Predicting bankruptcy in European e-commerce sector

Janda, Karel and Moreira, David

Charles Universisty, University of Economics, Prague

11 October 2016

Predicting bankruptcy in European e-commerce sector

Karel Janda David Moreira***

1 Introduction

In the emergent and fast pace business environment of e-commerce retailers, a predictive analysis about the likelihood of a bankruptcy is crucial for the adoption of the necessary measures to prevent this event. According to Witzany (2009), it is a compromise between the most advanced mathematical modelling techniques and the demand for a practical implementation. From the retail banking perspective, the exposure to financial lending is by far the most important material risk. Teplý and Pečená (2010) state that the credit risk exposure is affected typically by underperforming loans and difficult cash collections especially after the financial crisis period. The financial institutions often rely on credit evaluation models or scorecards to evaluate the credit worthiness of businesses, and to predict the likelihood of a bankruptcy event. These credit scorecards are at the core of bank financial activities because they support the underwriter's decisions when evaluating if the customer is likely to bankrupt or not. Based on this credit underwriting evaluation, the financial institutions decide if it will provide the loan, the related loan interests based on the associated risks, the maturities, and the credit value boundaries.

The aim of this research is to empirically demonstrate with a real case-study, evidences about the shortage of experience within credit underwriters when evaluating e-commerce retailers, and recommend improvements on their credit evaluation scorecards. To address this purpose, the authors developed different econometric techniques such as two-step cluster, logistic regression, classification tree, discriminant analysis, and roc curves, to predict the bankruptcy likelihood of an e-commerce business in Europe. A second purpose of this paper is to compare the accuracy of the econometric classification models and identify which one has the smallest rate of statistical failure, and the greatest easiness of practical use. The predictive models included relevant financial accounting ratios of solvability, productivity, and liquidity from e-commerce retailers registered in Europe, and the population in analyses was collected from the Bureau Van Dijk database (Pagell, Halperin, 1999). The choice of the financial ratios is based on relevant literature and existing credit evaluation scorecards used internally by several financial institutions in Europe.

---

[*] Prof. Ing. Karel Janda, M.A. Dr. Ph.D.: Department of Banking and Insurance, University of Economics, Prague, W. Churchilla 4, 130 67 Praha 3, Czech Republic and Institute of Economic Studies, Charles University, Opletalova 26, 110 00 Praha 1, Czech Republic, Karel-Janda@seznam.cz.

** Ing. David Moreira: Department of Banking and Insurance, University of Economics, Prague, W. Churchilla 4, 130 67 Praha 3, davidmoreiragaia@gmail.com.

[#] This project has received funding from the European Union's Horizon 2020 Research and Innovation Staff Exchange programme under the Marie Sklodowska-Curie grant agreement No 681228.We also acknowledge support from the Czech Science Foundation (grant 15-00036S) and from University of Economic, Prague (institutional support IP100040). Karel Janda acknowledges research support provided during his long-term visits at Australian National University and McGill University. The views expressed in the paper are those of the authors and not necessarily those of our institutions. All remaining errors are solely our responsibility.

From the perspective of the business decision makers, there are many important choices that managers have to do with a direct impact on their operations and financial statements, which may lead either to success or to bankruptcy (Janda, Rojcek, 2014, Janda 2009, 2011). Profit orientated shareholders, along with high leverage requirements, fast growth, and uncertain demand, are typical factors that influence the risk management of e-commerce retailers (Pavel, Sičáková, 2013). Since the bankruptcy probability is usually the most common risk criteria of the financial decision makers, the literature developed predictive algorithms that may be applied in real working practice by credit underwriters, and businesses willing to do credit self-assessment to avoid risk, uncertainty, and the probability of bankruptcy.

The most common statistical approaches to build an internal corporate bankruptcy evaluation are based on static learning models prepared to process large databases, and able to continuously include other economic variables such as industry behaviour, legal changes, or even interest rates (Gama, 2010). In this sense, the authors of this paper recommended in the conclusion part the addition of online metrics to better predict a bankruptcy event of an online retailer. The conceptual definition of bankrupt probability is described by the econometrics models used for classification of businesses into “good” and “bad” classes (Hand and Henley, 1997). The first known credit analysts using a consistent framework to evaluate the creditworthiness of businesses were the underwriters sent to the II World War. When these credit analysts left their companies to join the battle, they created tailor made scorecards and other evaluation requirements to be followed during their absence. In the 1950s, the combination of Bill Fair expertize of business along with Earl Isaac, who was a successful mathematician with forward computing skills, leaded to the creation of the first modern credit score model using advanced predictive statistics. They have used econometric techniques such as logistic regression and multivariate analysis, both applied to predict credit bankruptcy and risk management. When Bill and Earl have created for the first time the credit scorecard in 1958 to predict the creditworthiness of individuals based on their past behaviour, the bank lending industry had a revolutionary upgrade (Rosenberger, Nash, 2009). Further developments have showed that the past track of payment behaviour is a clear factor in classification between performing and non-performing loans. Durand in 1941 brought this methodology to the credit underwriting evaluation after following-up the first approach of differentiation between groups developed by Fisher's studies related to the classification problems of varieties of plants in 1936. Many researches made during the last decades were built up from the work contribution of Durand, regarding the selection of relevant risk indicators. The researches made by Witzany (2010a,b, 2011) described in the works about credit scoring functions (2011), and credit risk management modelling (2010a,b), were important contributions to the development of the predictive bankruptcy topic developed in this paper. Jakubík and Teplý (2011) constructed a new indicator named the JT index that evaluates the economy's financial stability and is based on a financial scoring model estimated on Czech corporate accounting data.

