A HYPOTHESIS TEST FOR THE LONG-TERM CALIBRATION IN RATING SYSTEMS WITH OVERLAPPING TIME WINDOWS

PATRICK KURTH, MAX NENDEL, AND JAN STREICHER

ABSTRACT. We present a statistical test that can be used to verify supervisory requirements concerning overlapping time windows for the long-term calibration in rating systems. In a first step, we show that the long-run default rate is approximately normally distributed with respect to random effects in default realization. We then perform a detailed analysis of the correlation effects caused by the overlapping time windows and solve the problem of an unknown distribution of default probabilities for the long-run default rate. In this context, we present several methods for a conservative calibration test that can deal with the unknown variance in the test statistic. We present a test for individual rating grades, and then pass to the portfolio level by suitably adapting the test statistic. We conclude with comparative statics analysing the effect of persisting customers and the number of customers per reference date.

1. Introduction

Financial institutions use statistical models to estimate the default risk of obligors in order to manage credit-risks. According to Basel II, banks are allowed to estimate risk parameters that are used to calculate regulatory capital with their own models. The legal framework for the use of such models in the internal ratings-based (IRB) approach is regulated in the Capital Requirements Regulation (CRR), see [5]. The CRR imposes specific requirements for the models, e.g., that “institutions shall estimate PDs by obligor grade from long run averages of one-year default rates”, see Article 180. In 2017, the European banking authority published the “guidelines on PD estimation, LGD estimation and the treatment of defaulted exposures” (EBA-GL) specifying the CRR requirement, see [10]. Paragraph 81 EBA-GL states that “institutions should calculate the observed average default rates as the arithmetic average of all one year default rates”. This requirement was additionally specified in [8] and in [9]. Here, the observed one-year default rate at a given reference date is defined as the percentage of defaulters in the following year, so that the observed long-run average default rate depends on the choice of the reference dates. Paragraph 80 EBA-GL allows institutions to choose “between an approach based on overlapping and an approach based on non-overlapping one-year time windows”. Overlapping one-year time windows occur when the time interval between two reference dates is less than one year. Due to computational simplicity, it is of course convenient to continue working with non-overlapping time windows and appropriately adjust the long-term default rate. In many cases, these type of adjustments are rather on the conservative side, see, e.g., [13] for an empirical study and [15, 18] for theoretical analyses in the context of asset correlation.

Date: December 15, 2023.

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1283/2 2021 – 317210226. The first and the third author are grateful for the support of the Landesbank Baden-Württemberg related to this work.
On the other hand, the approach, using overlapping time windows, provides more information on defaults due to potential short-term contracts, which cannot be observed during one-year periods. It is therefore favored by most financial institutions that handle portfolios with only few observed defaults. Another advantage of this approach lies in the fact that the bias caused by a specific choice of reference dates can be reduced, e.g., when calculating the long-run average default rate as the arithmetic mean of the one-year default rates on a quarterly basis.

Paragraph 87 EBA-GL requires institutions to use a statistical test of calibration at the level of rating grades and for certain portfolios. Classically, the literature dealing with calibration tests assumes a binomial-distributed default rate, see, e.g., [7] and [17] for a discussion of different hypothesis tests in this context. We also refer to [6] for the consideration of different PDs within the same rating grade and [3, 4] for a modified binomial test accounting correlated defaults. However, when considering overlapping time windows, the assumption of a binomial distribution can no longer be maintained when considering overlapping time windows.

More generally, Monte Carlo methods can be used to construct tests for distributions that cannot be determined analytically. For the use of Monte Carlo methods, precise knowledge of the distribution of the probabilities of default within the portfolio is essential. However, in our case, these probabilities are unknown and estimated by the underlying model, so that they cannot not be used to determine the test statistic. On the other hand, analytical tests are desirable since they satisfy requirements such as replicability and reproducibility. We refer to [1, 2] for hypothesis tests in analytic form that take into account correlation effects, replacing the i.i.d. assumption by a conditional i.i.d. assumption. In this context, we also refer to [16] for a one-observation-based inference on the adequacy of probability of default forecasts with dependent default events.

In this paper, we therefore present a statistical test that can be used to verify supervisory calibration requirements in the case of overlapping time windows. A major challenge lies in the analysis of the correlation effects that are caused by the overlapping time windows. On the level of individual ratings grades, the variance of the test statistic is already determined by the null hypothesis. This, however, is not the case on a portfolio level, which is why we focus in great detail on a conservative estimate for the variance by solving a related minimization problem. We then present a conservative calibration test that can deal with the unknown variance in the test statistic.

The rest of the paper is structured as follows. In Section 2.1, we introduce the terminology and notation to for the formulation of the hypothesis test in Section 2.2. In Section 2.3, we show that the long-run default rate is approximately normally distributed with respect to random effects in default realization. Thereafter, we focus on the analysis of the correlation effects that arise due to the overlapping time periods in Section 2.4 and Section 2.5. Taking these correlation effects into account is essential for determining the variance of the long-run default rate, see Section 2.6. In Section 3 we formulate the hypothesis test in detail, first at the level of individual rating grades, cf. Section 3.1 and then at portfolio level, cf. Section 3.2. We conclude with a discussion on the parameters of the test and further considerations in Section 4. A closed form solution to the minimization problem related to the estimation of the variance is derived in the Appendix A. In the Appendix B, we propose an alternative way to estimate the variance without solving and optimization problem.
2. Setup and preliminaries

2.1. Setup and notation. In this section, we introduce the terminology and notation used in this paper and give a formal description of the statistical test that is studied in this work. We begin with the general setup by considering the default state of an individual obligor. Throughout, a default state over a one-year time horizon is described by a Bernoulli-distributed random variable $x \sim B(1, p)$, where

$$ p \in \{\text{PD}_1, \ldots, \text{PD}_m\}, $$

$m \in \mathbb{N}$, and $0 < \text{PD}_1 < \ldots < \text{PD}_m < 1$ are the default probabilities of the rating grades in the underlying master scale. In the following, we also use the notation $\text{PD}_{\min}$ and $\text{PD}_{\max}$ for the default probabilities $\text{PD}_1$ and $\text{PD}_m$, respectively.

Next, we introduce the long-run default rate for a given history of reference dates $\text{RD}_1 < \ldots < \text{RD}_N$ with $N \in \mathbb{N}$. For all $t = 1, \ldots, N$, we are given a number $n_t \in \mathbb{N}_0$ of customers within the portfolio at the reference date $\text{RD}_t$, and write

$$ n_{\min} := \min \{n_t \mid t = 1, \ldots, N\} \setminus \{0\} \quad \text{and} \quad n_{\max} := \max \{n_t \mid t = 1, \ldots, N\} $$

for the minimal and maximal number of existing customers at any reference date throughout the history of the portfolio, respectively. Throughout, we assume that $n_{\max} \neq 0$ or, equivalently, that $n_t \neq 0$ for some $t = 1, \ldots, N$.

Let $q \in \mathbb{N}$ be the number of reference dates within a one-year time-horizon starting from an arbitrary reference date. To be more precise, we assume that $\text{RD}_{t+1} = \text{RD}_t + \frac{1}{q}$ for all $t = 1, \ldots, N - 1$.

According to the EBA guidelines [10], the long-run default rate should be computed at least on a quarterly basis, i.e., $q = 4$, or on an annual basis, i.e., $q = 1$, performing however an analysis of the possible bias that occurs due to the negligence of quarterly data. The main interest of our analysis lies in the case $q > 1$ leading to correlation effects caused by overlapping time windows. For $s, t = 1, \ldots, N$ with $s > t$, we define $w_{t,s} := \max\{0, \text{RD}_t + 1 - \text{RD}_s\}$ for the size of the overlap of the observation periods.

Since new customers are usually added over time and business relationship ends for others, assign to each customer during the history of reference dates a number from 1 to $M \in \mathbb{N}$, and we define $\Lambda_t \subseteq \{1, \ldots, M\}$ as the set of obligors at reference date $\text{RD}_t$ for all $t = 1, \ldots, N$. In particular, $\bigcup_{t=1}^{N} \Lambda_t = \{1, \ldots, M\}$, $|\Lambda_t| = n_t$, and $n_{\max} \leq M$. For $t = 1, \ldots, N$, we then define the one-year default rate among the considered customers at reference date $\text{RD}_t$ by

$$ X_t = \frac{1}{n_t} \sum_{j \in \Lambda_t} x_{j,t}, $$

where $x_{j,t} \sim B(1, p_{j,t})$ is the one-year default state and $p_{j,t} \in \{\text{PD}_1, \ldots, \text{PD}_m\}$ the probability of default over a one year time horizon of customer $j \in \Lambda_t$ at the reference date $\text{RD}_t$. Since $n_t = 0$ cannot be excluded, we use the convention $\frac{0}{0} := 0$.

