Bank of Russia

Minimum sample size definition for the purpose of the loss provisions' extrapolation under the presence of default correlation

Working Paper Series

No. 128 / May 2024

Henry Penikas

Henry Penikas

Bank of Russia, Research and Forecasting Department, penikasgi@mail.cbr.ru

Author is grateful to Ksenia Yudaeva, Alexander Morozov, Andrey Sinyakov (Research and forecasting department of the Bank of Russia), Evgeny Rumyantsev (Financial stability department of the Bank of Russia), Ruslan Bulatov, Ivan Cherkasov, Olga Kadreva (Banking regulation and analytics department of the Bank of Russia), Yury Festa for the discussion of the preliminary results; to Vasily Dudkin for drawing attention to the question; to anonymous reviewer for the feedback onto the earlier version of the manuscript; to Sergey Kozhevnikov (Research and forecasting department of the Bank of Russia) for the preparation of the English version of the manuscript; to Kirill Anikeev (Research and forecasting department of the Bank of Russia), Artur Zmanovsky (Monetary policy department of the Bank of Russia), Dmitry Smirnov (Public relations department of the Bank of Russia) for help in shaping the layout of the wording.

Cover image: `Shutterstock/FOTODOM`

Address: 12 Neglinnaya street, Moscow, 107016
Tel.: +7 495 771-91-00
Fax: +7 495 621-64-65
Website: www.cbr.ru

Minimum sample size definition for the purpose of the loss provisions' extrapolation under the presence of default correlation

Henry Penikas

Bank of Russia, Research and Forecasting Department, penikasgi@mail.cbr.ru

May 30, 2024

Abstract

In 2016, the Bank of Russia developed two ordinances to set forth a procedure using a limited sample of loans to conclude whether the level of loss provisions in the portfolio of uniform loans is sufficient or not and whether the bank's capital is adequate.

The existing procedure of reserve sufficiency evaluation previews as a rule considering only a part of the loan portfolio and transfer (extrapolation) of the provision thus assessed for the overall portfolio. Moreover, the acting approach to define the minimum loan sample size assumes the absence of the default correlation.

Author's contribution is the application of the known, though often neglected properties of the distribution of the sum of the correlated Bernoulli events (trials) for the novel task, namely for the provision extrapolation, which did not consider the possibility of the default correlation existence. As a result, we prove that its presence requires larger minimum sample size of loans compared with the instances of its absence. More specifically, we justify how the minimum sample size of loans depends upon the absolute and relative differences in default rates (provision rates, rate of legal violations) within two samples, upon the required significance levels and the levels of statistical power.

Key words: credit risk, probability of default, correlated Bernoulli trials.

JEL codes: C58, E58, G18, G21, G28.

1. Introduction

1.1. Motivation and research objective

In 2016-17 Bank of Russia (BoR) introduced a novel banking supervision approach which enables to increase the supervision efficiency by economizing resources spent for the surveillance of the similar loan agreements, see the principal legal regulation Bank of Russia (2017b), equivalent draft legislation Bank of Russia (2017a) and the approach description in the paper (Pozdyshev, 2017, pp. 14, 17).

The foundation of the supervision efficiency rise is the review of the portion of the loan portfolio. When reviewing the portion justified by the relevant legislation, the conclusion about the provisions' sufficiency or insufficiency are *extrapolated* (transferred, aggregated) at the entire portfolio level. Such transfer is justifiable when the loan comprises of the *homogeneous* loans. We will explain such homogeneity criteria below in subsection 1.2.

The above mentioned loan loss extrapolation approach is based upon the probability theory principles. We discuss them in greater detail in section 2. In short, we are interested in the minimal sample size sufficient enough to reject the null hypothesis. In other words, if we knew the information on all the objects (loans), we could unanimously (without a statistical deviance) conclude which the amount of provisions should be and hence whether the existing amount is sufficient or not. If we dealt with a portion of loans only, we may provide answers to the previous questions with a certain degree of uncertainty (with statistical error). Simply saying, if verifying only one in thousand agreements, one may get materially mistaken (the error is likely to be less than by thousand times, but still sizeable in terms of the provision amount). This is why, *given the distribution* of the variable of interest - in our case, it is the credit risk amount and its reflection in the amount of provisions (formed and required) - the probability theory helps us to define how many of loan agreements out of a thousand is sufficient to survey so that the arisen statistical error is still negligible (in fact, the degree of negligence is manageable what as an extension we get to in the very end of the paper, in subsection 7.2).

The above described Bank of Russia approach on loan provision extrapolation is a canonical example of the world best audit practice. Is presumes started the inspection with the minor part of objects (contracts, deals). In case of problems or deviations from the inspectors' expectations, the size of surveyed items is expanded up to the full coverage.

Though the task of extrapolation (of loan provisions - in the case of Bank of Russia) is globally recognized, the conventional approach is suitable for only a part of cases. The key element of the approach is its *assumption about the distribution* of an attribute of interest. In all cases which the author came across and which are described in the literature review section the *normal* distribution is assumed (i.e., that the data follows the bell-shaped form of the distribution coined by Carl Friedrich Gauss). Indeed, the distribution takes the normal (Gaussian, bell-shaped) form as a result of the law of large numbers' application (when we repeat our experiment multi-fold). However, it is true if and only if the objects (trials) of interest are *independent*.

When there is a relationship (not necessary a legally binding one of control or ownership; it is sufficient to be exposed to the common macroeconomic factors; for instance, small enterprises are more risky than the largest enterprises in the country) in-between objects (agreements, loans), then even under the law of large numbers the distribution of the sum in no longer Gaussian (unimodal), but it now has two modes (hikes). Such a drastic change in the distribution shape materially changes the findings over the sufficient minimal sample size. As a findings' preview, we argue that such positive relationship requires much larger minimal sample size.

This is why the **research objective** is the justification of which minimal sample size is sufficient given various parameter combinations including the mean default (provision) rates, as well as the default correlations. To meet the objective we use the known properties of the probability distribution of the correlated Bernoulli trials. The problem setting is novel in itself. The findings are of use for the supervision practice of the Bank of Russia, as well as for the practice of the internal and external audits in banks and corporations.

1.2. *Loan portfolio homogeneity (stability and granularity) criteria*

Loan provision extrapolation is viable if and only if the loan book is homogeneous. Otherwise, we come to an typical obstacle of comparing two ratios (fractions) for the two samples. We take the simple mean of ratios in one case and derive the ratio of numerators' total over the denominators' total.[1]

For the purpose of the current study, we take the loan homogeneity as granted. We did not challenge such homogeneity criteria - did not wish neither to confirm, nor to revise those - though it may be subject of a separate study. For the purpose of the existing research we limit ourselves to the references on the documents which suggest when loans are considered to be homogeneous within the research context, i.e., within the Bank of Russia documents: Bank of Russia (2017a,b).

All criteria can be broken down into two groups. The group of major criteria recommends which loans should be excluded, while the minor one - which ones should be included. Homogeneity criteria are the former ones, thought the latter ones are important to ensure the completeness criterion.

Documents contain the following homogeneity (stability and granularity) criteria when met can imply the application of the *provision extrapolation* principle:

1. Extremely large loans are excluded when they exceed 5% of the capital, (Bank of Russia, 2017b, par. 1.7);

2. Extremely small loans are excluded when they are less than RUB one thousand, (Bank of Russia, 2017a, par. 2.2);

3. Loans with the 100% loan loss provision (reserve) are excluded, (Bank of Russia, 2017b, par. 2.3.2), (Bank of Russia, 2017a, par. 2.2);

4. Loans surveyed should fully reflect the credit risk structure (Bank of Russia, 2017b, par. 2.3.8), i.e., they should be homogeneous in terms of the credit risk, (Bank of Russia, 2017a, par. 2). Such a requirement might be treated as having a complex loop cycle. When requiring loans to be homogenous in terms of credit risk, we implicitly assume that such evaluation is being done in correct manner or is proportionately biased. Otherwise, if the credit risk bias is not proportionate by loans (e.g., by same size loans), then the loans stop being proportionate. Hence, they cannot be considered together. Then the provision adjustment can neither be adequate. When dealing with such a loop cycle, one has to always decide which point in the loop should be the departing one. As a rule, the bank provision estimate is the baseline one. Using other estimates and other re-estimation approaches can form a separate research agenda.

5. Loans should be comparable (homogeneous, have similar parameter values) within such groups, analytical indicators, as delinquent loans duration (Bank of Russia, 2017a, par. 2.4), loan categories: mortgage, auto, consumer, other (Bank of Russia, 2017b, par. 2.3.5.1).

6. Loans to individuals and to small businesses cannot be considered homogeneous, (Bank of Russia, 2017b, par. 2.3.5.4), (Bank of Russia, 2017a, par. 2.6).

Though the legislation contains a recommendation for the loan book to be homogeneous, it is also advised by (Bank of Russia, 2017b, par. 1.7) to cover the major share of the book, i.e., so that the total size of loans covered exceeded 70% of all the assets. When the latter criterion is not met, the full coverage should be done and every loan agreement should be checked, as the homogeneity criteria may result in tiny portfolios and by total amount may not cover the most part of the total assets.