The financial wealth evaluation of an e-commerce retailer, or the prediction of its financial bankruptcy, is commonly based on the econometric methodologies of discriminant analysis and logistic regression. The methods of discriminant analysis subdivided the dimensions of univariate and multivariate modelling. William Beaver in 1968 has published a study developing a univariate model using one variable to classify the failure of a company when any of the following situations happened: inexistence of funds to pay stock dividends, bank account overdrawn, bond default, and bankruptcy. The research made by William Beaver compared a group of firms that failed with another group of successful firms within the same industry cluster. The predictions and forecasts of the financial failure were calculated taking in consideration ratios of cash-flow/total debt, net income/total assets (return on assets), and total debt/total assets (debt ratio). The multivariate model to predict bankruptcies was developed by Eduard I. Altman and Thomas P. McGough, and was based on five financial ratios weighted to maximize the predictive accuracy of the model producing a discriminant analysis score named “Z score”. With the beginning of the fast increase of credit and debit card users, it was necessary to develop an integrated decision making system made by computing processes. At that time, the credit scorecard was proving to be more accurate predictor of bankruptcy than any other judgemental process based on opinions. The linear programming and logistic regression were applied by Steenackers and Goovaersts during the 1980s into the credit evaluation and since then they turned to become one of the most used econometric models in financial institutions. During the last years, many machine learning and artificial intelligence techniques have been also applied to build and predict the credit bankruptcy models such as neural networks, nonparametric statistical models, K-nearest neighbour, and classification trees (Thomas, 2000). Currently other powerful systems and intensive research is being done using advance classification algorithms such as Bayesian network classifiers, support vector machines and colony algorithms (Crook, 2007, Henley, 1997). The regression and multivariate adaptive regression splines are also being included in some related researches.

In this paper we conduct the research using algorithms that predict a bankruptcy event and build explanatory decision support models to evaluate the financial worthiness of an e-commerce retail company in Europe. Such a new approach to a modelling of complex systems is provided for example by Stádník and Miečienskiené (2015). These decision support models are organized in tables and developed with the outcomes of the statistical models of clustering, logistic regression, discriminant analysis, classification trees, and roc curves.

Internet companies use mainly the worldwide computer network to carry on their business. They are currently more than doubling in size year over year, and transforming the communications and the commerce establishments. The unique capabilities of the internet and the respective e-commerce advances have led to a redesign of almost all related industry standards and international legal framework. The internet retailers have exclusive web structures with an unlimited potential to reach customers. With the development of the e-commerce field the global economy can offer now more products and services to companies and individuals, while contributing to the growth of supply chain networks, communications, and flow of capitals. With the related technologic advances, the online international transactions and payments are more secure and they are leading customers to shop without the typical brick and mortar constrains. Along with the development of internet, there are existing unique and new strategic alliances between different high-tech companies supplying online metrics, marketing, products, or services. Therefore the interconnection and cooperation among other related parties became a common practice. The European e-commerce sector has been characterized by an average business growth of more than 30% annually however most of the online retailers are under pressure to start making higher profits. To survive in this competitive environment they are driven to cut costs, to minimize expenses, to strive for sustainable expansion, and to forecast realistic business plans. To address these business survival requirements, internet companies have to obtain the capacity to access finance, to mitigate the perception of being an industry with higher financial risks, and to be able to prove the creditworthiness. Taking in consideration the evaluation assessments made by the credit underwriters, it is very important to keep the track of all the accounting ratios that typically make part of a credit scorecard evaluation. In the aftermath of the financial crisis in 2009, the improved expectations about the future direction that the economy was following (Teply, Tripe, 2015), triggered the inexperienced bank underwriters and credit evaluators to misleading assessments between solvent and insolvent online companies and credit risks. In the internet e-commerce environment, many online businesses are likely to have profitable projects and forward technological plans able to prospectively pay-off their liabilities. Notwithstanding, the lenders who had short time experience in providing credit lines to these online firms, are being very cautious and prudent avoiding any sort of uncertainty. The shift from the slower traditional brick & mortar retailers to the fast growing online field has leaded the underwriters towards a perception of high risk lending (Enoch, Green, 1997). The deterioration of e-commerce companies is often caused by their exponential growth and subsequently their incapacity to prove their solvency trustworthiness, and feed the operations with increasing access to higher credit limits. One of the main factors that are binding internet firms to access finance at the pace they commonly need is the lack of experienced specialized underwriters in this sector of financial and bank activity. Related to this topic, many researchers identified that the deterioration of the credit evaluation standards before the financial crisis was caused by the absence of underwriters able to evaluate accordingly the existing financial risk, permitting a large number of retail loans to be approved without enough certainty. The current deficient situation among the e-commerce businesses accessing finance is also caused largely by the same reason, however now the evaluators narrowed the access to finance requesting large collaterals to back the loans, along with high interest rates. The e-commerce companies have more difficulty to compete against the offline retailers often because their poor financial evaluation is a reflex of unskilled credit underwriters unable to distinguish between a good or a bad creditor, creating a situation of a loan non-performing sector. The banking expansion strategy in Europe is being followed by credit evaluation teams inherited from the periods of time when the computational information systems were poor. Therefore many loans are not being approved or extended due being oddly evaluated (Lucas, 2004).

