Warsaw School of Economics
Institute of Econometrics
Department of Applied Econometrics

Department of Applied Econometrics Working Papers
Warsaw School of Economics
Al. Niepodleglosci 164
02-554 Warszawa, Poland

Working Paper No. 10-06

Application scoring: logit model approach and the divergence method compared

Izabela Majer
Warsaw School of Economics

Application scoring: logit model approach and the divergence method compared

Izabela Majer
Warsaw School of Economics
im25961@sgh.waw.pl

Abstract

This study presents the example of application scoring. Two methods are considered: logit model approach and the divergence method. The practical example uses contemporary data on loan applications from the Polish bank. The constructed scoring models are validated on the hold-out sample. Both types of models seem to be acceptable and have high discriminatory power. The prediction accuracy measures indicate that the scoring based on divergence method is better than the one founded on logit model approach.

Keywords: credit scoring, logit model, divergence method, credit risk, classification
JEL codes: G21, C10

1. Introduction

Application scoring models are used by loan institutions to evaluate creditworthiness of potential clients applying for credit product. The aim of scoring models is to classify applicants into two groups: the ones who will not default and the ones who will default. Application scoring models take into account all relevant information about applicants that is known at the application date and reported in an application form (i.e. demographic characteristics such as age, education, income as well as employment, marital and accommodation status). Application scoring models are used in retail and small business segments as they enable the automation of creditworthiness evaluation process and help making quick and objective credit decisions.

Out of the variety of methods for scoring models, in this study we focus on two methods: logit model approach and the divergence method. Logit model is a widely used statistical parametric model for modelling binary dependent variable and is supported by statistical tests verifying estimated parameters. The divergence method is a kind of optimisation method, not supported by econometric theory and statistical testing.

The aim of the study is to show how the scoring model can be constructed. In section 2 we present two methods used for building scoring models: logit model approach and the divergence method. Section 3 provides detailed data description. In section 4 the dependencies between explanatory variables and their association with the dependent variable are examined. Sections 5 and 6 present the models constructed with the use of both logit approach and the divergence method. In section 7 the resulted models are evaluated in terms of their predictive power. Section 8 concludes the report.

2. Theoretical background

Preliminary action undertaken in the model building process is to collect the appropriate data set and to divide it into base sample (used for model building) and hold-out sample (used for model validation). An important aspect of scoring model building is to define the dependent variable. In most cases the dependent variable is a binary one which distinguishes between two groups of applicants, defaulted and non-defaulted ones, however, the definition of default may vary between models. Let us denote by Y the dependent dummy variable which equals 0 for non-defaulted applicants and 1 for defaulted ones, and by Y_j the value of the variable Y for the *j*-th applicant.

The next step is to select a set of predictors which are significantly associated with dependent variable and are possibly free of near multicollinearity. Depending on the type of analysed variables we can use:

- Pearson's linear correlation coefficients (two quantitative variables),
- Yule's coefficients of association (two dummy variables), or
- significance tests for difference in means of a given quantitative variable in the population of applicants characterised by different values of dummy variable (a quantitative variable and a dummy one).

These measures of dependencies between variables are described in details for example by Gruszczynski [1999].

As to the decision about the modelling approach, one can choose among great variety of methods. The scoring methods were reviewed for example by Hand and Henley [1997], Janc and Kraska [2001], Baesens et al. [2003] and Matuszyk [2004]. An overview of scoring methods can be also found in whitepapers of Fair Isaac [2003] and Fractal [2003].

In this study we focus on logit model and divergence method as they represent two different approaches to scoring model building.

The logit model is a widely used statistical parametric model for modelling binary dependent variable. The logit model is described in details for example in Greene [1997], Gourieroux [2000] and Gruszczynski [2002]. We assume that an unobservable variable Y^* determines the value of observable dependent dummy variable Y e.g. as follows:

$$Y_j = \begin{cases} 1 & dla \quad Y_j^* > 0 \\ 0 & dla \quad Y_j^* \le 0 \end{cases}.$$

The unobservable variable Y^* depends on applicants' characteristics as well as on an error term, in the following way:

$$Y_j^* = \beta' x_j + \varepsilon_j ,$$

where β is a parameters' vector, x_j is a vector of the values of explanatory variables for the *j*-th applicant and ε_j is an error term. We assume that the distribution of ε_j is logistic. This implies that the probabilities of default and non-default are equal respectively to:

$$P(Y_j = 1) = F_L(\beta' x_j) = (1 + e^{-\beta' x_j})^{-1} ,$$

$$P(Y_j = 0) = 1 - F_L(\beta' x_j) = 1 - (1 + e^{-\beta' x_j})^{-1} .$$

The parameters of logit model are usually estimated with the use of maximum likelihood method.

The second approach, i.e. the divergence method represents a kind of optimisation method, not supported by econometric theory and statistical testing. In divergence method, each characteristic is defined by the group of attributes. Each attribute is assigned a score equal to a *weight of evidence* (*WOE*) calculated according to the following formula:

$$WOE_{ij} = \ln\left(\frac{n_{ij|0}}{n_0} \Big/ \frac{n_{ij|1}}{n_1} \right),$$

where $n_{ij|0}$ is a number of non-defaulted applicants characterised by *j*-th attribute of *i*-th characteristic, $n_{ij|1}$ is a number of defaulted applicants characterised by *j*-th attribute of *i*-th characteristic, n_0 and n_1 are total numbers of respectively non-defaulted applicants and defaulted ones. The discriminatory power of particular characteristic is measured and compared with the use of *information values* (*IV*) and *divergences* (*DIV*) defined as:

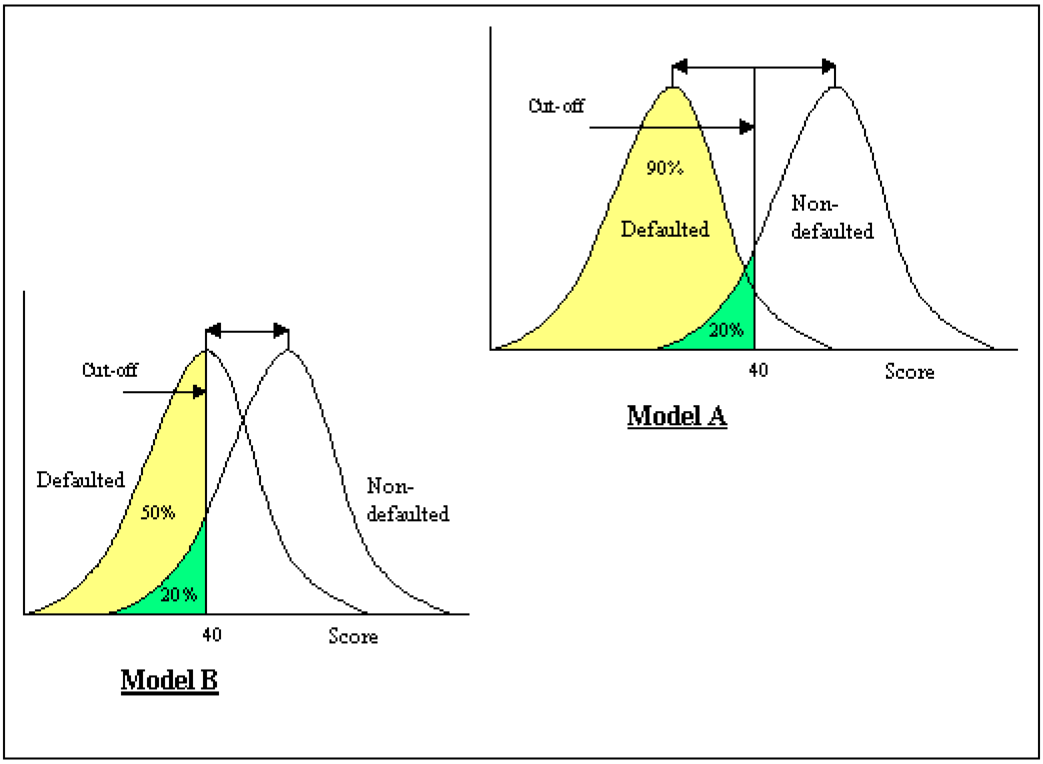
$$IV_i = \sum_j \left(\frac{n_{ij|0}}{n_0} - \frac{n_{ij|1}}{n_1} \right) \cdot WOE_{ij},$$

$$DIV_i = \frac{(\mu_{i0} - \mu_{i1})^2}{0.5(\sigma_{i0}^2 + \sigma_{i1}^2)},$$

where IV_i stands for information value of *i*-th characteristic, DIV_i stands for divergence of *i*-th characteristic, μ_{i0} and μ_{i1} are means of weights of evidence calculated for *i*-th characteristic for respectively non-defaulted applicants and the defaulted ones, σ_0^2 and σ_1^2 are variances of weights of evidence calculated for *i*-th characteristic for respectively non-defaulted applicants and defaulted ones.

The aim of this method is to built a model which will result in score distributions for defaulted and non-defaulted applicants as far apart as possible. The idea is presented in figure 1. There are two models: A and B. For both models the cut-off on 40 results in 20% of non-defaulted applicants who are rejected. However, model A ensures rejecting 90% of defaulted applicants whereas model B rejects merely 50% of defaulted applicants. That is why model A definitely outperforms model B. The distance between score distributions is measured by divergence. The total scoring for a given applicant is calculated as a sum of weights of evidence assigned to respective attributes of characteristics included in a model. The aim is to find a subset of characteristics which will result in the highest value of divergence. In this study this is done by exhaustive searching (i.e. we calculated divergences for all possible subsets of characteristics). The ratios used in divergence method are described for example by Hand and Adams [2000] and by Janc and Kraska [2001].

Figure 1. Divergence method – comparison of two models.



As soon as a scoring model is built the cut-off value has to be chosen. Applicants with the score higher than the cut-off will be accepted whereas those with lower score will be rejected. The decision on the cut-off level is critical as this sets the level of risk acceptable by decision maker.

The validation for the hold-out sample is the last step of the analysis. There is a variety of performance measures that can be used to evaluate the quality of scoring models. One can distinguish between performance measures depending on a cut-off point and those which depend only on the distribution of scores for defaulted and non-defaulted applicants (e.g. K-S statistic and Gini coefficient). The measures commonly used for evaluation of scoring models are presented for example by Kraft et al. [2002], Gruszczynski [2002] and Wilkie [2004].

3. Data description

The scoring models presented in this study were built on the basis of data supplied by one of the banks operating in Poland. Data set consists of 500 credit applications from August 2004 to May 2005. The information from application forms consists of 21 characteristics of applicants as well as the date of application. The applications are for new clients of the bank as well as those who have already used bank's products. The rejected applications are not available.

Data set on the credit performance covers time since May 2004 to July 2005 (snapshot data at the end of each month). During the analysed period some clients used more than one credit product, e.g. credit card and mortgage loan. Credit performance information is a kind of aggregate referring to all products used by a client. For each month the maximum number of days past due and the total amount owed to the bank is available. For example, if a client had payment delay of 20 days on mortgage and payment delay of 10 days on credit card in our data set there was information about 20 days past due. Data was selected in such a way that in the first month of credit performance period all the clients had no payment delay. For some clients credit performance data was also available for months preceding the application date (because they used other products of the bank).

Because of the very low default rate, some simplifying assumptions were unavoidable. The defaulted client is defined as a client who during the time period May 2004 – July 2005 had at least one payment delay of more than 30 days. However, default could have happened before the application date. Number of defaulted clients amounted to 250. The relevant applications cover time period August 2004 – May 2005 (among them 56 applications were submitted after 1st January, 2005). Other clients are regarded as non-defaults. The relevant application forms were submitted in August 2004.

Due to the fact that the initial data set consisted of a small number of applications, it seemed reasonable to increase the number of applicants by 50 randomly selected clients. As a result we come up with the sample of 550 applicants covering 50 pairs of applicants with the same profile (the same attributes of all characteristics as well as the same credit performance).

Next step was to select the hold-out sample by randomly selecting 20% of applications (110 applications, out of which 55 were defaulted clients and 55 non-defaulted ones) to be used for the validation. Remaining 440 applications were used as the base sample for models' construction.

The confidentiality of data does not allow for disclosing the names of applicant's characteristics (from the application form). Therefore the variables used in the models are coded. Some of data collected in application forms are the dates, other are either quantitative or qualitative variables. Table 1 presents the list of characteristics collected from application forms. In case of qualitative variables the number of attributes is given.

Table 1. Characteristics used in application forms.