The realized default rate on a reference date $\text{RD}_t$ is denoted by $\text{DR}_t$, i.e., $\text{DR}_t$ is the realization of the random variable $X_t$ for $t = 1, \ldots, N$. Moreover, we define

$$ R_N := \{t = 1, \ldots, N \mid n_t > 0\} $$

as the set of indices for reference dates, where the portfolio contains at least one customer, and denote its cardinality by $R(N)$. Since we assume that $n_{\max} \neq 0$, it follows
that $R(N) > 0$, and we define the realized long-run default rate as

$$LRDR := \frac{1}{R(N)} \sum_{t=1}^{N} \text{DR}_t,$$

which is a realization of the long-run default rate

$$Z := \frac{1}{R(N)} \sum_{t=1}^{N} X_t. \tag{2}$$

We emphasize that the long-run default rate is the arithmetic mean of the one-year default rates and not the arithmetic mean of the individual default states of all customers for all reference dates, which intuition might suggest. Due to the chosen convention $\bar{a} = 0$, it follows that $X_t = 0$ if $n_t = 0$, i.e., reference dates, where the portfolio contains no customers are completely disregarded in the computation of the long-run default rate. Nevertheless, these reference dates have to be included into the time line since they describe an overlapping period.

2.2. **Formal description of the test.** We now give an overview over the hypothesis test presented in this paper. The aim is to formulate a statistical test for comparing the realized long-run default rate LRDR with the estimated long-run default rate LRCT, which is called the long-run central tendency, both on the level of individual rating grades and on portfolio level. We use the random variable $Z$ with expected value $\mu$ as the test statistic. We formulate the following hypothesis test with

- null hypothesis $H_0 : \mu = LRCT$
- alternative hypothesis $H_1 : \mu \neq LRCT$.

Based on this, we determine values $k, K \in [0, 1]$ and consider the test function

$$\varphi(z) := \begin{cases} 0, & \text{if } z \in [k, K], \\ 1, & \text{otherwise.} \end{cases} \tag{3}$$

If $\varphi(LRDR) = 0$, the null hypothesis based on the available data is retained, whereas it is rejected in favor of the alternative hypothesis if $\varphi(LRDR) = 1$. We point out that the distribution of the test statistic $Z$ does not only depend on $\mu$ but also on other parameters that are not determined by the null hypothesis. The main focus of the paper is to postulate distributional assumption for $Z$ such that its distribution is determined by the null hypothesis. In this context, we first show that $Z$ is approximately normally distributed, i.e., $Z \sim \mathcal{N}(\mu, \sigma^2)$. In a second step, the focus lies on a conservative estimate for the variance $\sigma^2$ based on the information given by the null hypothesis. A major challenge lies in the fact that the consideration of overlapping time windows leads to unavoidable correlation effects that rule out independence assumptions on the random variables $X_1, \ldots, X_N$.

2.3. **Distribution of the long-run default rate.** As we have seen in the previous section, the long-run calibration test requires the formulation of a distribution assumption for the long-run default rate. However, the definition of the long-run default rate as the arithmetic mean of the one-year default rates leads to difficulties in the derivation of an analytical description of the distribution based on the Bernoulli-distributed default states of the individual obligors. The aim of this section is to show that, despite the correlation effects and the possibly varying number of obligors with different
PDs at each reference date, the long-run default rate is still approximately normally distributed.

In order to simplify notation, we define $y_{j,t} := x_{j,t}$ if $j \in A_t$ and $y_{j,t} := 0$ otherwise, for all $t = 1, \ldots, N$. Then,

$$Z = \frac{1}{R(N)} \sum_{t=1}^{N} \frac{1}{n_t} \sum_{j=1}^{M} y_{j,t} = \sum_{j=1}^{M} Y_j$$

with $Y_j = \sum_{t=1}^{N} \frac{1}{R(N)} \frac{1}{n_t} y_{j,t}$.

In the sequel, we will show that $Z$ is approximately normally distributed. For this, we assume that default states of different obligors are independent of each other, from which the independence of the family $(Y_j)_{j=1,\ldots,M}$ follows. This assumption is standard and sufficiently conservative for the calibration of rating models.

In order to apply the Lindeberg-Feller central limit theorem (CLT), which is a generalization of the classical CLT, we first show that the variances of the random variables $(Y_j)_{j=1,\ldots,M}$ cannot become arbitrarily small. In fact, by neglecting the covariance between the default states of an obligor at different reference dates,

$$\sigma^2(Y_j) = \frac{1}{R(N)^2} \sum_{t=1}^{N} \left( \frac{1}{n_t^2} \sigma^2(y_{j,t}) + 2 \frac{1}{R(N)^2} \sum_{s=t+1}^{N} \frac{1}{n_t n_s} \text{Cov}(y_{j,t}, y_{j,s}) \right)$$

$$\geq \frac{1}{R(N)^2} \sum_{t=1}^{N} \frac{1}{n_t^2} \sigma^2(y_{j,t})$$

for all $j = 1, \ldots, M$.

Moreover, for each obligor $j = 1, \ldots, M$, there exists at least one index $t = 1, \ldots, N$ with $\sigma^2(y_{j,t}) > 0$ since $\bigcup_{t=1}^{N} A_t = \{1, \ldots, M\}$. Hence, for each obligor $j = 1, \ldots, M$,

$$\sigma^2(Y_j) \geq \frac{1}{R(N)^2} \frac{1}{n_{\max}} \min_{k=1,\ldots,m} \{ (\text{PD}_k (1 - \text{PD}_k)) \} =: v_{\min}.$$

Note that $v_{\min}$ is independent of the customer $j = 1, \ldots, M$ and $v_{\min} > 0$. Since, up to now, we only consider finitely many customers, we extend the family $(Y_j)_{j=1,\ldots,M}$ by a sequence of independent random variables $Y_{M+1}, Y_{M+2}, \ldots$ with $\sigma^2(Y_j) \geq v_{\min}$, e.g. $Y_j \sim N(\mu_j, v_{\min})$ with arbitrary drift $\mu_j \in (0, 1)$ for $j \geq M + 1$.

Now, the central limit theorem shall be applied to the family $(Y_j)_{j\in\mathbb{N}}$. In the classical version by Lindeberg & Lévy, the CLT states that a proper renormalization of the arithmetic mean is approximately normally distributed if the underlying sequence of random variables is independent and identically distributed. Its generalized version by Lindeberg & Feller also applies to random variables that are not identically distributed as in our case. It states that a proper renormalization of the arithmetic mean of a family of independent random variables converges in probability against a normally distributed random variable if this family satisfies the so-called Lindeberg condition, see [14, Definition 15.41]. In our case, the Lindeberg condition applies if, for all $\varepsilon > 0$,

$$\lim_{k \to \infty} \frac{1}{s_k^2} \sum_{j=1}^{k} \mathbb{E} \left( (Y_j - \mathbb{E}(Y_j))^2 \cdot 1_{\{|Y_j - \mathbb{E}(Y_j)| > \varepsilon s_k\}} \right) = 0,$$

where

$$s_k = \sqrt{\sum_{j=1}^{k} \sigma^2(Y_j)}.$$
That is, the Lindeberg condition requires that the underlying sequence of random variables does not exhibit arbitrarily large deviations from the expected value.

In the following, we verify the Lindeberg condition in the our setup. To that end, let \( \varepsilon > 0 \). Since, for all \( j = 1, \ldots, M \), the random variable \( Y_j \) only takes values in \((0, 1)\),

\[
\sigma^2(Y_j) \geq v_{\text{min}} > 0, \quad \text{and} \quad s_k \to \infty \quad \text{as} \quad k \to \infty,
\]

there exists some \( k_0 \in \mathbb{N} \) such that, for all \( k \in \mathbb{N} \) with \( k \geq k_0 \), \( j \in \mathbb{N} \), and \( \omega \in \Omega \),

\[
|Y_j(\omega) - \mathbb{E}(Y_j)| \leq \varepsilon \cdot s_k.
\]

Hence, for all \( k \in \mathbb{N} \) with \( k \geq k_0 \),

\[
\frac{1}{s_k^2} \sum_{j=1}^{k} \mathbb{E}\left( (Y_j - \mathbb{E}(Y_j))^2 \cdot 1_{\{|Y_j - \mathbb{E}(Y_j)| > \varepsilon s_k\}} \right)
= \frac{1}{s_k^2} \sum_{j=1}^{k_0} \mathbb{E}\left( (Y_j - \mathbb{E}(Y_j))^2 \cdot 1_{\{|Y_j - \mathbb{E}(Y_j)| > \varepsilon s_k\}} \right)
\leq \frac{1}{s_k^2} \sum_{j=1}^{k_0} \mathbb{E}\left( (Y_j - \mathbb{E}(Y_j))^2 \right) \to 0 \quad \text{as} \quad k \to \infty,
\]

i.e., the Lindeberg condition is satisfied. By the central limit theorem, cf. [14, Theorem 15.44], we obtain that

\[
\lim_{k \to \infty} \mathbb{P}\left( \frac{1}{s_k} \sum_{j=1}^{k} (Y_j - \mathbb{E}(Y_j)) \leq z \right) = \Phi(z) \quad \text{for all} \quad z \in \mathbb{R},
\]

where \( \Phi \) is the cumulative distribution function for the standard normal distribution. Since the number of customers \( M \) is large but fixed throughout our analysis, we may therefore assume that

\[
Z \sim \mathcal{N}(\mu, \sigma^2) \quad \text{(4)}
\]

with \( \mu = \sum_{j=1}^{M} \mathbb{E}(Y_j) \) and \( \sigma = s_M \), suppressing the dependence of \( \mu \) and \( \sigma \) on the number of customers \( M \).