[1]Author expresses special thanks to Ruslan Bulatov, Ivan Cherkasov, Olga Kadreva for the recommendation to discuss the loan portfolio homogeneity criteria!

1.3. Manuscript composition

To describe how we arrived at the findings, the paper is structured in the following way. The next section 2 contains the literature review. It describes the problem of the minimal sample size definition with application to various domains of human activity from medicine to finance. The subsequent section 3 presents a data case to evidence why it is important to account for the default correlation when auditing the loan portfolio. Then section 4 lays down the research methodology as to how non-parametrically (not via analytical derivation of formulas) we obtain the answer to the posed question on the minimal sample size definition depending upon different parameter combinations. Major findings are disclosed in section 5. They are followed by the relevant robustness checks in section 6 to find out which input parameters impact the results to a greater degree when defining the minimal sample size. Conclusion and methodological issues finalize the paper in the final section 7.

2. Literature review

The task of determining the minimum sample size is relevant to sociology, mathematical statistics, medicine, financial and IT audit. In what follows we describe in detail the valuable results in the above-mentioned areas for the task of extrapolating provisions.

2.1. Sociology

To answer the question of how many observations (objects, agreements) are sufficient to consider for a reliable estimate of the parameter under study, it is sufficient to take typical *representatives* from the sample. This brings the task of creating a representative sample. It is of particular relevance in social surveys and research (for example, Davydov (1990); Novikov (2001).

In sociology, there are a number of approaches to building a representative sample. The baseline sample is *random* when data on each k-th person (the tenth, one hundredth, etc.) needs to be collected. Since public opinion can significantly depend on major observable factors (gender, age, region, type of settlement, income), the working approach is to first consider all these factors and then random sampling is then performed for subgroups (strata) that are already selected. This is how we come to the task of obtaining a *stratified* sample. If the analysis is based on two factors (region and type of settlement), there will be a two-tier random stratified sample. If these two factors sufficiently distinguish consumers, the sample obtained from all the consumers (the general population) will be representative, that is will reflect the range of opinions among *all* the consumers.

Therefore, when the sociological approach is applied to financial tasks, it may be appropriate to distinguish financial strata (by industry and borrower region, for example). This will make the resulting sample representative. If it shows the shortage of provisions of x per cent – as a result of the representative nature of the sample – it means that the total loan book under study is short of provisions by the same x per cent.

2.2. Mathematical statistics

Determining the minimum sample size to test a statistical hypothesis may seem a thoroughly studied topic, especially for variables in the form of proportions (p fractions), see, for example, Demidenko (2007), (Hogg et al., 2015, p. 327, formula 7.4-3), Bush (2015); Raos (2021).

The basic solution is to find a sample size when the standard error of the estimated ratio (fraction) $\sqrt{p \cdot (1-p)/n}$ is within certain error ϵ. Then, consistent with a conservative approach, $\hat{p} = 0.5$ и $\epsilon = 0.05$, and the minimum sample is equal to 100 observations, (Raos, 2021, p. 294).

$$n \geq \frac{p \cdot (1-p)}{\epsilon^2} = \frac{0.5 \cdot (1-0.5)}{0.05^2} \approx 100, \tag{1}$$

where n is the sample size (number of observations/objects); p is the true value of the share.

However, this approach does not take into account the distribution of the share (the confidence interval around it). Then, assuming a normal approximation, Gaussian distribution quantile (1) добавляется квантиль гауссовского распределения $N_{\alpha/2}$ is added to formula (1) to test the hypothesis that $\hat{p} \geq p$.

$$\hat{p} \geq p + N_{\alpha/2} \cdot \frac{p}{\sqrt{n}}, \tag{2}$$

where $N_{\alpha/2}$ is the quantile of normal distribution with significance level αin a test of the two-sided hypothesis of parameter inequality (for 95% confidence level ($\alpha = 1 - 0.95$) $N_{0.025} = 1.96$), (Raos, 2021, p. 294); p - is the true value of the share (to make conservative estimates of the minimum sample size, the value 0.5 is taken, which requires the maximal sample size, all else being equal); $\hat{p}$ is the value of the share which can be distinguished from the true value from the sample size n and at significance level α.

From formula (2) follows the basic property of statistical hypotheses that an increase in sample n reduces the standard error and almost always allows to prove the presence of a significant difference (a check of the research hypothesis), although in fact there may not be differences. This negative property leads to Demidenko (2016) recommending the use of the ROC, rather than t-statistics or, as the author calls it, the *d-value* (instead of the p-value).

The paper (Raos, 2021, pp. 291-297, subsections 16.1-16.2) describes how to determine the minimum sample size to obtain a reliable estimate of share p; it also discusses the limitations of this approach. Focus should now turn to Figure 16.2, which illustrates the logic of determining the minimum sample size by creating artificial data. Let us assume that we have two shares: true (50%) and sampling (60%). The problem is to determine the number of observations which is sufficient to make a valid conclusion that the sample share is different from the true one. A 'validity conclusion' here means that the confidence interval for the sample share should not include the true value in most cases. Such a majority of cases are also called statistical *power*. According to the source, the common threshold of statistical power is 80%. Then, essentially the following formula (3) should be used:

$$n \geq (N_{1-\alpha/2} + N_{1-\gamma})^2 \cdot \frac{\hat{p} \cdot (1 - \hat{p})}{(\hat{p} - p)^2} = (1.96 + 0.84)^2 \cdot \frac{0.5 \cdot (1 - 0.5)}{(0.5 - 0.6)^2} \approx 200, \tag{3}$$

where $1 - \gamma$ is the required statistical power.

Following the discussion of Figure 16.2, the author concludes that such an approach is insufficient since it fails to take into account the variation of the true value of the share. This is why the author's recommendation is to ensure that in most cases the lower bound of the confidence interval (CI_low) for the sample value of the share not only excludes the true (point, average) value of the share but exceeds the upper bound of the confidence interval (CI_up) for the distribution of the true value of the share.

At the same time, Raos (2021) highlights the problem that, in general, the sample value is unknown until obtained. This creates a closed cycle: the minimum sample size depends on the sample value against which the true one is checked, but the sample value proper is determined by which sample (including sample size) it is measured.

As applied to the extrapolation task, the approach described above applies as follows. Let us assume that the bank under study has submitted the estimate for its provisions as a share of the loan portfolio. At the same time, the comparable parts of other banks' statements show that the value of provisions (share) is significantly different. For example, it is significantly higher than that of this bank, that is, there is the risk that the bank under study has underestimated the provisions. Then the task of extrapolation amounts to determining how many agreements need to be randomly selected to reliably confirm that the share of provisions of an individual bank is distinctly different (lower) than that of other banks. Then the share of provisions in a particular bank is the true share (p) within the more general setting of the above task, and the share of provisions in other banks is the sample share ($\hat{p}$). If we succeed in confirming what is explained below, no further provisioning is needed. Otherwise, further provisions are needed.

2.3. Medical research

In medical research, the task of determining the minimum sample size is the basis for clinical drug trials, (Lachin, 2011, pp. 85-118, Ch. 3). University Hospitals Bristol NHS Foundation Trust has set the minimum numbers of patients, but they are non-regulatory (UHBristol, 2009, p. 2).

Similar to the description in Raos (2021), where it is important to compare the bounds of two, rather than one, confidence intervals (as discussed above), the *t-value* concept enters medical research in line with the recommendation of taking a sample two to three times more than the one that follows just from t-statistics (further details are in (Kulinskaya et al., 2008, pp. 4-5).

To determine the minimum sample size, a ready-made calculation module may be useful. It is available in the public domain: `https://www.dartmouth.edu/~eugened/power-samplesize.php` - and it is based on Demidenko (2007). The following paper says that a similar module is available in R code: `https://search.r-project.org/CRAN/refmans/WebPower/html/wp.logistic.html`.

2.4. Financial audit

The processes of financial audit, IT audit of financial companies or financial (forensic) investigation are completely based on the classic mathematical statistics problem of determining the minimum sample size.

Notably, this issue is covered by a separate standard (Standard 530 Russian Ministry of Finance (2021)) (excerpts from it are shown in Annex A.1). This standard, general in nature, lays out the principles which such a sample should meet. However, the exact number of agreements to be considered is determined by each auditor. Of use can be the heuristic rules of Herbert Arthur Sturges from equation (4) or of Benford from formula (5). The latter procedure should not be confused with the Newcomb-Benford law, although Benford corresponds to one person and this law is also applied in the financial audit to detect fraud, see Nigrini (2012).

$$n = 1 + [3,322 \cdot log_{10}(N)], \tag{4}$$

where "n is the number of intervals; N is the general population; log_{10} is the decimal logarithm; [...] is the operation of singling out the whole part (the quotient)" , (Zverev, 2018, p. 45).

$$N = [2^{n-1}], \tag{5}$$

where "N is the minimum number of elements of the aggregate; n is the required number of strata; [...] the operation of singling out the whole part (the quotient)" , (Zverev and Nikiforov, 2019, p. 73).