Characteristic code	Description	Number of attributes
Application date	Date	
Char. 1	Date	
Char. 2	Date	
Char. 3	Date	
Char. 4	Date	
Char. 5	Continuous variable	
Char. 6	Continuous variable	
Char. 7	Discrete variable	
Char. 8	Discrete variable	
Char. 9	Continuous variable	
Char. 10	Continuous variable	
Char. 11	Qualitative variable	
Char. 12	Qualitative variable	4
Char. 13	Qualitative variable	2
Char. 14	Qualitative variable	3
Char. 15	Qualitative variable	7
Char. 16	Qualitative variable	8
Char. 17	Qualitative variable	13
Char. 18	Qualitative variable	10
Char. 19	Qualitative variable	10
Char. 20	Qualitative variable	9
Char. 21	Qualitative variable	2

Source: Own analysis.

4. Associations between variables

The association between dummy variables and the variable Y has been verified with the use of Yule's coefficients of association as well as with the independence test for dummy variables based on Chi-square statistic. Table 2 presents the variables which are statistically significantly associated with the variable Y. It should be noted that each qualitative variable has been transformed into the set of binary variables, each representing one attribute.

Table 2. Qualitative variables statistically significantly associated with the dependent variable.

Characteristic code	Attributes	Variable code	Yule	Chi ^ 2
Char. 17	Attribute 1	X1	-0.200	17.520
	Attribute 2	X2	0.199	17.368
Char. 18	Attribute 1	X3	-0.125	6.855
	Attribute 2	X4	0.112	5.483
Char. 20	Attribute 1	X5	-0.535	126.082
Char. 15	Attribute 1	X6	-0.265	30.826
Char. 12	Attribute 1	X7	-0.140	8.627
Char. 21	Attribute 1	X8	-0.158	11.051
Char. 11	Attribute 1	X9	0.105	4.827
	Attribute 1	X10	0.173	13.244
	Attribute 1	X11	-0.111	5.404
	Attribute 1	X12	-0.096	4.088
	Attribute 1	X13	-0.148	9.591
	Attribute 1	X14	-0.182	14.545
Char. 16	Attribute 1	X15	-0.137	8.310

Source: Own calculations.

Table 3. Analysis of association between quantitative variables and the dependent variable.

Codes of characteristics used to create the variable	Variable code	*U* statistic
Variables significantly associated with *Y*		
Application date, Char. 4	X16	-7.957
Application date, Char. 2	X17	-6.605
Char. 8	X18	-5.950
Application date, Char. 1	X19	-5.279
Char. 9	X20	-4.652
Char. 10	X21	-4.326
Char. 6, Char. 8, Char. 9, Char. 10	X22	4.206
Char. 7	X23	-3.805
Char. 6, Char. 8	X24	3.096
Char. 8, Char. 9, Char. 10	X25	-2.607
Variables insignificantly associated with *Y*		
Char. 5	-	0.789
Char. 6	-	0.483
Application date, Char. 3	-	0.207
Char. 6, Char. 9, Char. 10	-	0.004

Source: Own calculations.

In case of quantitative variables (continuous or discrete ones) the verification of association with variable *Y* was based on significance tests for difference in means of a given variable in the population of defaulted and non-defaulted applicants. The test is based on normally distributed statistic *U*. Table 3 presents the values of statistic *U* for all analysed quantitative variables.

Next, all the continuous and discrete variables have been also transformed into dummy variables. For each such variable the set of binary ones has been constructed. The associations between the newly created dummy variables and the dependent variable (Yule's coefficients and Chi-square statistics) are presented in Table 4.

The collinearity of explanatory variables was analysed only for the variables which were significantly associated with *Y*. For each pair of quantitative variables we calculated the value of Pearson's linear correlation coefficient (see Table 5). The correlation coefficients which statistically significantly differ from zero are marked with blue background, while the ones which indicate a kind of dangerous collinearity are marked with bold lettering. In case of 12 pairs of quantitative variables, the correlation coefficients are higher than 0.3, whereas in case of 4 pairs of variables they are even higher than 0.85.

Table 4. Dummy variables statistically significantly associated with the dependent variable created by the transformation of quantitative variables.

Char. codes	Interval	Variable code	Yule	Chi ^ 2	Char. codes	Interval	Variable code	Yule	Chi ^ 2
Char. 6	Interval 1	X26	0.120	6.346	**Char. 8**	Interval 1	X56	0.273	32.881
	Interval 2	X27	0.099	4.350		Interval 2	X57	-0.138	8.344
	Interval 3	X28	-0.118	6.159		Interval 3	X58	-0.138	8.327
	Interval 4	X29	0.104	4.774	**Application date, Char. 4**	Interval 1	X59	0.286	36.103
Char. 5	Interval 1	X30	0.152	10.128		Interval 2	X60	0.175	13.469
	Interval 2	X31	0.112	5.551		Interval 3	X61	0.117	6.005
	Interval 3	X32	-0.117	6.051		Interval 4	X62	-0.152	10.101
	Interval 4	X33	-0.099	4.274		Interval 5	X63	-0.161	11.403
	Interval 5	X34	-0.104	4.774		Interval 6	X64	-0.160	11.206
Char. 6, Char. 8	Interval 1	X35	0.1337	7.8658		Interval 7	X65	-0.096	4.050
						Interval 8	X66	-0.170	12.712
Char. 6, Char. 8, Char. 9, Char. 10	Interval 1	X36	-0.123	6.605	**Application date, Char. 2**	Interval 1	X67	0.417	76.356
	Interval 2	X37	-0.161	11.353		Interval 2	X68	0.175	13.480
	Interval 3	X38	0.221	21.560		Interval 3	X69	0.160	11.282
	Interval 4	X39	-0.119	6.239		Interval 4	X70	0.139	8.451
	Interval 5	X40	0.164	11.826		Interval 5	X71	-0.329	47.619
	Interval 6	X41	0.156	10.653		Interval 6	X72	-0.351	54.272
Char. 6, Char. 9, Char. 10	Interval 1	X42	0.173	13.095		Interval 7	X73	0.098	4.199
	Interval 2	X43	-0.143	8.971	**Application date, Char. 1**	Interval 1	X74	0.199	17.366
	Interval 3	X44	0.126	7.040		Interval 2	X75	0.172	12.975
	Interval 4	X45	0.170	12.750		Interval 3	X76	-0.134	7.944
Char. 8, Char. 9, Char. 10	Interval 1	X46	-0.156	10.669		Interval 4	X77	-0.148	9.612
	Interval 2	X47	0.392	67.675	**Application date, Char. 3**	Interval 1	X78	0.365	58.761
	Interval 3	X48	-0.172	12.997		Interval 2	X79	-0.181	14.421
	Interval 4	X49	-0.248	26.954		Interval 3	X80	-0.113	5.570
Char. 10	Interval 1	X50	0.683	205.023		Interval 4	X81	-0.146	9.378
	Interval 2	X51	-0.286	36.061		Interval 5	X82	0.104	4.774
	Interval 3	X52	-0.311	42.464	**Char. 8**	Interval 1	X83	0.260	29.689
	Interval 4	X53	-0.280	34.510		Interval 2	X84	-0.229	23.106
	Interval 5	X54	-0.194	16.604	**Char. 7**	Interval 1	X85	0.178	13.910
	Interval 6	X55	-0.147	9.542		Interval 2	X86	-0.123	6.668

Source: Own calculation.

