

## Pawel Siarka

Wrocław University of Economics

e-mail: pawel.siarka@ue.wroc.pl

ORCID: 0000-0001-7488-0782

---

# BACKTESTING ANALYSIS. HOW TO ASSESS THE QUALITY OF PD MODELS IN RETAIL BANKING

---

## ANALIZA BACKTESTING. JAK OCENIĆ JAKOŚĆ MODELI PD W BANKOWOŚCI DETALICZNEJ

---

DOI: 10.15611/pn.2019.8.18

JEL Classification: C13, C16, D81, G21, G32, G33

**Summary:** The paper refers to the probability of default model validation procedure in retail banking. The author presents the idea of backtesting analysis focusing on sensitivity analysis of capital requirements under stress scenarios. The paper addresses statistical methods which can be applied in credit risk management under the backtesting exercise in retail banking. The advantages and drawbacks of specific approaches are discussed. Furthermore, the outcomes of the empirical implementation of selected methods are presented. The author considers the impact of positive asset correlation on various validation approaches, where no correlation is assumed, and proves that the zero-correlation assumptions may result in a more prudent approach. This finding was confirmed by the empirical analysis performed for retail portfolios. The research concerned PD parameters calculated for car and mortgage loans. The backtesting results revealed that PD forecasts created for mortgage portfolios underestimated credit risk during the crisis period which started in 2008. However, car loan portfolio credit risk predictions appeared to be robust.

**Keywords:** backtesting, model validation, probability of default.

**Streszczenie:** W niniejszym artykule odniesiono się do zagadnienia weryfikacji jakości modeli służących do szacowania prawdopodobieństwa niewypłacalności w bankowości detalicznej. Autor przedstawił koncepcję analizy *backtesting* w świetle wrażliwości wymogów kapitałowych w odniesieniu do testowania warunków skrajnych. W artykule odniesiono się do zagadnienia weryfikacji jakości prognoz modeli służących do szacowania prawdopodobieństwa niewypłacalności. Przedstawiono i omówiono wyniki wybranych metod. Autor omówił również wpływ dodatniej korelacji aktywów na uzyskane wyniki. Wykazał, że założenie zerowej korelacji może skutkować bardziej konserwatywnymi wynikami. Ustalenie to potwierdzono przez analizę empiryczną przeprowadzoną dla portfeli detalicznych. Badanie dotyczyło parametrów PD szacowanych dla portfeli kredytów samochodowych oraz hipotecznych. Otrzymane wyniki wykazały, że prognozy PD opracowane dla portfela kredytów

hipotecznych niedoszacowują ryzyko kredytowe. Prognozy ryzyka kredytowego dla portfela kredytów samochodowych okazały się trafne.

**Słowa kluczowe:** ryzyko kredytowe, prawdopodobieństwo niewypłacalności, PD.

## 1. Introduction

The Basel Committee on Banking Supervision is known as the primary global standard setter for the prudential regulation of banks. Their recommendations impacts the world banking system substantially. One of the main goals of the BCBS is to improve the financial market stability. The first fundamental regulation was Basel I [Basel Committee on Banking Supervision 1988] which incorporated principles regarding the methodology of capital requirements calculation. The recommendation was to protect the banks against bankruptcy caused by the uncontrolled growth of credit risk. However despite this big leap in applied methodology, the so-called ‘basic method’ did not include a statistical approach for credit risk prediction. For this reason, there was still no compliance between the bank’s economic capital and the actual credit risk. This has changed along with the implementation of Basel II [Basel Committee on Banking Supervision 2006]. Within the framework of Basel II, banks were allowed to estimate capital requirements for credit risk using statistical models. Nowadays, depending on the applied approach (foundation or advanced), the banks rely on advanced techniques dedicated to risk management process.

One of the fundamental elements affecting the credit risk capital requirements is the probability of default (PD). This is calculated within a 12-month horizon [Basel Committee on Banking Supervision 2006] and should reflect the long-term expected value of portfolio default rate. Hence, unlike the IFRS9 (International Financial Reporting Standards) standards where PD is estimated using the Point-In-Time [ISAB 2014] approach, capital requirements are based on the Through-The-Cycle methodology giving a long-term risk estimate. The validation process of PD models refers in particular to the verification of discrepancies between actual and forecasted PD values. This kind of ex-post procedure is called backtesting analysis and according to the Basel Committee should be a vital component of any efficient credit risk system.

In the literature there are many studies referring to the problem of PD model quality assessment [Jackson, Perraudin 2000] as well as the problem of PD model implementation. Kupiec [1995] presented one of the leading approaches based on the binomial test. However, he pointed out that this particular method may not be the most efficient due to stringent assumptions made by its author. Coppens [Coppens, Gonzalez, Winkler 2000] referred to the impact of borrower’s asset correlation on backtesting analysis outcomes. He underlined that dependencies between borrower

defaults may significantly affect the loss distribution function. Hence, conclusions drawn based on binomial statistical tests may not be valid any more. Tasche and Pluto [2005] also addressed this problem. An overview of the statistical methods which can be leveraged for asset correlation estimation was presented by Siarka [2014]. Tasche [2003] presented a general form of PD model validation using the traffic light approach. Similarly, Blochwitz [Blochwitz et al. 2005] focused his attention on the problem of the graphical presentation of forecast quality. Engelman [2006] has also significantly contributed to the area of backtesting analysis. He presented a series of statistical methods which can be used to assess the PD forecasts.

