NBP Narodowy Bank Polski

Natalia Nehrebecka

Probability-of-default curve calibration and validation of internal rating systems

IFC Biennial Basel Conference 8-9 September 2016

Contents of the presentation

I

Introduction and literature review

Aim and motivation

- The *default probability* model and the *master scale* are known as the rating system.
- This article with the issue of **rating system calibration**, i.e. allocation of rating classes to entities in order to ensure that the calibration power of the division created is as high as possible.
  - **Two main research questions** will be addressed.
    - The first one verifies if there is a calibration method that gives estimations of probabilities of significantly better calibration quality in logistic regression.
    - The second question concerns rating system structure: does number of rating classes really impact calibration quality?
- The subject matter of this article is important and actual, as there is no consensus made among practitioners regarding the selection of calibration methods and manners of testing them. Comparison of methods constitutes a significant added value.

Literature review

- To transform a credit score into a probability of default (*PD*):
  - **The first one** includes methods approximate the conditional (on default and non-default) score distributions into parametric distributions
    - Dey, 2010; Bennett, 2003; Krężołek, 2007; Tasche 2006; Tasche 2008; Tasche 2009
  - **The second one** includes methods for dummy variable (default or non-default) models
    - Tasche, 2009; Neagu, Keenan, 2009; Koenker, Yoon, 2009 ; Neagu, Keenan, Chalermkraivuth, 2009; Zadrozny, Elkan, 2002; Van der Burgt, 2008

- The calibration of the scoring system which is another important task in scoring model validation
  - **The first group of tests** can only be applied to one single rating grade over a single time period *(binomial test Clopper and Pearson, binomial test Agresti and Coulla, binomial test Wald , corrected binomial test Wald, binomial test Wilson, corrected binomial test Wilson, one-factor-model, moment matching approach and granularity adjustment)*
  - **The second group of tests** provide more advanced methods that can be used to test the adequacy of the default probability prediction over a single time period for several rating grades (*Spiegelhalter test, Hosmer-Lemeshow test, Blöchlinger test*).

II

Data description

Data sources

Companies	Financial institutions	Court
Financial statement data	**Prudential Reporting**	**Judicial events**
• *AMADEUS (Bureau van Dijk)* • *Notoria OnLine*	• *NB300 (Narodowy Bank Polski)*	• *The National Court Register*
2007 – 2012		

- The preliminary stage was the implementation of the ***scoring model*** based on the Nehrebecka approach (2015).
  - The score distribution was rescaled so that the values fell between 0 and 1.

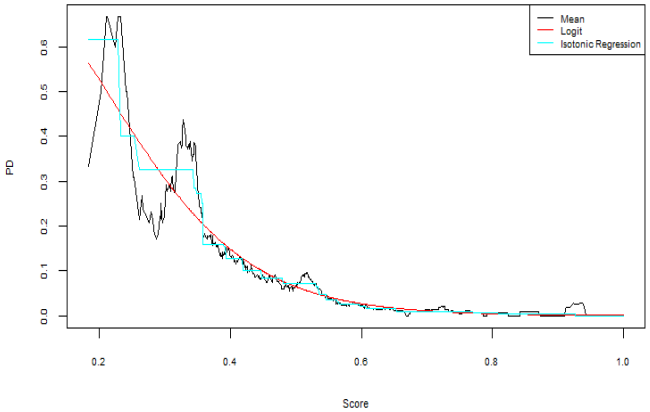
- A **master scale** was used in the article which was used in KBC bank (9 classes) in 2011, for corporate clients in Pekao bank (9 classes) in 2013, in Millenium bank (14 classes) in 2011, and in ING bank (19 classes) in 2012.

