. | 0.00113 | 0.00113 | 0.00425 | 0.00163 | 0.00243 |
| B3 | 1st Quart. | 0.00001 | 0.00001 | 0.00120 | 0.00036 | 0.00051 |
|    | Median     | 0.00015 | 0.00015 | 0.00269 | 0.00115 | 0.00168 |
|    | 3rd Quart. | 0.00113 | 0.00113 | 0.00554 | 0.00227 | 0.00339 |
| B4 | 1st Quart. | 0.00001 | 0.00004 | 0.00030 | 0.00015 | 0.00021 |
|    | Median     | 0.00015 | 0.00035 | 0.00112 | 0.00073 | 0.00104 |
|    | 3rd Quart. | 0.00113 | 0.00206 | 0.00597 | 0.00362 | 0.00537 |

Table 5: Mean absolute deviations for the B-scenarios
In the following some important results/remarks on Table 4 and 5 are reported.

1. There are identical values in the respective third column (s_i) of Table 4 and 5 over all subscenarios within A or B, because the basis here is the full model with all regressor variables.

2. The mean absolute deviations in the A-scenarios are bigger than in the B-scenarios, which is probably mainly caused by the different dimensions of the true default probabilities (cf. Figure 1). In comparison with the estimation based on the full information, the mean deviations (referring to the corresponding quantiles of the descriptive statistics) are higher mainly up to the factor 12 in the A-scenarios and up to the factor 7 in the B-scenarios. In relation to the very small default probabilities, this is not necessarily a problem if the estimated default probabilities are transformed into an ordinal scale with rating classes. Thereby, such small default probabilities typically constitute the best rating class.

3. Within all A-scenarios A4 contains the biggest deviations regardless of the aggregation method. In A4 the smallest amount of information is used, especially the consensus information (6 out of 9 regressor variables) is not identically to the full model as in scenarios A2 and A3. This is also the case in scenario A1.

4. Referring to the aggregation, the A-scenarios as well as the B-scenarios show higher absolute deviations for the arithmetic mean and the Z-score aggregation in comparison with the geometric mean. The Z-score aggregation shows higher mean absolute deviations than the arithmetic mean in the A-scenarios and smaller ones than the arithmetic mean in the B-scenarios.

The absolute values of deviation still hide the direction of the error that is made. For getting an insight on this issue, the rates of underestimated default probabilities are investigated. This underestimation rate is defined as

\[ u_i \text{ def } = \frac{|A_i|}{m}, \]

with \( A_i \text{ def } = \{ j : \hat{\pi}(j)(I_i) < \pi(I_i) \} \). Analogously, the underestimation rates \( u_{\cap i}, \tilde{u}, \bar{u}_{Gi}, \) and \( \bar{u}_{Si} \) are defined with the sets

\[ A_{\cap i} \text{ def } = \{ j : \hat{\pi}(I_i^\cap) < \pi(I_i) \}, \]
\[ \bar{A}_{Gi} \text{ def } = \{ j : \bar{\pi}(j)_{Gi} < \pi(I_i) \}, \]
\[ \bar{A}_{Si} \text{ def } = \{ j : \bar{\pi}(j)_{Si} < \pi(I_i) \}. \]

Figure 2 illustrates all these underestimation rates by means of hexbinplots. Hexbinplots are a kind of aggregated scatterplot. Where there is no data, there is no hexbin. The more data there is in a hexbin, the darker it is.

The main target here is an insight on aggregation issues, therefore the single rating agencies are not considered for the problem of underestimation. Apart from that, the graphs for the single rating agencies do not provide any additional or new insights for the basic problem of aggregation. For the hexbinplots in Figure 2 the following observations can be made.

1. Note the difference between the empirical distributions of the true default probabilities in the A- and B-scenarios for the following remarks. Having a big amount of low default probabilities, leads to high absolute frequencies near zero along the horizontal axes. The
main target of the hexbinplots is the illustration of the underestimation rates depending on the true default probabilities. Therefore, every hexbinplot has to be read mainly in a „vertical“ direction.

2. If there is no tendency of overestimation or underestimation, the empirical distributions of the underestimation rates are mainly concentrated around 0.5. This can bee seen if the models are estimated with full information ($u_i$). In the case of concentrations near zero as well as one, the estimation based on different regressor variables leads to an overestimation or an underestimation of the true default probabilities resp.

3. In the A- and B-scenarios there is a tendency for overestimation of the low default probabilities (underestimation rates near zero) and an underestimation for the higher default probabilities (underestimation rates near one).

4. From the aggregation perspective, there seems to be no preferable kind of aggregation, regarding under- or overestimation. The plots for $u_i$, $\bar{u}_i$, $\bar{u}_G$, $\bar{u}_S$ are quite similar (except for the case of identity of full information and consensus information, i. e. $I_i^\cap = I_i$).

