VĚDECKÉ SPISY VYSOKÉHO UČENÍ TECHNICKÉHO V BRNĚ

Edice Habilitační a inaugurační spisy, sv. 709

ISSN 1213-418X

Michal Karas

UTILIZING ACCOUNTING AND MACROECONOMIC VARIABLES IN THE PREDICTION OF SMES DEFAULT

BRNO UNIVERSITY OF TECHNOLOGY
Faculty of Business and Management
Department of Finance

Ing. Michal Karas, Ph.D.

UTILIZING ACCOUNTING AND MACROECONOMIC VARIABLES IN THE PREDICTION OF SMES DEFAULT

VYUŽITÍ ÚČETNÍCH A MAKROEKONOMICKÝCH UKAZATELŮ V PREDIKCI ÚPADKU MSP

HABILITATION THESIS SUMMARY
COURSE: EKONOMIKA A MANAGEMENT

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ NAKLADATELSTVÍ VUTIUM

BRNO 2021

KEYWORDS
small and medium enterprises; default prediction; Cox model; macroeconomic variables; financial ratios

KLÍČOVÁ SLOVA
malé a střední podniky; predikce úpadku; Coxův model; makroekonomické ukazatele; finanční ukazatele

MÍSTO ULOŽENÍ
Vysoké učení technické v Brně, Fakulta podnikatelská, Oddělení pro tvůrčí činnost a doktorské stadium, Kolejní 2906/4, 612 00 Brno

ISBN 80-214-6010-2
ISSN 1213-418X

CONTENTS

1 ABOUT AUTHOR

Michal Karas is a Professor Assistant in the Department of Finance at Faculty of Business and Management Brno University of Technology, where he specialises in financial management and corporate default prediction models. He received his master's degree in corporate finance in 2010 from the Faculty of Faculty of Business and Management, Brno University of Technology and his PhD in Corporate finance from the Faculty of Business and Management, Brno University of Technology in 2013. He focuses on corporate default prediction model and related issues. His research primarily aims to understand what drives corporate default and how the threat of the default could be spotted in financial statements and above that role is played by environmental factors.

Michal Karas participated in 2 projects supported by Technology Agency of the Czech Republic.

His H-index according to Web of Science is 7. His work has been published in one monograph (*Predicting Corporate Default and Mergers and Acquisitions Success*, published in 2021) and a number of scientific journals (5 papers published in scientific journal with impact factor – Thomson Reuters; 12 papers published in scientific journals indexed in Scopus).

He serves on the Editorial Board of two international scientific journals: *Financial and credit systems: prospects for development (ISSN 2310-8770)* and *Research on Enterprise in Modern Economy theory and practice (ISSN 2084-6495).* He develops and teaches a course on *Financial management of small company*, covering topic of financing business, preparing and evaluating financial plan, for bachelor degree students as well as *Financial management, Rating and business valuation* for master degree students. He supervised 134 master and bachelor thesis, one of supervised master thesis on topic of Predicting corporate default was awarded by CAFIN (Czech association for financial management) – Diploma thesis of the year competition and Atlas Copco – Atlas Copco Diploma thesis competition.

2 INTRODUCTION

The prediction of small and medium enterprise (hereinafter referred as SMEs) default represents a research gap in the current state of the art in the default prediction literature. This gap is a consequence of the opinion that the default prediction model can be effectively applied for prediction default in case of different business segments, periods and industry branches. Many studies show that this is not the case and thus it drives the effort to create new models, while there are also studies claiming the opposite. Predicting SME default is one of these issues. The need to adopt a special approach for the SME segment in assessing default risk is especially obvious from the limited access of SMEs to external funds, which affects their capital structure, related working capital issues and investment decisions. This limited access can be considered a market failure since credit is not provided to otherwise financially healthy businesses. This results from the application of the same metrics for SMEs and large businesses, leading to inappropriate assessment of the related credit risk. A better understanding of the default risk factor of SMEs could help with the adoption of policies to alleviate this unfavourable situation.

The aim of this work is to derive a default prediction model for SMEs combining macroeconomic and firm-specific (mainly accounting) types of indicator and assess the importance of incorporating macroeconomic factors for prediction purposes. New prediction models for SMEs should reflect the specifics of this segment of businesses, while not only adopting traditional approaches on a sample of SMEs. The traditional approach of utilising accounting data as the main sources of information for assessing default risk seems to be of limited potential, and this potential seems already exhausted by existing studies. A quest for potentially utilisable information for prediction purposes must lead towards different ways, while respecting the SME segment specifics, such as lack of capital market information. On the other hand, utilising information from the external environment seems promising. There have been attempts to employ both macroeconomic indicators together with accounting information in the prediction of SMEs default but the approach in this work differs. The macroeconomic variables were adopted in a different and more flexible manner. The current approach in addressing the issue is to adopt the macroeconomic indicator as a baseline hazard rate. Such an approach allows the utilising of only one macroeconomic indicator at a time. When utilising more indicators, an artificial indicator must be formed. The novelty of the approach adopted in this work lies in taking the advantage of the Cox regression model's specific feature, that the estimation of the model's coefficient is possible even if the baseline hazard rate is left unspecified. In such an approach, the macroeconomic indicators could enter the model as independent variables, while the baseline hazard rate is left unspecified. Such an application, however, cannot be meaningfully adopted for a single country dataset, which was the case of other authors' works and therefore, the research presented throughout this work is focused on a panel of 28 countries. Focusing on such a panel not only allows the application, but also leads to obtaining sufficient variability of macroeconomic data and consequently benefits model robustness.

The work is organised as follows. First, a review of the current state of the art in the prediction model is presented. Later, the aim of the work, together with the research hypotheses and the methodology adopted to verify the research hypothesis, is described. A description of the methods employed in the presented research follows. Afterwards, the results are presented, followed by hypothesis verification, discussion of the results, while the work ends with the conclusion and notes on the contribution of the thesis.