Table 5. Pearson's linear correlation coefficients for quantitative explanatory variables.

	X16	X17	X18	X19	X20	X21	X22	X23	X24	X25
X16	1	0.151	0.287	**0.462**	0.213	0.123	-0.134	0.227	-0.141	-0.055
X17		1	0.152	0.089	0.261	0.074	-0.026	0.088	-0.007	0.048
X18			1	0.112	**0.871**	0.027	**-0.315**	**0.895**	**-0.429**	**-0.308**
X19				1	0.124	0.170	-0.129	0.021	-0.087	0.106
X20					1	0.040	-0.270	**0.748**	**-0.350**	-0.220
X21						1	**-0.312**	0.010	0.039	**0.902**
X22							1	-0.271	**0.925**	-0.153
X23								1	**-0.370**	-0.268
X24									1	0.233
X25										1

Source: Own calculation.

The analysis of collinearity for pairs of binary variables (qualitative variables as well as transformed quantitative ones) is based on Yule's coefficients of association. In Table 6 we present the matrix of the coefficients calculated for qualitative variables.

Table 6. Yule's coefficients of association for qualitative explanatory variables.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15
X1	1	**-0.65**	0.27	-0.28	0.12	0.07	0.00	0.08	0.04	-0.03	0.07	0.01	0.06	0.00	0.15
X2		1	-0.10	-0.07	-0.14	-0.10	-0.09	-0.06	0.02	0.04	-0.07	0.04	-0.07	-0.02	-0.14
X3			1	-0.11	0.06	0.03	-0.06	**0.32**	-0.02	-0.01	0.03	-0.03	-0.02	0.02	0.22
X4				1	-0.01	-0.01	0.00	-0.03	-0.02	-0.04	-0.02	-0.04	-0.03	-0.06	0.05
X5					1	0.18	0.09	0.19	0.03	-0.10	0.05	0.06	0.09	0.14	0.07
X6						1	0.08	0.00	-0.10	0.01	-0.05	0.07	0.06	0.19	0.01
X7							1	0.02	0.02	0.07	-0.03	-0.04	0.05	0.06	-0.04
X8								1	0.09	-0.06	0.00	0.11	0.05	0.08	0.18
X9									1	-0.09	-0.14	-0.09	-0.06	-0.12	0.04
X10										1	-0.09	-0.06	-0.04	-0.08	-0.02
X11											1	-0.09	-0.07	-0.12	-0.03
X12												1	-0.04	-0.08	0.08
X13													1	-0.06	0.10
X14														1	-0.05
X15															1

Source: Own calculation.

The statistically significantly different from zero coefficients are marked with blue background, the ones indicating significant association are marked with bold lettering. As we can see, only 2 pairs of variables are significantly associated.

The Yule's coefficients of association were also constructed for pairs of transformed quantitative variables and for pairs consisting of a transformed quantitative variable and a qualitative one. Due to the high dimension of this matrix we present only the pairs of variables for which the values of Yule's coefficient are higher than 0.3 (Table 7).

Table 7. Yule's coefficients of association for chosen pairs of transformed quantitative variables and for chosen pairs consisting of an transformed quantitative variable and a qualitative one.

Variable		Yule	Variable		Yule
X5	X50	-0.379	X50	X55	-0.302
X5	X59	-0.376	X50	X67	0.304
X5	X67	-0.796	X56	X57	-0.436
X6	X56	-0.761	X56	X74	0.322
X6	X57	0.343	X56	X83	0.872
X6	X58	0.313	X57	X58	-0.308
X6	X83	-0.876	X57	X83	-0.466
X9	X58	0.381	X58	X83	-0.328
X50	X51	-0.455	X58	X84	0.740
X50	X52	-0.436	X27	X28	-0.350
X50	X53	-0.345	X74	X83	0.330

Source: Own calculation.

We have also analysed the relationships between quantitative variables and qualitative ones verifying statistical significance of difference in means of a quantitative variable for the groups of clients with different values of a dummy (qualitative) variable. The directions of associations are presented in Table 8. Those which are statistically significant (i.e. difference in means differs statistically from zero) are marked with a blue background.

Table 8. Relationships between quantitative variables and qualitative ones.

	X16	X17	X18	X19	X20	X21	X22	X23	X24	X25
X1	+	+	+	-	+	-	+	+	+	-
X2	-	-	-	-	-	-	-	-	-	-
X3	+	+	+	-	+	-	-	-	-	-
X4	-	-	-	-	-	-	+	-	+	-
X5	+	+	+	+	+	+	-	+	-	+
X6	+	+	+	+	+	+	-	+	-	-
X7	+	+	+	+	+	+	-	+	-	+
X8	+	+	-	+	+	+	-	-	-	+
X9	-	+	-	-	+	-	+	-	+	+
X10	-	-	-	+	-	-	+	-	-	-
X11	-	+	-	-	+	-	+	-	+	-
X12	+	+	+	+	+	+	-	+	-	-
X13	+	+	+	+	+	+	-	+	-	+
X14	+	+	+	+	+	-	-	+	-	-
X15	+	+	-	+	+	+	+	-	+	+

Source: Own calculation.

To sum up, the results of the analysis of variable selection show that some variables have no influence on *Y*. Moreover, some of the variables significantly associated with *Y* cannot be included into the model because of high degree of collinearity with other explanatory variables.

5. Logit model

To adequately specify the logit model we targeted the subsets of explanatory variables, not significantly pair-wise associated / correlated. For each subset the parameters of logit model have been estimated. The variables with low t-ratio have been excluded. Table 9 presents the subsets: explanatory variables originally included in a given model are marked as X, variables finally included in a given model are marked with green background, variables included in the best model (i.e. in the one with the highest value of likelihood ratio) are marked with blue background.