2 Findings and evidences from retail banking and credit insurance field

The authors of this paper reviewed several credit scoring reports released by financial institutions and credit insurance companies. The credit reports in review were developed by the following credit financial institutions: HSBC, Dun & Bradstreet, Bisnode, Euler Hermes, Atradius, Credito Y Caucion, Credit Reform, OeKB. The scorecards in review were generally structured by:

- Company historical information – identification, registration, ownership, stock capital, turnover, number of employees, activities, management, board administration.
- Payment behaviour - annually, periodically.
- Financial statements – comparison year-over-year, and ratios of profitability, liquidity, activity, leverage.
- Negative business history – insolvencies, liquidation, executions, debts.
- Observations – press releases, sale of equity participation (among other specific events).

The most typical approach is holistic rather than statically quantitative, as it was identified by the authors of this paper when observing the credit reports released by these financial institutions. These credit evaluation models were in general static and did not adapt when the population in analyse changed, which means that the parameters of credit evaluation do not adjust according to the industry (Thomas, 2010). This misclassification may be responsible for later increasing costs due to potential credit delinquency or to healthy loans not given to costumers. The credit evaluators generally were relying on out-of-date population that lacked meaningful response, target, or data with quality. This approach led to a lack of confidence when analysing a population sample that should be diverse enough to represent different types of repayment behaviour.

3 Research methodology and database

The methodology considered in this research includes different statistical approaches applied in the credit evaluation field such as two steps cluster, logistic regression, discriminant analysis, decision tree, and ROC curve. We also evaluated the best fitting approach that may be more suitable according to each evaluation determinants. The research summarized in this paper gathers a combination of detailed information related to accounting and financial ratios collected in Bureau van Dijk's (Amadeus) database (Klapper, Allende, Sulla, 2002). There exist a large number of papers about the credit assessments of brick and mortar retailers, however just few attempts were made about European e-commerce companies as for example the successful billion euro e-businesses Asos, Zalando, Vente Privee, among others.

The data used to model the statistical approaches was collected from the last available financial and accounting ratios at the database till the year 2014. To each e-commerce company was assigned a bankruptcy outcome of "0" if not bankrupted, and "1" if bankrupted. The rate of bankruptcy within this sample is 48%, and the number of European companies making part of the dataset is 124. The table 1 contains the explanation of each individual variable referring to the accounting ratios of solvency, and profitability. The framework of this methodology evolves a unidimensional analysis with a binary output predicting the bankruptcy outlook of e-commerce retailers in Europe. Our approach includes different classification experiments to model and predict this probability. We have considered the statistical classification method named twostep cluster analyses to explore within the dataset which were the predictors with more importance and contribution to the model. This method clusters the similarity and statistical power between the predictors supporting the authors in the selection of the independent variables. The other statistical classification approaches included in the paper and also widely accepted in financial credit field were logistic regression, decision tree, discriminant analyses, and ROC curves to measure the accuracy of the models. The following table describes financial calculation of the predictive variables applied in the statistical methods developed in this paper.

Tab. 1: Predictive variables

X	Variable explanation	Calculation in %
X_1	Solvency ratio asset based	$Solvency\ Asset = \frac{Shareholders\ funds}{Total\ assets} * 100$
X_2	Return on capital employed	$ROCE\ (P\&L) = \frac{P\&L\ b.\ tax + Interest\ paid}{Shareholders\ funds + Non-current\ liabilities} * 100$
X_3	Return on equity using P&L	$ROE\ (P\&L) = \frac{P\&L\ before\ tax}{Shareholders\ funds} * 100$