3 THE NEXUS BETWEEN MACROECONOMIC DEVELOPMENT AND BUSINESS DEFAULT

This chapter deals with the results of a literature review on the direct and indirect relationships between macroeconomic conditions and various company default prediction issues. In particular it looks at the relationship between macroeconomic conditions and the probability of default, or rather the macroeconomic conditions and their influence on the firm-level determinants of default. Attention is also paid to a strand of literature suggesting that model effectiveness is bound by the prevailing macroeconomic conditions, while under alternative conditions its effectiveness is degraded. The rest of the chapter is dedicated to the application of macroeconomic factors in predicting business default, with a special focus to SME applications.

3.1 THE RELATIONSHIP BETWEEN PROBABILITY OF DEFAULT AND MACROECONOMIC CONDITIONS

From a general perspective, Allen Saunders (2004), who provided a literature overview on how to incorporate systemic influences into risk measurement, notes that there is historical **evidence showing that default and credit events multiply in the time of time of distressed macroeconomic conditions**. The relationship between macroeconomic conditions and the probability of default has been addressed by several studies (e.g. Fama, 1986, Wilson, 1997, Carey, 1998), where the authors concluded that default rates **increase when the economy turns down**.

Koopman and Lucas (2005) have pointed out that the Altman-type models emphasise the cross-sectional rather than the timeseries dimension of the sample when distinguishing between ‘good’ from ‘bad’ companies, while stressing that the dynamic behaviour of credit risk has become increasingly important over the past few years. From the perspective of macroeconomic factors, the study mentions that it is generally thought that the **systematic risk factors correlate with macroeconomic conditions**, while further mentioning that default rates tend to be higher in times of recession. Focusing on an exceptionally long data sample, the authors, among others, concluded that cyclical co-movements between GDP and business failure mainly arise at the longer frequency. Carling et al. (2007) were concerned with the survival time to default for borrowers in the business loan portfolio of a major Swedish bank. The main result of their study is that macroeconomic variables have a significant **explanatory power for firm default risk in addition to a number of common financial ratios**. The authors further found a duration dependency, which implies that binary default models are inappropriate, as the idiosyncratic risk factors need to be complemented with information on survival time to obtain consistent default risk estimates.

As noted by Carling et al. (2007), the importance of macroeconomic effects for firm default risk is currently an underexplored topic in the empirical literature. However, while there has been some improvement in recent years, Carling’s statement still partly holds.

1.1 THE RELATIONSHIP BETWEEN FIRM-LEVEL DETERMINANTS OF DEFAULT AND MACROECONOMIC CONDITIONS

Interesting studies can be found dealing with the influence of macroeconomic conditions on firm performance factors (or rather determinants of default). Among others, the leverage factor (i.e. the extent to which the business uses debt financing) receives much attention in these studies. The importance of leverage factors as a determinant of corporate default was highlighted recently by Traczynski (2017) who showed that the only two risk factors that can explain default risk across all industry sectors are financial leverage and market return volatility. In this perspective, Cathcart et al. (2020) add that for unlisted businesses financial leverage might be the most important predictor of financial distress. This is in line with the general expectation noted by Zavgren (1985) and Stiglitz (1972) following from the high proportion of debt in present in the capital structure of distressed businesses. The influence of macroeconomic variables of monetary policy on corporate leverage has been analysed by Azofra et al.

(2020), who focused on the question of whether this influence of leverage is shaped by the presence of bank debt. The authors pointed out that a lot of research is primally concerned with understanding the different firm characteristics that explain how firms shape their capital structures over time, while macroeconomic factors have received comparatively little attention. Gungoraydinoglu and Öztekin (2011) analysed the determinants of capital structure and found that the firm-level covariates drive two-thirds of the variation in capital structure across countries, while the country-level covariates explain the remaining one-third.

1.2 THE APPLICATION OF MACROECONOMIC FACTORS IN THE COURSE OF PREDICTING BUSINESS DEFAULT - THE HAZARD MODEL APPROACH

Shumway (2001) was among the first to attempt to model default probability with respect to time. Shumway employed a hazard model on a sample of NYSE or AMEX traded firms, covering the period from 1962 to 1992. As a baseline hazard rate, Shumway (2001) used the firm's age, defined as the logarithm of the number of days the business has been listed on the NYSE.

As noted by Gupta et al. (2018), since Shumway's seminal work, the use of the hazard rate modelling technique has become popular in bankruptcy prediction studies. The hazard model was further applied, for example, in the paper of Chava and Jarrow (2004), Hillegeist et al. (2004), Nam et al. (2008), Nouri and Soltani (2016), and Campbell et al. (2008).

Chava and Jarrow (2004) focused on US companies traded on the AMEX, NYSE or NASDAQ from 1962 to 1999, while they employed the variables of Shumway (2001) and Altman (1968), Zmijewski (1984). The models were re-estimated to cover the period 1962-1991 and to test the period of 1992-1999. Results confirmed the superiority of Shumway's model, i.e. the hazard approach model. Chava and Jarrow (2004) demonstrated the importance of the industry effect for hazard rate modelling, the industry effect was incorporated in to the model both in terms of intercept and slopes.

Hillegeist et al. (2004) on a sample of **listed businesses**, from 1980 to 2000, compared the information content of the discrete hazard model and the re-estimated Altman (1968) and Ohlson (1980) models using the hazard model approach with a Merton's approach model. According to their results, the market-based Merton's approach model provides significantly more information about the probability of bankruptcy than do either of the popular accounting-based measures (Altman's or Ohlson's accounting variables). The recent percentage of bankruptcies was employed as a baseline hazard rate in this study.

Hillegeist et al. (2004) also adjusted the scores for industry effect, in line with the methodology proposed by Fama and Fench (1997), i.e. the decomposition of bankruptcy scores into industry means and deviations.

