THE ISSUE OF PD ESTIMATION
– A PRACTICAL APPROACH

Paweł Siarka

Abstract. The issue of estimating the probability of default constitutes one of the foundations of risk systems applied in modern banking. The Basel Committee pays a lot of attention to ways of its estimation and validation. This paper discusses statistical methods enabling PD estimations with consideration of the retail character of a credit portfolio. The author refers to the issue of defining default and to the way of calculating the number of days in arrears. This paper presents the results of research studies obtained on the basis of retail credit portfolio. For selected sub-portfolios, the author makes a comparison of the probability of default, which enables the explicit risk assessment.

Keywords: credit risk, probability of default, credit scoring.

JEL: C51, C53, E58, G2, G33.

1. Introduction

Credit activity is inseparably connected with credit risk. Credit risk always comes into being where losses caused by cessation of credit servicing by a client are covered from own capital of a company taking the responsibility for risk. In order to assess credit risk, practitioners use a series of statistical methods allowing for diagnosing the current level of risk and estimating its level in the upcoming future. The multiplicity of statistical models results from various assumptions accepted in the course of their construction. It results also from various aims set for analysts responsible for the area of risk. The process of credit risk management is such a complicated issue that even its quantification offers many difficulties. If somebody wants to get to know the real credit risk of a financial institution, a series of parameters should be defined. The most important of them include the loss forecast horizon, value of the probability of default, value of a recovery rate because of credit recovery proceedings, correlation of borrower assets, price of overdue receivables obtained on the market of purchase and sale of re-

_____

Paweł Siarka
  Paweł Siarka, Wrocław School of Banking, Fabryczna Street 29/31, 53-110 Wrocław, Poland.
  E-mail: pawel.siarka@i4fs.pl

ceivables. Only cognition of all elements being included in the full model of credit risk assessment allows for a comprehensive view of the studied issue. It is also impossible to get to know the risk scale without understanding the principles of calculating particular elements of the applied statistical model.

One of the essential elements of the majority of models concerning credit risk assessment is the determination of the borrower's value of the probability of default (PD). This standardised measure of risk from the <0, 1> range allows for a synthetic comparison of risk levels – especially in the area of portfolios referring to the same credit product. Within the framework of Basel II (Basel Committee..., 2006), the Basel Committee introduced a global standard referring to the loss forecast horizon, including the PD probability. Currently, banks are obliged to create forecasts in the time horizon of one year. Such an approach enables the risk analysis in the background of a full reporting year of a financial institution. The estimation of risk in the time horizon of one year has the advantage of allowing for testing the sensitivity of the annual financial result, which takes into account provisions covering the risk.

Borrower probability of default is one of the fundamental measures of credit risk assessment. Its aim is to find an answer to the question about the forecast of proportion of loans (expressed in percent), which are not going to be serviced any more in the time horizon of one year.

Constant monitoring of the PD enables the current analysis of factors that stimulate the risk level. Therefore, it supports the process of quality assessment of the conducted risk policy. Based on PD changes, it is possible to analyse the effects of the changes in internal risk management policies. In this way, early warning systems are created which allow, much in advance, drawing bank analysts' attention to the negative results of decisions made before.

The probability of default (PD) constitutes the basic component of a model applied in order to calculate the capital requirements in accordance with Basel II. Minimum capital requirement is constructed in such a way that it informs on the level of unexpected losses – losses for which the probability of occurrence is low but still possible. The possibility of getting to know potential losses developed as a result of credit activity and maintenance of capital on this level aims at the minimisation of the danger of the bank's bankruptcy.

The issue of estimating the probability of default is the subject of many studies. Usually they concern companies assessed by rating agencies. Among others, Blochwitz and Hohl (2001), as well as Jafry and Schuermann (2004), have been dealing with relations between the obtained rating

and the probability of default. Cantor and Falkenstein (2001), as well as Cantor and Mann (2003) in later studies, presented the analysis of the influence of macroeconomic factors on PD. Heitfield (2005) described in his paper the issue of sensitivity of rating assessments to changes of economic environment. In reference to retail loans, Allen et al. (2003) dealt with the issue of estimating PD. In addition, Sabato (2006) in his study drew attention to the issue of estimating PD in the case of retail portfolio with low risk (low-default probability).

In the studies of Duffie and Singelton (2003), one may find a description of methods applied in the process of estimating PD probability. Hastie (2001), in turn, presented extensively the classification methods. Gruszczynski (2001) described a series of useful approaches to the issue of estimating the probability of default. In addition, Saunders (2002) presented in his book the wide range of methods used by practitioners in the process of estimating PD.