The literature suggests that, in standard cases, a sample size of at least 30 is required in order to achieve useful results with the approximation of a normal distribution, see, for example, [12]. In our setting, however, the situation is more complex due to the specific calculation method of the long-run default rate. As we will see in Section 4.3, the convergence depends both on the number of reference dates \( N \) and the number of customers \( n_t \) per reference date \( RD_t \), hence indirectly on \( M \), but not solely on \( M \). As a consequence, a large number of customers does not necessarily imply a good approximation, e.g., if we have \( N = 2 \) reference dates and \( M = 1000000 \) customers, but only one customer at the second reference date, i.e., \( n_2 = 1 \). For more details, we refer to Section 4.3, where we focus on the convergence for different combinations of \( N \) and numbers of customers \( n_t \) per reference date \( RD_t \).

2.4. Covariance between Default States. Pursuant to Paragraph 80 EBA-GL [10], when calculating the long-run default rate, institutions may choose between overlapping time windows of one year and non-overlapping time windows. According to Paragraph 78 EBA-GL, the reference dates used should include at least all quarterly reference
dates. The overlap of the time windows of the one-year default rates leads to correlations between them, which affects the variance of the long-run default rate and thus the acceptance ranges of the test.

In order to be able to describe the distribution of the test statistic as precisely as possible, in Section 2.5 below, we thus focus on the analysis of the covariances between the default rates, and start by considering the covariance between two default states of individual obligors for overlapping observation periods in this subsection. To that end, we consider two reference dates \( \text{RD}_t \) and \( \text{RD}_s \) with \( \text{RD}_t < \text{RD}_s < \text{RD}_t + 1 \) and an obligor who can be observed in the period \([\text{RD}_t, \text{RD}_s + 1]\). The default state in the period \([\text{RD}_t, \text{RD}_t + 1]\) is described by a random variable \( x_t \in \{0, 1\} \), the default state in the period \([\text{RD}_s, \text{RD}_s + 1]\) by a random variable \( x_s \in \{0, 1\} \). That is, a realization of \( x_t \) or \( x_s \) answers the question whether or not the obligor under consideration defaulted in time period \([\text{RD}_t, \text{RD}_t + 1]\) or \([\text{RD}_s, \text{RD}_s + 1]\), respectively.

In order to compute the covariance between the two default states \( x_t \) and \( x_s \), a detailed knowledge of the time of default is necessary, where, in the case of more than one default event, we focus on the time of the temporally first default. The time of first default is described by a random variable \( T \in [\text{RD}_t, \infty) \), which creates a link between default status and time of default, see, e.g., [11]. While \( T \) describes the timing of the first default starting at time \( \text{RD}_t \), we model the timing of the first default after the default described by \( T \) and after \( \text{RD}_s \) using a random variable \( T_s^* \in [\text{RD}_s, \infty) \), with \( T_s^* > T \). We then define

\[
T_2 := \begin{cases} 
T_1, & \text{if } T_1 \geq \text{RD}_s \text{ and } T_s^* > \text{RD}_s + 1, \\
T_s^*, & \text{otherwise.}
\end{cases}
\]

If besides \( T \) more than one default is observed in the interval \([\text{RD}_s, \text{RD}_s + 1]\), \( T_2 \) describes the time of the second default, otherwise \( T_2 \) is used to model the time of the first default starting from \( \text{RD}_s \). We now divide the considered time period into three disjoint intervals

\[
I_1 := [\text{RD}_t, \text{RD}_s], \quad I_2 := [\text{RD}_s, \text{RD}_t + 1], \quad \text{and} \quad I_3 := (\text{RD}_t + 1, \infty],
\]

and look at the following disjoint events

\[
E_1 := \{T_1 \in I_1\}, \quad E_2 := \{T_1 \in I_2\}, \quad \text{and} \quad E_3 := \{T_1 \notin I_1 \cup I_2\}.
\]

Moreover, we define

\[
E_4 := \{T_2 \in I_2\}, \quad E_5 := \{T_2 \in I_3\}, \quad \text{and} \quad E_6 := \{T_2 \notin I_2 \cup I_3\}.
\]

In order to simplify notation, we set \( p_t := P(x_t = 1) \) and \( p_s := P(x_s = 1) \) and assume that

\[
P(x_t = 1 \mid E_5) = p_t \quad \text{and} \quad P(x_s = 1 \mid E_1) = p_s.
\]

In a first step, we focus on the probability \( P(E_1) \). For this, we assume that the probability of default in a certain time interval depends on the length of this interval and on the general creditworthiness of the obligor, i.e., we assume that

\[
P(T_1 \in I \mid x_t = 1) = f_1(|I|)
\]

for every interval \( I \subseteq [\text{RD}_t, \text{RD}_t + 1] \) and a non-negative and non-decreasing function \( f_1 \) with \( f_1(0) = 0 \) and \( f_1(1) = 1 \). This ensures that the probability of observing a default in a given time interval increases if the length of the interval increases and that the
probabilities of default are identical for intervals of the same length. We deduce
\[ P(E_1 \mid x_t = 1) = 1 - P(E_2 \mid x_t = 1) = 1 - f_1(w_{t,s}), \]
and therefore, using Bayes’ theorem,
\[ 1 - f_1(w_{t,s}) = P(E_1 \mid x_t = 1) = P(x_t = 1 \mid E_1) \cdot \frac{P(E_1)}{P(x_t = 1)} = \frac{P(E_1)}{p_t}. \]
Rearranging terms, we find that
\[ P(E_1) = (1 - f_1(w_{t,s}))p_t \]
and, analogously,
\[ P(E_2) = f_1(w_{t,s})p_t. \]
Hence,
\[ \mathbb{E}[x_t \cdot x_s] = P(x_t = 1, x_s = 1) = P(x_t = 1, x_s = 1 \mid E_1) \cdot P(E_1) + P(x_t = 1, x_s = 1 \mid E_2) \cdot P(E_2) + P(x_t = 1, x_s = 1 \mid E_3) \cdot P(E_3) \\
\quad = P(x_s = 1 \mid E_1) \cdot P(E_1) + P(E_2) = p_s (1 - f_1(w_{t,s}))p_t + f_1(w_{t,s})p_t \\
\quad = p_t p_s + f_1(w_{t,s})p_t(1 - p_s). \]
Analogously, one obtains
\[ P(E_4) = f_2(w_{t,s})p_s \quad \text{and} \quad P(E_5) = (1 - f_2(w_{t,s}))p_s, \]
for a non-negative and non-decreasing function \( f_2 \) with \( f_2(0) = 0 \) and \( f_2(1) = 1 \), and we find that
\[ \mathbb{E}[x_t \cdot x_s] = P(x_t = 1, x_s = 1) = P(x_t = 1, x_s = 1 \mid E_4) \cdot P(E_4) + P(x_t = 1, x_s = 1 \mid E_5) \cdot P(E_5) + P(x_t = 1, x_s = 1 \mid E_6) \cdot P(E_6) \\
\quad = P(E_4) + P(x_t = 1 \mid E_5) \cdot P(E_5) = f_2(w_{t,s})p_s + p_t (1 - f_2(w_{t,s}))p_s \\
\quad = p_t p_s + f_2(w_{t,s})p_s(1 - p_t). \]
Equating the previous expectations results in
\[ f_1(w_{t,s}) = f_2(w_{t,s}) \frac{p_s(1 - p_t)}{p_t(1 - p_s)}, \]
i.e., \( f_1 \) is a linear transform of \( f_2 \). To determine the covariance between \( x_t \) and \( x_s \), knowledge of at least one of the functions \( f_1 \) or \( f_2 \) is required. We assume that the distribution of the time of default within an observation year is uniform. We postulate this assumption for \( x_s \), since it gives preference to the more recent information at RDs over the information at RDt, which lies further in the past. We therefore set \( f_2(w_{t,s}) = w_{t,s} \) and end up with
\[ \text{Cov}(x_t, x_s) = \frac{1}{n_t n_s} \sum_{i \in \Lambda_t} \sum_{j \in \Lambda_s} \text{Cov}(x_{i,t}, x_{j,s}) = w_{t,s} p_s(1 - p_t). \] (5)

### 2.5. Covariance between Default Rates

Having previously examined the covariance between obligor default states, we extrapolate this result to the covariance of default rates. To do this, we consider a sample of debtors on reference dates RDt and RDs with RDs < RDt + 1 and \( t, s = 1, \ldots, N \). During the transition from the first to the second reference date, debtors can be removed from monitoring due to defaults, terminated business relationships, or migrations to other rating systems, and new debtors can be added on the second reporting date due to new business or migrations to the rating system under consideration. In addition, there are debtors who can be observed on both reference dates. The number of these so-called persisting customers with respect to reference dates RDt and RDs is denoted by \( |\Lambda_t \cap \Lambda_s| =: k_{t,s} \in \mathbb{N} \). The one-year default rates on the two reference dates are then given by \( X_t = \frac{1}{n_t} \sum j_{\in \Lambda_t} x_{j,t} \) with \( x_{j,t} \sim \text{B}(1, p_{j,t}) \) and \( X_s = \frac{1}{n_s} \sum j_{\in \Lambda_s} x_{j,s} \) with \( x_{j,s} \sim \text{B}(1, p_{j,s}) \). Hence,
\[ \text{Cov}(X_t, X_s) = \frac{1}{n_t n_s} \sum_{i \in \Lambda_t} \sum_{j \in \Lambda_s} \text{Cov}(x_{i,t}, x_{j,s}). \]

Again, assuming that default states for pairs of different customers are independent, using Formula (5), we find that
\[ \text{Cov}(X_t, X_s) = \frac{1}{n_t n_s} \sum_{j \in \Lambda_t \cap \Lambda_s} \text{Cov}(x_{i,t}, x_{j,s}) = \frac{w_{t,s}}{n_t n_s} \sum_{j \in \Lambda_t \cap \Lambda_s} p_{j,s}(1 - p_{j,t}), \] (6)
where \( w_{t,s} \) describes, as before, the size of the overlap between the observation periods starting from the reference dates RDt and RDs.