X_4	Return on assets using P&L	$ROA\ (P\&L) = \frac{P\&L\ before\ tax}{Total\ assets} * 100$
X_5	Return on equity using net income	$ROE\ I. = \frac{P\&L\ period\ (net\ income)}{Shareholders\ funds} * 100$
X_6	Return on capital employed (income)	$ROCE\ I. = \frac{P\&L\ for\ period + Interest\ paid}{Shareholders\ funds + Non-current\ liabilities} * 100$
X_7	ROA return on assets using net income	$ROA\ I. = \frac{P\&L\ for\ period\ (net\ income)}{Total\ assets} * 100$
X_8	Profit margin	$Profit\ M. = \frac{P\&L\ before\ tax}{Operating\ revenue\ (turnover)} * 100$
X_9	Gross margin	$Gross\ M. = \frac{Gross\ profit}{Operating\ revenue\ (turnover)} * 100$
X_{10}	EBITDA margin	$EBITDA\ M. = \frac{Operating\ P\&L + Depreciations\ \&\ Amortizations}{Operating\ revenue\ (turnover)} * 100$
X_{11}	EBIT earning before interests and taxes	$EBIT\ M. = \frac{Operating\ P\&L\ (EBIT)}{Operating\ revenue\ (turnover)} * 100$
X_{12}	Cash-flow and operational revenue	$CF\ O.R. = \frac{Cash\ flow}{Operating\ revenue\ (turnover)} * 100$
X_{13}	Solvency ratio liability based in %	$\text{Solvency L.} = \frac{Shareholders\ funds}{(Non-current\ liabilities + Current\ liabilities)} * 100$
X_{14}	Cost of employees and operational revenue	$\text{Cost E.} = \frac{Cost\ of\ employees}{Operating\ revenue\ (turnover)} * 100$

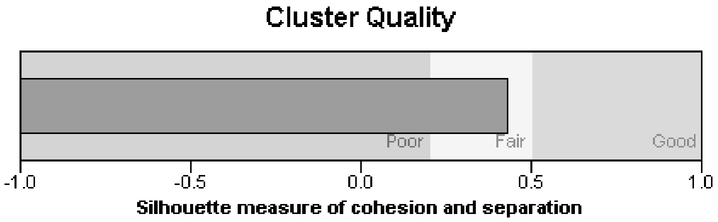
Source: Bureau Van Dijk database

4 The statistical findings are summarized in this section with an interpretation of the correspondent observations.

4.1 Twostep Cluster

The bankruptcy grouping and segmentation observed in the sample were analysed using the twostep cluster method. The main advantage of this procedure is that it does not require a matrix of distances between all pairs of cases and rather requires only categorical or continuous data. It produces solutions based on a mixture of different variables for different number of clusters. The predictors were also clustered by hierarchical degree of importance (Terziovski, 2008). Through this process, the financial ratios (clusters) were compared and it was possible to identify which predictor had the higher contribute to the model. Summing up the outcomes, the strongest variables were the financial ratios ROA using P&L before tax, EBIT margin, and Profit. The variables that had the least importance to the clustered model were Solvency, ROE using P&L, and in the last place ROCE using P&L.