Nam et al (2008) extended the work of Shumway (2001) and Hillegeist et al. (2004) by presenting a duration model with time-varying covariates and a baseline hazard function incorporating macroeconomic dependencies. The Nam et al. (2008) study was conducted on a sample of Korean listed businesses from 1991 to 2000. They applied the change in interest rates suggested by Hillegeist et al. (2001) and the volatility of the foreign exchange rate. The advocacy behind the choice of foreign exchange rate was the author's suspicion "*that the Asian economic crisis is triggered by a drastic deficiency of foreign exchange, especially in the case of Korea*" see Nam et al. (2008). Moreover, they raise doubts about the utilisation of recent default rates, as suggest by Hillegeist et al. (2001) as the baseline hazard rate, arguing that it "*can be interpreted as an actual realisation of the unconditional baseline hazard rate in the previous period. Moreover, this autoregressive specification would have no forecasting power given unexpected macroeconomic shocks…*"

1.3 APPLICATION OF MACROECONOMIC VARIABLES TO THE SME BUSINESS SEGMENT IN THE COURSE OF DEFAULT PREDICTION

It is worth mentioning that most of the applications were done on a sample of listed and thus large businesses and as noted by Filipe et al. (2016), most European SMEs are small and do not satisfy the entry requirements of stock exchanges. Only a limited number of papers, such as Holmes et al (2010), Gupta et al. (2015), El Kalak and Hudson (2016) and Gupta et al. (2018), have dealt with the application of the hazard model for SME default modelling.

Gupta et al. (2015) argued that the SME segment is not homogenous, while there is a large diversity in terms of capital structure, firm size, access to external finance, management style, number of employees and others. Gupta et al. (2015) further highlight that heterogeneity has been neglected by empirical studies on SME credit risk. The authors applied the discrete-time duration-dependent hazard rate on a large sample of UK nonfinancial SMEs from the period 2000-2009, while adopting the European Union definition of SMEs. Their model was separately developed for micro, small and medium businesses, while by comparing the model's version, their results suggest that the micro business segment should be treated separately from the whole SME segment. Gupta et al. (2015) used the logarithm of the firm's age, insolvency rate and industry "weight of evidence" variables to control for both survival time and macroeconomic conditions.

El Kalak and Hudson (2016) applied the same approach as Gupta et al. (2015) on the sample of US SMES from the period 1980-2013, while the SBA (Small Business Administration) definition was adopted. El Kalak and Hudson (2016) confirmed the conclusion of Gupta et al. (2015) on the necessity of treating micro businesses separately from the rest of the SMEs segment, due to different (i.e. lower) survival probabilities. On the other hand, El Kalak and Hudson (2016) point out that Gupta's approach of utilising the insolvency rate variable as the baseline hazard rate distorts the baseline hazard idea.

4 AIM OF THE WORK AND METHODOLOGY ADOPTED

The aim of the work is to verify the extent to which the prediction accuracy of the probability of default of SMEs could be increased by the addition of macroeconomic variables to a set of accounting variables.

4.1 ADDRESSED RESEARCH GAPS

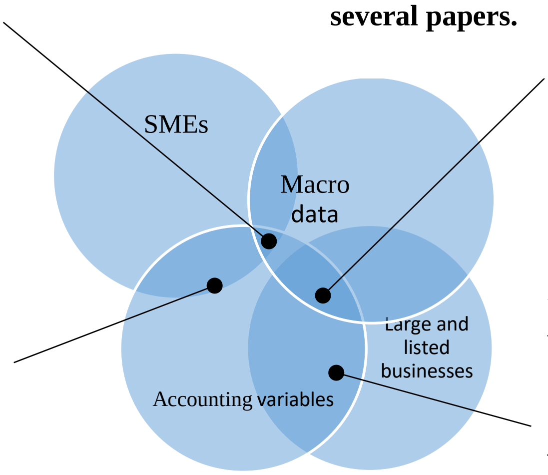
In the current state of the art on default prediction, there is a clear **discrepancy** in the attention paid to large and listed businesses and the attention paid to SMEs, while the specifics of the SME segment poses many **challenges to the modelling process**. I will try to summarise the main issues.

- Research on business default prediction begins with the accounting-based models of Altman type (see Altman, 1968), while in recent years most of the research has been on structural model approaches of the Merton type (see Trujillo-Ponce et al., 2013), with the prevailing consensus being on the superiority of structural approach models. It is worth mentioning that the structural approach has two main drawbacks: **first,** the default probability under this approach is based on the volatility of assets and asset value, which **are unobservable** and have to be estimated (see Agarwal and Taffler, 2008). **Second**, this approach is applicable for listed business only, whereas most European SMEs do not satisfy the entry requirements of stock exchanges, and thus such an approach is practically inapplicable for the SME segment (see Filipe et al., 2016).
- The accounting models of the Altman type are often criticised due to their static nature or for not respecting the **multiperiod nature of the default process** (e.g. Shumway, 2001, Berent et al., 2017).
- There is a sharp contrast in the number of studies pointing out the **specifics of SMEs** (in comparison with large businesses), such as the financial constraints issue, meaning that the SMEs have to face and having implications on its weaker position (e.g. Beck et al, 2006, Jin et al.., 2018, North et al, 2010) and the number of studies addressing **default prediction issues, especially in the SME segment.** *In other words, there is a consensus about the specificity of the SME segment, however with a week reflection in the default prediction literature*. Nevertheless, several studies on the default prediction issues of SMEs can be found (e.g. Edminster, 1972, Altman, Sabato, 2007, Altman, Sabato, 2010, Holmes et al., 2010, Gupta et al, 2015, El Kalak, Hudson, 2016 and Gupta et al., 2018).
- In the case of SMEs, firm-level information utilisable for default prediction is rather limited in comparison to large and listed business, due to the lack of financial market data, meaning that accounting data is the main source of utilisable information. On the other hand, there is no such limitation for the availability of macroeconomic data. There are several studies pointing out that the combination of these two types of data can result in a synergic effect (e.g. see Trujillo-Ponce et al., 2013, Agarwal and Taffler, 2008, Tinoco and Wilson, 2013), however these studies addressed the issue on a sample of large and listed businesses.
- There are studies showing that systematic risk factors are correlated with macroeconomic conditions (e.g. Koopman, Lucas, 2005) or that the default rates tend to be higher when under distressed macroeconomic conditions (Fama, 1986, Wilson, 1997, Carey, 1998). **Based on this, it can be argued that macroeconomic factors have an impact on the default probability of businesses or rather could be an important distress predictor.**
- Several studies dealing with accounting type models demonstrated that the application of the model under alternative conditions (different countries, industries or periods) is connected with a drop-in model accuracy (see, for instance, Platt, Platt, 1990, Grice, Dugan, 2001, Wu, Gaunt, Gray, 2010). **These studies indirectly showed that environmental factors may have an impact on the accuracy of the accounting model.**

The motivation for focusing on this particular topic and the potential **research gap** can be summarised in the following Venn diagram. The size of the overlap of the sets represents the number of papers devoted to the given area and not the influence of one feature on another.