### 2.6. Variance of the Long-Run Default Rate

Based on the considerations for estimating the covariance of default rates in overlapping periods, we now discuss the variance of the long-run default rate. For the random variable \( X_t \) with \( t = 1, \ldots, N \), we define \( \mathbb{E}[X_t] =: \mu_t \) and \( \sigma(X_t) =: \sigma_t \). For the long-run default rate \( Z \), which is approximately normally distributed with \( Z \sim \mathcal{N}(\mu, \sigma^2) \), we get
\[ \mu = \frac{1}{R(N)} \cdot \sum_{t=1}^{N} \mu_t = \frac{1}{R(N)} \cdot \sum_{t=1}^{N} \frac{1}{n_t} \sum_{j \in \Lambda_t} p_{t,j}. \]
and
\[
R(N)^2 \cdot \sigma^2 = \sum_{t=1}^{N} \sigma_t^2 + 2 \sum_{t=1}^{N-1} \sum_{s=t+1}^{N} \text{Cov} (X_t, X_s)
\]
\[
= \sum_{t=1}^{N} \sigma_t^2 + 2 \sum_{i=1}^{q-1} \sum_{t=1}^{N-i} \text{Cov} (X_t, X_{t+i}),
\]
where the second term contains the covariances caused by overlapping time periods.

Using Equation (6), we end up with
\[
R(N)^2 \cdot \sigma^2 = \sum_{t=1}^{N} \frac{1}{n_t} \sum_{j \in \Lambda_t} \lambda_t \cdot \lambda_t (1 - \mu) \sum_{i=1}^{q-1} \sum_{j \in \Lambda_t \cap \Lambda_{t+i}} p_{t+i,j} (1 - p_{t,j})
\]
\[
= \sum_{t=1}^{N} \sigma_t^2 + 2 \sum_{i=1}^{q-1} \sum_{t=1}^{N-i} \text{Cov} (X_t, X_{t+i}),
\]
where
\[
\lambda_t := 2(q - i) \frac{1}{q} \sum_{i=1}^{N-i} \frac{1}{n_t \cdot n_{t+i}}.
\]

For the final calibration test, it will be crucial to estimate the variance of the long-run default rate by a term that solely depends on the expected value \( \mu \), the number of debtors per reference date, and debtor-independent variables, see Section 3.2.

### 3. Hypothesis Test for Long-Term Calibration

The aim of this section is to formulate a statistical test for comparing the realized long-run default rate with the estimated long-run default rate.

#### 3.1. Statistical Test per Rating Grade

We start with the simplest case, and consider a portfolio consisting of only one rating grade with an associated probability of default, which we call PD\(_{\text{grade}}\). That is, all considered obligors have the same unknown probability of default, which was estimated by PD\(_{\text{grade}}\). We recall that the measured long-run default rate, denoted by LRDR, is the realization of a random variable \( Z \) for which approximately \( Z \sim N(\mu, \sigma^2) \) holds. We consider the hypothesis test, described in Section 2.2, with

- null hypothesis \( H_0 : \mu = \text{PD}_{\text{grade}} \) and
- alternative hypothesis \( H_1 : \mu \neq \text{PD}_{\text{grade}} \).

From equation (8), we get
\[
R(N)^2 \cdot \sigma^2 = \sum_{t=1}^{N} \sigma_t^2 + \mu (1 - \mu) \sum_{i=1}^{q-1} \lambda_i,
\]
where
\[
\lambda_i := \frac{2(q - i)}{q} \sum_{i=1}^{N-i} \frac{k_{t,t+i}}{n_t \cdot n_{t+i}}.
\]

Moreover,
\[
\sum_{t=1}^{N} \sigma_t^2 = \sum_{t \in R_N} \frac{1}{n_t} \cdot \mu (1 - \mu) = \mu (1 - \mu) \sum_{t \in R_N} \frac{1}{n_t},
\]
which implies that
\[
\sigma^2 = \frac{\mu (1 - \mu)}{R(N)^2} \left( \sum_{t \in R_N} \frac{1}{n_t} + \sum_{i=1}^{q-1} \lambda_i \right).
\]
Therefore, assuming the validity of the null hypothesis, $\sigma$ is known. In order to define the limits of the acceptance range for a significance level $\alpha \in (0, 1)$, we choose $k, K \in [0, 1]$ in such a way that

$$E[\varphi(Z)]|_{H_0} \leq \alpha,$$

where $\varphi$ is given by (3). To that end, let

$$k = \text{PD}_{\text{grade}} + \phi^{-1}\left(\frac{1}{2}\alpha\right) \cdot \sqrt{\frac{\text{PD}_{\text{grade}}(1 - \text{PD}_{\text{grade}})}{R(N)^2} \left(\sum_{t \in R_N} \frac{1}{n_t} + \sum_{i=1}^{q-1} \lambda_i\right)},$$

and

$$K = \text{PD}_{\text{grade}} + \phi^{-1}\left(1 - \frac{1}{2}\alpha\right) \cdot \sqrt{\frac{\text{PD}_{\text{grade}}(1 - \text{PD}_{\text{grade}})}{R(N)^2} \left(\sum_{t \in R_N} \frac{1}{n_t} + \sum_{i=1}^{q-1} \lambda_i\right)}.$$

Then, the calibration test passes if

$$\text{LRDR} \in [k, K].$$

To see the influence of persisting customers on the acceptance range and the density function of the long-run default rate, we refer to Section 4.1 and Section 4.2.

3.2. Statistical Test on Portfolio Level. We now consider a portfolio of heterogeneous obligors, i.e., each obligor has its own individual probability of default. For this purpose, we denote the estimated PD of customer $j$ on reference date $\text{RD}_t$ by $\hat{p}_{t,j}$, the mean PD based on the rating model examined on the reference date $\text{RD}_t$ by $c_{\text{PD}}$ for $t = 1, \ldots, N$, and the long-run central tendency by

$$\text{LRCT} = \frac{1}{N} \cdot \sum_{t=1}^{N} \hat{p}_{t,j} = \frac{1}{N} \cdot \sum_{t=1}^{N} \frac{1}{n_t} \sum_{j \in \Lambda_t} \hat{p}_{t,j}.$$  

While on the level of rating grades, it is not an exception that the number of customers at a reference date $\text{RD}_t$ might be zero, i.e., $n_t = 0$, on a portfolio level, this rarely happens in practice. In this section, we therefore restrict our attention to the case, where $n_t > 0$ for all $t = 1, \ldots, N$ and therefore $R(N) = N$.

We consider the hypothesis test from Section 2.2 with

null hypothesis $H_0 : \mu = \text{LRCT}$ and
alternative hypothesis $H_1 : \mu \neq \text{LRCT}$.

As in the case of individual rating grades, the test statistic depends on $\mu$ and $\sigma$ but, in this case, Equation (8) can no longer be reduced to a term depending only on $\mu$. More precisely, if the correctness of the null hypothesis is assumed, the distribution parameters are only partially determined. If, however, we replace $\sigma$ by an expression $\sigma_{\text{min}}(\mu)$ with $\sigma \geq \sigma_{\text{min}}(\mu)$, the null hypothesis implies a concrete distribution for the underlying test statistic. This results in a narrower confidence interval and therefore a greater likelihood of committing a type I error. At the same time, the probability of a type II error is reduced, comparable to lowering the significance level.