We also estimated a logit model using only dummy variables (i.e. transformed quantitative as well as qualitative ones). Initial subset of explanatory variables included only those not significantly associated with each other. For the final model the value of loglikelihood is –96.1 while the likelihood ratio equals to 0.68. Table 10 presents the estimation results.

Table 9. Analysed subsets of explanatory variables and variables finally included in various logit models.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
X1	X	x	x	x	x		x		x		x				x	x			x		x		x	x		x		x				
X2						x		x		x		x	x	x			x	x		x		x			x		x		x	x	x	x
X3	X		x		x	x			x	x			x	x	x				x	x			x	x	x				x		x	
X4	X	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
X5	X	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
X6	X	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
X7	X	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
X8		x		x			x	x			x	x				x	x	x			x	x				x	x	x		x		x
X9	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
X10	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
X11	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
X12	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
X13	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
X14	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
X15	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
X16					x	x	x	x	x	x	x	x							x	x	x	x		x	x	x	x					
X17	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
X18																							x	x	x	x	x	x	x	x		
X19	x	x	x	x									x	x	x	x	x	x					x					x	x	x	x	x
X20	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x									x	x
X21															x	x			x	x	x	x	x	x	x	x	x	x	x	x	x	x
X22			x	x					x	x	x	x	x				x															
X24	x	x			x	x	x	x						x	x	x		x	x	x	x	x									x	x
X25	x	x	x	x	x	x	x	x	x	x	x	x	x	x			x	x														
Logarithm of likelihood function	-181		-181		-181				-181				-184		-185		-185		-185				-186	-186				-186	-186	-187	-187	
Likelihood ratio index	0.407		0.407		0.406				0.406				0.396		0.395		0.395		0.392				0.39	0.392				0.39	0.39	0.39	0.388	

Source: Own calculation.

Table 10. Estimation results for the logit model.

Explanatory variable	Estimated parameter	Standard error	t-ratio	prob
Constant	7.037	1.401	5.020	0.000
X71	-4.135	0.742	-5.570	0.000
X51	-3.481	0.688	-5.060	0.000
X72	-2.139	0.443	-4.830	0.000
X53	-7.251	1.574	-4.610	0.000
X52	-3.475	0.827	-4.200	0.000
X54	-4.444	1.141	-3.890	0.000
X62	-2.333	0.614	-3.800	0.000
X5	-4.703	1.379	-3.410	0.001
X55	-2.241	0.683	-3.280	0.001
X10	3.390	1.145	2.960	0.003
X63	-2.442	0.826	-2.960	0.003
X4	3.831	1.391	2.750	0.006
X79	-1.698	0.648	-2.620	0.009
X75	1.105	0.519	2.130	0.033
X66	-2.489	1.190	-2.090	0.037

Source: Own calculation.

Both likelihood ratio test and Wald test rejected the hypothesis that all estimated parameters except for the constant are zero ($LR = 417.76$; $W = 90.55$). All estimates have expected signs.

Table 11 presents marginal effects and elasticities of the probability of default calculated for the applicant characterised by average values of all explanatory variables (in case of binary variables we used the probability that a given variable is equal to 1).

Table 11. Marginal effects and elasticities of the probability of default for the logit model.

Explanatory variable	Marginal effect	Elasticity
X4	0.434	0.045
X5	-0.662	-1.553
X10	0.446	0.085
X51	-0.613	-0.179
X52	-0.607	-0.165
X53	-0.697	-0.232
X54	-0.597	-0.073
X55	-0.465	-0.054
X62	-0.488	-0.099
X63	-0.490	-0.064
X66	-0.487	-0.039
X71	-0.693	-0.308
X72	-0.486	-0.251
X75	0.244	0.085
X79	-0.388	-0.069

Source: Own calculation.

Table 12 presents the comparison of marginal effects for pairs of explanatory variables. Element in the *i*-th row and *j*-th column is the ratio of the estimated parameter for variable in *i*-th row head to the estimated parameter for variable in *j*-th column head.

Table 12. Comparison of marginal effects for pairs of explanatory variables in the logit model.

	X5	X10	X51	X52	X53	X54	X55	X62	X63	X66	X71	X72	X75	X79
X4	-0.81	1.13	-1.10	-1.10	-0.53	-0.86	-1.71	-1.64	-1.57	-1.54	-0.93	-1.79	3.47	-2.26
X5		-1.39	1.35	1.35	0.65	1.06	2.10	2.02	1.93	1.89	1.14	2.20	-4.26	2.77
X10			-0.97	-0.98	-0.47	-0.76	-1.51	-1.45	-1.39	-1.36	-0.82	-1.58	3.07	-2.00
X51				1.00	0.48	0.78	1.55	1.49	1.43	1.40	0.84	1.63	-3.15	2.05
X52					0.48	0.78	1.55	1.49	1.42	1.40	0.84	1.62	-3.14	2.05
X53						1.63	3.24	3.11	2.97	2.91	1.75	3.39	-6.56	4.27
X54							1.98	1.90	1.82	1.79	1.07	2.08	-4.02	2.62
X55								0.96	0.92	0.90	0.54	1.05	-2.03	1.32
X62									0.96	0.94	0.56	1.09	-2.11	1.37
X63										0.98	0.59	1.14	-2.21	1.44
X66											0.60	1.16	-2.25	1.47
X71												1.93	-3.74	2.44
X72													-1.94	1.26
X75														-0.65

Source: Own calculation.

Due to the fact that the model was estimated on the balanced sample there is no need to adjust the constant and cut-off can be set on 0.5 (according to the standard prediction rule).

6. Divergence method

First step in scorecard building process was to calculate information values, weights of evidence and contributions for each attribute of each characteristic. Table 13 presents information values and divergences for all analysed characteristics.

Table 13. Information values and divergences.

Characteristic	Information value	Divergence
Char. 20	4.357	1.607
Char. 10	3.462	1.422
Application date, Char. 2	2.197	2.667
Char. 8, Char. 9, Char. 10	1.176	1.224
Application date, Char. 4	1.118	1.132
Char. 6, Char. 8, Char. 9, Char. 10	0.721	0.718
Application date, Char. 3	0.680	0.723
Char. 11	0.602	0.581
Char. 6, Char. 9, Char. 10	0.484	0.499
Application date, Char. 1	0.482	0.504
Char. 8	0.428	0.424
Char. 5	0.364	0.374
Char. 9	0.356	0.371
Char.15	0.288	0.301
Char. 17	0.263	0.255
Char. 6	0.204	0.205
Char. 6, Char. 8	0.179	0.181
Char. 7	0.156	0.156
Char. 8	0.119	0.111
Char. 21	0.102	0.103
Char. 12	0.098	0.093
Char. 16	0.085	0.086
Char. 19	0.049	0.049

Source: Own calculation.