Fig. 1: Twostep Cluster – measure of the quality of predictors



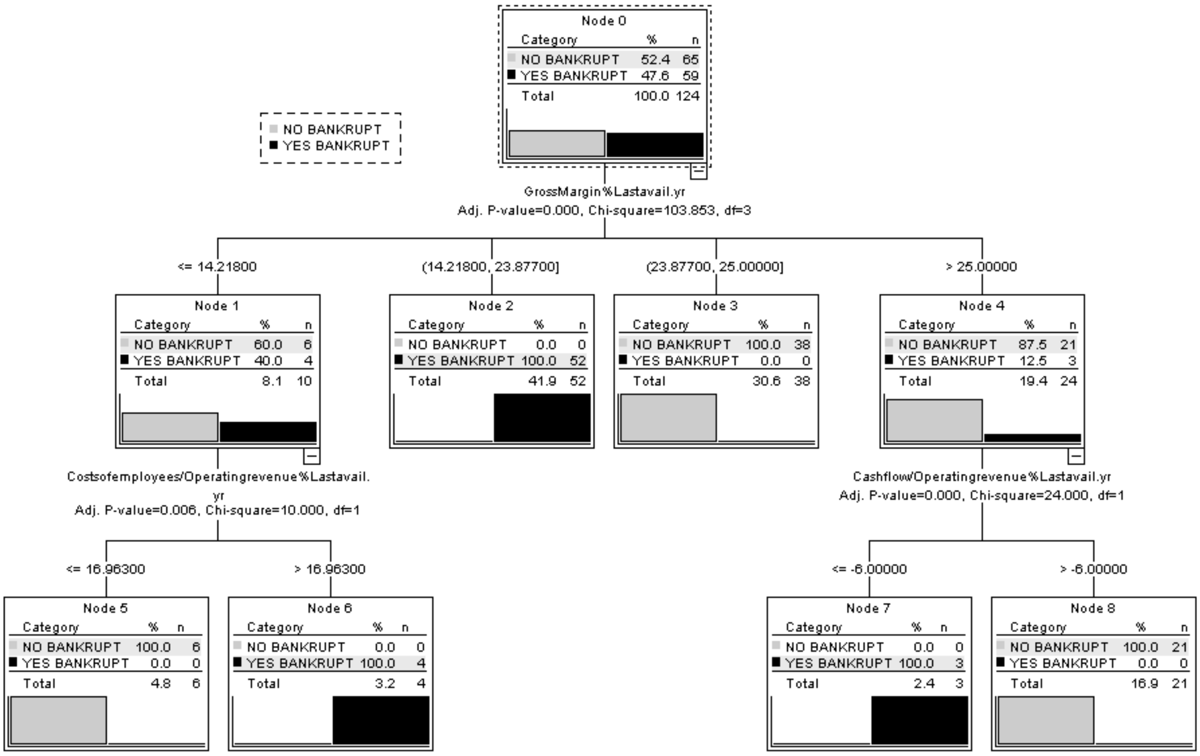
Source: authors

Overall this approach was fairly successful because it was able to prove statistically the significance of the variables to the binary outcome *Yes bankrupt* and *No bankrupt*. Furthermore the hierarchical clustering of the predictors guided the authors in ranking their significances to the bankruptcy predictions.

4.2 Decision tree

The datamining model decision tree was applied to predict the bankruptcy of internet e-commerce companies and it led to a bankruptcy classification cases into groups, and predicted outcomes through a ramification of independent values (financial ratios). This decision tree model may have real working practice application to categorize the online companies according to whether or not they represent a financial risk. This model validates tools for confirmatory and exploratory classification analysis. It provides features that allow the identification of homogeneous groups of predictors (variables) with high or low risk of bankruptcy (Janda and Rakicova, 2014), and makes easier the prediction of bankruptcies of internet businesses. The method applied to develop the decision tree and to evaluate the credit bankruptcy probability is named Chi-squared automatic interaction detection (CHAID). The advantage of using the CHAID approach is that at each step it chooses the financial ratio (independent variable or predictor) that has the strongest interaction with the bankruptcy probability (dependent variable). In the case that the categories of the financial ratios are not significantly different relatively to the probability of bankruptcy (dependent variable), the CHAID method merges them. Although 15 financial ratios (independent variables) were introduced in the model, at the end only three were included in the final version of the decision tree model. The variables gross margin, cost of employees/operating revenue, and cash-flow/operating revenue were the only ones being included by the model, mainly because they had a significant contribution to the decision tree classification model.

Fig. 2: Classification Decision Tree



Source: authors

The CHAID process demonstrated that the ratio of gross margin is the best predictor of bankruptcy. Within the range between 14.218 % and 25.000 % of gross income category, the node 3 and 4 had just the node 0 as the only significant predictor of bankruptcy therefore they were considered the terminal nodes. The node 1 corresponds to the lowest gross margin category and its best predictors are the nodes 5 and 6, which are related to the ratio cost of employees/operation revenue. The node 4 is related to the category of e-commerce companies with a gross margin percentage of more than 25.000 %. The node 4 classifies the majority of 87.5% businesses that did not bankrupt, and its best predictor of bankruptcy is the category cash-flow/operating revenue although representing solely 12.5% of companies bankrupted. Considering that the tree was created with the CHAID method, the chi-square values having the highest significance level are the ones with less than 0.0001 % value on all branches in the model. In this respect, only the category cost of employees/operation revenue has a chi-square with a value of 0.006 therefore with a lower significance. Observing the classification results is possible to evaluate that the model works perfectly with the standard error at the value less than 0.0001, and with an overall classification fitting of 100%.

4.3 Logistic Model

The logistic model is extensively applied in the financial field to classify subjects based on values of a set of predictor variables, as in the model proposed by Witzany (2009) to estimate the Loss Given Default (LGD) correlation. The predictive probability of bankruptcy using the logistic model is expressed by this formula:

$$P(X) = \frac{1}{1 + e^{-(\propto + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots, \beta_\infty X_\infty}} \quad (1)$$

The terms ∞ and β_i in this model represent unknown parameters that the authors estimated based on the dataset obtained from the Bureau Van Dijk on the X_s and on *P(X)* (X=probability of bankruptcy). Taking this in consideration, a credit underwriter or an e-commerce company willing to make a self-assessment could use the determined values of X_1 through X_k for a prediction of bankruptcy probability. Thus, if an individual responsible for predicting the bankruptcy probability in the online retail field would start applying this logistic model, it would just have to change these X_i to its own financial ratios to obtain the desired bankruptcy likelihood.