The key challenge is to derive an analytic expression for $\sigma_{\text{min}}$ that depends only on PD-independent model parameters, such as $N$, $M$, $n_t$, etc., and satisfies $\sigma_{\text{min}}(\mu) \leq \sigma$ under the null hypothesis for all possible combinations of $p_{t,j} \in [\text{PD}_{\text{min}}, \text{PD}_{\text{max}}]$. Since we aim to minimize the distance between $\sigma_{\text{min}}(\mu)$ and $\sigma$, it is sensible to define $\sigma_{\text{min}}$ via
the solution to a minimization problem with cost functional given by the right-hand side of (8) under the side condition

\[
\mu = \frac{1}{N} \cdot \sum_{t=1}^{N} \frac{1}{n_t} \sum_{j \in \Lambda_t} p_{t,j} \quad \text{and} \quad p_{t,j} \in [\text{PD}_{\text{min}}, \text{PD}_{\text{max}}].
\]  

(9)

Observe that the right-hand side of (8) is, in general, neither convex nor concave with respect to the \( p_{t,j} \). This together with the high dimensionality of the problem, makes the numerical or analytical computation of solutions is rather involved. In order to circumvent this issue, we therefore replace the right-hand side of (8) by a suitable linear cost functional, which turns the minimization into a linear program that can be solved analytically, up to a sort algorithm, and therefore also numerically with great efficiency.

First, note that mixed terms, i.e., products of PDs of the same customer at different reference dates, always appear in the sums

\[
\sum_{j \in \Lambda_t \cap \Lambda_{t+i}} p_{t+i,j} (1 - p_{t,j})
\]

caused by the covariances. Hence in Formula (8), we want to substitute the subtrahend (blue) by an affine linear function with respect to the multiplier (orange) in a conservative way. On a portfolio level, it is appropriate to assume that an obligor’s PD remains constant on average over a one year time horizon, i.e., we assume

\[
\sum_{j \in \Lambda_t \cap \Lambda_{t+i}} p_{t+i,j} (1 - p_{t,j}) \approx \sum_{j \in \Lambda_t \cap \Lambda_{t+i}} p_{t+i,j} (1 - p_{t+i,j}).
\]  

(10)

For most portfolios, this is not only a plausible but also a fairly conservative assumption regarding the portfolio variance since, for a given vector \((x_1, \ldots, x_n)\) with \(0 < x_i < 1\) for \(i = 1, \ldots, n\), the sum

\[
\sum_{i=1}^{n} x_i (1 - y_i)
\]

is minimized for \(x_i = y_i\), under the assumption that, for each \(i = 1, \ldots, n\), there exists some \(j = 1, \ldots, n\) with \(x_i = y_j\). Certainly, even more conservative assumptions can be made at this point with \(\text{PD}_{\text{max}}\) being the most conservative subtrahend possible. The impact of such a choice is discussed in Section 4.6, where we indicate that the acceptance ranges hardly change when using \(\text{PD}_{\text{max}}\) instead of \(p_{t+i,j}\). We point out that the subsequent discussion applies also to this choice without limitations.

From Equation (8), we get

\[
N^2 \cdot \sigma^2 = \sum_{t=1}^{N} \frac{1}{n_t^2} \sum_{j \in \Lambda_t} p_{t,j} (1 - p_{t,j})
\]

\[
= \sum_{t=1}^{N} \frac{1}{n_t^2} \sum_{j \in \Lambda_t} g_t(p_{t,j}) + 2 \sum_{i=1}^{q-1} \sum_{t=1}^{N-i} \frac{1}{n_t \cdot n_{t+i}} \sum_{j \in \Lambda_t \cap \Lambda_{t+i}} g_i(p_{t+i,j})
\]

with

\[
g_i(x) := \frac{q - i}{q} \cdot x (1 - x) \quad \text{for} \quad x \in (0, 1) \quad \text{and} \quad i = 0, \ldots, q - 1.
\]
Now, we replace each of the functions $g_i$ by an affine linear function $f_i$ such that the functions $g_i$ and $f_i$ coincide on $\text{PD}_{\text{max}}$ and $\text{PD}_{\text{min}}$. Since $g_i$ is concave and $f_i$ is affine linear, this implies that $g_i(x) \geq f_i(x)$ for all $x \in [\text{PD}_{\text{min}}, \text{PD}_{\text{max}}]$. To be precise, the functions $f_i$ are given by

$$f_i(x) := \alpha_i \cdot x + c_i \quad \text{for all } x \in (0, 1)$$

with

$$\alpha_i := \frac{q - i}{q}(1 - \text{PD}_{\text{max}} - \text{PD}_{\text{min}}) \quad \text{and} \quad c_i = \frac{q - i}{q}\text{PD}_{\text{min}}\text{PD}_{\text{max}}$$

for all $i = 0, \ldots, q - 1$. We therefore end up with the estimate

$$N^2 \cdot \sigma^2 = \sum_{t=1}^{N} \frac{1}{n_t^2} \sum_{j \in \Lambda_t} g_0(p_{t,j}) + 2 \sum_{i=1}^{q-1} \sum_{t=1}^{N-i} \frac{1}{n_t \cdot n_{t+i}} \sum_{j \in \Lambda_t \cap \Lambda_{t+i}} g_i(p_{t+i,j})$$

$$\geq \sum_{t=1}^{N} \frac{1}{n_t^2} \sum_{j \in \Lambda_t} f_0(p_{t,j}) + 2 \sum_{i=1}^{q-1} \sum_{t=1}^{N-i} \frac{1}{n_t \cdot n_{t+i}} \sum_{j \in \Lambda_t \cap \Lambda_{t+i}} f_i(p_{t+i,j})$$

$$= \sum_{t=1}^{N} \sum_{j \in \Lambda_t} \alpha_{t,j} p_{t,j} + C,$$

where

$$C = c_0 \cdot \sum_{t=1}^{N} \frac{1}{n_t} + 2 \cdot \sum_{i=1}^{q-1} c_i \cdot \sum_{t=1}^{N-i} \frac{k_{t,t+i}}{n_t \cdot n_{t+i}}$$

and

$$\alpha_{t,j} = \frac{1}{n_t^2} \alpha_0 \mathbb{1}_{\Lambda_t}(j) + \frac{2}{n_t} \cdot \sum_{i=1}^{q-1} \frac{\alpha_i}{n_{t-i}} \mathbb{1}_{\Lambda_{t-i} \cap \Lambda_t}(j)$$

with $\Lambda_{t-i} := \emptyset$ if $t - i < 1$ for $t = 1, \ldots, N$ and $j = 1, \ldots, M$. We have therefore reduced the original problem to

$$\text{minimize } \sum_{t=1}^{N} \sum_{j \in \Lambda_t} \alpha_{t,j} p_{t,j} \quad \text{under the side condition } (9). \quad (12)$$

As shown in the Appendix A, the solution $p(\mu)$ to the minimization problem (12) can be easily computed analytically and numerically. We denote the minimal value by $m(\mu)$, and define $\sigma^2_{\text{min}}(\mu) = \frac{m(\mu) + C}{N^2}$. For the modified random variable $Z^* \sim \mathcal{N}(\mu, \sigma^2_{\text{min}}(\mu))$, we are now able to define the limits $k, K \in [0, 1]$ of the acceptance range for a significance level $\alpha \in (0, 1)$ in such a way that $\mathbb{E}(\varphi(Z^*))|_{H_0} \leq \alpha$. For this we define

$$k := \text{LRCT} + \phi^{-1} \left( \frac{1}{2} \alpha \right) \cdot \sigma_{\text{min}}(\text{LRCT})$$

and

$$K := \text{LRCT} + \phi^{-1} \left( 1 - \frac{1}{2} \alpha \right) \cdot \sigma_{\text{min}}(\text{LRCT}).$$

The calibration test passes if

$$\text{LRDR} \in [k, K].$$
4. Discussion and further considerations

4.1. Effect of Persisting Customers on the Variance of $Z$. In this section, we aim to highlight the effects caused by persisting customers on the distribution of $Z$. We calculate the density function of the long-run default rate using a sample portfolio in a rating grade with PD = 0.02 and set $q = 4$. The number of reference dates is $N = R(N) = 32$, i.e., we consider a portfolio with a relatively short history of eight years. The number of customers is set to be constant over time with $n = 50$. We furthermore assume that the ratio of persisting customers per time interval is constant over time. We have $k_{t,t+1} = 45$ for all $t \leq 31$, $k_{t,t+2} = 40$ for all $t \leq 30$ and $k_{t,t+3} = 35$ for all $t \leq 29$. Only about 70% of the persisting customers are still in the considered rating grade after 3 quarters, which represents a relatively high fluctuation.

These specifications determine the density function of the long-run default rate (orange). In comparison, one can see the density functions of the long-run default rate with a maximum and minimum number of persisting customers in green and blue, respectively. We have

$$\sigma_{\text{orange}}^2 \approx 4.13 \cdot 10^{-5}, \quad \sigma_{\text{blue}}^2 \approx 1.23 \cdot 10^{-5}, \quad \text{and} \quad \sigma_{\text{green}}^2 \approx 4.71 \cdot 10^{-5}.$$ 

Figure 2 shows that that correlation effects caused by overlapping time windows should not be neglected since, otherwise, the acceptance range would be way too tight, see also Example 4.2 below.

![Density functions of long-run default rates](image)

**Figure 2.** Density functions of long-run default rates

4.2. Effect of Persisting Customers on Acceptance Range. We examine the influence of the number of persisting customers on the width of the acceptance range. Again, we choose $q = 4$ and, for the sake of simplicity, we consider the case at the level of the individual grades choosing $PD_{\text{grade}} = 0.02$. The proportion of persisting customers is usually lower at the level of individual grades than in the overall portfolio,
since a rating migration automatically means that the customer in question is no longer in the sample for the individual grade but is still in the overall portfolio.