5. As already mentioned above, nearly all scenarios show higher absolute deviations and higher error rates for the arithmetic mean in comparison with the geometric mean. This fact becomes clear, taking a look on the underestimation rates. For positive values, the arithmetic mean is an upper bound for the geometric mean (cf. Hardy, Littlewood & Pólya (1988, Theorem 9, p. 17)). If there is a tendency for overestimating the low default probabilities—which form the majority in the scenarios here (cf. Figure 1)—an aggregation by the geometric mean acts like a correction under such conditions. As a
consequence, this is one possible explanation for the corresponding lower error rates in Tables 2 and 3.

6. The outcome of the aggregation based on the Z-score, depends on the properties of the logit function $G$ in (2). Because of its S-shaped form, the logit function has a convex part and concave part as well. By Jensen’s inequality (cf. Hardy et al. (1988, Theorem 90, p.74)), the aggregation by the arithmetic mean is an upper bound for the Z-score aggregation if the logit function is convex and a lower bound if the logit function is concave. The logit function is strictly increasing and has an inflection point at zero. As a consequence, all estimated default probabilities smaller than 0.5 are transformed within the convex part of the logit function. Having a vast majority of low default probabilities means, that the Z-scores are transformed mainly within the convex part of the logit function. This is a possible explanation for the similarity of outcomes between the geometric mean and the Z-score aggregation, because the Z-score aggregation works as a correction quite similar to the geometric mean as described above, especially in the B-scenarios. The A-scenarios contain a bigger portion of default probabilities (cf. Table 1) that are transformed within the concave part. This is a possible explanation, that nearly all A-scenarios show higher mean absolute deviations with the Z-score aggregation than with the arithmetic mean.

5 Conclusion

From the theoretical considerations and the simulations above, there is no preferable strategy to get a more precise rating by aggregating single ratings. Whether using the single ratings nor their aggregated forms necessarily lead to improvements. Only the consensus rating provides an advantage herein, which is expectable from the theoretical considerations. But the implementation of a consensus rating remains a theoretical issue as already described. Comparing only the „real“ aggregated ratings, the aggregation by the geometric mean performs slightly better than the arithmetic mean, but this may be due to the problem of overestimation of low default probabilities in the simulation framework used here. Therefore, a generalization seems not to be appropriate. Aggregating the ratings via the Z-score performs quite similar as the geometric mean with respect to the error rates. But in the case of default probabilities about 0.5, the mean absolute deviations overrun those of the arithmetic mean as well as the geometric mean. Such effects are probably driven by the relations between the different mean concepts in combination with the properties of the logit function. For practical purposes, default probabilities above 0.5 are not so important. Therefore, one would recommend here, to aggregate with the geometric mean. But as already said, a generalization is not indicated due to the outcomes above.

Every combination of model, estimation method and an arbitrary aggregation method causes different outcomes and shows its individual interplay. Furthermore, the quality measures used here (mean absolute deviation, error rate, underestimation rate) do not indicate any practical use of the aggregated estimated default probabilities. Taking into account any time dynamics, as it is considered in Grün et al. (2013), the whole issue becomes more complicated. Based on this one-period simulation study it is not to be expected to match better results in the sense of preciseness in a multi-period case. The bottom line is: From a theoretical point of view and looking at the simulation results, an aggregation of single ratings to get a higher precision is an extremely questionable issue.
Further research should concentrate on a formal proof that achieving a higher precision by means of rating aggregation is an impossible issue. In doing so, some different levels of consideration can be taken into account. First: Does the probability theory provide any approach to disprove the reasonable combination of different conditional probabilities resulting from different models? The second level is the estimation problem: Does the estimation theory provide any approach to disprove the reasonable merging of different estimates from different estimation methods?

References


42/04  S. Huschens: Faktorstruktur und Marktmodelle.  

43/04  S. Huschens, G. Stahl: Granularität dominiert Korrelation.  


45/05  S. Höse, K. Vogl: Predicting the Credit Cycle with an Autoregressive Model.


47/08  S. Höse, S. Huschens, R. Wania: Rating Migrations.  


51/10  S. Höse, S. Huschens: Confidence Intervals for Quantiles of a Vasicek-distributed Credit Portfolio Loss.

52/10  D. Tillich: Risikomaßzahlen für Kreditportfoliotranchez.  

53/10  S. Huschens: Kann es Rückzahlungswahrscheinlichkeiten von 100% geben?  


57/12  S. Höse, S. Huschens: Credit Portfolio Correlations and Uncertainty.  

58/12  S. Fischer: Ratio calculandi periculi - Ein analytischer Ansatz zur Bestimmung der Verlustverteilung eines Kreditportfolios


60/14  C. Lehmann, D. Tillich: Consensus Information and Consensus Rating – A Note on Methodological Problems of Rating Aggregation  

61/15  C. Lehmann, D. Tillich: Applied Consensus Information and Consensus Rating – A Simulation Study on Rating Aggregation