The aim of this paper is the presentation of methods enabling the estimation of the probability of default in the context of portfolios including retail credits. Thus, a great part of this paper concerns PD models in which individual borrower characteristics play a key role. The author's aim is also to present the results of analysis conducted within the framework of a retail credit portfolio. Selected methods were used as comparison of risk levels in two separate groups. It allowed for the verification of the hypothesis about the presence of higher risk in group A than in group B.

This paper consists of an introduction, after which one presents the definition of default, constituting the basis of setting PD. In the further part, the author reports a series of methods aimed at the estimation of the probability of default. Then, one can find an example illustrating the practical application of a method based on mortality tables in the process of assessing risk of retail loan portfolios. Next, the author includes conclusions from the conducted research studies.

2. Borrower default definition

In a further part of this paper, the event of default is understood in accordance with the definition accepted by the Basel Committee. We assume that default takes place in relation to a given borrower if there has occurred one of two following events:

1. Any exposure having the character of credit is more than 90 days overdue, while exposure is claimed as overdue if the overdue amount exceeds the threshold amount.

2. Due to internal analysis, one claims that probably the borrower does not completely fulfil their credit obligation.

The threshold amount, in excess of which one recognizes the given credit as default, is accepted at the level of 50 EUR for use of further considerations.

The second condition allows for the classification of given receivables as overdue although delay in payment may come to less than 90 days. This condition in the case of retail portfolio aims at recognition as default of all these loans for which one has not noted repayment of the first two instalments. The experiences of many banks indicate that the vast majority of these borrowers have taken loans with the aim of not repaying them. Thus, quicker classification of these credits as default enables the earlier beginning of credit recovery proceedings.

Therefore, a key element of the assessment of default is the number of days in arrears of repayment. In the 1990s, many banks accepted the methodology describing the moment of occurrence of arrears as a moment in which one observed arrears on the given loan, and after that time, one did not make full repayment of arrears. Such an approach causes the number of days in arrears to be determined from the first date of arrears in the case when the borrower pays their instalments with a significant delay (that is, e.g. two months of delay), Thus, a borrower who had a problem with repaying only the first and second instalment, and then repays the full instalment every month, according to this method after 12 months was delayed in repayment for one year. It is worth paying attention to the fact that the debtor had arrears to the bank equal to two instalments. A different approach to the process of establishing the date of occurrence of arrears is the method consisting in determining the date of arrears of the oldest unpaid instalment. In other words, loan instalments are treated as "separate" receivables, and the oldest overdue credit instalment indicates the moment from which we calculate arrears. It is assumed in this method that every payment is booked to cover the oldest arrears. In practice, one takes into consideration not only the principal maturity date, but also the balance of overdue interests, overdue penalty interests and calculated costs. The first approach was applied for many years and it was a result of IT systems limitations and the lack of computerisation. The second approach, much more reasonable, is currently dominating in modern banking. Actually, this principle of determining the number of days in arrears is applied in the further part of this paper.

3. Methods of estimating the probability of default

Among the methods applied in the process of estimating PD, so-called probability models are particularly significant from the practical point of view. Within the framework of this approach, the PD is a function of the arguments which are the characteristics of borrowers. Thus, the probability of the event consisting in that *i*-th borrower becomes insolvent is modelled according to the general rule:

$$P_i = F(\boldsymbol{x}), \tag{1}$$

where P_i is the probability of default of *i*-th borrower, $F()$ is a function of vector $\boldsymbol{x}$ (borrower characteristics).

One of the simplest approaches to the issue of estimating the probability of default is the linear model, presented by the following formula:

$$P_i = F(\boldsymbol{x}_i^T \boldsymbol{\beta}) = \boldsymbol{x}_i^T \boldsymbol{\beta}, \tag{2}$$

where $\boldsymbol{\beta}$ is a vector of model parameters. The subject of modelling here is the probability of default, thus the following condition shall be met:

$$0 \le \boldsymbol{x}_i^T \boldsymbol{\beta} \le 1. \tag{3}$$

Because the values of the function of linear regression often exceed a range relevant to the measure of probability, the following form of linear model of probability is accepted:

$$P_i = F(\boldsymbol{x}_i^T \boldsymbol{\beta}) = \begin{cases} 0 & \boldsymbol{x}_i^T \boldsymbol{\beta} \le 0 \\ \boldsymbol{x}_i^T \boldsymbol{\beta} & 0 \le \boldsymbol{x}_i^T \boldsymbol{\beta} \le 1 \\ 1 & \boldsymbol{x}_i^T \boldsymbol{\beta} \ge 1 \end{cases}. \tag{4}$$