The number of reference dates is $N = R(N) = 60$ and the number of customers is constant, i.e., $n_t := n = 50$ for $t = 1, \ldots, 50$. The extreme cases regarding persisting customers now represent the two scenarios in which, on the one hand, all customers remain the same over the entire history and, on the other hand, none of the customers in a specific quarter occurs in the previous quarter or in the following quarter.

The first case implies $k_{t,t+1} = n$ for $t \leq N - 1$, $k_{t,t+2} = n$ for $t \leq N - 2$ and $k_{t,t+3} = n$ for $t \leq N - 3$. Thus, for two one-year default rates $X_t$ and $X_s$ where $1 \leq t < s \leq 50$, we have

$$\text{Cov} (X_t, X_s) = \frac{1}{n} \cdot w_{t,s} (1 - \text{PD}_{\text{grade}}) \cdot \text{PD}_{\text{grade}}.$$  

For the second case, we have $k_{t,t+1} = 0$ for $t \leq N - 1$, $k_{t,t+2} = 0$ for $t \leq N - 2$ and $k_{t,t+3} = 0$ for $t \leq N - 3$, implying

$$\text{Cov} (X_t, X_s) = 0$$

for two one-year default rates $X_t$ and $X_s$. Since the covariance between default rates is always 0, in this case, the distribution of the long-run default rate is the same as if correlation effects due to overlapping time windows were ignored completely. Note that such a scenario is extremely unlikely, especially for a large number of ratings, and is therefore purely hypothetical since customers would constantly switch rating grades, which would imply a high level of instability in the rating system.

In Figure 3, we can see that the width of the acceptance range to a given level $\alpha$ for a portfolio only consisting of persisting customers is almost doubled compared to a portfolio with no persisting customers with respect to a three months time horizon.

![Figure 3. Effect of proportion of persisting customers on acceptance range per confidence level $\alpha$](image-url)
4.3. Some Thoughts on the Rate of Convergence. In view of Section 2.3, we briefly discuss the rate of convergence and mention a few particularities that need to be taken into account concerning the asymptotic behaviour of the long-run default rate $Z$. When using the classical central limit theorem, it is common to specify a minimum number of random variables ensuring a reasonable approximation of the normal distribution. In our case, additional factors need to be taken into account since, for example, a large number of customers, i.e., a large number $M$ in

$$Z = \frac{1}{N} \sum_{t=1}^{N} \frac{1}{n_t} \sum_{j=1}^{M} y_{j,t} = \sum_{j=1}^{M} Y_j,$$

does not directly lead to a good approximation. This stems from the calculation logic of the long-run default rate. For a simple illustration, we assume that we are in the setting, where all customers have the same PD with $PD = 0.01$. For $N = 1$ and $M = 1000$ or, analogously, $N = 1000$ and $n_t = 1$ for $t = 1, \ldots, 1000$, the normal approximation is way better than for $N = 2$ with $n_1 = 1$ and $n_2 = 999$ since, in the first case, $P(Z \geq 0.5) \approx 0$ while, in the second case, $P(Z \geq 0.5) > 0.01$. Thus, the rate of convergence depends on the number of customers, the number of reference dates, and the number of customers per reference date. While the previously described scenarios are uncommon at portfolio level, at the level of individual rating grades, constellations can arise in which, on certain reference dates, there is only one customer in the corresponding rating grade.

In the sequel, we elaborate more on this particularity and aim to provide a rule of thumb, i.e., conditions for $N$, $n_{\text{max}}$, and $n_{\text{min}}$ that imply a satisfactory normal approximation. To that end, we consider a synthetic portfolio. For the sake of simplicity, we assume that $q = 1$ and simulate a convolution of weighted binomial distributions. We study the difference between the distribution functions of the long-run default rate on the level of a rating grade with $PD = 0.01$ when using a simulation on the one hand and using the normal approximation on the other hand in different scenarios. We assume
the number of reference dates is \( N = 56 \). The number of customers \( n_t \) per reference date \( RD_t \) and thus the number of all customers \( M \) in the portfolio varies between the considered scenarios.

**Figure 5.** Distribution functions: simulation and normal approximation with \( n_t \leq 5 \) for \( t = 1, \ldots, 8 \), and \( n_t = 100 \) for \( t = 9, \ldots, 56 \)

In Figure 5, there are eight reference dates where only one to five customers are in the portfolio. On the other reference dates, we assume the number of customers to be
Figure 7. Distribution functions: simulation and normal approximation with $n_t = 1$ for $t = 1,\ldots,8$, and $n_t = 10$ for $t = 9,\ldots,56$

Figure 8. Distribution functions: simulation and normal approximation with $n_t = 2$ for $t = 1,\ldots,8$, and $n_t = 20$ for $t = 9,\ldots,56$

constant with $n_t = 100$. We observe large differences in the two distribution functions, especially in the tails, which are particularly relevant for the test. If we now increase the number of customers on the critical reference dates (those with $n_t \leq 5$) to 10, we observe, in Figure 6, that the distribution functions are almost the same. Hence, the poor convergence in Figure 5 is caused by the few reference dates with very few
customers. Furthermore, we see that, in this special case, $\frac{n_{\text{min}}}{n_{\text{max}}} \geq \frac{1}{10}$ seems to be an appropriate bound in order to obtain a satisfactory approximation.

For a second comparison, we now reduce the number of customers significantly to $1 \leq n_t \leq 10$. Again, the number of customers equals one only on the first eight reference dates. In Figure 7, we still see major differences in the distribution functions. Doubling the number of customers leads to Figure 8, where we already obtain useful results.

On the other hand, if the number of customers is very low, e.g., $n_t = 2$ for half of all reference dates, we still get a good approximation, cf. Figure 9. From our examples and statements from literature, we conclude that $N \geq 30$, $n_{\text{min}} \geq 2$, and $\frac{n_{\text{min}}}{n_{\text{max}}} \geq \frac{1}{10}$ seem to be useful conditions to guarantee a satisfying normal approximation.

4.4. An alternative way to bound the variance. In the Appendix B, we propose a different way to estimate the variance in an even more conservative way. If one accepts the associated stricter test as a conservative check of the calibration, one can use

$$\sigma^2_{\text{alt}}(\mu) := \frac{1}{N} (C + K_1) \mu - \frac{K_2}{N^2}$$

as the variance of the test statistic with parameters $K_1$ and $K_2$ depending only on the underlying portfolio, see Appendix B for the details.

In this case, we do not have to solve an optimization problem, and can determine the variance directly using the portfolio-specific parameters. However, we additionally need conservative estimates for two parameters that are not directly determined by the null hypothesis. Again, we refer to the Appendix B for the details. We briefly discuss the consequences of this alternative method at this point.

The determination of $\sigma^2_{\text{alt}}(\mu)$ is based on a successive estimation of the variance of the test statistic against the maximum conservative value in each single step. This procedure can therefore be understood as a maximally conservative solution to the test.
problem. In the case of a portfolio with approximately constant size, approximately constant distribution of default probabilities over time, and no other special features, the conservatism can often be tolerated. In case of build-up or run-down portfolios or portfolios with a focus on high investment grade obligors, the risk of a type I error may become disproportionately large and the test may therefore not be suitable. This can be easily illustrated by the following two points. The estimation of quadratic terms on the left-hand side of (19) is done using PD_{max} regardless of how many obligors belong to the associated rating grade. For portfolios with mainly obligors with good ratings, this estimate is certainly too conservative. Moreover, the parameters $K_1$ and $K_2$ depend on $\min \left\{ \frac{1}{n_t} \mid t = 1, \ldots, N \text{ and } n_t > 0 \right\}$ and $\min \left\{ k_{t,t+i} \mid t = 1, \ldots, N - i \right\}$, i.e., for portfolios where the number of customers or the number of persisting customers changes strongly over time, a lot of information is lost with this method. Thus, the estimate on the variance of the test statistic in Section 3.2 is much less conservative and suitable for significantly more portfolios.

4.5. **Additional Conditions on the Rating Distribution.** In this section, we give an outlook and offer suggestions as to which requirements can be placed on rating distributions. The implementation of some suggestions is briefly outlined. The rating distribution of $p(\mu)$ obtained by minimizing (12) tends to be U-shaped, cf. Figure 10, with all but at most one rating in the best or the worst rating grade. A U-shaped rating distribution, i.e., a distribution almost exclusively between the two extreme grades is highly unusual, both in the case of a large number of ratings and in a portfolio that hardly has any customers in the lower rating grade.

![Figure 10. Rating distribution of rating model and rating distribution after minimization without additional assumptions on the distribution](image-url)

Moreover, high values of PD_{max} lead to a narrowing of the acceptance range. Depending on the portfolio, this reduction can be disproportionately strong if, for example, the bad rating grades in low-default portfolios are heavily underrepresented. Hence, the question may be raised why the rating distribution, estimated by the rating model,
is not used directly. As already mentioned, this would, however, require assumptions that go far beyond the null hypothesis, and specifying the rating distribution in that way, would automatically determine the null hypothesis. However, it may be sensible to place additional yet conservative conditions on the rating distribution in (9).