Aim of the work. Predicting default, especially for SMEs, while respecting the specifics of this segment of business. Utilising mainly firm-specific data and macroeconomic data – **very limited research so far, increasing attention of**

The utilization of macroeconomic data in prediction of default of large and listed business, together with firm-specific variables – **several papers.**



Predicting default especially for SMEs, while respecting the specifics of this segment of business. Utilising mainly firm-specific data – **several papers, however currently attaining increasing attention of researchers.**

Predicting default of large businesses, based on accounting variables – **currently well explored, with limited space to contribute.**

Note: *SMEs and large businesses are presented as disjoint sets as the definition of SMEs clearly distinguishes between these two groups.*

The potential contribution to the current state of the art in default prediction of this work is mainly given by focusing on SME default from the perspective of utilising both firm-specific type variables and macroeconomic type of variables, while the research is conducted on a large sample of SMEs, which should result in a more robust model. The robustness of the model should be further enhanced by respecting the heterogeneity of this segment (the difference between small and medium business), industry specifics, and the multiperiod nature of the default process in the modelling phase.
To the best of my knowledge, this issue has not been addressed so far to this extent in the literature.

4.2 THE RESEARCH HYPOTHESES AND THEIR VERIFICATION

To fulfil the aim of the work, which is verifying the extent to which the prediction accuracy of probability of default of SMEs could be increased by the addition of macroeconomic variables to a set of firm-specific variables, the following research hypothesis was formulated:

> **Null hypothesis:** *The model combining a set of macroeconomic and* firm-specific *variables will achieve a significantly* higher discrimination power**, in terms of AUC, than a model utilising only a set of* firm-specific *variables, while not employing a set of macroeconomic variables.*
>
> **Alternative hypothesis:** *The model combining a set of macroeconomic and firm-specific variables will not achieve a significantly* higher discrimination power**, in terms of AUC, than a model utilising only a set of firm-specific variables, while not employing a set of macroeconomic variables.*

Note: *The **difference between compared** AUCs will be evaluated using the DeLong test (see DeLong et al., 1988); **The **discrimination power** of the model will be assessed in terms of Area Under Curve (AUC) estimated under the assumption of binomial distribution.

Verifying the above-stated hypotheses is complicated by the very nature of the researched phenomenon, which is the accuracy of the model. Model accuracy is a feature of a model as a whole, whereas the

accuracy (among other reasons) depends on the choice of model's variables. Selecting the optimal set of model variables is often done in terms of stepwise procedures, the aim of which is to ensure that model contains only significant variables. **The core of the problem regarding the verification of the research hypothesis lies in the fact that it cannot be ensured that after changing the set of potential variables (i.e. adding macroeconomic variables to a set of firm-specific variables), the original set of firm-specific variables will remain unchanged in the newly formed model.** Thus, the potential difference between the original model and the newly formed model will be simultaneously affected by two effects:
1) The effect of changing the set of variables.
2) The effect of adding the macroeconomic variables.

To be able to isolate these two effects, a set of three models were formulated. The research hypothesis testing was done in terms of comparing models' out-of-sample accuracy (in terms of AUC). The specifics of the derived models are as follows:
1) **Model 1** – the model was derived in a stepwise manner with a full set of variables (both firm-specific variables and macroeconomic variables).
2) **Model 2** – the model was derived only on a set of firm-specific variables, which were included in model 1, while the variables were forced to enter into the model, i.e. the stepwise procedure was not applied.
3) **Model 3** – the model was derived from a full set of firm-specific variables in a stepwise manner.

The purpose of deriving models 2 and 3 was for comparison – to analyse the extent to which the macroeconomic variables improve the accuracy of the model solely based on firm-specific variables.
From a statistical point of view, all the created models are nested. Demler et al. (2012) stressed that application of DeLong's test on nested models and in-sample testing will result in weakening the power of the test. To avoid this, the assessment of the research hypotheses will be based only on the results gained on the test samples (i.e. out of the sample).
Furthermore, the two versions of the firm-specific variables model make it possible to distinguish between three types of effect:
1. Effect of adding macroeconomic variables to what is otherwise the same set of firm-specific variables (by comparing models 1 and 2).
2. Effected gain by the combination of macroeconomic variables and accounting variables and solely firm-specific variables (by comparing models 1 and 3).
3. Effect of changing the set of accounting variables (by comparing models 2 and 3).

The research hypothesis will be verified (i.e. accepted or rejected) under the following schema:
The **hypothesis** will be accepted (or rather not rejected) when the three following conditions are met simultaneously:
1) The AUC value achieved in out-of-sample testing (on the test sample) of model 1 will be higher than the AUC value achieved on the same sample by model 2.
2) The AUC value achieved in out-of-sample testing (on the test sample) of model 1 will be higher than AUC value achieved on the same sample by model 3.
3) All the mentioned differences proved statistically significant in terms of DeLong et al (1988) at least at a 5% significance level.

Not meeting a single one of the mentioned conditions would result in not accepting the presented hypothesis.