The main purpose of this paper is to verify whether the proposed methods can be leveraged for the quality assessment of PD models in the field of retail banking. The author referred to the Basel Committee regulations which underline the necessity of backtesting procedure implementation. He also showed that underestimation of PD may lead to severe consequences. The other purpose of this paper is to present the empirical outcomes achieved for the implemented methods based on data acquired from a retail bank.

The paper starts with an introduction. In the next part the author presents the one-factor model recommended by the Basel Committee. The model is shown on the basis of a minimum capital requirements calculation. The role of the PD parameter in the process of unexpected losses estimation was also emphasized. A further part of this paper includes a review of the statistical methods leveraged for the PD models' quality assessment. The final part of this paper contains the empirical study, conclusions and some remarks concerning the implementation of backtesting procedures in the banking industry.

## 2. Probability of default and capital requirements

The Basel Committee in its consultative documents underline the key role of the probability of default. The recent economic downturn which started in 2008 revealed that the underestimation of PD may lead to spectacular bankruptcy under severe macroeconomic conditions. The crucial role of the PD parameter can be clearly explained in terms of the one-factor model proposed by Merton [1974] and then recommended by the Basel Committee in 2006. According to this approach, the value of borrower assets triggers an event of default. The lower the value of the company's assets, the lesser the incentive for the shareholders to continue the business. Under the one-factor model the value of the borrower assets depends on two factors:

$$X_i = Y\sqrt{\rho} + Z_i\sqrt{1-\rho}, \quad (1)$$

where  $Y$  is a market factor common for all borrowers, therefore affects each individual assets value ( $X_i$ ). Unlike the market factor,  $Z_i$  is a variable relevant only to a given

$i$ -th borrower, hence it represents idiosyncratic risk. The coefficient  $\rho$  reflects vulnerability of borrower assets to the changes of market factor ( $\sqrt{\rho}$ ) and idiosyncratic factor ( $\sqrt{1-\rho}$ ) respectively.

Assuming that the default is triggered always when the value of borrower assets drops below a certain threshold limit, a conditional probability of default (conditional on market factor) can be derived. However, the asset threshold limit needs to be estimated first. This value is derived based on an inverse normal cumulative distribution function applied for the portfolio's expected probability of default. There are many techniques for the estimation of the PD's expected value, while the most popular in the banking industry are based on historical default rates. However Pluto and Tasche [2005] show that a simple average may often underestimate its value.

It is convenient for further considerations to define auxiliary variable  $Z_L$  which takes one in the case of the default of  $i$ -th borrower and zero in other cases. Then, conditional probability (conditional on the market factor) of default can be presented in the following form [Vasicek 1991]:

$$P(L_i = 1|Y) = N\left(\frac{N^{-1}(PD) + Y\sqrt{\rho}}{\sqrt{1-\rho}}\right), \quad (2)$$

where  $PD$  is the expected value of the unconditional probability of default.  $N^{-1}(PD)$  is an inverse function of normal cumulative distribution.

For the portfolio which consists of  $n$  borrowers, one can define variable  $L$  as the average rate of defaulted loans:  $L = \frac{1}{n} \sum_{i=1}^n L_i$ . Thus, assuming that default probabilities of all loans are equal to  $PD$ , we can estimate the probability of any loss (portfolio loss) i.e. the probability that  $L$  will not exceed  $x$ :

$$P(L \leq x) = N\left(\frac{\sqrt{1-\rho}N^{-1}(x) - N^{-1}(PD)}{\sqrt{\rho}}\right). \quad (3)$$

Under the New Basel Capital Accord, the above-mentioned formula is recommended for the calculation of minimum capital requirements. The bank's capital  $K$  should cover at least the unexpected losses estimated according to the following formula [Basel Committee on Banking Supervision 2006]:

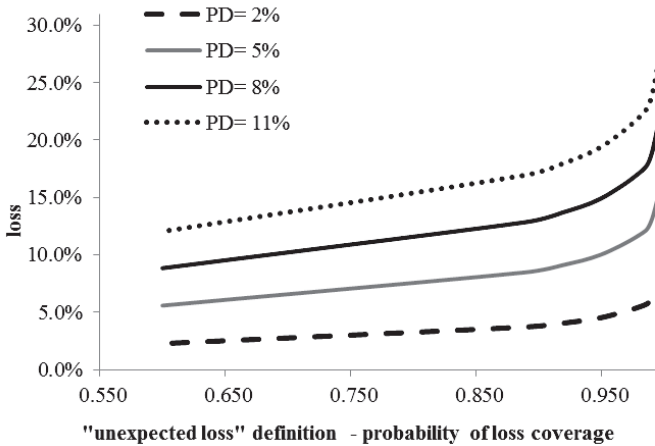
$$K = \left[ N\left(\frac{N^{-1}(PD) + N^{-1}(0.999)\sqrt{\rho}}{\sqrt{1-\rho}}\right) - PD \right] LGD, \quad (4)$$

where  $LGD$  (Loss Given Default) is an expected value of the unrecovered part of the defaulted portion.