For example, one could demand that ratings exhibit a conservative distribution, given by a function \( \varphi_1: \{PD_{m-n+1}, \ldots, PD_m\} \to [0, 1] \) across the \( n \) worst rating grades with

\[
\mathbb{P}(p_{t,j} = PD_k \mid p_{t,j} \geq PD_{m-n+1}) = \varphi_1(PD_k) \quad \text{for} \quad k \geq m - n + 1.
\]

For most portfolios a conservative choice of \( \varphi_1 \) would be \( \varphi_1(PD_k) = \frac{1}{n} \), cf. Figure 11.

Another approach would be to trust the distribution estimated by the rating model to some extent, and require that there is at least a fixed fraction \( \delta \in [0, 1] \) of the number of ratings per grade given by this distribution with density \( \varphi_2: \{PD_1, \ldots, PD_m\} \to [0, 1] \).

Then, in addition to (9), one might demand

\[
n(PD_k) \geq n_{\min}(PD_k) := \left\lceil \frac{N}{\sum_{i=1}^{N} n_i \cdot \varphi_2(PD_i) \cdot \delta} \right\rceil,
\]

where \( n(PD_k) \) is the number of the ratings in rating grade \( k \), cf. Figure 12.

![Figure 11. Rating distribution of rating model and rating distribution after minimization with assumption of uniform distribution for bad rating grades](image)

We conclude this section with a modified version of previously described approach using a uniform distribution among the \( n \) worst rating grades. As before, we consider the \( n \) worst rating grades \( \{PD_{m-n+1}, \ldots, PD_m\} \) with \( PD_{m-n+1} > \mu \). Our aim is to formulate the requirement of a uniform distribution in such a way that we can again use the method for minimizing the variance already presented. To that end, we define

\[
PD := \frac{1}{n} \sum_{i=m-n+1}^{m} PD_i.
\]

We follow the idea from Section 3.2, and proceed exactly as in the Appendix A with \( PD \) instead of \( PD_{\max} \). The effect on the variance or, almost equivalently, the acceptance
range is illustrated in Figure 13. On the one hand, this approach avoids unnecessary high conservatism. On the other hand, it still sufficiently conservative and simple to implement.

4.6. Impact of Simplification on the Acceptance Range. In this section we discuss, how different choices of affine linear functions for the simplification of the variance effect the acceptance range. We focus on the impact when choosing either the constant function $g = \text{PD}_{\text{max}}$ or identity function as in (10).
We start by computing the reduction of the acceptance range due to choosing the affine linear function as the constant $PD_{\text{max}}$ using a synthetic portfolio. We choose the number of customers per reference date to be constant with $n = 1000$, let the number of reference dates be $N = 60$, and $q = 4$. We define $PD_{\text{max}} = 0.2$, $PD_{\text{min}} = 0.0003$, and set $\mu = 0.01$. We recall that the components of the minimal solution $p(\mu)$ of (12) consist only of $PD_{\text{max}}$, $PD_{\text{min}}$ and $\mu_{k_0 + 1}$, where the latter is uniquely determined by the minimization problem.

We now choose $g = PD_{\text{max}}$, i.e., the most conservative function possible. Proceeding exactly as in Section 3.2 and in the Appendix A, we simply get different values for $\alpha_i$ and $c_i$ with $i = 0, \ldots, q - 1$, namely,

$$\alpha_i = \frac{q - i}{q} (1 - PD_{\text{max}})$$

$$c_i = 0.$$ 

Estimating the variance as in Section 3.2, once with $g = PD_{\text{max}}$ and once with $g = \text{id}$, we get

$$\frac{|K(PD_{\text{max}}) - k(PD_{\text{max}})|}{|K(\text{id}) - k(\text{id})|} \approx 0.994,$$

where $K(PD_{\text{max}})$ and $K(\text{id})$ denote the upper and lower bound of the acceptance range for $g = PD_{\text{max}}$ and $g = \text{id}$, respectively. In this particular setting, we see that the choice of $g$ does not have a substantial influence on the size of the acceptance range.

The difference in the size of the acceptance range is largely influenced by the contributions of the obligors with $PD = PD_{\text{min}}$. By Theorem 1, large values of $PD_{\text{max}}$ lead to an increased number of pairs $(t, j)$ with $p_{t,j} = PD_{\text{min}}$. Hence, the size of the quotient in (13) is reduced for very large values of $PD_{\text{max}}$ and very small values of $\mu$. For instance, by redefining $PD_{\text{max}} = 0.45$ and $\mu = 0.0006$ we get

$$\frac{|K(PD_{\text{max}}) - k(PD_{\text{max}})|}{|K(\text{id}) - k(\text{id})|} \approx 0.778.$$

As a result, relevant differences in the acceptance ranges mostly occur in settings with very large values of $PD_{\text{max}}$ and values of $\mu$ that are close to $PD_{\text{min}}$. In practice, the maximally conservative choice of $g = PD_{\text{max}}$ is therefore not recommendable, since it assumes that on each reference date all persisting customers have been in the worst rating grade, independent of their current rating. Concluding, it is recommended to choose $g = \text{id}$ rather than $g = PD_{\text{max}}$.

APPENDIX A. Minimization Problem

Let $n \in \mathbb{N}$ and $\alpha_i \geq 0$ for $i = 1, \ldots, n$. In this section, we compute the minimum of the function

$$f : \mathbb{R}^n \to \mathbb{R}, \quad x = (x_1, \ldots, x_n) \mapsto f(x) := \sum_{i=1}^{n} \alpha_i x_i$$

under side conditions

$$\sum_{i=1}^{n} \beta_i x_i = m \quad \text{and} \quad c_1 \leq x_i \leq c_2 \quad \text{for} \quad i = 1, \ldots, n$$

with given constants $\beta_i > 0$ for $i = 1, \ldots, n$, $m > 0$, and $0 < c_1 < c_2$. We sort the ratios $(\alpha_i/\beta_i)_{i=1,\ldots,n}$ in descending order and denote by $\alpha_{(k)}$ and $\beta_{(k)}$ are the numerator and denominator of the $k$-th largest element of these ratios. In case of equal values
among the ratios \((\alpha_i/\beta_i)_{i=1,...,n}\), the one with smaller index \(i\) is listed first. Moreover, \(x_{(k)}\) denotes the variable of \(f\) with coefficient \(\alpha_{(k)}\). We define
\[
k_0 := \max \left\{ k \left| \sum_{j=1}^{k} c_1 \cdot \beta_{(j)} + \sum_{j=k+1}^{n} c_2 \cdot \beta_{(j)} \geq m \right. \right\}.
\]

**Theorem 1.** The cost function in (14) is minimized under (15) for
\[
x_{(k)} = \begin{cases} 
c_1, & k \leq k_0, 
c_{k_0+1}, & k = k_0 + 1, 
c_2, & k_0 + 2 \leq k \leq n,
\end{cases}
\]
where the value \(c_{k_0+1}\) is uniquely determined by (15) through
\[
\mu_{k_0+1} = \frac{m - \sum_{j=1}^{k_0} c_1 \cdot \beta_{(j)} - \sum_{j=k_0+2}^{n} c_2 \cdot \beta_{(j)}}{\beta_{(k_0+1)}}.
\]