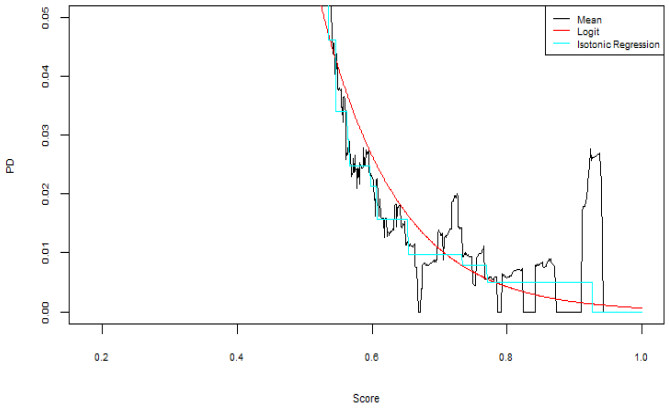
III

Calibration and verification using test for the whole rating system

Calibration

- To transform a credit score into a probability of default (*PD*):
  1. Quasi-moment-matching method [Tasche, 2009]
  2. Methods of approximating parametric distribution (Skewnormal distribution; Scaled beta distribution; Asymmetric Laplace distribution; Asymmetric Gauss distribution)
  3. Regression method and others (Approach based on ROC and CAP curves; Logit and probit model, complementary log-log (CLL) function, cauchit function; Broken curve model; Isotonic regression)





Distribution of units on rating classes for different calibration methods and calibration tests for the whole rating system based on a **master scale KBC bank**

Verification using test for the whole rating system

Method	Spiegelhalter	Spiegelhalter p-value	Hosmer-Lemeshow	Hosmer-Lemeshow p-value	Birer Score	Blochlinger Chi2	Blochlinger p-value
QMM	-0,3921	0,6525	9,2857	0,2328	0,0008	0,9929	0,6087
Skewnormal - MLM1	-0,6614	0,7458	36,4415	0,0000	0,0003	0,7943	0,6722
Skewnormal - MLM2	-0,4636	0,6785	8,7482	0,1882	0,0005	0,7990	0,6706
Skewnormal - MM	-0,3718	0,6450	7,3424	0,1964	0,0006	0,6332	0,7286
Beta - MLM1	-0,6189	0,7320	34,8075	0,0000	0,0003	0,6180	0,7342
Beta - MLM2	-0,3421	0,6339	4,3948	0,4941	0,0009	0,7244	0,6961
Beta - MM1	-0,1602	0,5636	5,3872	0,3705	0,0026	0,9845	0,6112
Beta - MM2	-0,4065	0,6578	5,6195	0,3450	0,0022	0,8965	0,6388
Laplace - MLM1	-0,6560	0,7441	27,3488	0,0000	0,0010	1,4238	0,4907
Laplace - MLM1	-0,6773	0,7509	26,6065	0,0000	0,0010	1,4856	0,4758
Gauss - MLM1	-0,4127	0,6601	9,9288	0,1926	0,0006	0,6840	0,7103
Gauss - MLM1	-0,3903	0,6518	9,8438	0,1976	0,0006	0,6352	0,7279
ROC – MSE	163,3894	0,0000	6641797	0,0000	0,0002	123,1925	0,0000