An interesting way of releasing from troublesome assumption referring to the narrow scope of the obtained results due to the specificity of the measure of probability is the approach within the framework of the probit model. It enables the estimation of the probability of default through the use of the normal cumulative probability distribution function, whose values are within the [0, 1] range. In this model, the probability of default determined on the basis of the above-mentioned function of *i* borrower comes to:

$$P_i = F(\boldsymbol{x}_i^T \boldsymbol{\beta}) = \int_{-\infty}^{\boldsymbol{x}_i^T \boldsymbol{\beta}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt. \tag{5}$$

Finding the inverse function to the cumulative distribution function, we receive expressions called probits in the following form:

$$x_i^T \boldsymbol{\beta} = F^{-1}(P_i) = G(P_i).$$ (6)

Another binomial model, which is seen the most often in banking practice, is the logit model. According to this approach, function $F()$ determined on a product of vectors $x_i^T \boldsymbol{\beta}$ is the logistic cumulative distribution function in the following form:

$$P_i = F(x_i^T \boldsymbol{\beta}) = \frac{1}{1 + \exp(1 - x_i^T \boldsymbol{\beta})} = \frac{\exp(x_i^T \boldsymbol{\beta})}{1 + \exp(x_i^T \boldsymbol{\beta})}.$$ (7)

Based on the inverse function for $F()$, one sets logits in the following form:

$$x_i^T \boldsymbol{\beta} = F^{-1}(P_i) = \ln \frac{P_i}{1 - P_i}.$$ (8)

The obtained logits are natural logarithms of a quotient of probability of default in relation to the probability of repaying the loan.

A different approach to modelling the probability of default is the use of the logarithmic-linear model, in which function $F()$ is as follows:

$$P_i = F(x_i^T \boldsymbol{\beta}) = \exp(x_i^T \boldsymbol{\beta}).$$ (9)

Similarly to the linear model, there occurs a problem with the scope of function values, which may take values above one. In order to prevent it, it is essential to keep this condition $x_i^T \boldsymbol{\beta} < 0$.

An alternative method of estimation of PD, which takes into account a vector of borrower characteristics, involves the use of the Burr distribution function (Burr, 1942). According to this approach, the probability of default may be modelled by the following relation:

$$P_i = F(x_i^T \boldsymbol{\beta}) = 1 - \frac{1}{(1 + (x_i^T \boldsymbol{\beta}^c)]^k},$$ (10)

where parameters c, k are positive and expression $x_i^T \boldsymbol{\beta}$ is not negative.

Analogously to the above-presented approaches, the Urban model (Aldrich, Nelson, 1984) can also be used. The proposed function $F()$ for given borrower characteristics is as follows:

$$P_i = F(x_i^T \boldsymbol{\beta}) = \frac{1}{2} + \frac{1}{\pi} \arctan(x_i^T \boldsymbol{\beta}).$$ (11)

The above-presented methods constitute a very significant support for analysts estimating the level of the probability of default. These approaches may be used both to estimate PD for a specific borrower and in reference to separate risk classes. The latter application has particularly significant meaning in the context of determining credit risk capital requirements. By dividing credit portfolio into several classes including loans with similar probability of default, banks may estimate PD values in particular groups based on the average values of borrower characteristics in a class. In this way, the obtained probabilities may serve the estimation of a value of unexpected losses of particular borrower groups, which determine the level of capital requirements (Basel Committee..., 2006). However, such an activity requires an adequately large set of historical data, which includes not only information on default, but first of all takes into regard borrower characteristics. This last condition causes the construction of this type of models to be impossible in many banks. Thus, bankers often use methods that allow estimating the probability of default for selected portfolios based only on the repayment history. In this case, the borrower characteristics mentioned before are not known or they have been already used on the stage of grouping loans in homogenous portfolios.