**Proof.** We show that for any \(y \in \mathbb{R}^n\) that fulfills (15), \(f(y) \geq f(x)\) holds if \(y \neq x\). Per construction of \(x\), there exists some \(\varepsilon > 0\) such that either
\[
\sum_{j=1}^{k_0} y_{(j)} \cdot \beta_{(j)} = \sum_{j=1}^{k_0} x_{(j)} \cdot \beta_{(j)} + \varepsilon \quad \text{and} \quad \sum_{j=k_0+1}^{n} y_{(j)} \cdot \beta_{(j)} = \sum_{j=k_0+1}^{n} x_{(j)} \cdot \beta_{(j)} - \varepsilon
\]
or
\[
\sum_{j=1}^{k_0+1} y_{(j)} \cdot \beta_{(j)} = \sum_{j=1}^{k_0+1} x_{(j)} \cdot \beta_{(j)} + \varepsilon \quad \text{and} \quad \sum_{j=k_0+2}^{n} y_{(j)} \cdot \beta_{(j)} = \sum_{j=k_0+2}^{n} x_{(j)} \cdot \beta_{(j)} - \varepsilon
\]
holds. Without loss of generality we regard the first case. Let \(\varepsilon = \sum_{i=1}^{k_0} \varepsilon_i\) such that
\[
y_{(i)} = x_{(i)} + \frac{\varepsilon_i}{\beta_{(i)}} \quad \text{for all } i = 1, ..., k_0
\]
and \(\varepsilon = \sum_{i=k_0+1}^{n} \varepsilon_i\) such that
\[
y_{(i)} = x_{(i)} - \frac{\varepsilon_i}{\beta_{(i)}} \quad \text{for all } i = k_0 + 1, ..., n.
\]
Then,
\[
\sum_{j=1}^{k_0} y_{(j)} \alpha_{(j)} = \sum_{j=1}^{k_0} \left( x_{(j)} + \frac{\varepsilon_j}{\beta_{(j)}} \right) \alpha_{(j)} \geq \sum_{j=1}^{k_0} x_{(j)} \alpha_{(j)} + \frac{\alpha_{(k_0)}}{\beta_{(k_0)}} \varepsilon
\]
and
\[
\sum_{j=k_0+1}^{n} y_{(j)} \alpha_{(j)} = \sum_{j=k_0+1}^{n} \left( x_{(j)} - \frac{\varepsilon_j}{\beta_{(j)}} \right) \alpha_{(j)} \geq \sum_{j=k_0+1}^{n} x_{(j)} \alpha_{(j)} - \frac{\alpha_{(k_0+1)}}{\beta_{(k_0+1)}} \varepsilon.
\]
Hence,
\[
\sum_{j=1}^{n} y_{(j)} \alpha_{(j)} \geq \sum_{j=1}^{n} x_{(j)} \alpha_{(j)} + \left( \frac{\alpha_{(k_0)}}{\beta_{(k_0)}} - \frac{\alpha_{(k_0+1)}}{\beta_{(k_0+1)}} \right) \varepsilon \geq \sum_{j=1}^{n} x_{(j)} \alpha_{(j)}
\]
For the second case, one proceeds exactly in the same way. \(\square\)
We now translate the previous theorem into the setting of Section 3.2. There, we aim to minimize the sum
\[ \sum_{t=1}^{N} \sum_{j \in \Lambda_t} \alpha_{t,j} p_{t,j} \]
under the side condition
\[ \mu = \frac{1}{N} \cdot \sum_{t=1}^{N} \frac{1}{n_t} \sum_{j \in \Lambda_t} p_{t,j} \quad \text{and} \quad p_{t,j} \in [\text{PD}_{\text{min}}, \text{PD}_{\text{max}}]. \]
We rewrite \( \mu = \sum_{t=1}^{N} \sum_{j \in \Lambda_t} \beta_{t,j} p_{t,j} \)
with \( \beta_{t,j} := \frac{1}{N} \cdot \frac{1}{n_t} \). We first sort the set of all possible tuples \((t, j)\) in ascending order, first with respect to \(j\) and then with respect to \(t\), i.e., we can identify any possible tuple \((t, j)\) with a natural number \(i = 1, \ldots, n := \sum_{t=1}^{N} n_t\), which simplifies the cost function to
\[ f: \mathbb{R}^n \rightarrow \mathbb{R}, \quad p = (p_1, \ldots, p_n) \mapsto f(p) := \sum_{i=1}^{n} \alpha_i p_i \]
with side conditions
\[ \sum_{i=1}^{n} \beta_i p_i = \mu \quad \text{and} \quad \text{PD}_{\text{min}} \leq p_i \leq \text{PD}_{\text{max}} \quad \text{for} \quad i = 1, \ldots, n. \]
Now, we are exactly in the setting of (14) and (15), and can apply Theorem 1.

**Appendix B. Test on Portfolio Level without Solving Minimization Problem**

We start by estimating the covariance between two default rates on reference dates \(\text{RD}_t\) and \(\text{RD}_s\). Using Equation (5), we find that
\[ \text{Cov} (X_t, X_s) = \frac{1}{n_t \cdot n_s} \sum_{j \in \Lambda_t \cap \Lambda_s} \text{Cov} (x_{j,t}, x_{j,s}) \]
\[ \geq \frac{1}{n_t \cdot n_s} \sum_{j \in \Lambda_t \cap \Lambda_s} p_{j,s} w_{t,s} (1 - \text{PD}_{\text{max}}) \]
\[ = \frac{k_{t,s}}{n_t \cdot n_s} w_{t,s} (1 - \text{PD}_{\text{max}}) \quad \frac{1}{k_{t,s}} \sum_{j \in \Lambda_t \cap \Lambda_s} p_{j,s} \]
\[ \geq \frac{k_{t,s}}{n_t \cdot n_s} w_{t,s} (1 - \text{PD}_{\text{max}}) \gamma \cdot E(X_s). \]
The parameter \(\gamma > 0\) is less or equal to the ratio of the average default risk of persisting customers in the portfolio and the average default risk of the entire portfolio, i.e.,
\[ E \left( \frac{1}{k_{t,s}} \sum_{j \in \Lambda_t \cap \Lambda_s} x_{j,s} \right) \geq \gamma \cdot E \left( \frac{1}{n_s} \sum_{j \in \Lambda_s} x_{j,s} \right). \]
In view of the fact that persisting customers should be the majority in the portfolio, it is reasonable to assume that \(\gamma \approx 1\). The value is to be estimated conservatively depending on the portfolio under consideration, i.e., in case of doubt, it should be chosen relatively small. If a portfolio consists only of persisting customers, \(\gamma\) is always
1, while the parameter for portfolios with a small proportion of persisting customers can be above or below 1. In the following, the parameter $\gamma$ is to be understood as an estimate that is independent of the reference date and conservative for the entire history. From Equation (8), we see that

$$\sigma^2 \geq \frac{1}{N^2} \sum_{t=1}^{N} \frac{1}{n_t^2} \sum_{j \in A_t} p_{t,j}(1-p_{t,j}) \geq C \cdot \mathbb{E}\left(\frac{1}{N} \sum_{t=1}^{N} X_t\right)$$

(19)

with $C = \frac{1-PD_{\text{max}}}{n_{\text{max}}}$. Hence, using Equation (6),

$$N^2 \cdot \sigma^2 = \sum_{t=1}^{N} \sigma_t^2 + 2 \sum_{t=1}^{N} \sum_{i=1}^{N} \text{Cov}(X_t, X_{t+i})$$

$$\geq \sum_{t=1}^{N} \sigma_t^2 + 2 \sum_{t=1}^{N} \sum_{i=1}^{q-1} \frac{k_{t,t+i}}{n_t \cdot n_{t+i}} \cdot \frac{q-i}{q} (1-PD_{\text{max}}) \cdot \gamma \cdot \mu_{t+i}$$

$$\geq \sum_{t=1}^{N} \sigma_t^2 + 2 \sum_{t=1}^{N} \min_{i=1, ..., N-i} \left( \frac{k_{t,t+i}}{n_t \cdot n_{t+i}} \right) \frac{q-i}{q} (1-PD_{\text{max}}) \cdot \gamma \sum_{t=1}^{N-i} \mu_{t+i}$$

$$\geq \sum_{t=1}^{N} \sigma_t^2 + K_1 \sum_{t=1}^{N} \mu_t - K_2$$

with

$$K_1 := 2 \sum_{i=1}^{q-1} \min_{i=1, ..., N-i} \left( \frac{k_{t,t+i}}{n_t \cdot n_{t+i}} \right) \frac{q-i}{q} (1-PD_{\text{max}}) \cdot \gamma$$

and

$$K_2 := 2 \sum_{i=1}^{q-1} \min_{i=1, ..., N-i} \left( \frac{k_{t,t+i}}{n_t \cdot n_{t+i}} \right) \frac{q-i}{q} (1-PD_{\text{max}}) \cdot \gamma \cdot i \cdot \mu_{\text{old}}$$

where $\mu_{\text{old}} \geq \max_{i=1, ..., q-1} \mu_i$. The term $\mu_{\text{old}}$ cannot be determined by the null hypothesis, thus a conservative value is to be determined. For portfolios consisting of customers with good credit ratings, the choice $\mu_{\text{old}} = PD_{\text{max}}$ is clearly not adequate. Using Equation (19), it follows that

$$N^2 \cdot \sigma^2 \geq c \sum_{i=1}^{N} \mu_i + K_1 \sum_{i=1}^{N} \mu_i - K_2 = N (C + K_1) \mu - K_2$$

with $c \geq \frac{1}{n_{\text{max}}} (1-PD_{\text{max}})$ and

$$\sigma^2 \geq \sigma^2_{\text{alt}}(\mu) := \frac{1}{N} (c + K_1) \mu - \frac{K_2}{N^2}$$

(20)

Moreover, for portfolios with a short history and low PDs, it is not always possible to choose $\mu_{\text{old}} = PD_{\text{max}}$ since $\frac{1}{N} (c + K_1) \mu - \frac{K_2}{N^2}$ could become smaller than 0. In order to be able to rule out this case, which also appears to be of little interest for applications of the test, we set an upper bound for $\mu_{\text{old}}$, i.e.,

$$\mu_{\text{old}} \leq \frac{N \cdot (c + K_1) \mu}{(q-1) \cdot K_1}.$$
Considering the random variable \( Z^* \sim \mathcal{N}(LRCT, \sigma^2_{alt}(LRCT)) \), we define the limits of the acceptance range as

\[
k := LRCT + \phi^{-1}\left(\frac{1}{2} \alpha\right) \sigma_{alt}(LRCT)
\]

and

\[
K := LRCT + \phi^{-1}\left(1 - \frac{1}{2} \alpha\right) \sigma_{alt}(LRCT).
\]

The calibration test passes if

\[
LRDR \in [k, K].
\]

REFERENCES