4.2.1 The form of the model and model versions

For evaluation of the research hypothesis, the method of Cox semiparametric model was adopted. The model was estimated in two forms – the initially estimated model and the model with interaction terms. The model was initially estimated in the form:

$$ln[h(t)] = ln[h_0(t)] + \gamma_1 IND_2 + \cdots + \gamma_3 IND_4 + \gamma_4 SB + \beta X_{i,t} \quad (1)$$

However, by analysing the initial results, it was found that an interaction term is probably missing (for details Kennedy, 2005). Thus, the interaction between categorial variables (Industry group, category of company, and OENEG) and the continuous variables (X) was added, under these assumptions the model takes form:

$$ln[h(t)] = ln[h_0(t)] + \gamma_1 \cdot IND_2 + \cdots + \gamma_3 \cdot IND_4 + \gamma_4 \cdot SB + \beta \cdot X_{i,t} + \delta_1 \cdot SB * X_{i,t} + \delta_2 \cdot IND * X_{i,t} + \delta_3 \cdot OENEG * X_{i,t} \quad (2)$$

where: *$h_0(t)$ – baseline hazard date, SB – small business dummy (1 – in case of small business, 0 – in case of medium business), IND – industry group, SB*X, IND*X, OENEG* X – interaction term (between OENEG indicator and continues variables), s, γ;β;δ – regression coefficients.*

4.2.2 The procedure of estimating the model

The initial step in deriving a hazard model lies in checking the p**otential differences in the survival curves** of different groups in the sample. The groups are commonly distinguished by adding a dummy variable (in this case presented, such a role was played by the variables of "SB" – differentiating between small and medium business, "IND" – for distinguishing different industries and OENEG variables). The hazard model assumptions are not violated unless the survival curves are not crossing each other, this can be usually verified by a graphical check of the estimated survival curves, using the Kaplan-Maier procedure, whereas the necessity of adding the mentioned dummy variables to the model is verified by log-rank test, which results examine the difference in survival curves.
The next step lies is applying the **initial discrimination procedure** (also referred as univariate discrimination), under such procedure there is a model created separately for each of the analysed continuing predictors by applying the Cox regression method (or generally a method which later serves for deriving the final model). The purpose of this step is to reduce the number of potential predictors and to keep only those predictors, which exhibit a significant estimate and exhibit an expected coefficient sign. This procedure is commonly employed in deriving the prediction model . Some researchers criticise the step of checking the expected, as the information on the expected sign is based on theory about the relation between the predictor and the dependent variable. In case that the log-rank test proves that there are differences in survival curves among the analysed groups, the initial discrimination procedure has to respect this and is done by utilizing the model in the form (2).

5 RESEARCH METHODS AND SAMPLES

In this section, the research sample and methods used for deriving the model and verification of the research hypothesis will be presented.

5.1 RESEARCH SAMPLE

The sample under analysis consists of 202,209 SMEs from EU 28 countries, covering the period from 2014-2019. Out of this, 59,709 went legally bankrupt within one year, while the financial statements from the prefinal period (a year prior to bankruptcy) were analysed. In the course of this study, a business is considered a small company if its operating revenue is lower than 1 mil. EUR, its total asset value is lower than 2 mil. EUR and the number of employees is lower than 15. The business is considered a medium company if it's not a small company and its operating revenue does not exceed 10 mil. EUR, its total asset value does not exceed 20 mil. EUR and the number of employees is lower than 150.

Table 1, Number of defaults per observed period

Year		**2015**	**2016**	**2017**	**2018**	**2019**	**Total**
Status	Non-default	2,217	4,238	30,973	104,528	544	142,500
	Default	7,917	16,302	21,086	14,362	42	59,709
Total		10,134	20,540	52,059	118,890	586	202,209

Source: Own calculation based on Amadeus database

As not all businesses have managed to publish their financial statements for 2019, the number of observations for this period is significantly lower. The sample was randomly divided into a learning part (70% of all observations) and a testing part (30%) and later the ROC curve method was adopted. This approach in relation to the hazard model was also employed e.g. by Gupta et al. (2015).

Studies on credit scoring often employ several generic terms used to describe the event, which is the later the subject of prediction, and this includes the following terms: *financial distress*, *default failure*, *business failure*, *bankruptcy* and *insolvency*. In the course of this work I employ the following default definition: "**Default** is a judicial decision declaring a company insolvent." In line with Gupta et al. (2015), I tend to differentiate between small and medium businesses, as SMEs are not a homogenous segment, and to control for that, a dummy variable (called "category of company") was added.

I further employ an industry dummy ("IND") to control for the industry effect. There are two reasons for that. The first is that the analysed data comes from businesses in different industries. The second is that it has been shown that industry-specific data plays a significant role in bankruptcy prediction (specifically in the case of the hazard model, see Chava and Jarrow, 2004 or for a more general perspective, see Grice and Dugan, 2001). Primarily, the NACE rev. 2 main section industry classification was employed, which is the European industry classification. There are 21 main sections under this classification. From a modelling perspective, this is to smooth the differentiation and thus we place industries into four groups. This grouping is inspired by Chava and Jarrow (2004).

5.2 FIRM-SPECIFIC POTENTIAL VARIABLES

Reviewing previous studies with static models might not be useful as the hazard approach analyses the significance of variables over more than just one specific period. For this reason, empirical studies dealing with the hazard approach and SMEs were reviewed. The information on the expected variable signs was drawn from these studies as well. In some cases, the authors stated the expected sign explicitly, while in others the signs were drawn from the final model details (i.e. parameter estimates published in the papers). The expected sign plays a significant role in selecting the variables of the model and the specific procedure will be described in the methodology section of this work.