Interestingly, 15 years ago, a US task force made an effort to unify all approaches to determining the minimum audit sample size (2008 Audit Sampling Guide Task Force). Among its findings is a paper by Stewart (2012), which considers minimum audit samples under different distributions. Conceptually, this paper aims to accomplish the same: the minimum sample size will be considered for distributions with no correlation of outcomes (defaults) and with this correlation.

In financial audit, recommended specific digital values of minimum sample are more common, see (AICPA, 1999, pp. 26-29, 51), (NAO, 2001, pp. 9, 19), than in medicine or sociology.

Bank of Russia Ordinance 4466-U Bank of Russia (2017b) and draft Bank of Russia Ordinance Bank of Russia (2017a). The extrapolation of provisions is a special task case in financial audit. There are two Bank of Russia documents (Bank of Russia Instruction Bank of Russia (2017b) and Bank of Russia draft regulations Bank of Russia (2017a)), which, albeit conceptually close, have differences. In what follows we explain the general scheme and these differences.

The two documents were developed in 2016; the first (4466-U) focused on the sufficiency of banks' capital (assets); the second (draft), on the sufficiency of provisions in uniform loan portfolios. The logic of both documents is similar. First, it is proposed that a *random* sample is taken of 100 agreements (Clause 2.3.9.1 of Bank of Russia (2017b), Clause 3.7 of Bank of Russia (2017a)). This corresponds to a conservative recommendation from mathematical statistics (see equation (1) from (Raos, 2021, p. 294)).

Thereafter, each agreement has to be checked, a list of discrepancies made and the percentage of the agreement amount determined that should be additionally covered by provisions due to the violation. For each agreement, the largest relative violation is taken. For such violations, the average one for the sample is considered. Here, Ordinance 4466-U and the draft ordinance are different in that 4466-U calculates the average violation amount, whereas the draft proposes that the weighted agreement amount average violation be considered. Let us compare equation (6) and (7).

$$S_{DRez} = \sqrt{\frac{\sum_{i=1}^{n}(DRez_i - \overline{DRez})^2 \cdot OD_i}{\sum_{i=1}^{n} OD_i}}, \tag{6}$$

where OD_i - "is the principal amount under the i-th loan as of the valuation date " (clause 3.9, Bank of Russia (2017a)).

$$S_{DRez} = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n}(DRez_i - \overline{DRez})^2}, \tag{7}$$

where n - "is number of loans in the sample (Clause 2.3.9.2, Bank of Russia (2017b)).

Based on the table in the Annex to the ordinances, it is possible to determine whether an additional number of agreements needs to be checked. If so, they are checked as well, and the average violation is recalculated based on the expanded sample. Otherwise, the estimate for the average violation obtained from the primary sample remains unchanged. Next, we build a confidence interval *on the assumption of a normal distribution* (that is, in the absence of the correlation of defaults) for the average violation size which is equal to the random one. We take the right bound of the obtained (calculated) confidence interval to apply this value to the whole sample. For example, with the average violation at 1%, approximately on the selected quantile, which is equal to 2 for simplicity (4466-U and the draft ordinance operate different levels and thus different multipliers: 1.96 and 1.65), which leads us to conclude that the provisions for the total loan portfolio should be higher: $1\% \cdot 2 = 2\%$.

The formal sequence of calculating the additional provisions under 4466-U and the draft ordinance is summarised for the reader's convenience in Annex A.2.

2.5. Conclusions from literature

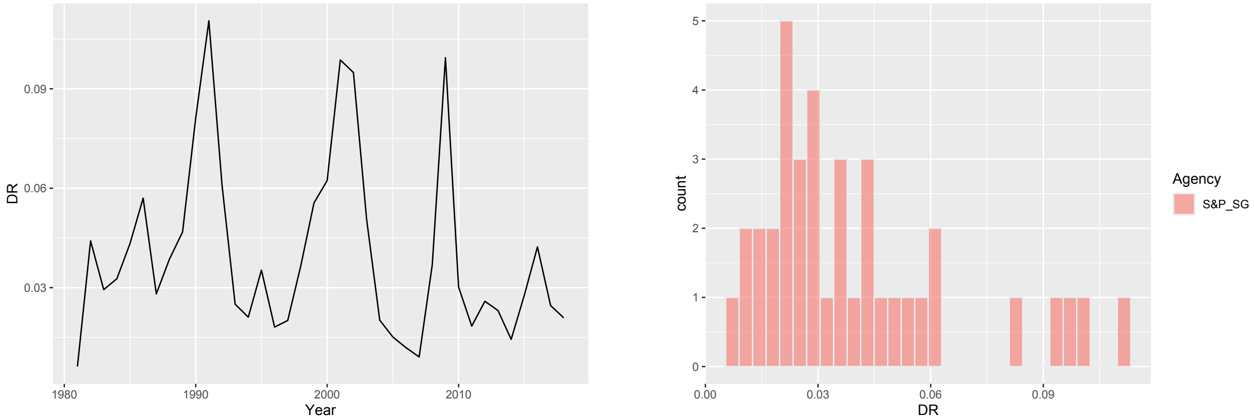
The task of determining the minimum sample size is well known and applies across various areas of activity. However, this work proposes a novel approach inasmuch as all the prior papers have assumed the lack of the correlation of outcomes (defaults, violations). This work aims to make the case for the minimum sample size given a correlation.

3. Data

To illustrate the importance of the correlation of defaults, let us turn to open data on borrowers with speculative grades, which are assigned by a global rating agency, in Figure 1. Over the past 40 years (before the beginning of the pandemic), the share of defaults (vertically) has been marked by almost a regular cycle (left of the figure). However, if the distribution of the share of defaults is considered in isolation from temporal dimension, we see the vivid bimodality of the distribution is in evidence (right of the figure). This means that the share of defaults is more likely to take on either small or large values. This distribution is misaligned with the assumption of Gaussian (normal, dome-shaped) distribution for the classic Bernoulli scheme involving a sequence of independent observations, of which each can take on a value of either zero or one. The reason for the difference from the Gaussian distribution is the correlation of outcomes (defaults in the case of the rating agency data). The Gaussian distribution assumes that there is no such correlation in the Bernoulli scheme.

To determine the correlation of defaults over a time series of data, it is sufficient to compare the factual variance of the fraction ($Var(DR)$) with the variance($\overline{DR} \cdot (1 - \overline{DR})$) in the absence of the correlation of outcomes (defaults), see formula (8) (the formula can be derived from the theoretical and probabilistic properties of the distribution of a random variables that is the sum of correlated Bernoulli outcomes; an example of a non-parametric method to arrive at the same result is discussed in Kruppa et al. (2018)).

Figure 1: Correlation of defaults, global rating agency's data (example).



Source: (S&P Global Ratings, 2019, p. 3).

$$\rho = \frac{Var(DR)}{\overline{DR} \cdot (1 - \overline{DR})}, \tag{8}$$

where $\overline{DR}$ is the average historical share of defaults; $Var(DR)$ is its variance.

Based on a global rating agency's data, in Figure 1, the result we obtain shows that the average historical share of defaults ($\overline{DR}$) is 4%, and its variance ($Var(DR)$) is 0.07%. The correlation of defaults (ρ) is thus approximately 2%. Despite the seemingly low correlation of defaults, it significantly increases the spread of values of the observed (realised) of default rate (DR).

4. Methodology

As a reminder, the procedure for determining the minimum sample size essentially consists in building confidence intervals for the two sought-after (comparison) values of the share (provisions, defaults, violations), for example for $p1$ and $p2$. A minimally sufficient sample is such in which most of the confidence intervals do not overlap.

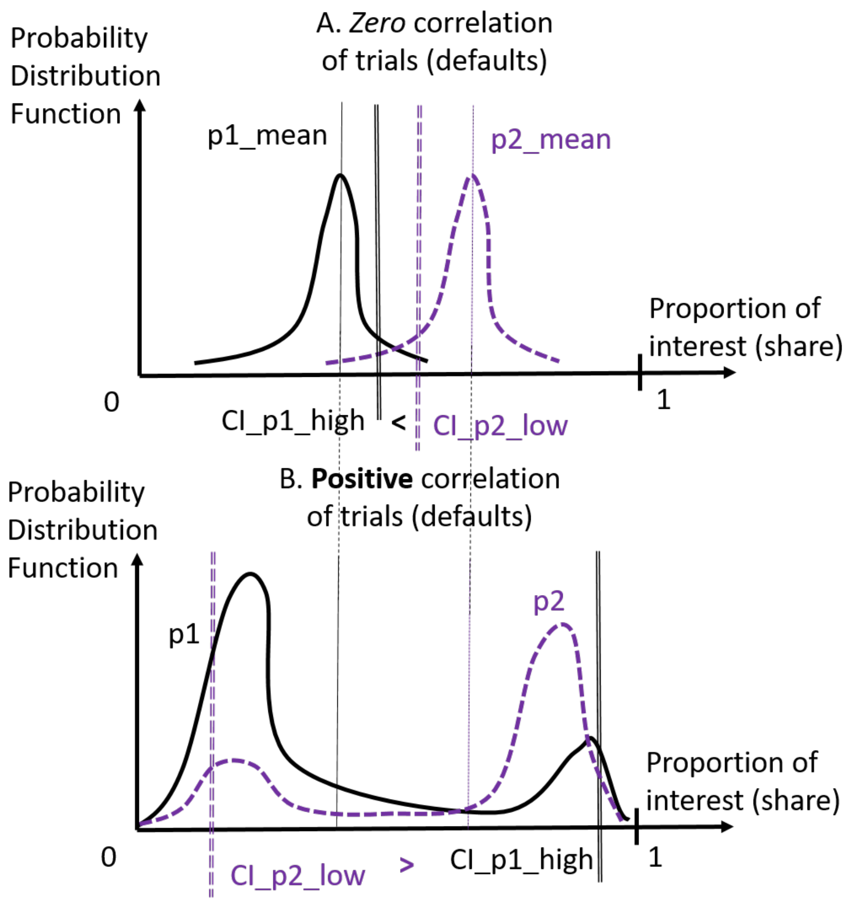
In the absence of the correlation of defaults, the schematic illustration of the above-described principle is shown in Part A of Figure 2. However, if there is a positive correlation of defaults, as in the example with data of a global rating agency in Section 3, the distributions under study become bimodal. Part B provides an example with the correlation of defaults for both comparison distributions, which is significantly close to one (resulting in almost all the realisations of the share - being zero or one with almost no intermediate values).