Tab. 2: Logistic predictors applied in the logarithm

∞ = 3.768	*P(X)* = Probability of bankruptcy of e-commerce retailers			
X_1 = -0.009	X_3 = 0.017	X_6 = 0.011	X_9 = -0.150	X_{12} = -1.041
X_2 = -0.002	X_4 = -0.413	X_7 = 0.0420	X_{10} = 0.693	X_{13} = -0.010
X_3 = 0.017	X_5 = -0.009	X_8 = 0.555	X_{11} = -0.545	X_{14} = -0.006

Source: authors

The bankruptcy probability in this model is coded as 1 if bankrupted and 0 if a company did not bankrupt. There are 15 independent variables included in this logistic model which are the estimation predictors of the unknown parameters. Since the probability of bankruptcy must lie within 0 and 1 the final result is a probability between 1% and 100%. The test applied to the model to assess its statistical fit was the Hosmer-Lemeshow goodness-of-fit. Furthermore was also made a diagnostic using a residual plot to observe the change in deviance versus the predictable probabilities, and the Cook's test to observe the distances versus predicted probabilities. The Hosmer-Lemeshow statistical analysis is considered to have a good fit if the significance value is more than 0.05, therefore as this test had the outcome value of 0.301 the model is considered adequately fitting the data. The statistical test Cox & Snell R^2 based on the log likelihood has a result of 0.538, and the Nagelkerke R^2 test of parameter estimates has the value of 0.718. These two results suggest that the dataset included in the model is useful to explain the bankruptcy probability. The Wald statistics identified the following individual parameters $X_4, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}$, as the most significant to the model. The use of this logistic model to predict the probability of bankruptcy whether by credit underwriters or by e-commerce companies is very accurate and convenient, however it requires at least an entrance level of accounting, financial, and statistical knowledge to successfully be applied.

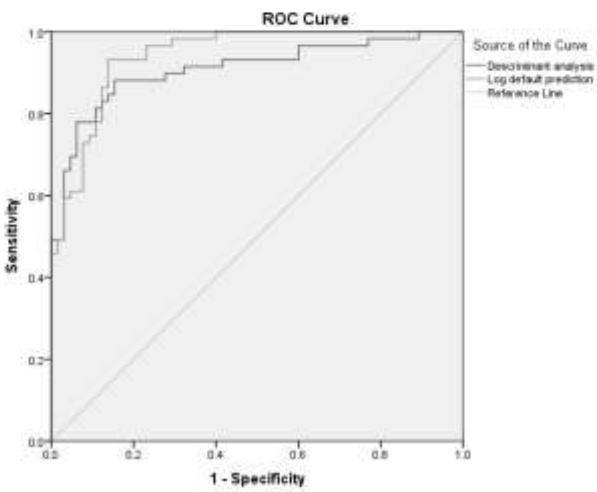
4.4 Discriminant Analysis

The discriminant analysis is a classification approach that models the value of a dependent categorical variable, based on its relationship with the predictors. It provides a powerful technique for examining differences between two or more groups of objects and several variables simultaneously. In the financial field, credit underwriters may use discriminant analysis to predict or explain which companies will be likely to bankrupt or not. The discriminant functions are combinations of the predictors and have the following formula expression:

$$d_{ik} = b_{0k} + b_{1k}x_{i1} + \ldots + b_{pk}x_{ip} \tag{2}$$

The predictive analysis was made using the IBM SPSS statistical software, and its first step was the assignment of a first function that divided the e-commerce companies into groups. There are several assumptions that have to be met to proceed with the use of a discriminant analysis such as: the predictors cannot be highly correlated, a normal distribution curve, the correlation between the predictors is constant, the variance and mean of a predictor are also not correlated. The Discriminant analysis assigned cases to the two groups, and for each case a classification grading is attributed individually by a hierarchical rank. The e-commerce companies which had the largest solvency ratio are the ones less likely to bankrupt. The tested values of solvency, ROCE (P&L), ROA (P&L), ROE, profit, gross margin, EBITDA, an EBIT, indicate that as higher are these financial ratios as wealthier are likely to be the e-commerce companies, and as less likely are to fall into bankruptcy. In the same fashion the opposite also holds true. The authors tested the equality of group means to display the outcomes of the unidirectional ANOVA method addressing each financial ratio as the factor. In this test, the maximum threshold of significance is the value of 0.10, therefore if the independent variable is lower than 0.10, this means that it has contributed to the discriminant model. The Wilks Lambda test measured the potential contribute of a financial ratio to the model. The smaller results indicate greater discrimination between the groups moreover the ratios of solvency, profit and margin were the ones with better performance. The results of the classification model show that 51 out of 59 e-commerce companies that bankrupted were correctly classified, corresponding to a successful percentage of 86.4%. By other hand 55 out of 65 companies that did not bankrupt were classified correctly at the level of 84.6%. Considering the overall statistical evaluation, the classification predicted correctly the bankruptcy of e-commerce retailers at the level of 85.5%.