One of the methods used to calculate the probability of default of portfolios based on historical data is the method based on mortality tables. It consists in the analysis of historical observations within a coherent portfolio. The idea of this approach consists in the determination of annual rates MMR (Marginal Mortality Rate), based on which one constructs PD forecasts. Rates $MMR_{i,j}$ are set in accordance with the following formula:

$$MMR_{i,j} = \frac{LD_{i,j}}{L_{i,j}} ,\qquad (12)$$

where $LD_{i,j}$ is a number of loans granted in j period, which defaulted in year i, $i \in <1,...,n>$ (n is last year of credit life). $L_{i,j}$ is the number of loans which were granted in j period and were not classified as default cases in i-th year since their granting. Index j means a period in which loans were granted. Usually one assumes annual periods for the analysis. However, one may assume a shorter period, that is, six months or quarters, which has a significant meaning in the case of a dynamically developing portfolio of loans with a relatively short period of their life. The best example is a portfolio of cash loans. In such a case, grouping portfolios according to the year of their start may cause a relatively high estimation error. It appears, how-

ever, that dynamically growing sales of loans causes the average value of outstanding to grow rapidly every month. This may cause a significant underestimation of risk rated. In such a case, one recommends grouping loans in quarters, which should eliminate this kind of estimation error.

Based on rate $MMR_{i,j}$, one determines one aggregated indicator designating the probability of default of a portfolio in year i since its granting. Rates $M\bar{M}R_i$ are set in accordance with the following formula:

$$M\bar{M}R_i = \sum_{j=1}^{k} MMR_{i,j} \cdot w_{i,j}, \qquad (13)$$

where k is the last number of period for which one conducts the MMR analysis, $w_{i,j}$ is a weight for j-th period determined in accordance with the following formula:

$$w_{i,j} = \frac{L_{i,j}}{\sum_{j=1}^{k} L_{i,j}}. \qquad (14)$$

The above-used weights certainly meet the assumption: $\sum_{j=1}^{k} w_{i,j} = 1$.

Rate $M\bar{M}R_i$ presents the value illustrating the proportion of loans which defaulted in i-th year since the moment of their granting. Based on these rates, it is possible to set PD forecast for a portfolio of loans with a various age structure.

Based on MMR rates, it is also possible to set SR rate (Survival Rate), constituting the complement to the unity of MMR in the following form:

$$M\bar{M}R_i = 1 - SR_i. \qquad (15)$$

Due to setting SR rates, it is possible to change the forecast horizon of the probability of default. Thus, instead of analysing the nearest year within the framework of risk assessment system, bank analysts may concentrate on the long-term results of the conducted risk policy lasting for several years. The probability of default in the horizon of m periods (years) is marked by CMR_m (Cumulative Mortality Rate) and set in accordance with the following formula:

$$CMR_m = 1 - \prod_{i=1}^{m} SR_i = 1 - (SR_1 \cdot SR_2 \cdot, ..., \cdot SR_m). \qquad (16)$$

4. Practical example of estimation of PD

In order to illustrate the above-mentioned method based on mortality tables, the author presents the results of the conducted research studies. The author analysed data obtained from a financial institution specialising in car loans. All loans were granted to individuals between 1998-2000. The purpose of the loans was the purchase of cars, which constituted collateral. In this analysis, two groups of loans were taken into account. The first of them included loans granted for the purchase of used cars (marked as A), the second one included loans for new cars (marked as B). In the first case, the transaction took place between two individuals, in the second case – usually at a car dealer. In the course of analysing borrower characteristics, the author noticed that persons taking loans for new cars were on average older had a higher net income and longer work experience. In addition, the fraction of single persons in this group was on a relatively low level. Thus, it was reasonable to conduct the analysis of the probability of default for both portfolios in order to verify the hypothesis about a significantly different risk level. Those results were supposed to serve the process of assessing portfolio profitability.

Table 1 and Table 2 present the values of parameters $MMR_{i,j}$ for portfolios observed in the course of the first four years of life. In the analysis, we took into account the division of portfolios into years of their start, from 1998 to 2000.

Table 1. MMR analysis for car loan portfolio (used cars)

		MMR				MMR	SR
		year of granting					
		1998	1999	2000			
year of life	1	3.5%	3.4%	3.7%		3.6%	96.4%
	2	3.1%	3.1%	3.9%		3.5%	96.5%
	3	3.1%	3.4%			3.3%	96.7%
	4	2.5%				2.5%	97.5%

Source: author's own work.

The MMR column presents the average values of rates $MMR_{i,j}$ for the subsequent years of credit life. Thus, $\overline{MMR}_1$ refers to the average value of the probability of default observed for loans in the course of the first year of their life. Based on the obtained results, we can notice that the probability of

default is the highest in the first year of credit life and came to 3.6% for the used car loan portfolio. Table 2, which refers to the portfolio of loans granted to the purchase of new cars, presents analogous data arrangement.