In that case, if one sample size is sufficient to reliably argue that $p1$ is statistically significantly different from $p2$ *in the absence of default correlation in* case A, then, in the presence of this correlation this conclusion is not robust in case B. The same sample size, *in the presence of default correlation* correlation (case B) enables us to confirm the hypothesis about a much larger difference between the parameters than the difference in its absence (case A). Considering the objective of this paper – determining the minimum sample size given the correlation of defaults – it can be expected that *growth in default correlation suggests the need for a larger sample to compare the same average values of the share (provisions, violations).* Towards this goal, we use a non-parametric approach to creating artificial data. As an example, let us consider:

To obtain the set objective, we use the non-parametric approach to generate artificial data. For illustration we consider:

- Two basic values of $p1$ (PD_true) against which to test the hypothesis: 10 and 50%. These can be understood as the **bank**'s values submitted by the bank under inspection.
- Two differences (deviations) dif from the base to the alternative value: 5 and 10%. Accordingly, the *alternative* shares $p2$ (PD_sample) are equal depending on the base value of the share: 15, 20

Figure 2: Concept of test: correlation of defaults increases spread and requires larger minimum sample.



and 55, and 60%. The difference can be understood as the average size of the violation found. The alternative share $p2$ is the full ratio for the provisions to be generated by the bank accounting for the violations identified, or it may be a rate specific to the industry or segment of borrowers.

- The four values of the correlation of defaults ρ are as follows: 0, 2.5, 5, 7.5%. We consider such low correlations for two reasons. First, they were observed in the actual data in Section 3. Second, the problem does not have a solution if the correlation values are large. Let us think of an extreme case of the correlation of defaults at 100%. Then the right quantile of the distribution with a lower value of the average share is one, and the left quantile of the distribution division with the higher value is zero. They will always overlap.

To create artificial data to rely on them in determining a sufficient minimum sample size, we use the following algorithm:

1. 1. Independently, two portfolios with number of borrowers N_loans. This provides us with two values (two realisations) of the DR share of defaults.

$$DR = \frac{\sum_{i=1}^{N} D_i}{N}, \tag{9}$$

where D_i is the default indicator (equal to zero or one) for the i-th borrower ($i = \overline{1;N}$);

   - To generate the correlation of defaults, we rely on the approach after Lunn and Davies (1998), presented by formula (10).

$$D_i = U_i \cdot Y + (1 - U_i) \cdot X_i, \tag{10}$$

where $Y \sim Bin(1, PD)$ - is the binomial (Bin) distributed system (common) factor for one realisation, where one realises with probability PD and zero otherwise; $X_i \sim Bin(1, PD)$ - is also a binomially distributed (but independent of Y) random value characterising the individual (idiosyncratic) factor $U_i \sim Bin(1, \rho)$ is a measure of the contribution of the system factor to the default of the borrower; it is also binomially distributed but its parameter of the correlation of defaults is ρ.

2. We then repeat operation (1) to create portfolios N_portf times. This brings us two DR distributions. In one distribution – *based on the bank's data* - we take the quantile from the top (CI_up, here and below we use notations from the code in Annex), in the second, the *alternative (by industry, by first sample)* from below (CI_low). We now take interest in cases of the upper quantile of the first distribution being below the lower quantile of the second ($CI\_low \geq CI\_up$).
3. We repeat iteration (2) N_ci times to ensure that the condition (the ratio of quantiles) is met at least $1 - \gamma$ times. For example, we take $\gamma = 0.2$ to ensure that the test power requirement is met (it is met in 80% of the cases, CI).
4. We implement iterations (1-3), gradually increasing the sample size N_sample (N_loans), and stop when the target test power is reached.
5. 5. The sample size corresponding to the algorithm stop is the sought-for minimal number of observations.

Reproducible code is shown in Annex A.4. This code allows to estimate the sufficiently small sample for any parameters of interest, not limited to those listed above, *provided that the problem has a solution (that is, for the correlation of events:* $0 \leq \rho << 1$*).*

5. Major findings

The key findings will be presented in stages. First, let us explain the logic of algorithm operation based on one example in Subsection 5.1, thereafter we present a summary of findings for a wide range of combinations of parameters in Subsection 5.2.

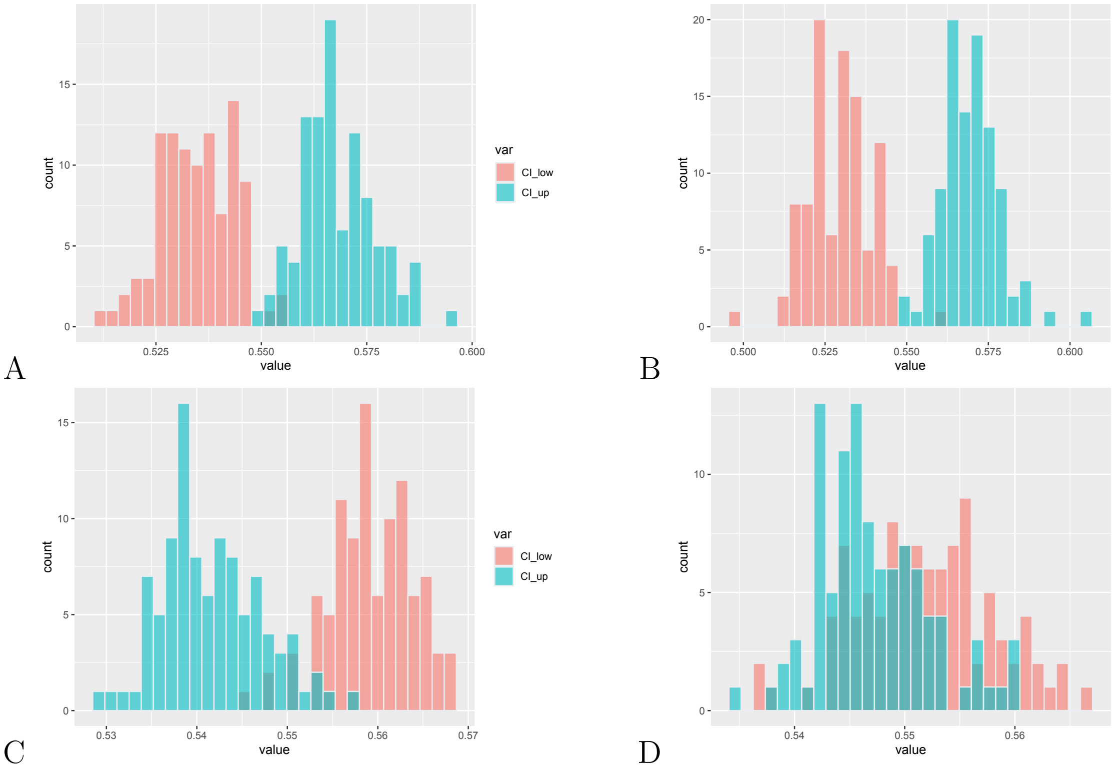
5.1. Discussion of one illustrative example

Let us show the logic of the algorithm to determine the sufficiently small sample size in Figure 3. The figures show two histograms of the quantile distribution (not the shares of default themselves): the red distribution is the lower quantile (CI_low) of the higher value $p1$; and the green one is the upper quantile (CI_up) of the lower value $p2$. The target situation is when the green distribution is on the left and the red one on the right. This situation to the greatest extent corresponds to Figure C, where no correlation of defaults is assumed with the minimum sample size of 500 observations. Here, 98% of the confidence intervals do not overlap (the upper distribution quantile with a lower share value is in 98% of cases to the left of the lower distribution quantile with the higher value).