4.5 ROC curves – comparison of predictive accuracy between the models logistic regression and discriminant analyses



The roc curve is a powerful technique to assess the performance of classification procedures of the two categories (bankrupt and not bankrupt), by which the subjects are classified. A credit underwriter willing to accurately classify e-commerce companies into groups, may apply the roc curve to evaluate the performance of these predictive classifications. Observing the distances from the central reference line, the discriminant analysis and the logistic binomial model are both performing well in terms of bankruptcy prediction. The comparison between these roc curves show that the logistic regression slightly performs better than the classification. The area under the curve for both models is significant because they have an asymptotic value of less than 0.05 therefore they are both correctly predicting the bankruptcy of e-commerce companies. Analysing the confidence test of the asymptotic interval at 95% level is possible to observe that the discriminant analysis is inferior to the logistic regression because the amplitude of its interval has comparatively a lower percentage of lower bound and upper bound. The areas under the curve of this graph provide guidance to the cut-off which determines if the results are positive or negative. This test is commonly used in finance to test the overall performance of a classifier on training datasets. The best performing models have the curve at the top left corner therefore the larger is the curve amplitude the better.

5 Conclusion

The emergent sector of e-commerce retail is booming all around the world. The new paradigm of retail businesses is shifting from a brick & mortar concept to an online mode denominated as e-commerce. Typically the online companies are characterized by fast growth rates, large web reach, technologic drive, intangible assets, and strong need of capital to fund the operations. Observing this reality from the financial perspective makes it possible to acknowledge that the access to finance is being constrained by the lack of experience among the credit underwriter's that did not follow the developments in this uprising online field.

The credit evaluation models rely often on out-of-date criteria and basic accounting ratios without taking in consideration the specificities associated with e-commerce businesses such as the online performance metrics. We collected also evidences that several credit scorecards used by large financial institutions do not consider econometric models to evaluate the creditworthiness of these companies. This paper addresses these findings developing statistical models and tests such as clustering twostep model, decision tree, logistic regression, discriminant analysis, and ROC curve. The predictive models may be applied in real working practice cases whether by credit underwriters, or by business users to predict the likelihood of a bankruptcy event. These classification models in general successfully predicted the e-commerce companies that bankrupted at the level of more than 85% however there are differences between the final outcomes, and the related easiness of its use. The decision tree is the model easier to interpret graphically and to use in reality, however lacks detail in terms of credit assessment. The logistic regression performed slightly better than the classification analysis when analysed by the roc curve test, and in terms of real practice application the predictive logistic logarithm have showed also high accuracy with the training dataset. Although the methods have different fitting characteristics, the logistic regression and the linear discriminant analysis are producing both linear decision thresholds.

The main differences between these two approaches are related to the assumption that in the linear discriminant models the results are observed through a Gaussian distribution with a similar covariance matrix in each class achieving this way a better performance comparatively to the logistic regression. Notwithstanding, in the case that the assumptions of Gaussian distribution are not met the logistic regression outperform the other model also because it is applicable to a wider range of financial research situations.

A possibility to extend the studies on this topic is to include a wider population of e-commerce companies, and to develop a new credit evaluation scorecard including the online metrics recommended in this paper.

5.1 Recommendations addressed to credit underwriters to improve the financial evaluation of e-commerce companies in Europe:

- Inclusion of statistical classification analyses to accurately predict the likelihood of a bankruptcy event. The holistic approach is proven to be inefficient.
- Develop a tailor-made credit evaluation scorecard including also online metrics such as: bounce rate, visitors, conversion rate, page views, time-on-site, returning visitors, average basket value, numbers of orders, online revenue, and number of transactions.

Acknowledgement

This project has received funding from the European Union's Horizon 2020 Research and Innovation Staff Exchange programme under the Marie Sklodowska-Curie grant agreement No 681228.We also acknowledge support from the Czech Science Foundation (grant 15-00036S) and from University of Economic, Prague (institutional support IP100040). Karel Janda acknowledges research support provided during his long-term visits at Australian National University and McGill University. The views expressed in the paper are those of the authors and not necessarily those of our institutions. All remaining errors are solely our responsibility.

Predicting bankruptcy in European e-commerce sector

Summary

In the current competitive and uncertain e-commerce environment, businesses have the need to predict in advance their likelihood of falling into bankruptcy. The central focus of this paper is to statistically model through different approaches the bankruptcy probability of e-commerce companies in Europe. The authors examine the econometric techniques twostep cluster, logistic regression, discriminant analysis, data mining tree, and roc curves to correctly classify these companies into bankrupt and not bankrupt. The paper finds evidences about the current credit underwriting inexperience among several financial institutions. The classification approaches included in this paper may be applied in real working practice whether by credit underwriters or by business decision makers. The research was developed using financial and accounting information available in the Bureau Van Dijk database. The

paper suggests further analytical developments in the field of predictive bankruptcies, and recommends improvements on the credit evaluation scorecards such as the inclusion of advanced online metrics to increase the accuracy of the creditworthiness evaluation of an e-commerce company.