Table 2. MMR analysis for car credit portfolio (new cars)

		MMRi				MMR	SR
		year of granting					
		1998	1999	2000			
year of life	1	2.22%	1.61%	1.63%		1.70%	98.30%
	2	1.87%	2.52%	1.69%		1.83%	98.17%
	3	2.09%	1.58%			1.80%	98.20%
	4	1.85%				1.85%	98.15%

Source: author's own work.

Figure 1 presents in a graphical way values of indicators $MMR_{i,j}$ set for the portfolio of loans granted to the purchase of used cars. The four-year time of observing a portfolio means that only for loans granted in year 1998, we observe all values (bars), whose amount reflects historical frequencies of default cases.
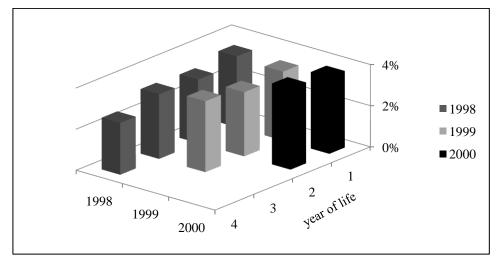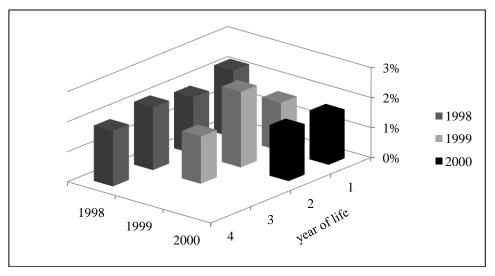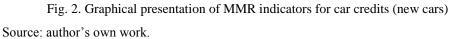

Fig. 1. Graphical presentation of MMR indicators for car credits (used cars)

Source: author's own work.

Figure 2 illustrates the historical frequency of default observed for the portfolio of loans granted to purchase new cars. Values decidedly lower

than those obtained for previous portfolio indicate the occurrence of the significant difference in credit risk levels.
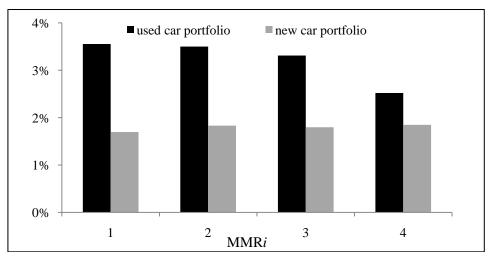

Fig. 2. Graphical presentation of MMR indicators for car credits (new cars)

Source: author's own work.

Fig. 3. Comparative analysis of MMR indicators for car credit portfolio
(used cars *vs*. new cars)

Source: author's own work.

Figure 3 presents the comparison of values of indicators $M\bar{M}R_i$ for both portfolios. Based on the obtained results, we can observe the significant difference between the values of default rates in both portfolios, regardless of the year they concern. Values $M\bar{M}R_i$ determined for used car loan portfolio on average come to 3.2%, which exceeds significantly the average value obtained for new car credit portfolio coming to 1.8%. On this basis, we may claim that the risk level of car loans granted to purchase new cars is characterised by a much lower risk level. In every year of credit portfolio life, this difference is significant, thereby confirming the thesis set before.

The obtained results may be applied to the risk assessment of a portfolio with any age structure. Let us consider for example that the management board of a financial institution expects the dynamic development of the market of sale of used cars. Thus, it was assumed that as a result of credit sale, the balance of the bank would contain loans whose amount is presented in Table 3 (line 2). Considering the above-mentioned assumptions, we decided to estimate the level of the probability of default of a portfolio. Taking into regard the time structure of loans and rates $M\bar{M}R_i$ set for particular years of credit life, the author carried out the estimation of the probability of default. That probability came to 3.29%. Knowing its value and the value of credit outstanding, as well as the expected recovery rate, the management board of the bank may determine the expected value of losses. The value of the expected loss shall be considered in the profit and loss account as the value of provisions for covering credit risk.

Table 3. Probability of default of a credit portfolio – analysis

	current portfolio			
	(date of granting)			
	2000	2001	2002	2003
age of loans [years]	4	3	2	1
number of good loans	6 000	7 000	8 500	10 100
weight of loans in portfolio	19%	22%	27%	32%
MMR	2.52%	3.31%	3.5%	3.55%
Portfolio PD	3.29%			

Source: author's own work.