The formula above (4) has a relevant interpretation in terms of possible economic downturns. It implies that banks should be able to absorb unexpected losses with its capital, which may occur once every one thousand years. Such an approach seems

to be quite stringent and conservative to the banking industry, however further research [Siarka 2015] revealed that stress test models disclosed some challenges which are not addressed by the Basel II formula.

Figure 1 presents the relation between unexpected portfolio loss and the probability representing the severity of market conditions; for example value 0.950 indicates losses which may appear five times every 100 years (1-0.950). Such an exercise was conducted for various levels of portfolio PD ranging from 2% to 12%.



**Fig. 1.** Presentation of portfolio losses for various PD values

Source: own study.

The results presented in Figure 1 clearly show that unexpected losses rise along with more severe market conditions, but what is more important, the growth rate changes significantly for higher probabilities representing economic deterioration. Hence a relatively small change in ‘conservativity’ level (e.g. from 0.99 to 0.999) has a huge impact on portfolio losses and capital requirements. It also can be noticed that the growth of PD value contributes to the more rapid increase of unexpected losses. For this reason the Basel Committee strongly recommends the regular monitoring of PD estimates.

### 3. Statistical methods under backtesting procedure

PD model validation is a fundamental procedure performed by banks to examine the adequacy of credit risk forecasts. A statistical binomial test is a useful tool leveraged for the quality assessment of PD estimates. This approach assumes that all defaults within a loan portfolio appear independently. According to this assumption, the number of bankruptcies in any portfolio consisting of  $N$  borrowers with a given PD,

may be modelled with a binomial random variable. Within the framework of a binomial approach, it is assumed that the observed default rates come from the defined distribution function. In fact a statistical hypothesis is tested, where the alternative hypothesis says that the observed actual values cannot be deemed as generated by a random variable with the assumed distribution function. Hence the model cannot be deemed as reliable.

In order to accept or reject the null hypothesis, the critical value shall be calculated using the following formula:

$$d_\alpha = \left\{ d : \sum_{j=d}^N \binom{N}{j} PD^j (1-PD)^{(N-j)} \leq 1-\alpha \right\}, \quad (5)$$

where  $\alpha$  is the statistical test significance level. Under this approach always when  $d$  (the number of empirically observed defaulted borrowers) is lower than critical value  $d_\alpha$ , there is no reason to reject the null hypothesis, so the forecasting model should be considered as reliable.

The presented approach can be simplified by applying the CLT (Central Limit Theorem). With respect to the CLT the critical value is calculated as follows:

$$d_\alpha = N \cdot PD + \Phi^{-1}(\alpha) \cdot \sqrt{N \cdot PD(1-PD)}, \quad (6)$$

where  $\Phi^{-1}$  is an inverse function of the standard normal cumulative distribution function.

As was emphasized above, the underlying assumption of the presented statistical test was the lack of positive correlation among the events of defaults in a portfolio (i.e. correlation equal to zero). However, many researchers deem that the value of borrower assets correlation which directly impacts the default event is non-zero (i.e. positive value). Some evidence supporting this thesis was presented by Rösch and Scheule [2004], Düllmann, Scheicher and Schmieder [2007], Lopez [2004], Siarka [2012] and Crook [2009]. Furthermore, the Basel Committee also agrees with this concept by implementing the asset correlation formula where its value depends on PD. For retail exposures the formula is as follows:

$$\rho = 0.03 \left( \frac{1 - e^{-ssPD}}{1 - e^{-ss}} \right) + 0.16 \left[ 1 - \left( \frac{1 - e^{-ssPD}}{1 - e^{-ss}} \right) \right]. \quad (7)$$

The positive assets correlation has a significant impact on the distribution function of portfolio losses. For this reason, Coppens [Coppens, Gonzalez, Winkler 2007] considered the implementation of asset correlation into the process of assessing the quality of PD forecasts. In his study he generalized the binomial test, which led him to a new formula. Under his approach the critical value for the statistical test was calculated as follows:

$$d_\alpha = q + \frac{2q-1}{2N} - \frac{q(1-q)}{\Phi\left(\frac{\sqrt{\rho}\Phi^{-1}(1-\alpha)+\Phi^{-1}(PD)}{\sqrt{1-\rho}}\right)} \cdot \frac{(1-2\rho)\Phi^{-1}(1-\alpha)-\sqrt{\rho}\Phi^{-1}(PD)}{2N\sqrt{\rho(1-\rho)}}, \quad (8)$$

where

$$q = \Phi^{-1}\left(\frac{\sqrt{\rho}\Phi^{-1}(\alpha)+\Phi^{-1}(PD)}{\sqrt{1-\rho}}\right). \quad (9)$$