If the sample is reduced to 200, we move from Part C to Part A – where the distributions are located in quite the opposite way, that is, all the confidence intervals are fully overlapped. It means that for the indicators of shares under study ($p1 = 50\%, p2 = 60\%$) a sample of 200 observations is not enough to argue that the indicators are statistically different. In practice, this suggests that following a review of 200 agreements, the bank cannot make the case that its provisions should be at 50% – vs the industry (subsample) average of 60%. Consequently, it is reasonable to require that the bank adjust its provisions upwards by 10pp on average, in addition to the existing average level of 50%.

If we assume in case C that there is a correlation of defaults and that it is similar in scale to the data of the global rating agency in Section 3 ($\rho = 2.5\%$), then we move from case C to D. We observe a deterioration in this case, which is however not critical. The target ratio of confidence intervals is in place only in 69% of cases, which is already less than the acceptable target power of 80% (see Section 2), that is, a 500 observation sample is insufficient to reasonably test the hypothesis about ratio of shares $p1$

Figure 3: Obtained minimum sample size (graphic illustration)



Note: $p1 = PD\_true = 50\%$ ((the green columns: the upper quantile is taken CI_up),
$p2 = PD\_sample = 60\%$ (the red columns, the lower quantile is taken – CI_low).
Sample size by subfigure: A, C - $N = 200$; B, D - $N = 500$.
Test power by subfigure: A - 0%; B - 0%; C - 98%; D - 69%
(acceptable result at power above 80%).

and $p2$. A priori, the 80% test power requirement given this combination of parameters will be met with a sample of 550 observations (see Table 1 of the Annex).

5.2. General findings

The obtained results are clearly summarised in Figure 4. The vertical value is the sufficiently small sample size, and the horizontal value is the ratio of deviation (discrepancy) dif to the value of the share (provision) in bank sample $p1$ (PD_true). The left part (A) of the figure contains the results for the absolute deviation (dif) of five percentage points, the right part (B), for a deviation of ten points. The dashed line corresponds to the assumption that the correlation of defaults is equal to 2.5% in both distributions; the solid line assumes no correlation of defaults. The details for a larger range of default correlations (including the combinations of its presence in one sample and absence in another) are shown in Table 1 in the Annex.

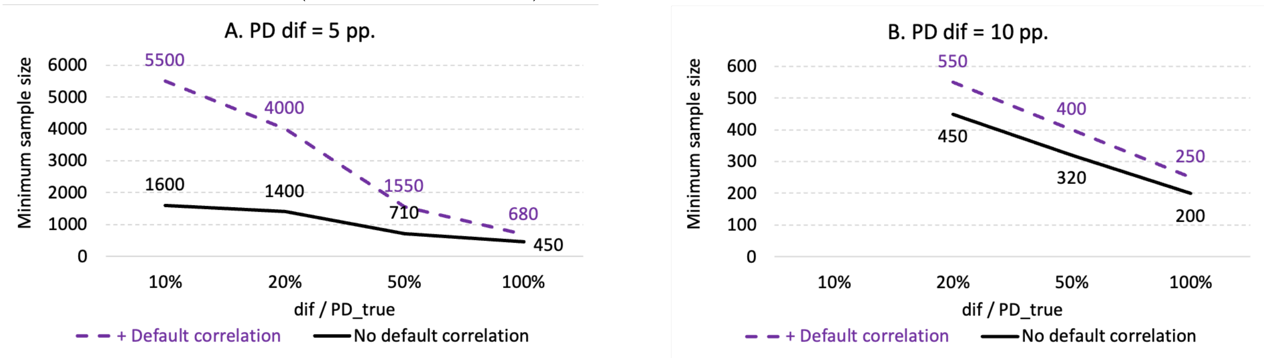
Therefore, the following key conclusions can be made from Figure 4. The less the minimum sample size,

- the greater the difference in *absolute* terms between the shares under check (dif). Notably, if such a difference is small relative to the share submitted by the bank (for example, 20%), then the *savings* the sample affords are significantly higher than for a high relative share (for example, 100%): cf. 4,000 and 550 against 680 to 250 (if the correlation is in place).

- The higher the *relative* discrepancy between the shares (dif/PD_true).

- The lower the correlation of defaults. Moreover, the effect is more significant (fewer observations are required) in the case of a large *relative* discrepancy than in the case of a small one (movement to

Figure 4: The minimum sample size decreases with growth in the ratio of the discrepancy to the bank's estimated share (of the provisions).



Note: author's calculations, details are presented in the Annex, see Table 1.

the right along the horizontal axis); as well as in the case of a larger *absolute* discrepancy (transition from Subfigure A to B).

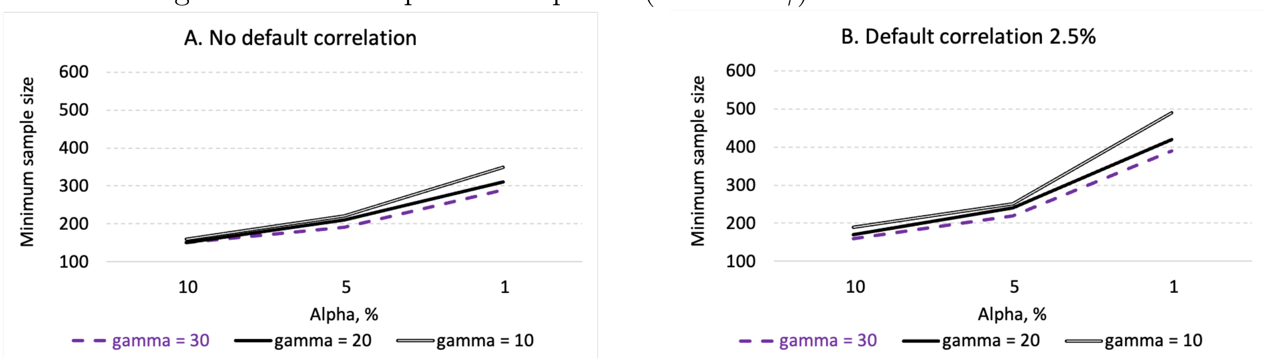
6. Robustness check

The results we present relied on a number of unchanged parameters:

- α is the significance value. The standard value of 5% is considered. Then, for a two-sided confidence interval, about two observations on the left and right are cut off by this interval.
- $1-\gamma$ is the test power (80% by default, based on Raos (2021)), where $\gamma = 0.2$.
- N_ci is the number of confidence intervals under study. The default value is 100. In order to meet the standard capacity requirement of 20%,a difference is needed for 20 intervals.
- N_portf is the number of artificial portfolios on which one confidence interval of N_ci, is determined including its upper or lower quantile. The default value is 100.
- N_step is the value fit step. The default value of 10 units was used for $p1 = 0.10$ and 50 units, for $p1 = 0.50$.

For the results of the study to find their application in supervisory practices, it is advisable to understand how the minimal size estimates change as specified parameters change. In what follows we discuss this in further detail.

Figure 5: Minimum sample size grows with increasing significance level (drop in α) and slightly increases with growth in the required test power (decrease γ).



Note: author's calculations, details are presented in the Annex, see Table 2.

In mathematical statistics, the key parameters are the level of significance and statistical power. Therefore, Figure 5 presents the calibration of the results for them (detailed results are available in Table 2).

It can be seen that a 10pp increase in required power does not significantly affect the sample size, although in general the larger required power $(1-\gamma)$ corresponds to the larger required minimum sample size. The level of significance better distinguishes the results. The requirement of non-rejection of the hypothesis with the greater accuracy of approximately 10pp (given lower α: from 10% to 1%) leads to almost twofold growth in the minimum sufficient sample size.

Unlike the significance and power parameters, the other parameters leave the results unaffected. Changing the number of portfolios under study N_portf and/or the number of confidence intervals N_ci only increases the time for calculations. For example, the search for the combination of 100–100 takes less than ten seconds, but that for 1,000-1,000 takes six minutes. Reducing the number of confidence intervals to 500 and keeping the number of portfolios at the level of two thousand requires twice as little time, that is three minutes.

At the same time, inaccuracy (the share of overlapping confidence intervals) declines from 20% to 5%. The implication is that it is possible to reduce the search step N_step step and obtain a more accurate value of the sufficient sample. While maintaining the fit step in some cases, the required sample size can vary by no more than one scale value, that is by the value of the specified fit step.

The search step itself defines the round-off limits for the sought-after number. In Table 1 of the Annex, for the highest correlation values (7.5-7.5%), a step of 500 observations is considered to ensure an acceptable search time for a sufficient sample size.

7. Findings and discussion

The task of determining the minimum sample size is typical of mathematical statistics, including cases where the sought-after indicator is considered as a proportion (ratio). However, this task has been traditionally solved on the assumption of that there is no correlation (provisions, defaults, discrepancies – for the financial industry). This is the first paper to show size requirements for the sufficiently small sample size given a correlation of defaults.

It must be acknowledged that the mathematical and statistical solution we have presented leaves open a number of issues, which, if resolved, would improve the result of the problem under study in terms of a better alignment between the interests and expectations of the involved parties – inspectors and those inspected.