The analysis of the probability of default may constitute also a significant element of a model of profitability assessment of particular loan portfo-

lios. The recognition of risk costs at the assessment of net cash flows generated by given credit portfolio shall be considered at each change of credit offer, as well as it shall be conducted periodically in order to assess bank profitability. However, PD values obtained in this way shall be verified within the framework of back-testing procedure. This analysis shall be conducted *ex-post*, and its aim is the assessment of accuracy of the created forecasts. One should not forget that banks function in a dynamic environment, where a series of macroeconomic factors has an influence on the average default rate. Thus, the constant control of forecast quality aims at the detection of the possible underestimation of the level of losses.

5. Conclusions

The issue of credit risk constitutes a key area for modern banking, within the framework of which quantitative methods measuring it are dynamically developing. Regulations such as Basel II motivate banks to develop new tools, as well as to use the existing ones. Models presented in this paper may be applied in the process of estimating the probability of default of particular borrowers and of homogenous credit portfolios. However, their use is conditioned by the possession of sufficient sets of historical data. Thus, it is so important to create in banks data warehouses, which collect systemically information on borrower characteristics and their credit history.

The results of research studies presented in this paper illustrate one of the more popular among bankers methods applied in the process of credit risk assessment. Based on the obtained results, the author established that a higher probability of default is present in the credit group marked as A. Comparative analysis of the probability of default in both groups enables determination of capital requirements with consideration of the estimated PD. We should also remember that the allocation of capital for covering risk is one of the key issues in the process of valuation of risk assets.

Literature

Aldrich J.H., Nelson F.D. (1984). *Linear probability, logit, and probit models*. Newgury Park. London.

Allen L., DeLong G.L., Saunders A. (2003). *Issues in the Credit Risk Modeling of Retail Markets*. NYU Stern School of Business Working Paper No. FIN-03-007. http://ssrn.com/abstract=412520.

Basel Committee on Banking Supervision (BCBS) (2006). International Convergence on Capital Measurement and Capital Standards. Bank for International Settlements. Basel.

Blochwitz S., Hohl S. (2001). *The Worst Case or What Default Rates Do We Have to Expect from the Rating Agencies?* Working Paper. Deutsche Bundesbank. Frankfurt.

Burr I.W. (1942). *Cumulative frequency functions*. Annals of Mathematical Statistics. Vol. 13.

Cantor R., Falkenstein E. (2001). *Testing for rating consistency in annual default rates*. Journal of Fixed Income. Vol. 11(2).

Cantor R., Mann C. (2003). *Are corporate bond ratings procyclical?* Moody's Special Comment. October.

Duffie D., Singleton K.J. (2003). *Credit Risk: Pricing Measurement and Management*. Princeton University Press. Princeton, NJ.

Gruszczyński M. (2001). *Models and Forecasts of Qualitative Variables in Finance and Banking*. Warszawa.

Hastie T., Tibshirani R., Friedman J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer. New York.

Heitfield H. (2005). *Studies on the Validation of Internal Rating Systems*. Working Paper 14. Basel Committee on Banking Supervision.

Jafry Y., Schuermann T. (2004). *Measurement, estimation and comparison of credit migration matrices*. Journal of Banking and Finance. Vol. 28(11).

Jajuga K. (1993). *Multivariate Statistical Analysis*. PWN. Warszawa.

Lopez J.A. (2004). *The empirical relationship between average asset correlation, firm probability of default, and asset size*. Journal of Financial Intermediation. No. 13. Pp. 265-283.

Merton R.C. (1974). *On the pricing of corporate debt: The risk structure of interest rates*. Journal of Finance. No. 29. Pp. 449-470.

Pluto K., Tasche D. (2005). *Thinking positively*. Risk. Vol. 18(8). Pp. 72-78

Rösch D., Scheule H. (2004). *Forecasting retail portfolio credit risk*. Journal of Risk Finance. Winter/Spring. Pp. 16-32.

Sabato G. (2006). *Managing Credit Risk for Retail Low-Default Portfolios*. http://ssrn.com/abstract=908373.

Saunders A. (2002). *Credit Risk Measurement: New Approaches to Value at Risk and Other Paradigms*. John Wiley & Sons.

Schuermann T., Hanson S. (2004). *Estimating Probability of Default*. Federal Reserve Bank of New York Staff Reports.

Svec M. (2009). *PD Backtesting Empirical Study on Credit Retail Portfolio*. CSOB Bank.

Tasche D. (2003). *A Traffic Lights Approach to PD Validation*. Working Paper. Deutsche Bundesbank.

Vasicek O. (1991). Working paper. KMV Corporation.

Vasicek O. (2002). *The distribution of loan portfolio value*. Risk. December.