Table 2, List of analysed ratios

Abbrev.	Description	Ex. sign	Abbrev.	Description	Ex. sign
C/TA	cash/total assets[3;4]	(-)	QA/TA	Quick Assets/total assets[4]	(-)
CA/CL	current assets/current liabilities[1;7;3]	(-)	QR	Quick Ratio; (current assets− inventory)/current liabilities[3;5]	(-)
CA/S	Current asset/sales[3]	(+)	RE/TA	retained earnings/total assets[8;7;6;4;3;5]	(-)
CashR	Cash Ratio; cash/current liabilities[5]	(-)	S/TA	sales/total assets[8;7;6]	(+)
CE/TL	Capital employed/total liabilities[1;4;3;5]	(-)	S/TTA	sales/tangible assets[5]	(-)
CL/E	Short term debt/equity book value[4;3;5]	(+)	SHP	Stock holding period; (stock × 365)/sales[5]	(+)
CL/TA	Current liabilities/total assets[3]	(+)	size	Ln (Total Assets/GDP price level index)[6]	(-)
DCP	Debtor collection period; (trade debtors × 365)/sales[5]	(+)	ST/TA	Stock/total assets[4]	(+)
EBIT/CE	Earnings before interest and taxes/capital employed[5]	(-)	St/WC	stock/working capital[1]	(+)
EBIT/S	Earnings before interest and taxes/sales[5]	(-)	T/TA	Taxes/total assets[4;5]	(-)
EBIT/TA	Earnings before interest and taxes/total assets [8;7;6]	(-)	TC/TA	Trade creditors/total assets[4;3]	(+)
EBITDA/IE	Earnings before interest taxes, depreciation and amortization/interest expenses[4;3;5]	(-)	TC/TD	Trade creditors/trade debtors[1]	(+)
EBITDA/TA	Earnings before interest, taxes, depreciation, and amortization/total assets[4;3;5]	(-)	TC/TL	Trade creditors/total liabilities[1;4]	(+)
FE/S	financial expenses/sales[5]	(+)	TCPP	Trade creditors payment period; (trade creditors × 365)/sales[5]	(+)
FE/TA	financial expenses/total assets[3;5]	(+)	TD/TA	Trade debtors/total assets[4]	(+)
IA/TA	Intangible assets/total assets[4]	(+)	TL/NW	Total liabilities/net worth[5]	(+)
Ln(age)	natural logarithm of age (no. of days)[7]	(-)	TL/QA	Total liabilities/quick assets[1]	(+)
log (CA/CL)	log (current assets/current liabilities)[4]	(-)	TL/TA	total liabilities/total assets[8;7;2;3]	(+)
NI/E	net income/equity[3;5]	(-)	TL/TTA	Total liabilities/tangible total assets[5]	(+)
NI/S	net income/sales[3;5]	(-)	WC/S	Working capital/sales[3]	(-)
NI/TA	net income/total assets[8;7;6;2]	(-)	WC/TA	Working capital/total assets[8;7;6;5]	(-)

Source: 1- Altman et al (2010); 2 - Campbell et al (2008); 3 - El Kalak and Hudson (2016); 4 - Gupta et al (2015); 5 - Gupta et al (2018); 6 - Hillegeist et al. (2004); 7 – Chava and Jarrow (2004); 8 - Shumway (2001)

5.3 MACRO-ECONOMIC POTENTIAL VARIABLES

In line with Nam et al. (2008), we employed macroeconomic variables to capture time-varying macro dependencies and beyond this, as this study deals with panel data, to capture differences between countries raising from different levels of economic development between European countries. The choice of potential macroeconomic variables was inspired by previous studies of hazard models and others dealing with default risk or SMEs' financial constraints, which is expected to reflect the specific features that SME survival is sensitive to. The data on macroeconomic variables was taken from the EUROSTAT database.

Table 3, Overview of hazard model literature employing macroeconomic variables

No.	Macro-economic variable	Literature	Ex. sign[3]
1	Exchange rate	Holmes et al (2010), Nam et al (2008)[1]	(+)
2	Interest rate	Christidis and Gregory (2010), Tinoco and Wilson (2013), Holmes et al (2010), Nouri and Soltani (2016), Hillegeist et al (2004)[2]	(+)
3	Gross Value Added (GVA) per employee	Holmes et al (2010)[4]	(-)
4	Personal Cost (PC) per employee	Holmes et al (2010)[4]	(+)
5	Inflation	Christidis and Gregory (2010), Nouri and Soltani (2016), Tinoco and Wilson (2013)	(+)
6	Employment	Holmes et al (2010)	(-)
7	GDP annual growth rate	Simons and Rolwes (2009), Nouri and Soltani (2016)	(-)
8	GDP per capita	Beck et al (2006)	(-)

Notes: 1 – exchange rate volatility, 2 – change of interest rate, expected sign (+) increase of the variable means increase in default probability, (-) otherwise, 4 – Holmes et al. (2010) used the sectoral wage and sectoral value added.

According to Holmes et al. (2010), the exchange rate factor might be particularly import for SME survival as they are more likely to "face competition from abroad and to be involved in exports and imports". Changes in exchange rate are expected to have an adverse effect on firm survival, as changes "imply a worsening of the competitive position relative to overseas competitors" (see Holmes et al., 2010). The exchange rate from local currency to EUR was employed in this study, while data was taken from the Amadeus database, which quotes the exchange rate based on data from the International Monetary Fund (IMF) website and the exchange rates are those for the closing date of the statement.

The interest rates influence the firm survival probability through the capital structure, low interest rates are incentives for firms' investments, and the expected return on investment is higher when interest rates are low than in the case the interest rates are high. On the other hand, high interest rates lead to rising costs on debt capital, and firms have to pay more to their lenders (see Tinoco and Wilson, 2010). Thus, the higher interest rates are expected to increase the firm's probability of failure. In this study, then, the yield on government bonds with a maturity of ten years was adopted as the interest rate variable, and such interest rates are used to define the Maastricht criterion on long-term interest rates.

Gross value added is expected to have a positive influence on the firm's survival (decreasing the probability of failure), since increasing GVA is associated with a growing market, while conversely, a wage increase (personnel cost) means a rise in costs, and thus is expected to increase the probability of failure (see Holmes at all, 2010).