These matters relate to the following areas:

1. Determining the correlation (of defaults);
2. Proportionate regulation;
3. Ways of measuring underestimates:
    - On an average or weighted average basis;
    - As a share of violations or the amounts of discrepancies and actual provisions;
4. Method of adjustment: on an average or quantile basis.

Below we make comments on the above points.

7.1. Determining the correlation (of defaults).

Equation (8) showed the way of measuring the correlation of defaults based on the share of defaults in the loan portfolio and its spread. However, it is important to understand that such a measurement takes several cuts at different times. This can be delivered by samples of different (non-overlapping, non-recurring) loans for different dates. Alternatively, but less preferably, one sample of loans, including bad loans, can be taken and its loan payment/default status can be reproduced for several dates. For such a subportfolio, it is possible to calculate the shares of defaults for each of the dates (it is not an issue that the direct composition of the subportfolio may change over time; the formula equally allows to estimate the correlation of default in this scenario since the share for the total subportfolio is considered).

Importantly, this work has studied the property of correlation of events. The term 'defaults' (the correlation of defaults) is introduced for clarity of the presentation of only one of the possible applications of the findings. However, the results apply to other measures expressed in terms of percentage, but such that not necessarily relate to defaults.[2] For example, the findings about the need for larger samples given a correlation extend to violations as well. The correlation between such violations can be naturally caused by uniform (systematic) decisions on individual matters by executives of the financial institution.

Therefore, a method needs to be developed for calculating the default correlation indicator and its appropriate application. The matter is that the bank changes its credit policy more often (than, for example, the rating agency its grading policy, as its data show in Section 3).[3] This gives the inspectors two options. First, if enough time steps have passed since a change in credit policy to measure the correlation of defaults, this correlation must be measured only for a period that is homogeneous in terms of default correlation (that is, after a change in credit policy). Second, if the time elapsed after the change is insufficient, it is recommended that the approach from stress testing be used, which consists in extending to the bank the estimates for default correlation from comparable segments of other banks or the financial industry.

As regards these measures of default correlation, it is helpful to remember that the indicator itself can change over time. Most often, it grows in times of crisis and declines in an upcycle. The first paper to record such an effect for the stock market was Longin and Solnik (2001). Therefore, in the first approximation, it is recommended to use the indicator averaged for a uniform period of time. In the absence of a sufficient data window for a particular bank, a conservative estimate for the industry should be taken.

7.2. Proportionate regulation.

Proportionate regulation has recently jumped up the global agenda, driven by innovations in banking regulation, especially in the aftermath of the 2007–2009 crisis. The regulation essentially aims to make certain that more complex financial institutions are more thoroughly audited, whereas those smaller in size or business complexity enjoy simplified procedures. This principle is reflected in the way systemically important credit institutions are designated as well as in the development of requirements for internal capital adequacy assessment processes (ICAAP).

The objective of the minimum sample size makes it possible to fully implement the principle of proportionate regulation by controlling parameter α. As noted above, the two regulatory documents (Ordinance 4466-U and the draft ordinance) use different levels that are however uniform for all organisations. It can therefore be recommended to differentiate α by degree of complexity and organisation size, that is, the more close is the organisation's position to a systemically significant one or the more complex its business, the lower its parameter α should be, that is, *all other things being equal*, the samples under review should be larger.

7.3. Methods of measuring underestimates

On an average or weighted average basis.

Ordinance 4466-U is based on the calculation of the average underestimated amount (violation), see Equation (7). The draft ordinance relies on the loan amount-weighted average, see Equation (6). In principle, both approaches are valid *when certain conditions are met.*

On the one hand, the calculation of the weighted average deviation is justified if loans in the general population are unevenly distributed by amount. However, if this is the case, such a portfolio may not be

[2] The author would like to thank Evgeny Rumyantsev for the time he took to discuss with me a certain point in this paper.

[3] The author would also like to thank Alexander Morozov for taking the time to discuss another drawing attention to this point.

uniform. It is important to note here that the logic of extrapolation is overall intended for substantially similar (uniform) loans. For example, it would be inappropriate to extrapolate the share of violations for individual loans to that for corporate loans, and vice versa. In the legal area, extrapolation is appropriate within the business segment of borrowers under study.

On the other hand, if the uneven size distribution of loans is not our concern and if it takes place in the general population, then before using the calculated weighted average it is necessary to make sure that the data are representative. Put another way, it is necessary to first compare the distribution of loan amounts in the general population and in the obtained sample (for example, based on the metrics by Kolmogorov–Smirnov or Cramer-von-Mises). If the distributions are statistically equal, it is indeed appropriate to extend the weighted average estimate to the general population. However, if the distributions are not equal, this transfer of the estimate is unjustified. This would require a separate adjustment method, which can be defined in an individual paper.

Therefore, from the practical point of view, it seems more justified to apply the average deviation, consistent with Ordinance 4466-U.

Through the share of violations or through total violations and the actual provision.

The logic of Ordinance 4466-U and the draft Ordinance is that the null hypothesis is checked, that is, the check is whether the average deviation indicator is equal to zero (for instance, whether 2% are statistically equal to zero). In actual life, however, an inspection aims to establish the true provision amount. In other words, the inspector checks another hypothesis about whether the bank's provision ratio (say, 5%) is equal to the sum of this rate (5%) and the identified shortage of provisions (2%), i.e., in our example, we wish to check whether $7\% = 2\% + 5\%$ are statistically equal to 5%.

For example, the inspector could find a 10% underestimate given the bank's ratio of 50%. The logic of the current approach is to check the hypothesis that a 10% underestimate is equal to zero. However, in actual fact the inspector attempts to establish whether it can be said that 50% and $10 + 50 = 60\%$ are equal. As seen from outside, it is tempting to say that permuting the summands of a sum does not change the value of the sum. Nevertheless, it changes in our problem. For example, a first-type check (that is, when dealing with a small (zero) value of the bank's estimate for provisions) assuming the non-existence of default correlation, 200 observations are required, see Table 1. At the same time, double the observations (450 pieces) are required to check the same absolute difference for a higher value of the bank's estimate (50%).

Therefore, to come to reliable conclusions, it can be recommended to compare not the value of identified violations with zero, but the value of the current amount of provisions and the shortage of provisions to be added. For practical inspection purposes, this means that the minimum sample should be about twice as large.

7.4. Method of adjustment.

The Bank of Russia's draft ordinance Bank of Russia (2017a) proposes that further upward adjustments be made to provisions based on the general population according to the value of the right edge (quantile) of the confidence interval close to the share of the underestimated amount of provisions for the sample created. This proposition is probably driven by conservative considerations.

However, it is appropriate to recall that the confidence interval exists not only for a random value reflecting the underestimated provision, but also for the base estimate of provisions. That same logic is behind the model by Vasicek (1987, 2002) in the internal ratings-based (IRB) approach. The average (mathematical expectation) of the distribution of loss shares is called *expected losses* and is taken as the value of provisions (deducted from the numerator of the capital adequacy ratio), while the difference between the quantile and the average is called *unexpected loss* and is added to the value of risk-weighted assets (ABP, RWA) (to the denominator of the ratio).

Therefore, it is advisable to discuss the most appropriate way of how to decompose the extrapolated provisions, for organisations with capital adequacy ratio constrains, into the expected and unexpected losses; and if to decomposed how to properly account for those decomposed elements within the capital adequacy ratio.

A. Annexes

A.1. Ministry of Finance of Russia Standard 2021 (extract), Russian Ministry of Finance (2021)

ISA 530: Audit Sampling (auditing standard):

"Extrapolation of misstatements

14. In conducting a detailed check, the inspector is obliged to extrapolate the discrepancies of the sample they have found to the whole general population (see para. A18–A20).

...

Extrapolation of misstatements (see Clause 14)

A18. In order to gain a more general understanding of the extent of misstatements, the inspector is obliged to extrapolate the misstatements to the general population; however, this may not be enough to determine the amount of misstatements.

A19. If a misstatement is found to be an anomaly, it can be excluded from the misstatements to be extrapolated to the general population. However, in extrapolating non-anomality misstatements, focus should turn to the effects of all the above misstatements if present.

A20. In testing controls, the extrapolation of deviations is not explicitly needed since the rate of deviation in the sample is also the rate of deviation extrapolated to the whole general population. ISA 330[3] provides guidance for the case when deviations are found from controls that the inspector intends to rely on.