Another attempt at incorporating asset correlation into the backtesting procedure was presented by Tasche and Pluto [2005]. They used a one-factor model to calculate the probability of  $k$  defaults. According to their approach the probability of observing less or equal to  $k$  defaults can be calculated according to the following formula:

$$P(X \leq k) = \int_{-\infty}^{\infty} \varphi(y) \sum_{i=0}^k \binom{N}{i} G(PD, \rho, y)^i (1-G(PD, \rho, y))^{N-i} dy, \quad (10)$$

where

$$G(PD, \rho, y) = \Phi\left(\frac{\Phi^{-1}(PD)-\sqrt{\rho}y}{\sqrt{1-\rho}}\right). \quad (11)$$

The methods shown above can be applied to any loan portfolio where individual PDs are equal for all the borrowers. However, banking professionals often struggle with the problem of quality forecast assessment where few portfolios are analyzed at the same time or just one portfolio, but with few forecasts created in time. Moreover, internal rating systems provide borrower segmentation resulting in various PDs assigned to selected segments. Then, the Hosmer-Lemeshow test can be leveraged to handle the multiple forecast problem. Under this approach it is assumed that the portfolio consisting of  $N$  loans is divided into separate  $K$  groups with given credit risk  $PD_k$ , hence the number of borrowers in group  $k$  is denoted as  $N_k$ . Assuming that the default events are independent, one can use the following statistic for forecasts' quality assessment:

$$S_K^{\chi^2} = \sum_{k=1}^K \frac{(N_k \cdot PD_k - d_k)^2}{N_k \cdot PD_k \cdot (1 - PD_k)}. \quad (12)$$

When the number of elements in each group increases, the distribution of the above-mentioned statistic converges to  $\chi^2$  distribution function with  $K-2$  degrees of freedom. The backtesting analysis according to the Hosmer-Lemeshow statistical test comes down to a comparison of the value of the test statistic with the critical value derived from  $\chi^2$  cumulative distribution function for given significance level  $\alpha$ . When the value of a test statistic is below the value derived from statistical tables, there is no reason to deem that the credit risk system provides inaccurate PD forecasts.

Another method which handles multiple forecast problem is based on the Brier Score statistics. According to this approach, the average difference between forecasted PDs and the observed actual rates of default are analyzed. The Brier Score value is calculated according to the following formula:

$$B = \frac{1}{k} \sum_{k=1}^K (p_k - PD_k)^2, \quad (13)$$

where  $k$  indicates a segment.

A Brier Score is a measure of consistency between the forecasts and empirically observed values. The credit risk rating system is deemed to be reliable when the  $B$  value is small. The values close to zero indicate that the forecasts are robust.

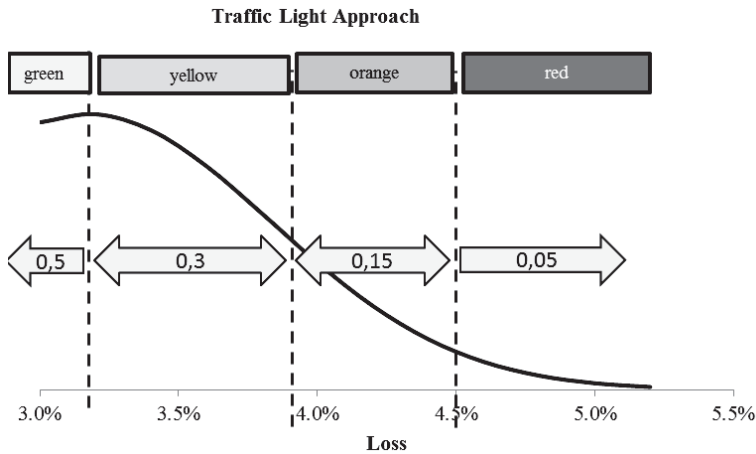
Another powerful and easy to interpret approach is called the TLA (Traffic Light Approach). This method was proposed by Tasche [2003] who applied the approach for the credit risk forecast assessment process. Previously this method was known as a popular tool for market and operational risk management. Tasche identified distinct PD ranges separated by the 'critical level' values. Firstly he specified the shape of the losses distribution function and then he derived the critical PD values for the arbitrarily set probabilities  $\alpha_{low}$  and  $\alpha_{high}$ . Within this approach the actual default rate should not exceed  $d_{low}$  level with the probability higher than  $\alpha_{low}$  and  $d_{high}$  with the probability higher than  $\alpha_{high}$ . Tasche proposed the limit values for PD based on a quintile 0.95 and 0.999, respectively.