Key words: e-commerce, Europe, bankruptcy, econometrics, prediction
JEL classification: G33

References

Beaver, Correia, McNichols (2011): Financial Statement Analysis and the Prediction of Financial Distress. Now Publishers Inc (2011)

Bielikova, Banyiova, Piterkova, (2014): Prediction Techniques of Agriculture Enterprises Failure. Procedia Economics and Finance. ECE, 2014

Camp, Filipe, Hammoudi, Piattini (2006): Enterprise Information Systems. Springer Science & Business Media, 2006

Enoch, Green, (1997): Banking Soundness and Monetary Policies. IMF Institute International Monetary Fund, 1997

Gibson (2010): Financial Reporting and Analysis: Using Financial Accounting Information. Cengage Learning. 06/04/2010, pp. 469-472

Gama, (2010): Knowledge discovery from data streams, London, Chapman & Hall. 2010

Gibson, (2010): Financial Reporting and Analysis: Using Financial Accounting Information, Cengage Learning, 2010

Jan Pavel, Emília Sičáková, (2013): Do E-Auctions Realy Improve the Efficiency of Public Procurement? The Case of the Slovak Municipalities. Prague economic papers : quarterly journal of economic theory and policy. 2013. v. 22, no. 1, p. 111--124. ISSN 1210-0455.

Jakubík, Petr -- Teplý, Petr (2011): The JT index as an indicator of financial stability of corporate sector. Prague Economic Papers. 2011. v. 20, no. 2, p. 157-176, 20 ISSN 1210-0455.

Janda,Karel --- Jakub Rojcek, (2014), Bankruptcy Triggering Asset Value: Continuous Time Finance Approach. In: Pinto, A A. - Zilbermand, D. Modeling, Dynamics, Optimization and Bioeconomics I. Springer Link, 2014. p. 357--382. ISBN 978-3-319-04849-9

Janda, Karel -- Rakićová, Anna (2014): Corporate Bankruptcies in Czech Republic, Slovakia, Croatia and Serbia. Mnichov. 2014.

Janda, Karel (2011). Inefficient Credit Rationing and Public Support of Commercial Credit Provision. Journal of Institutional and Theoretical Economics, vol. 167 no. 2, pages 371-391.

Janda, Karel (2009). Bankruptcies With Soft Budget Constraint. Manchester School, vol. 77, no.4, pages 430-460.

Klapper, Allende, Sulla, (2002): Small- and Medium-Size Enterprise Financing in Eastern Europe. World Bank eLibrary, 2002

Lucas, (2004): Updating scorecards: removing the mystique. Readings in Credit Scoring: Foundations, Developments, and Aims. Oxford University: New York, 2014, pp. 93-109.

Pagell, Halperin, (1999): International Business Information: How to Find It, How to Use It, Global Professional Publishi, 1999, page 144

Rosenberger, Nash, (1946): The Deciding Factor: The Power of Analytics to Make Every Decision a Winner. Wiley, pp. 7-8

Tagarelli, Andrea (2012): XML Data Mining: Models, Methods, and Applications: United States of America. Information Science Reference. 2012 pp. 471-473

Stádník, B. Miečienskiené, A. (2015): Complex Model of Market Price Development and its Simulation, Journal of Business Economics and Management, Volume 16, Issue 4, 2015, pp: 786-807

Teplý, Petr -- Pečená, Magda Credit risk and financial crises. Prague: Karolinum Press, 2010. ISBN 978-80-246-1872-2.

Teplý, Petr -- Tripe, David (2015): The TT Index as an Indicator of Macroeconomic vulnerability of EU new Member States. Ekonomický časopis = Journal of economics. 2015. v. 63, no. 1, p. 19--33. ISSN 0013-3035.

Terziovski, (2008): Energizing Management Through Innovation and Entrepreneurship: European Research and Practice. Routledge, 2008.

Thomas, (2010): Consumer finance: challenges for operational research. Journal of the Operational Research Society. 2010, pp.41-52.

Thomas, Edelman, Crook, (2002): Credit Scoring and Its Applications. SIAM, 2002

Witzany, Jiří (2010a): Credit risk management and modelling. Praha: Oeconomica, 2010. ISBN 978-80-245-1682-0.

Witzany, Jiri. (2010b) On Deficiencies and Possible Improvements of the Basel II Unexpected Loss Single-Factor Model. Czech Journal of Economics and Finance, Vol. 60, No. 3, Pages: 252–268.

Witzany, Jiří (2011): Definition of Default and Quality of Scoring Functions. Bulletin of the Czech Econometric Society. 2011. v. 18, no. 28, p. 33--52. ISSN 1212-074X.

Witzany, Jiří (2009): Estimating LGD Correlation. Financial Risk Management. 2010. v. VII, no. 4, p. 73--83. ISSN 0972-916X.

Witzany, Jiří (2009): Loss, Default and Loss Given Default Modeling. IES Working Paper. 2009. no. 9. URL: http://ies.fsv.cuni.cz/sci/publication/show/id/3650/lang/en, P. 1.