A.2. Extrapolation of provisions in absence of default correction based on BoR Ordinance 4466-U Bank of Russia (2017b) and BoR draft ordinance Bank of Russia (2017a): detailed description of logic

$$DRezGr = \overline{DRezK} \cdot GrC, \tag{11}$$

where $DRezGr$ "is the amount of additional provisions for a loan group " (clause 2.3.16, Bank of Russia (2017b)); $\overline{DRezK}$ is the adjusted amount of provisions related to sample insufficiency (clause 2.3.15, Bank of Russia (2017b)); GrC "is the total principal amount of the loan group in rubles".

$$\overline{DRezK} = (1 + \Pi) \cdot \overline{DRez}, \tag{12}$$

where Π "is the average-size adjustment (in percentage terms) for the additional provisions" (clause 2.3.15, Bank of Russia (2017b)); $\overline{DRez}$ "is the average amount (in percentage terms) of additional provisions for loans in the sample" (clause 2.3.14, Bank of Russia (2017b)).

$$\Pi = \frac{1.96}{\overline{DRez}} \cdot \frac{S_{DRez}}{\sqrt{n}}, \tag{13}$$

where S_{DRez} "is the standard deviation of individual size (in percentage terms) of additional provisions for loans in the sample " (clause 2.3.9.2, Bank of Russia (2017b)). The assumption is that 5% will be cut off from both sides.

$$\overline{DRez} = \frac{\sum_{i=1}^{n} DRez_i \cdot SZ_i}{\sum_{i=1}^{n} SZ_i}, \tag{14}$$

where $DRez_i$ is the total amount (in percentage terms) of the upward adjustment of provisions for the i-th loan taking into account all the violations identified (clause 2.3.13, Bank of Russia (2017b)); SZ_i - "outstanding debt as at valuation date".

$$\Pi_{DRez} = 1.645 \cdot \frac{S_{\overline{DRez}}}{\overline{DRez}}, \tag{15}$$

where Π_{DRez} is the average-size adjustment (in percentage terms) of the upward adjustment of provisions for the subsample (clause 4.10, Bank of Russia (2017a)). The assumption is that 10% will be cut off from both sides.

The logic of formula (13), (15) corresponds to formula (2).

$$S_{\overline{DRez}} = \sqrt{\frac{S^2_{DRez}}{n_1} \cdot (1 - \frac{n_1}{N})}, \tag{16}$$

where OD_i " is the principal amount under the i-th loan as at valuation date " (clause 4.9, Bank of Russia (2017a)).

$$S_D = \sqrt{\frac{D \cdot (1 - D)}{n} \cdot (1 - \frac{n}{N})}, \tag{17}$$

where n "is the number of loans in the sample" (clause 4.6, Bank of Russia (2017a)); N "is the number of loans in the group of portfolios"; $D = \frac{n_1}{n}$ - "is the number of loans in the portfolio group " (clause 4.5, Bank of Russia (2017a)); where n_1 "is the number of loans in the subsample";

A.3. Granular estimates of the minimal sample size

Table 1: Minimal sample size grows when the default correlation rises to a greater extent when the provision rate is higher (PD_sample), than when it is lower (PD_true).

$PD\_true = 0.10, PD\_sample = 0.20$

Def.cor., %	0.0	2.5	5.0	7.5
0.0	200	220	230	260
2.5	230	250	260	300
5.0	380	420	550	640
7.5	1070	1370	2100	3500

$PD\_true = 0.10, PD\_sample = 0.15$

Def.cor., %	0.0	2.5	5.0	7.5
0.0	710	770		
2.5	1210	1550		

$PD\_true = 0.50, PD\_sample = 0.60$

Def.cor., %	0.0	2.5	5.0	7.5
0.0	450	500	650	1100
2.5	500	550	850	1550
5.0	600	750	1300	3000
7.5	850	1250	2700	9500

$PD\_true = 0.50, PD\_sample = 0.55$

Def.cor., %	0.0	2.5	5.0	7.5
0.0	1600	3000		
2.5	3000	5500		

Min set - minimal sample size; Breach, % - percent of the confidence intervals (one minus the statistical power); time, sec - required time to find the sufficient minimal sample size.

Table 2: Minimal sample size grows when the confidence level rises (decline in α) and immaterially increases when we raise the needed power threshold (decrease in γ).

			α, %		
			10	5	1
No default correlation					
γ, %	30	Min set	150	190	290
		Breach, %	17	20	28
		time, sec	4	5	8
	20	Min set	150	210	310
		Breach, %	18	15	15
		time, sec	4	5	8
	10	Min set	160	220	350
		Breach, %	8	10	9
		time, sec	4	6	10
Default correlation 2.5%					
γ, %	30	Min set	160	220	390
		Breach, %	22	24	24
		time, sec	4	6	12
	20	Min set	170	240	420
		Breach, %	14	16	19
		time, sec	4	7	14
	10	Min set	190	250	490
		Breach, %	5	6	9
		time, sec	5	7	17

Note: Min set - minimal sample size; Breach, % - percent of the confidence intervals (one minus the statistical power); time, sec - required time to find the sufficient minimal sample size.

A.4. R Code

```
#========================================
# simulating CI for proportion (EXTRAPOLATION ) #
#========================================



#/////////////////////////////////////////////////////////////////////////////////////
# ========== PARAMETERS (CAN BE CHANGED) ===========
#/////////////////////////////////////////////////////////////////////////////////////

# ========== LOAN PORTFOLIO PARAMETERS ===========

# sample size search
N_start <-     000
N_step <-     500

#-----------------------------

# number of portfolios
N_portf <- 100
```

```
# provision rates (EL, PD, DR)
PD_sample      <-       0.55   # e.g., industry average # CI_low
PD_true      <-       0.50   # e.g., reported by bank # CI_up

# Default correlations
Rho_sample      <-       0.05   # e.g., industry average
Rho_true      <-       0.05   # e.g., reported by bank

#========================== CI PARAMETERS ===========
# number of CIs
N_ci  <- 100

# ========== GENERAL PARAMETERS ===========
# statistical significance level (for CI)
alpha   <- 0.05

# statistical power (usually 80% is used, hence gamma is 1-0.8)
gamma <- 0.20

# ===== OTHER PARAMETERS (assume no changes to be done) ======
# values inputted by default when creating an object
input  <- 0
trials       <- 1
start_count  <- 1
two     <- 2

#/////////////////////////////////////////////////////////////////////////////////////////
#============ MAIN CODE (NO CHANGES) ==================
#/////////////////////////////////////////////////////////////////////////////////////////
##############################################
start.time <- Sys.time()

#========= generate auxiliary variables =============

two_sided <- alpha / two

sign_level1 <-             two_sided
sign_level2 <- start_count - two_sided

#================================
# initial values

N_loans <- N_start

Prop_breach <- 1

#------------------------
# cycle to search the optimal sample size
#------------------------

# vector with DR realisations per one (single) CI
CI_single_sample    <- rep(input, N_portf )
```

```
CI_single_true       <- rep(input, N_portf )

#==========================================

 while( Prop_breach > gamma    ) {


#//////////////////////////////////////////////////////////////////////////////////////

N_loans <- N_loans + N_step

# vector with lower CI boundaries
CI_low    <- rep(input, N_ci )
CI_up    <- rep(input, N_ci )

# vector with mean DRs
DR_mean_sample    <- rep(input, N_ci )
DR_mean_true     <- rep(input, N_ci )

#------------------------
#------------------------


# cycle to fill in lower CI boundary
for (j in start_count:N_ci) {



#------------------------
# cycle to fill in DR values per CI
for (k in start_count:N_portf) {



#------------------------

# individual factor
Y1      <- rep(input, N_loans)
Y2      <- rep(input, N_loans)

# correlation parameter to simulate dependent outcomes
U1  <- rep(input, N_loans)
U2  <- rep(input, N_loans)

# outcome (default / non-default)
X1 <- rep(input, N_loans)
X2 <- rep(input, N_loans)

#------------------------

#=================================
# generate a single SAMPLE loan portfolio (single DR value)
# assumed DR (EL) value, or industry typical
```

```
# expect it to be higher than the true value reported by bank


#------ traditional Gaussian (bell shaped) distribution (NO correlation) --
# Defaults <- rbinom(N_loans, trials, PD_sample)
# DR         <- mean(Defaults)

#-------------------------------------------
#------ Lunn, Davies (1997) model
#-------------------------------------------

# common (systemic) factor
CF01 <- rbinom(trials, trials, PD_sample)
CF1    <- rep(CF01, N_loans)

# individual (idiosyncratic) factor
 Y1 <- rbinom(N_loans, trials, PD_sample)

# default correlation
 U1 <- rbinom(N_loans, trials, Rho_sample)
#-------------------------------
# final outcome (D / ND status)
  X1   <- U1 * CF1  + (1 - U1) * Y1

#-------------------------------

# Def <- sum(X1)

CI_single_sample[k]   <- sum(X1) / N_loans

#================================
# generate a single TRUE loan portfolio (single DR value)
# reported by bank


#------ traditional Gaussian (bell shaped) distribution (NO correlation) --
# Defaults <- rbinom(N_loans, trials, PD_true)
# DR <- mean(Defaults)

#-------------------------------------------
#------ Lunn, Davies (1997) model
#-------------------------------------------

# common (systemic) factor
CF02 <- rbinom(trials, trials, PD_true)
CF2    <- rep(CF02, N_loans)

# individual (idiosyncratic) factor
 Y2 <- rbinom(N_loans, trials, PD_true)

# default correlation
 U2 <- rbinom(N_loans, trials, Rho_true)
#-------------------------------
```

```
# final outcome (D / ND status)
  X2   <- U2 * CF2  + (1 - U2) * Y2
#--------------------------------

#Def <- sum(X2)

CI_single_true[k]  <- sum(X2) / N_loans


#==================================

}

#------------------------


# lower boundary of SAMPLE CI
CI_low[j]    <- quantile(CI_single_sample, sign_level1 )
DR_mean_sample[j]  <-     mean(CI_single_sample)

# upper boundary of TRUE CI
CI_up[j]     <- quantile(CI_single_true, sign_level2 )
DR_mean_true[j]  <-     mean(CI_single_true)


#####################
}
#------------------------
#------------------------

# hist(CI_low)

#==============================
# count POWER (number of CI lower boundary exceeding the point est.)

breach <- ( CI_low < CI_up )

   N_breach <- sum(breach)
# N_breach

# portion of breaches (want it to be low than gamma parameter)
Prop_breach <- N_breach / N_ci

#/////////////////////////////////////////////////////////////////////////////////////

print(N_loans)



}


#//////////////////////////////////////////////////////////////////////////////////////////
```

```
#============ OUTPUT ===================
#////////////////////////////////////////////////////////////////////////////////////////////

######################
df <- data.frame(var = c(rep('CI_low', N_ci), rep('CI_up', N_ci) ), value = c(CI_low, CI_up)

#plot multiple histograms
ggplot(df, aes(x=value, fill=var)) +  geom_histogram( color='#e9ecef', alpha=0.6, position='i


###############################

end.time   <- Sys.time()
time.taken <- end.time - start.time

######################

N_loans
Prop_breach
time.taken

#=========================================
```