The divergence method amounts in fact to finding the combination of characteristics giving the highest divergence for the model as a whole. 24 subsets of variables were found as not collinear. For each such subset the combination of characteristics with the highest divergence value has been chosen. Table 14 presents the selected subsets of characteristics and the characteristics composing the most predictive combinations (i.e. the ones with the highest value of divergence among all the combinations of characteristics of a given subset). The characteristics primarily chosen to the model are marked with X. Those finally included are marked with green background. The variables included in the best model (i.e. with the highest value of divergence) are marked with the blue background.

Table 15 presents the scores associated with attributes of characteristics included in the model – after the appropriate scaling to ensure a total scoring of any applicant is not lower than 0 and not higher than 100.

Setting the cut-off point was the final step. The weighted average of mean score for defaulted clients and mean score for non-defaulted ones was equal 42.03. On the other hand, the number of scores for which the Mahalanobis distances between mean score for defaulted clients and mean score for non-defaulted ones amounted to 40.66. The criterion applied to the cut-off choice was minimization of the percentage of incorrectly classified defaulted clients (misclassification matrices are presented in Table 16). So, finally we set the cut-off on 42.03.

Table 14. Subsets of characteristics and characteristics finally included in alternative models – divergence method.

Characteristic	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Char. 6									X	X					X	X					X	X		
Char. 5											X	X					X	X					X	X
Char. 6, Char. 8													X	X					X	X				
Char. 8	X	X			X	X															X	X	X	X
Char. 7			X	X			X	X	X	X	X	X	X	X	X	X	X	X	X	X				
Char. 6, Char. 9, Char. 10	X	X	X	X																				
Char. 6, Char. 8, Char. 9, Char. 10					X	X	X	X																
Char. 9									X	X	X	X	X	X										
Char. 8, Char. 9, Char. 10															X	X	X	X	X	X	X	X	X	X
Char. 10									X	X	X	X	X	X										
Application date, Char. 2	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Char. 20	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Char. 15	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Char. 19	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Char. 12	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Char. 18	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Application date, Char. 1	X		X		X		X		X		X		X		X		X		X		X		X	
Application date, Char. 4		X		X		X		X		X		X		X		X		X		X		X		X
Char. 16	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Application date, Char. 3	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Char. 17	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Char. 21	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Char. 11	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Divergence	**3.91**	**4.35**	**3.91**	**4.35**	**3.98**	**4.41**	**3.98**	**4.41**	**6.30**	**7.04**	**6.43**	**7.14**	**6.22**	**7.01**	**4.94**	**5.43**	**5.09**	**5.50**	**4.84**	**5.32**	**4.94**	**5.43**	**5.09**	**5.50**

Source: Own calculation.

Table 15. Scorecard from the divergence method – scores associated with attributes of characteristics.

Characteristic	Attribute	Score	Characteristic	Attribute	Score
Char. 5	Attribute 1	**1.27**	Application date, Char 3	Attribute 1	**0.31**
	Attribute 2	**3.35**			
	Attribute 3	**0.92**		Attribute 2	**5.99**
	Attribute 4	**7.61**		Attribute 3	**5.25**
	Attribute 5	**4.67**		Attribute 4	**3.98**
	Attribute 6	**3.90**		Attribute 5	**6.18**
	Attribute 7	**4.84**		Attribute 6	**8.64**
	Attribute 8	**5.75**		Attribute 7	**5.05**
	Attribute 9	**6.70**	Application date, Char 2	Attribute 1	**-4.77**
	Attribute 10	**4.08**		Attribute 2	**-0.37**
	Attribute 11	**5.89**		Attribute 3	**-1.82**
	Attribute 12	**3.14**		Attribute 4	**9.51**
	Attribute 13	**2.38**		Attribute 5	**8.01**
Char. 9	Attribute 1	**1.37**		Attribute 6	**1.83**
	Attribute 2	**2.85**		Attribute 7	**6.43**
	Attribute 3	**5.35**		Attribute 8	**3.35**
	Attribute 4	**4.72**	Application date, Char 4	Attribute 1	**3.21**
	Attribute 5	**5.95**		Attribute 2	**-2.58**
Char. 10	Attribute 1	**0.29**		Attribute 3	**0.75**
	Attribute 2	**10.32**		Attribute 4	**2.14**
	Attribute 3	**12.36**		Attribute 5	**3.78**
	Attribute 4	**28.56**		Attribute 6	**4.61**
	Attribute 5	**26.45**		Attribute 7	**7.06**
	Attribute 6	**8.28**		Attribute 8	**8.10**
Char. 21	Attribute 1	**5.45**		Attribute 9	**6.69**
	Attribute 2	**3.39**		Attribute 10	**12.29**
Char. 18	Attribute 1	**3.92**	Char. 11	Attribute 1	**2.32**
	Attribute 2	**5.27**		Attribute 2	**-1.09**
	Attribute 3	**3.50**		Attribute 3	**5.93**
	Attribute 4	**-0.80**		Attribute 4	**4.95**
	Attribute 5	**4.34**		Attribute 5	**4.08**
Char. 19	Attribute 1	**5.10**		Attribute 6	**3.28**
	Attribute 2	**3.88**		Attribute 7	**4.76**
	Attribute 3	**4.84**		Attribute 8	**3.32**
	Attribute 4	**3.73**		Attribute 9	**6.54**
	Attribute 5	**4.70**		Attribute 10	**11.62**
	Attribute 6	**2.97**		Attribute 11	**7.79**
	Attribute 7	**3.14**		Attribute 12	**2.46**

Source: Own calculation.

Table 16. Misclassification matrices for different cut-off rules.

Cut-off: Mahalanobis distance				Cut-off: Weighted average			
	$\hat{Y}=1$	$\hat{Y}=0$	Total		$\hat{Y}=1$	$\hat{Y}=0$	Total
$Y=1$	197	23	**220**	$Y=1$	202	18	**220**
$Y=0$	18	202	**220**	$Y=0$	22	198	**220**
Total	**215**	**225**	**440**	Total	**224**	**216**	**440**

Source: Own calculation.

7. Predictive accuracy of constructed scoring models

The predictive accuracy was verified with the use of the hold-out sample of 110 applicants (55 defaulted clients and 55 non-defaulted ones).