For the TLA approach where the binomial distribution function of the number of defaults was assumed, the critical value can be derived using the following formula:

$$p_\alpha = PD + \Phi^{-1}(\alpha) \sqrt{\frac{PD(1-PD)}{N}}. \quad (14)$$

The above-mentioned approach was improved by Blochwitz [Blochwitz et al. 2005] by the implementation of four ranges where each range was marked with a distinct colour. Every successive colour indicates a higher deviation of the empirically observed default rate from the expected value resulting from PD forecast. Therefore, the green area corresponds to a high quality forecast. Yellow indicates a moderate deterioration, but still acceptable forecast quality, While orange area signals some quality issues which need to be examined. Red suggests that the statistical model is not reliable as the forecasts generated by this model are of poor quality. Blochwitz proposed the above-mentioned PD ranges based on arbitrarily set probabilities of a wrong decision (considering the model as unreliable while it is still correct). The following probabilities were suggested: 0.5, 0.3, 0.15, and 0.05. Figure 2 presents the idea of four colours leveraged under the TLA approach.





**Fig. 2.** Graphical presentation of the decision areas in accordance with the Traffic Light Approach

Source: own study.

Figure 2 presents the graphical implementation of the Traffic Light Approach under the backtesting procedure. The ranges corresponding to an acceptable quality of PD forecast are green and yellow, and the conditional acceptance range is marked as orange. Always when the empirically observed default rate hit this area, it is necessary to examine the reasons behind the relatively large inaccuracy. Once the empirical value gets closer to the red area, immediate action should be taken by credit risk managers, as it is the alarm zone. It should be deemed that PD forecasts were too low in relation to the actual credit risk. For this reason, internal proceedings should be initiated to explain the observed discrepancy. Such a proceeding needs to be a part of the emergency plan within a comprehensive risk management system.

The model validation procedure outcomes depend on underlying assumptions such as non-zero borrower asset correlation. This is why it is worth to compare the results obtained under the zero correlation assumption and positive asset correlation. This issue was already discussed by Coppens [Coppens, Gonzalez, Winkler 2007] who examined statistical test critical values derived for various levels of correlation. The results were then used for the verification of the null hypothesis, i.e. the model provides high quality PD forecasts. His approach was the same as the one presented in this paper i.e. the empirical values of credit risk below the critical value let him accept the null hypothesis and consider the model as reliable. Coppens noted that the critical values for non-zero asset correlation were always higher than those derived for zero-correlated assets. Hence it turned out that the assumption regarding independent defaults imposes more stringent quality requirements on PD models. This results from the fact that the positive asset correlation means a higher dispersion of the losses distribution function and higher critical value, therefore the higher the asset correlation is assumed, the easier the model can pass the quality tests. Thus the

zero-correlation assumption may be considered as a conservative approach within the process of PD models' validation.

#### 4. The empirical results of the backtesting analysis

In order to illustrate the above presented methods, a backtesting analysis was performed for a Polish branch of a leading global bank specialising in consumer finance. In Poland the bank was active in the car and mortgage loan fields, therefore the credit risk forecast quality assessment process was conducted for these two products. The mortgage portfolio incorporated 64,237 loans. For each month in 2008, PD forecasts were made by the internal credit risk team. The second examined portfolio incorporated 20,000 consumer loans to finance car purchases. The validation period covered the year 2002, where twelve monthly forecasts were provided.

Table 1 presents the PD forecasts and the actual values for the mortgage portfolio. It can be noted that except for December 2008, the PD forecasts were below the actual values. Furthermore, the discrepancy increased during the second half of the year.

**Table 1.** Comparison of PD forecasts and actual values – mortgage loan portfolio

	Date											
	2008-01	2008-02	2008-03	2008-04	2008-05	2008-06	2008-07	2008-08	2008-09	2008-10	2008-11	2008-12
PD forecast	0.76%	0.78%	0.82%	0.86%	0.89%	0.92%	0.97%	1.03%	1.09%	1.20%	1.29%	1.37%
PD	0.75%	0.79%	0.85%	0.93%	1.04%	1.17%	1.14%	1.21%	1.20%	1.26%	1.38%	1.30%

Source: own study.

Table 2 shows a comparison of the PD forecasts and the actual values for the car loan portfolio. This time the actual values were higher than its expectations only during a few months of the second half of 2002.

**Table 2.** Comparison of PD forecasts and actual values – car loan portfolio

	Date											
	2002-01	2002-02	2002-03	2002-04	2002-05	2002-06	2002-07	2002-08	2002-09	2002-10	2002-11	2002-12
PD forecast	1.80%	1.85%	1.90%	1.92%	1.93%	1.96%	1.99%	2.04%	2.06%	2.10%	2.22%	2.41%
PD	1.66%	1.70%	1.62%	1.57%	1.76%	1.98%	1.95%	2.14%	2.20%	2.21%	2.20%	2.28%

Source: own study.

Based on the data contained in Table 1, the backtesting analysis was performed using a binomial statistical test. For the purposes of the analysis, the test significance level was set at 5%. Figure 3 presents the backtesting analysis outcomes derived for June 2008. The left part contains all the input data and assumptions necessary to calculate the test statistic. The graph presented on the right side of the table shows the theoretical distribution function of the number of defaulted loans (under the assumption that the PD forecast was correct).