ROC – AUC	-0,6517	0,7427	60,3700	0,0000	0,0001	8,0220	0,0181
Logit	-0,3861	0,6503	8,1855	0,3165	0,0009	0,9083	0,6350
Logit - Platt	-0,4056	0,6575	8,5176	0,2892	0,0008	0,8725	0,6464
Logit - Box-Cox	-0,2960	0,6164	3,6196	0,7280	0,0013	0,9905	0,6094
Probit	-0,4829	0,6854	15,1820	0,0337	0,0006	0,8206	0,6634
Probit – Platt	-0,4965	0,6902	15,8352	0,0267	0,0005	0,8435	0,6559
Probit - Box-Cox	-0,2144	0,5849	3,6169	0,6058	0,0014	0,8675	0,6481
CLL	-0,3632	0,6418	7,5552	0,3734	0,0010	1,0587	0,5890
CLL - Platt	-0,3846	0,6497	7,4704	0,3816	0,0010	1,1238	0,5701
CLL - Box-Cox	-0,1353	0,5538	4,4086	0,4922	0,0015	0,9481	0,6225
Cauchit	-1,3317	0,9085	23,3999	0,0000	0,0019	4,7105	0,0949
Cauchit – Platt	-1,3336	0,9088	23,3383	0,0000	0,0019	4,8233	0,0897
Cauchit- Box-Cox	-0,5178	0,6977	3,8313	0,5739	0,0010	0,8677	0,6480
Isotonic	-0,3861	0,6503	8,1855	0,3165	0,0009	0,9083	0,6350
Logit (Broken curve)	-0,3937	0,6531	0,0000	1,0000	0,0010	0,8410	0,6567
Logit - Platt (Broken curve)	-0,3965	0,6541	7,6606	0,3635	0,0009	1,0528	0,5907
Logit - Box-Cox (Broken curve)	-0,3010	0,6183	3,4684	0,6282	0,0009	0,6681	0,7160
Probit (Broken curve)	-0,4131	0,6602	16,1469	0,0238	0,0009	1,3129	0,5187
Probit - Platt (Broken curve)	-0,4305	0,6666	16,5331	0,0207	0,0009	1,1963	0,5498
Probit - Box-Cox (Broken curve)	-0,3116	0,6223	3,4907	0,7452	0,0009	0,6617	0,7183
CLL (Broken curve)	-0,3679	0,6435	7,6705	0,3625	0,0009	1,0228	0,5997
CLL - Platt (Broken curve)	-0,3869	0,6506	7,9413	0,3378	0,0009	1,0793	0,5830
CLL - Box-Cox (Broken curve)	-0,2952	0,6161	3,5131	0,7422	0,0009	0,8016	0,6698
Cauchit (Broken curve)	-0,3242	0,6271	20,5298	0,0001	0,0006	1,3413	0,5114
Cauchit - Platt (Broken curve)	-0,3383	0,6324	20,3797	0,0001	0,0006	1,3899	0,4991
Cauchit- Box-Cox (Broken curve)	-0,3491	0,6365	4,9884	0,4173	0,0009	0,9112	0,6341