Table 17 presents misclassification matrices for both models. When comparing both models it is easily visible than in case of the logit model we come up with both type I and type II errors higher than in case of the divergence model.

Table 17. Misclassification matrices – hold-out sample.

Logit model				**Divergence method**			
	$\hat{Y}=1$	$\hat{Y}=0$	Total		$\hat{Y}=1$	$\hat{Y}=0$	Total
$Y=1$	46	9	**55**	$Y=1$	49	6	**55**
$Y=0$	8	47	**55**	$Y=0$	4	51	**55**
Total	**54**	**56**	**110**	Total	**53**	**57**	**110**

Source: Own calculation.

The measures of predictive accuracy presented in Table 18 clearly indicate the dominance of the divergence approach (please refer to appendix for details on mathematical formulae of presented predictive accuracy measures).

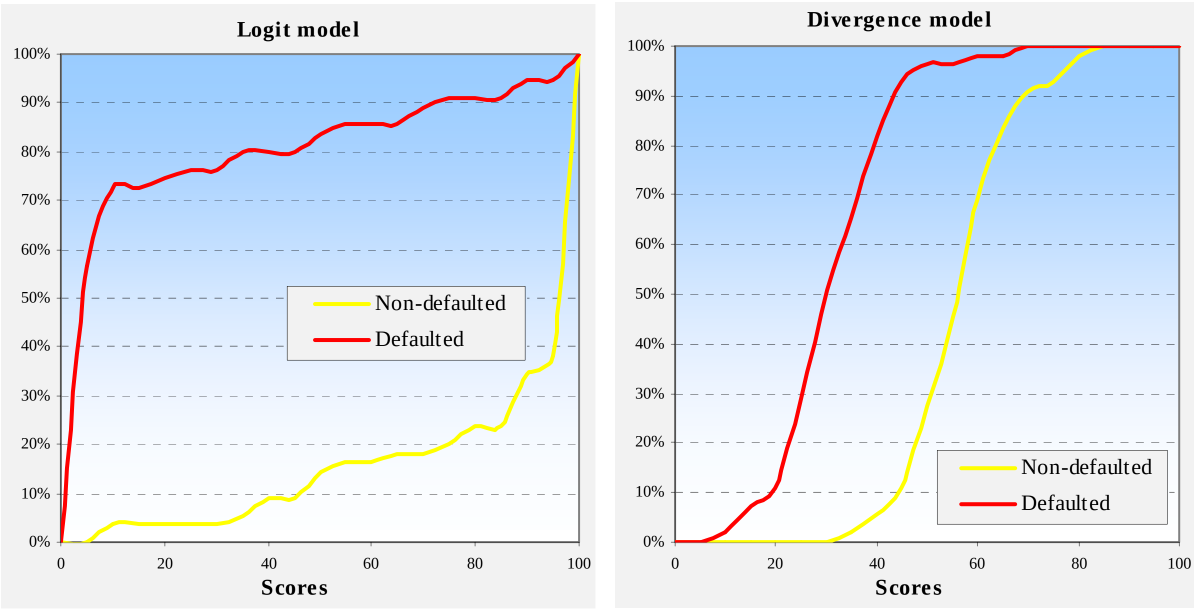
Table 18. Measures of predictive accuracy – hold-out sample.

Predictive power measure	**Logit model**	**Divergence method**
Total percentage of correctly classified	84.5%	90.9%
Percentage of correctly classified non-defaulted applicants	85.5%	92.7%
Percentage of correctly classified defaulted applicants	83.6%	89.1%
Odds ratio	30.028	104.125
Mean difference	6.593	3.988
K-S statistic	0.745	0.855
Gini coefficient	0.874	0.896
Information value	7.602	9.551

Source: Own calculation.

Figure 2 presents score distributions for both scoring models constructed in this study. Distributions for defaulted applicants are marked with red lines whereas distributions for non-defaulted applicants are represented by yellow ones. In case of the logit model score intervals for defaulted applicants and non-defaulted ones are almost overlapping, contrary to the divergence model. For the divergence approach the maximum score assigned to a defaulted applicant amounted to 66.5 whereas the minimum score assigned to a non-defaulted applicant was equal 30.7.

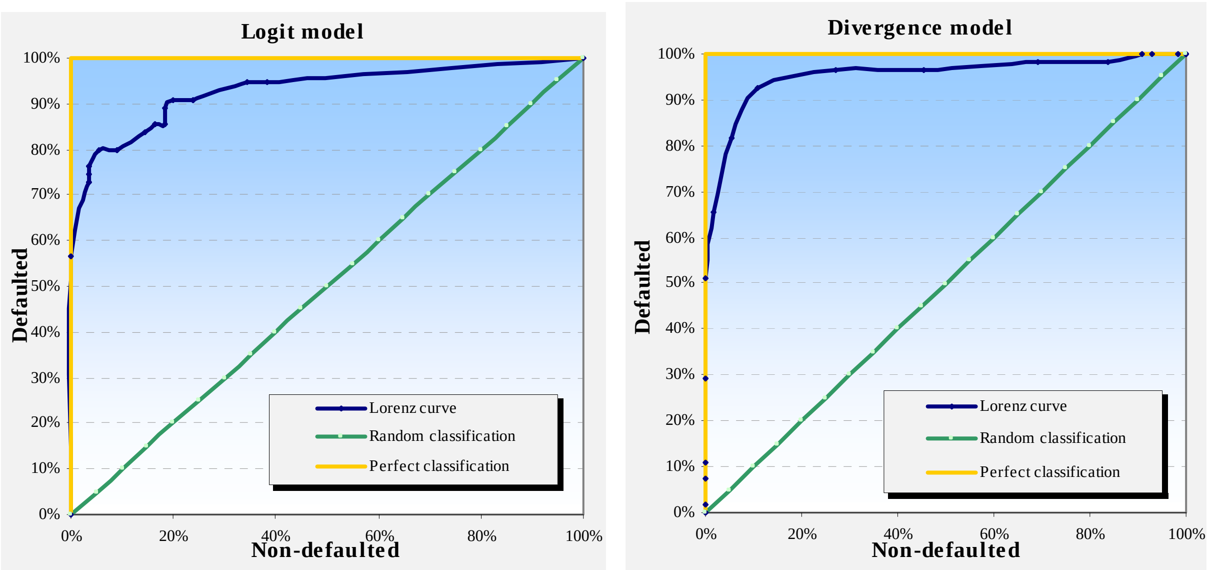
Figure 2. Score distributions.



Source: Own calculation.