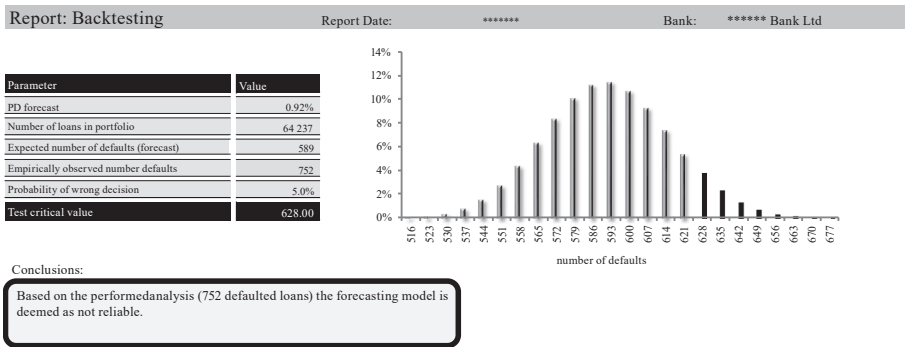


Fig. 3. The results of the backtesting analysis for PD forecast – mortgage portfolio – June 2008

Source: own study.

The backtesting analysis applied for June 2008 revealed the PD forecast (0.92%) was significantly below the actual value. According to the forecast, it was expected that 589 loans would default within the given horizon. In fact 752 default were observed. The statistical test critical value was estimated at 630 loans. As the empirically observed number of defaulted loans appeared to be higher than the critical value, the forecast needs to be deemed as not reliable.

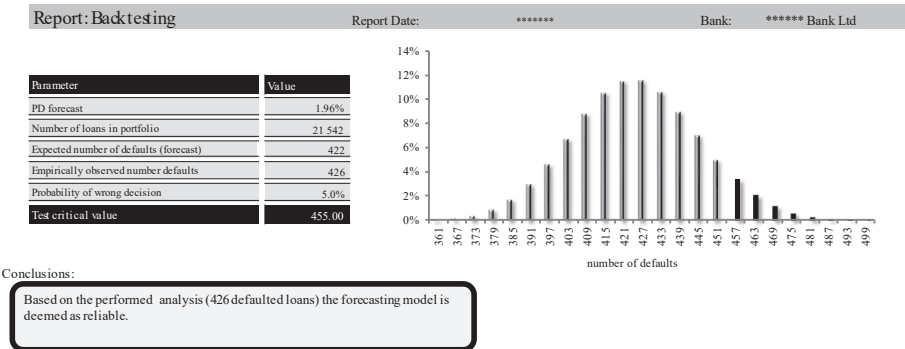


Fig. 4. The results of backtesting analysis for a PD forecast of car portfolio – June 2002

Source: own study.

The analogous analysis was performed for the car loan portfolio. The validated forecast concerned June 2002 and the results are presented in Figure 4. According to the PD forecast the credit risk should be observed at 1.96%, which gives 422 defaults. The empirically observed number of defaulted loans turned out to be four loans higher. However this discrepancy appeared to be not statistically significant, as the acceptable limit of the defaulted loans (critical value) was equal to 455, hence much above the observed defaults. Therefore the forecast should be considered as reliable at the confidence level  $\alpha = 5\%$ .

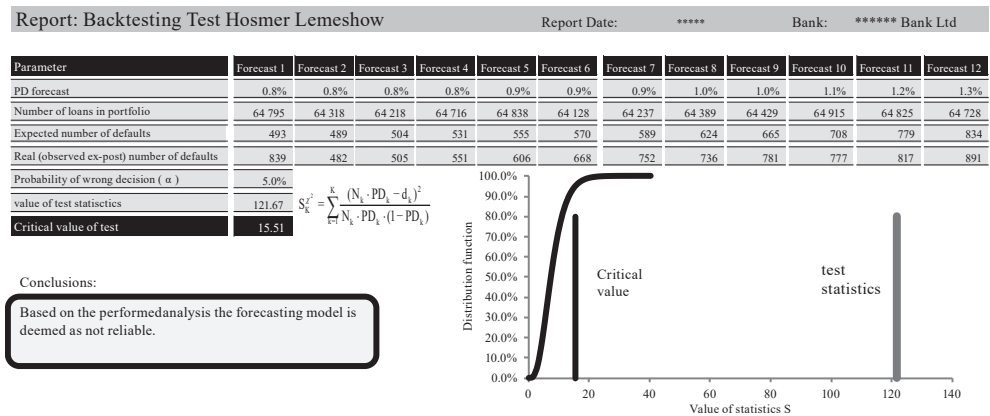
Table 3 presents all the backtesting results performed for both the portfolios and the given months. The validation outcomes were coded as “ok” when the forecast appeared to be reliable according to the binomial test. When the forecasts were too low comparing to the actual values, the code ‘bad’ was assigned. The analysis revealed that in 2008 the mortgage forecasts could be considered as reliable only for five months (January, February, March, October, and December). For comparison, all the car loan portfolio forecasts were assessed as accurate.