Inflation is expected to affect the probability of a firm's default indirectly by serving as an incentive to invest savings, rather than see their purchase power erode further in the future through inflation. Thus, inflation increases the risk-taking capacity of investors and by that lowers default probability (Tinoco and Wilson, 2013 or Qu, 2008). However, as acknowledged further by Qu (2008), the direction of inflation's effect on default probability has not been unequivocally established due to the complexity of inflation's effect on the economy. Mare (2012) noted that high inflation rates are a sign of weak macroeconomic conditions under which there is also a high number of bank crises. In this work, I adopted the Harmonised Index of Consumer Prices (HICP), specifically the annual average rate of change as an inflation rate. Within this study, based on the above-mentioned arguments, it is expected that an increase in inflation rate is related to an increase in the firm's hazard probability.

The employment rate is expected to lower the probability of failure. Employment is a proxy for demand, so the higher employment is, the higher the expected demand (see Holmes et al., 2010). The employment rate was taken from the EUROSTAT database and refers to the percentage of employed people between the ages of 15 and 64 expressed as a share of the total population.

Studies on SMES are often regarded as vulnerable to economic environment changes. Simons and Rolwes (2009) reported a significant negative relationship between GDP growth and the firm default rate.
Beck et al. (2006) found that businesses in countries with higher levels of financial intermediary development, more liquid stock markets, more efficient legal systems and higher GDP per capita, report lower financing obstacles. Ullah (2019) highlights that “among all the business environment constraints affecting firm growth, financial constraint has been identified as one of the most detrimental growth obstacles.” The reason why Gupta et al. (2015) suggest treating small businesses separately from the media business was their lower survival probability. GDP per capita might serve as a proxy of financial obstacles a business has to face in their country, while on the other hand growth obstacles seem to indirectly affect survival probability. For these reasons, a negative relation between GDP per capita and firm survival might be expected.

5.4 CLASSIFICATION METHODS USED IN DEFAULT PREDICTION MODEL

To derive the model, the Cox semiparametric proportional model approach was employed, and this approach was adopted by Lando (1998) who was the first to model default with the Cox model. Further seminal work in this field was done by Shumway (2001), who demonstrated the superiority of the hazard model approach in predicting business default over the static approach model (i.e. not considering the multiperiod nature of the data). The superiority of the hazard approach was confirmed by other authors, e.g. Chava and Jarrow (2004) and Berent et al. (2017). Study of Berent et al. (2017) highlights the need to treat default as a multiperiod process, as “the real economy as well as firms are driven by multi-period processes”, which suggests the employment of Cox’s hazard model approach.
According to Gupta et al. (2015): “the discrete hazard modelling technique is well suited to analyse data that consists of binary dependent variables and exhibit both time-series and cross-sectional characteristics, such as bankruptcy data.”
On the other hand, other opinions can be found, and for example Gupta et al. (2018) mentioned that “this growing popularity of hazard models in bankruptcy prediction seems to be trend or momentum driven, rather than being based on a strong theoretical underpinning”.
Despite criticism, the Cox approach seems to be flexible and appropriate to the multiperiod nature of the data, which is the main reason for employing that approach in course of this work.
The model was originally developed by Cox (1972), whereas the general formula of the Cox model is:

$$\lambda(t;z) = exp(z\beta)\lambda_0(t) \tag{3}$$

The main problem behind the Cox model is the relationship between the distribution of failure time (t) and variable z. B is the parameter vector and $\lambda_0(t)$ is the baseline hazard function for the standard set of conditions z=0, while λ0(t) might be replaced by any known function h(zβ), see Cox (1972). The Cox proportional hazard model could be expressed also in log form (see Landau and Everitt, 2004):

$$ln[h(t)] = ln[h_0(t)] + \beta_1 X_1 \dots + \beta_q X_q \tag{4}$$

where *h0(t) is the baseline hazard function;* “being the hazard rate for individuals with all explanatory variables equal to zero, this function is left unspecified. The estimated cumulative baseline hazard can be estimated from sample data and is often useful” (Landau and Everitt, 2004). The advantage of the Cox semiparametric hazard model is that its estimation is possible even after leaving the baseline hazard function unspecified, which “offers a considerable advantage when we cannot make a reasonable assumption about the shape of the hazard” (see Cleves et al., 2008, p. 129).
The applications of the hazard model are most often inspired by the seminal paper of Shumway (2001), which showed that a discrete-time hazard model is equivalent to a multiperiod logit model, while authors tend to specify the baseline hazard rate.

Shumway (2001) specified the applied hazard model as follows:

$$\phi(t;x;\theta_1;\theta_2) = \frac{1}{1+exp\left(g(t)'\theta_1 + x'\theta_2\right)} \tag{5}$$

where: ϕ *is the hazard function, g(t) is the natural logarithm of the number of days the business was listed on NYSE,* θ_1, θ_2 *– regression parameters, x is explanatory variable.*

Generally, there are two main approaches to the specification of the baseline hazard rate. The first is the use of time dummies, as shown by Beck et al (1998), or employing macroeconomic variables, as suggested by Nam et al. (2008), who argue that indirect measures such as time dummies are less effective in capturing time-varying macro dependencies. Gupta et al (2015) followed this suggestion of Nam et al. (2008) and to accommodate the macroeconomic impact firms have to face, they construct the baseline hazard rate including the insolvency risk variable. According to El Kalak and Hudson (2016) this approach distorts the idea of the baseline hazard rate.

In this paper, we use the Cox semiparametric model, while leaving the baseline hazard rate unspecified and employ the macroeconomic variables as explanatory variables. This approach is different from the other studies mentioned (e.g. Nam et al., 2008). The main difference is that under this approach the macroeconomic variables influence the hazard rate through a shift in baseline hazard (as other explanatory variables), which seems to be useful as the analysis deals with panel data.

6 THE MAIN RESULTS OF THE HABILITATION THESIS

In this chapter the results of deriving the new default prediction model will be presented. The results will be compared to those obtained with the model of Altman and Sabato (2007), which was derived especially for SMEs. To make the comparison more efficient, the model was tested both with the original setting and with re-estimated coefficients. The model's re-estimation was done on the learning samples.