References

AICPA (1999). Audit sampling (1999): Audit and accounting guide. American Institute of Certified Public Accountants. Audit Sampling Task Force. Industry Developments and Alerts. 334. URL: https://egrove.olemiss.edu/cgi/viewcontent.cgi?article=1333&context=aicpa_indev#page9, open access, accessed on Feb. 7, 2024.

Bank of Russia (2017a). Draft ordinance of the Bank of Russia "about the approach to assess the correctness of the provision computation for the portfolio of homogenous loans by using the extrapolation method". [in Russian] https://www.garant.ru/products/ipo/prime/doc/71504728/, open access, accessed on Feb. 7, 2024.

Bank of Russia (2017b). Ordinance dated Jul. 12, 2017 no. 4466-u "about the method of the evaluation by the Bank of Russia of the bank's property adequacy assessment to reconcile its liabilities". [in Russian] https://www.cbr.ru/Queries/UniDbQuery/File/90134/349, open access, accessed on Feb. 7, 2024.

Bush, S. (2015). Sample size determination for logistic regression: A simulation study. Communications in Statistics - Simulation and Computation, 44(2):360–373. https://doi.org/10.1080/03610918.2013.777458, restricted access.

Davydov, A. A. (1990). Subsample representativity. Consultations, 1:115–121. [in Russian] URL: https://www.hse.ru/data/2010/09/08/1221348941/P¥PѳPўCЌPұP«Pў%20PґPҳP»CЃPҳPҷPҳP„CЂPѳCЂPҷPўP„P«CҒCЂCÆ%20PўCЌPçP«CЃPәPҷ.pdf, open access, accessed on Feb. 7, 2024.

Demidenko, E. (2007). Sample size determination for logistic regression revisited. Statistics in Medicine, 26(18):3385–3397. https://doi.org/10.1002/sim.2771, restricted access.

Demidenko, E. (2016). The p-value you can't buy. The American Statistician, 70(1):33–38. https://doi.org/10.1080/00031305.2015.1069760, open access, accessed on Feb. 7, 2024.

Hogg, R. V., Tanis, E. A., and Zimmerman, D. L. (2015). Probability and Statistical Inference. Pearson, 9th edition. https://www.amazon.com/Probability-Statistical-Inference-Robert-Hogg/dp/0321923278, restricted access; https://faculty.ksu.edu.sa/sites/default/files/677_fr37hij.pdf, open access, accessed on Feb. 15, 2024.

Kruppa, J., Lepenies, B., and Jung, K. (2018). A genetic algorithm for simulating correlated binary data from biomedical research. Computers in Biology and Medicine, 92:1–8. https://doi.org/10.1016/j.compbiomed.2017.10.023, restricted access.

Kulinskaya, E., Morgenthaler, S., and Staudte, R. G. (2008). Meta Analysis. A Guide to Calibrating and Combining Statistical Evidence. Wiley. https://www.wiley.com/en-us/Meta+Analysis%3A+A+Guide+to+Calibrating+and+Combining+Statistical+Evidence-p-9780470985526, restricted access.

Lachin, J. M. (2011). Biostatistical methods. The assessment of relative risks. Wiley, 2nd edition. https://www.doi.org/10.1002/9780470317051, restricted access.

Longin, F. and Solnik, B. (2001). Extreme correlation of international equity markets. Journal of Finance, LVI:649–676. https://doi.org/10.1111/0022-1082.00340, restricted access.

Lunn, A. D. and Davies, S. J. (1998). A note on generating correlated binary variables. Biometrika, 85:487–490. https://www.jstor.org/stable/2337376, restricted access.

NAO (2001). A practical guide to sampling. national Audit Office; URL: https://www.nao.org.uk/wp-content/uploads/2001/06/SamplingGuide.pdf, open access, accessed on Feb. 7, 2024.

Nigrini, M. J. (2012). Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection. John Wiley and Sons, New York. https://www.wiley.com/en-us/Benford%27s+Law%3A+Applications+for+Forensic+Accounting%2C+Auditing%2C+and+Fraud+Detection-p-9781118152850, restricted access.

Novikov, S. V. (2001). Stratified sample in the sociological studies. Issues of research methodology and method, 1:1–5. [in Russian] URL: https://www.hse.ru/data/2010/09/08/1221348961/РЉР«РўРҷРәР«Рў%20РғСЂСЃРҽСЂРҷСЋРҷСҘРҷСЃР«РўРҽР„Р„РҽСЅ%20РўСЌРҫР«СЃРәРҽ%20РЎ%20СҒР«СҘРҷР«РҺР«РӯРҷСЉРҳСҒРәРҷСЖ%20РҷСҒСҒРҺРҳРұР«РЎРҽР„РҷСЅСЖ.pdf, open access, accessed on Feb. 7, 2024.

Pozdyshev, V. A. (2017). Banking regulation in the years of 2016-2017: major changes and development perspectives. Russian Journal of Money and Finance, 1:9–17. [in Russian] URL: https://rjmf.econs.online/archive/2017/1/bankovskoe-regulirovanie-v-2016-2017-godakh-osnovnye-izmeneniya-i-perspektivy-razvitiya/, open access, accessed on Feb. 13, 2024.

Raos (2021). Chapter 16. Design and sample size decisions. https://statmodeling.stat.columbia.edu/wp-content/uploads/2021/01/raos_chapter16.pdf, open access, accessed on Jan. 10, 2024.

Russian Ministry of Finance (2021). International standard of audit 530 «audit sample». [in Russian] Introduced for the application in the Russian Federation by the ordinance of the Russian Ministry of Finance dated Jan. 09, 2019 No. 2n; https://minfin.gov.ru/ru/document/?id_4=116595, open access, accessed on Feb. 7, 2024.

S&P Global Ratings (2019). 2018 annual global corporate default and rating transition study. URL: https://www.spratings.com/documents/20184/774196/2018AnnualGlobalCorpo-rateDefaultAndRatingTransitionStudy.pdf, limited access (registration is required), accessed on December 20, 2019.

Stewart, T. R. (2012). Technical notes on the AICPA audit guide audit sampling. https://us.aicpa.org/content/dam/aicpa/publications/accountingauditing/keytopics/downloadabledocuments/sampling_guide_technical_notes.pdf, open access, accessed on Feb. 7, 2024.

UHBristol (2009). How to: Set an audit sample & plan your data collection. University Hospitals Bristol Clinical Audit Team – Version 3. URL: https://uhbristol.nhs.uk/files/nhs-ubht/5%20How%20To%20Sample%20Data%20Collection%20and%20Form%20v3.pdf, open access, accessed on Feb. 7, 2024.

Vasicek, O. A. (1987). Probability of loss on loan portfolio. https://www.moodysanalytics.com/-/media/whitepaper/before-2011/02-12-87-probability-of-loss-on-loan-portfolio.pdf.

Vasicek, O. A. (2002). The distribution of loan portfolio value. https://www.bankofgreece.gr/MediaAttachments/Vasicek.pdf; open access.

Zverev, E. (2018). Isa 530 «audit sample»: approaches to select elements of the general sample. IFRS and ISA in the credit institution, 69(3):37–46. [in Russian] URL: https://www.iia-ru.ru/inner_auditor/publications/articles/analitika-i-rabota-s-dannymi/msa-530-auditorskaya-vyborka-sposoby-otbora-elemen/, open access, accessed on Feb. 7, 2024.

Zverev, E. and Nikiforov, A. (2019). Expert subsample: formation for the large sample. IFRS and ISA in the credit institution, 72(2):68–74. [in Russian] URL: https://www.iia-ru.ru/inner_auditor/publications/articles/analitika-i-rabota-s-dannymi/ekspertnaya-vyborka-formirovanie-dlya-bolshoy-sovo/, open access, accessed on Feb. 7, 2024.