**Table 3.** The results of the backtesting analysis on the basis of the binomial test for a mortgage and car portfolio

	Date											
	2008-01	2008-02	2008-03	2008-04	2008-05	2008-06	2008-07	2008-08	2008-09	2008-10	2008-11	2008-12
Mortgage loans portfolio	ok.	ok.	ok.	bad	bad	bad	bad	bad	bad	ok.	bad	ok.
	2002-01	2002-02	2002-03	2002-04	2002-05	2002-06	2002-07	2002-08	2002-09	2002-10	2002-11	2002-12
Car loans portfolio	ok.	ok.	ok.	ok.	ok.	ok.	ok.	ok.	ok.	ok.	ok.	ok.

Source: own study.

In the course of further analysis, the credit portfolios were subjected to a complementary backtesting procedure using the Hosmer-Lemeshow test. Under this approach it was assumed that the default rates in every month were independent. The test statistic was calculated only for these months where the forecasts appeared to be below the actual values. This approach was applied to provide a higher degree of conservatism, as it focuses only on the risk underestimation events. The rationale behind this method results in the assumption that any overestimation cases are not that harmful for the bank.



**Fig. 5** The results of the backtesting analysis – mortgage loan portfolio, the Hosmer-Lemeshow test

Source: own study.

Figure 5 presents the results of the backtesting analysis performed for the mortgage portfolio. The table in the figure contains the PD forecasts and the actual number of defaults used for the test statistic calculation. The graph illustrates the cumulative distribution function of the test statistic under the assumption that the forecasts were accurate. The value of the test statistic was equal to 121.67, while the test critical value for significance level of 5% was equal to 15.51. Therefore, based on the Hosmer-Lemeshow test results, it can be deemed that at least one of the examined forecasts was inaccurate. Hence, this method also confirmed that the credit risk projections of the mortgage portfolio need to be assessed as poor quality.

Figure 6 illustrates the results of the backtesting analysis performed for the car loan portfolio. The value of the test statistic was estimated at 4.59 while the critical value was equal to 5.99. Hence, it can be concluded that the forecasts are of high quality which confirms the binomial test's earlier findings

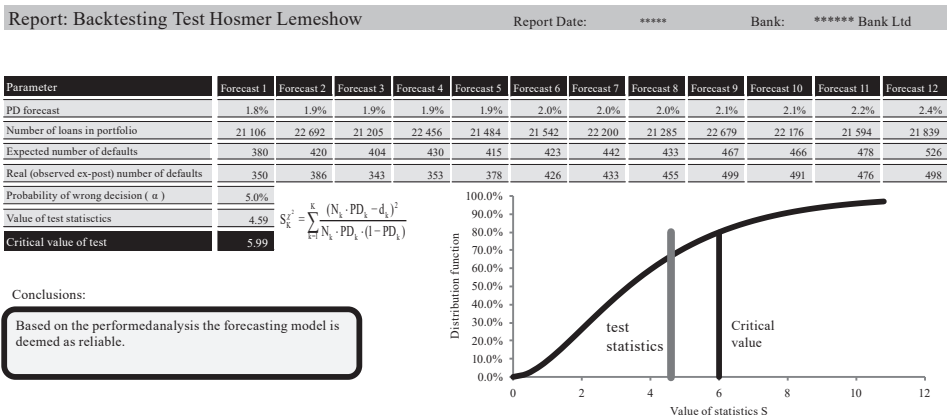


Fig. 6. The results of the backtesting analysis – car loan portfolio, the Hosmer-Lemeshow test

Source: own study.

The backtesting analysis outcomes show that the results obtained with various statistical methods are consistent. The PD forecasts made for the car loan portfolio were deemed as reliable, while for the mortgage loan portfolio they appeared to be much below expectations. In 2008, some forecasts were too low compared to the actual values, and the discrepancy could not be explained by normal course PD volatility. Facing such findings, the managers responsible for the risk management process should thoroughly review the statistical model used for forecasting. The review should cover the model assumptions' verification, the coefficient estimation process, the optimal model selection and its implementation. The reason for the poor model accuracy could be an earlier change in the bank's credit risk policy or changes in the macroeconomic variables which were not considered by the model's developers. In 2008 a global financial crisis took place and many financial institutions reported an uncontrolled spike in credit risk trends.

## 5. Conclusions

The PD estimation process is one of the fundamental elements of any credit risk management system in contemporary banking. Highly accurate forecasting allows the banks to prepare for expected and unexpected losses. Any mistakes made at this stage may lead to severe consequences such as the loss of control over risk and in extreme cases – bankruptcy. Hence for this reason the backtesting procedure should be performed on a regular basis.

In this paper it was proved that capital requirements calculated according to the IRB approach recommended by the Basel Committee are highly sensitive to changes of the probability of default. A relatively small underestimation of PD or asset correlation may result in a significant decrease of minimal capital requirements. Consequently, the bank may keep insufficient capital, which under severe macroeconomic conditions may lead to insolvency.