Figure 3 presents Lorenz curves for both models (blue lines) as well as for hypothetical models resulting in perfect (yellow lines) and random (green lines) classifications of applicants. The figures show that both models are definitely better than random classification of applicants. However, in this study the divergence approach outperforms the logit model.

Figure 3. Lorenz curves.



Source: Own calculation.

8. Conclusions

The study presents an example of application scoring. On the basis of the same data set we constructed two scoring models, one based on the logit approach and the second on the divergence approach.

The validation on the hold-out sample shows that both models are acceptable and have high discriminatory power. However, in this study the model constructed with the use of divergence method outperforms the one resulted from the logit approach. Obviously, such results are restricted to the particular data set.

References

A Fair Isaac Whitepaper [2003], *A discussion of data analysis, prediction and decision techniques*, A Fair Isaac Whitepaper, http://www.fairisaac.com

A Fractal Whitepaper [2003], *Comparative analysis of classification techniques*, http://www.fractalanalytics.com

Allen L., DeLong G., Saunders A. [2004], *Issues in the credit risk modeling of retail markets*, Journal of Banking and Finance 28, p. 727-752.

Baesens B., Van Gestel T., Viaene S., Stepanowa M., Suykens J., Vanthienen J. [2003], *Benchmarking state-of-the-art classification algorithms for credit scoring*, Journal of the Operational Research Society 54, p. 627-635.

Foster D.P., Stine R.A. [2004], *Variable selection in data mining: Building a predictive model for bankruptcy*, Department of Statistics, The Wharton School of the University of Pennsylvania, http://www-stat.wharton.upenn.edu/~bob/research/bankrupt.pdf

Gourieroux Ch. [2000], *Econometrics of qualitative dependent variables*, Cambridge University Press.

Greene W.H. [1997], *Econometric analysis*, Prentice Hall, Upper Saddle River, NJ.

Gruszczynski M. [1999], *Scoring logitowy w praktyce bankowej a zagadnienie koincydencji*, Bank i Kredyt, 5, p. 57-62.

Gruszczynski M. [2002], *Modele i prognozy zmiennych jakosciowych w finansach i bankowosci*, Oficyna Wydawnicza SGH, Warszawa.

Hand D.J., Adams N.M. [2000], *Defining attributes for scorecard construction in credit scoring*, Journal of Applied Statistics 27, 5, p. 527-540.

Hand D.J., Henley W.E. [1997], *Statistical classification methods in consumer credit scoring:a review*, Journal of the Royal Statistical Society, Series A ,160, 3, p. 523-541.

Janc A., Kraska M. [2001], *Credit-scoring. Nowoczesna metoda oceny zdolnosci kredytowej*, Biblioteka Menedzera i Bankowca, Warszawa.

Jozwiak J., Podgorski J. [2000], *Statystyka od podstaw*, PWE, Warszawa.

Kraft H., Kroisandt G., Muller M. [2002], *Assessing the discriminatory power of credit scores*, http://edoc.hu-berlin.de/series/sfb-373-papers/2002-67/PDF/67.pdf

Landao D. [2004], *Credit risk modeling: theory and applications*, Oxford Princeton University Press.

Lasek M. [2002], *Data Mining. Zastosowanie w analizach i ocenach klientow bankowych*, Biblioteka Menedzera i Bankowca, Warszawa.

Matuszyk A. [2004], *Credit scoring. Metoda zarządzania ryzykiem kredytowym*, CeDeWu, Warszawa.

McNab H., Wynn A. [2000], *Principles and practice of consumer credit risk management*, The Chartered Institute of Bankers, Canterbury.

Olivier R.M., Wells E. [2001], *Efficient frontier cutoff policies in credit portfolios*, Journal of the Operational Research Society 52, p. 1025-1033.

Wilkie A.D. [2004], *Measures for comparing scoring systems*, in: Thomas L.C., Edelman D.B., Crook J.N., *Readings in credit scoring foundations, developments and aims*, Oxford University Press 2004.

Appendix: Measures of predictive accuracy

Odds ratio:

$$OR = \frac{n_{11} \cdot n_{00}}{n_{01} \cdot n_{10}},$$

where n_{11} and n_{00} are the numbers of correctly classified defaulted applicants and non-defaulted ones, respectively, whereas n_{01} and n_{10} are the appropriate numbers of incorrectly classified applicants.

Mean difference:

$$MDIF = \frac{(\overline{U}_o - \overline{U}_1)}{D(U)},$$

where $\overline{U}_o$ is the mean scoring for non-defaulted applicants, $\overline{U}_1$ is the mean scoring for defaulted applicants, and $D(U)$ is the scoring standard deviation calculated as:

$$D(U) = \frac{\left(n_0^2 D_0(U) + n_1^2 D_1(U)\right)^{1/2}}{n_0 + n_1},$$

where

$$D_0(U) = \left[\sum_i i^2 \frac{n_0(i)}{n_0} - \overline{U}_0^{\,2}\right]^{1/2},$$

$$D_1(U) = \left[\sum_i i^2 \frac{n_1(i)}{n_1} - \overline{U}_1^{\,2}\right]^{1/2},$$

$n_0(i)$ and $n_1(i)$ are the numbers of non-defaulted applicants and defaulted ones, respectively, with the scoring = i, while n_0 and n_1 are total numbers of non-defaulted and defaulted applicants, respectively.

Information value:

$$IV = \sum_i \left(\frac{n_0(i)}{n_0} - \frac{n_1(i)}{n_1}\right) \cdot \ln\left(\frac{n_0(i)}{n_0} \bigg/ \frac{n_1(i)}{n_1}\right).$$

K-S statistic:

$$KS = \max_i \left(N_1(i) - N_0(i)\right),$$

where

$$N_0(i) = \sum_{j=0}^{i} \frac{n_0(j)}{n_0},$$

$$N_1(i) = \sum_{j=0}^{i} \frac{n_1(j)}{n_1}.$$

Gini coefficient:

$$GINI = \frac{A}{A+B} = 2A,$$

where A is the area between Lorenz curve for actual model and Lorenz curve for random classification of applicants, and B stands for area between Lorenz curve for perfect discrimination of applicants and Lorenz curve for the actual model. Each point of the Lorenz curve is related to a given scoring (i). The horizontal axis represents percentage of non-defaulted applicants with scoring not higher than i, i.e. $N_0(i)$, and the vertical axis represents percentage of defaulted applicants with scoring not higher than i, i.e. $N_1(i)$.

Lorenz curve