Distribution of units on rating classes for different calibration methods and calibration tests for the whole rating system based on a **master scale ING bank**

Verification using test for the whole rating system

Method	Spiegelhalter	Spiegelhalter p-value	Hosmer-Lemeshow	Hosmer-Lemeshow p-value	Birer Score	Blochlinger Chi2	Blochlinger p-value
QMM	-0,1322	0,5526	22,4454	0,0328	0,0002	0,2005	0,9046
Skewnormal - MLM1	-0,5891	0,7221	84,8333	0,0000	0,0001	0,2728	0,8725
Skewnormal - MLM2	-0,2972	0,6168	10,4497	0,4020	0,0001	0,2080	0,9012
Skewnormal - MM	-0,1897	0,5752	8,5445	0,4803	0,0001	0,2414	0,8863
Beta - MLM1	-0,5537	0,7101	108	0,0000	0,0001	0,3310	0,8475
Beta - MLM2	-0,0889	0,5354	9,3763	0,4033	0,0003	0,2010	0,9044
Beta - MM1	0,4464	0,3277	24,3487	0,0020	0,0010	0,2922	0,8641
Beta - MM2	0,1112	0,4557	25,4699	0,0025	0,0009	0,2592	0,8784
Laplace - MLM1	-0,2934	0,6154	44,3942	0,0000	0,0001	0,9417	0,6245
Laplace - MLM1	-0,3121	0,6225	40,4449	0,0000	0,0001	0,9879	0,6102
Gauss - MLM1	-0,2089	0,5827	7,9660	0,7163	0,0001	0,2202	0,8957
Gauss - MLM1	-0,1779	0,5706	7,9472	0,7180	0,0001	0,2128	0,8991
ROC – MSE	163,6131	0,0000	143246075	0,0000	0,0002	120,7016	0,0000
ROC – AUC	-0,6358	0,7376	68,1846	0,0000	0,0001	3,1924	0,2027
Logit	-0,1258	0,5500	23,3483	0,0249	0,0002	0,2018	0,9040
Logit - Platt	-0,1497	0,5595	21,4656	0,0440	0,0002	0,1984	0,9056
Logit - Box-Cox	0,0217	0,4914	17,7018	0,0602	0,0003	0,2399	0,8870
Probit	-0,3158	0,6239	132,6172	0,0000	0,0001	0,1735	0,9169
Probit – Platt	-0,3369	0,6319	155,0342	0,0000	0,0001	0,1751	0,9162
Probit - Box-Cox	0,1954	0,4225	19,6716	0,0201	0,0004	0,2643	0,8762
CLL	-0,0581	0,5232	30,1901	0,0015	0,0002	0,2600	0,8781
CLL - Platt	-0,0869	0,5346	31,6973	0,0009	0,0002	0,2503	0,8824
CLL - Box-Cox	0,1245	0,4504	18,1938	0,0330	0,0004	0,2389	0,8874
Cauchit	-0,8665	0,8069	75,8417	0,0000	0,0006	0,5387	0,7639
Cauchit – Platt	-0,8704	0,8080	74,7127	0,0000	0,0006	0,5816	0,7477
Cauchit- Box-Cox	-0,5498	0,7088	29,6169	0,0001	0,0003	0,2567	0,8795
Isotonic	-0,1258	0,5500	23,3483	0,0249	0,0002	0,2018	0,9040
Logit (Broken curve)	-0,1230	0,5489	0,0000	1,0000	0,0003	0,3179	0,8530
Logit - Platt (Broken curve)	-0,0612	0,5244	16,0861	0,1380	0,0001	0,2550	0,8803
Logit - Box-Cox (Broken curve)	-0,0232	0,5092	6,4494	0,8418	0,0001	0,2384	0,8876
Probit (Broken curve)	-0,0805	0,5321	26,9415	0,0292	0,0001	0,2626	0,8770
Probit - Platt (Broken curve)	-0,1025	0,5408	34,4180	0,0030	0,0001	0,2582	0,8789
Probit - Box-Cox (Broken curve)	0,0575	0,4771	3,5298	0,9396	0,0001	0,2880	0,8659
CLL (Broken curve)	-0,0324	0,5129	6,7619	0,8180	0,0001	0,2290	0,8918
CLL - Platt (Broken curve)	-0,0568	0,5226	6,9662	0,8018	0,0001	0,2304	0,8912
CLL - Box-Cox (Broken curve)	0,0708	0,4718	4,1293	0,9027	0,0001	0,4228	0,8095
Cauchit (Broken curve)	-0,1263	0,5503	16,5599	0,0110	0,0001	1,2777	0,5279
Cauchit - Platt (Broken curve)	-0,1433	0,5570	16,3799	0,0119	0,0001	1,2273	0,5414
Cauchit- Box-Cox (Broken curve)	-0,0485	0,5193	6,8013	0,5582	0,0001	0,3755	0,8288

Conclusions

- The use of several tests allows us to take into account different definitions of high-quality calibration. The results obtained were not unambiguous, however they do allow us to answer the basic research question.

  - First, that there are methods which deliver considerably better calibrated estimates of probability in comparison with logistic regression estimators. The difference observed is relatively large and concerns calibration of the system as a whole.

  - The second research question concerned the quality of calibration for different numbers of classes in the master scales. Using four different approaches (master scales including from 9 to 19 classes), whose sources were bank reports on risk, a decrease in quality of calibration of the whole rating system was noted in conjunction with an increase in the number of classes. This dependence is dictated first and foremost by the wider ranges of probability for particular classes, and by the properties of the statistical tests applied.

DZIĘKUJĘ BARDZO!