One of the underlying assumptions of credit risk models refers to the borrower asset correlation. The higher the asset correlation the larger the unexpected loss, and therefore the capital requirements. However under the backtesting exercise, the overestimation of asset correlation gives more space for incorrect model assessment (e.g. the TLA approach). It was shown that when the correlation coefficient is set to zero, the PD volatility is lower and therefore it results in a more conservative approach. Finally, the PD confidence interval is narrower and also the areas under the TLA approach are also narrower.

The example presented in this paper was used to illustrate the application of the statistical methods under the backtesting exercise. The outcomes obtained for the car loan portfolio revealed that forecasts can be deemed as accurate. The mortgage portfolio forecasts turned out to be inaccurate showing some serious underestimations. It also should be noted that the mortgage analysis concerns the year 2008, which was the beginning of the global economic downturn. Therefore, backtesting analysis may be considered as an early warning system giving signals for statistical model recalibration.

## Bibliography

- Basel Committee on Banking Supervision, 1988, *International Convergence of Capital Measurement and Capital Standards*, Basel.
- Basel Committee on Banking Supervision (BCBS), 2006, *International Convergence on Capital Measurement and Capital Standards*, Bank for International Settlements, Basel.
- Blochwitz S., Hohl S., Tasche D., Wehn C.S., 2004, *Validating Default Probabilities on Short Time Series*, Deutsche Bundesbank, Working Paper.
- Blochwitz S., Hohl S., Tasche D., Wehn C.S., 2005, *Reconsidering Ratings*, Deutsche Bundesbank, Working Paper, May.

- Braun A., Rymaszewski P., Schmieser H., 2010, *A Traffic Light Approach to Solvency Measurement of Swiss Occupational Pension Funds*, Working Papers on Risk Management and Insurance, no. 74, University of St. Gallen.
- Coppens F., Gonzalez F., Winkler G., 2007, *The Performance of Credit Ratings Systems in the Assessment of Collateral Used in Eurosystem Monetary Policy Operations*, Occasional Paper Series, European Central Bank.
- Crook J., Bellotti T., 2009, *Asset Correlations for Credit Card Defaults*, Working Paper, Credit Research Centre, University of Edinburgh.
- Düllmann K., Scheicher M., Schmieder C., 2007, *Asset Correlations and Credit Portfolio Risk – An Empirical Analysis*, Deutsche Bundesbank Discussion Paper, series 2, no. 13.
- Emmer S., Tasche D., 2005, *Calculating Credit Risk Capital Charges with the One-Factor Model*, <http://arxiv.org/abs/cond-mat/0302402v5>.
- Engelman B., 2006, *The Basel II Risk Parameters: Estimation, Validation, and Stress Testing*, Springer.
- Ferguson T.S., 1967, *Mathematical Statistics. A Decision Theoretic Approach*, Academic Press, London.
- Hamerle A., Liebig T., Scheule H., 2004, *Forecasting Credit Portfolio Risk*, Deutsche Bundesbank, Working Paper, May.
- IASB, 2014, IFRS Standard 9: Financial Instruments.
- Jackson P., Perraudin W., 2000, *Regulatory implications of credit risk modelling*, Journal of Banking & Finance, 24.
- Jajuga K., 1993, *Multivariate Statistical Analysis*, PWN, Warszawa.
- Kupiec N.H., 1995, *Techniques for verifying the accuracy of risk measurement models*, Journal of Derivatives, 3, pp. 73-84.
- Lopez J.A., 2004, *The empirical relationship between average asset correlation, firm probability of default, and asset size*, Journal of Financial Intermediation, 13, pp. 265-283.
- Merton R.C., 1974, *On the pricing of corporate debt: The risk structure of interest rates*, Journal of Finance, 29, pp. 449-470.
- Pluto K., Tasche D., 2005, *Thinking positively*, Risk, 18(8), pp. 72-78.
- Rösch D., Scheule H., 2004, *Forecasting retail portfolio credit risk*, Journal of Risk Finance, Winter/Spring, pp. 16-32.
- Schuermann T., Hanson S., 2004, *Estimating Probability of Default*, Federal Reserve Bank of New York Staff Reports.
- Siarka P., 2012, *Assets Correlation. The Issue of Estimating the Losses on the Example of Mortgage Portfolio*, Bank & Credit, National Bank of Poland, Warsaw.
- Siarka P., 2014, *Asset correlation of retail loans in the context of the New Basel Capital Accord*, The Journal of Credit Risk, 10 (3).
- Siarka P., Chan L., 2015, *Comprehensive capital analysis and review stress tests: Is regression the only tool for loss projection?*, Journal of Risk Model Validation, 9 (3), pp. 71-99.
- Svec M., 2009, *PD Backtesting Empirical Study on Credit Retail Portfolio*, CSOB Bank.
- Tasche D., 2003, *A Traffic Lights Approach to PD Validation*, Deutsche Bundesbank, Working Paper.
- Tasche D., Pluto K., 2005, *Estimating Probabilities of Default for Low Default Portfolios*, Working Paper, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=635301](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=635301).
- Vasicek O., 1991, *Working Paper*, KMV Corporation.
- Vasicek O., 2002, *The Distribution of Loan Portfolio Value*, Risk.