6.1 SURVIVING TIMES OF SMALL AND MEDIUM COMPANIES

The underlying idea of the Cox regression is to analyse the time to an event (in this case the default of a company). From this perspective, a closer look at the survival time of the companies under analysis is useful. The survival time was analysed separately for small and medium companies, as a different survival time is expected, especially from the perspective mentioned by Gupta et al. (2015). In the case of small businesses, the highest default rate could be observed in the first five years of their existence, and at the end of this period only 66% still survived, while after the next five years, only 38% of the businesses had survived (i.e. to 10 years after establishing the business).

The median survival time for small businesses is 9.31 years. In the case of the media business, the situation is significantly different, where 99% of the businesses survive till the end of the first five-year period, while at the end of the second five-year period (i.e. 10 years after the business was established), 97% of the businesses are still active. The median survival time of the media business is 30.00 years.

The data in studies are left censored, which means that the data does not cover the whole period of business life or rather the time the observation of the business enters the study is not the time of its establishment. For this reason, further analysis will deal with the time of study.

6.2 INITIAL STEP OF DERIVING THE HAZARD MODEL - SURVIVAL FUNCTION COMPARISON

The initial step in deriving the model was to analyse whether there is a difference between the different groups of businesses analysed. The variables of the category of companies and industry served to denote subgroups in which a difference in survival is expected. The following figure represents the survival functions for small and medium businesses, confirming the conclusion of Gupta et al. (2015), for the sample under analysis, about the heterogeneity of the SME group.

Table 4, Log-rank test results

Categorial variable	Chi-Square	df	Sig.
IND	32,737.478	4	0.00000
SB	196,408.359	1	0.00000
OENEG	45,270.951	1	0.00000

Source: Own calculation based on Amadeus database

6.3 INITIAL DISCRIMINATION ANALYSIS

Out of the 42 tested variables, only 25 exhibited significant coefficient estimates and at the same time enjoyed the expected sign (10 ratios were excluded as nonsignificant, another 7 ratios were excluded due to not having the expected signs, and 4 ratios were excluded due to both being nonsignificant and not having the expected signs). The following table show the list of variables excluded due to not meeting the expected sign or not having a significant estimate.

Table 5, Initial discrimination analysis - list of excluded variables (insignificant coefficient estimate)

Abbreviation	Exp. sign	B	SE	Wald	df	Sig.	Exp(B)
CA/CL	(-)	-0.0024	0.0012	3.8217	1	0.050591	0.9976
EBIT/CE	(-)*	0.0033	0.0049	0.4554	1	0.499784	1.0033
EBIT/TA	(-)	-0.0146	0.0080	3.3713	1	0.066341	0.9855
log (CA/CL)	(-)	-0.0821	0.7655	0.0115	1	0.914579	0.9212
NI/E	(-)*	0.0001	0.0042	0.0011	1	0.973931	1.0001
S/TA	(+)	0.0042	0.0024	3.0885	1	0.078846	1.0042
S/TTA	(-)	-0.00003	0.0000	2.2453	1	0.134024	1.0000
St/WC	(+)*	-0.00001	0.0000	0.6088	1	0.435229	1.0000
TD/TA	(+)*	-0.0005	0.0189	0.0007	1	0.978570	0.9995
TL/QA	(+)	0.0001	0.0002	0.1850	1	0.667133	1.0001

Note: *the estimated variable sign is not meeting the expectation. Source: Own calculation based on Amadeus database. Note: B – estimated coefficient, SE – standard error, df – degrees of freedom.

The expected sign is (-) if a higher value of the ratios is expected to be related to a lower probability of default, while the opposite is expected in the case of a (+) sign. Among the excluded variables, a significant portion is represented by profitability ratios (EBIT/CE, EBIT/TA, and NI/E), liquidity ratios (CA/CL, log(CA/CL)) and asset management ratios (S/TA, S/TTA).

Two of the profitability ratios EBIT/CE (i.e. EBIT over capital employed) and NI/E (net income over equity) exhibit an unexpected sign. As both indicators deal implicitly with capital structure, an analysis of the capital structure differences between defaulting and non-defaulting groups of business could provide some explanation. For this purpose, the value of the TL/TA (total liabilities over total assets) indicator was analysed. The mean value of the indicator on a sample of non-defaulting businesses was 0.66 while the mean value of the same indicator on a sample of defaulting businesses was 2.02 (see for details see appendix), which means that on average defaulting companies in the sample should suffer from negative equity. Such a phenomenon could explain the unexpected sign of the corresponding variables.

Another variable exhibiting an unexpected sign is the ratio of trade debtors over total assets (TD/TA).

Table 6, Initial discrimination analysis - list of excluded variables (significant coefficient estimate)

Variable	expected sign	B	SE	Wald	df	Sig.	Exp(B)
CL/TA	(+)	-0.0120	0.0021	33.0347	1	0.000000	0.9881
RE/TA	(-)	0.0108	0.0016	44.9099	1	0.000000	1.0108
TC/TA	(+)	-0.0614	0.0168	13.3989	1	0.000252	0.9404
TC/TL	(+)	-0.2587	0.0207	155.9551	1	0.000000	0.7721
TL/TA	(+)	-0.0113	0.0016	47.9542	1	0.000000	0.9888
WC/S	(-)	0.0016	0.0003	21.0759	1	0.000004	1.0016
WC/TA	(-)	0.0112	0.0021	29.0394	1	0.000000	1.0113

Source: Own calculation based on Amadeus database. . Note: B – estimated coefficient, SE – standard error, df – degrees of freedom.

Many authors consider total indebtedness (TL/TA) one of the most significant indicators of bankruptcy, including Cathcart et al. (2020), according to whom financial leverage might be the most important predictor of financial distress of unlisted business. Furthermore Zavgren (1985) and Stiglitz (1972) suggest that a high proportion of debt is present in the capital structure of a distressed business. In the case presented, the TL/TA achieved a coefficient estimated with a negative sign, whereas the magnitude of the coefficient is relatively low. A possible explanation could be the financial constraints aspect, often mentioned in relation to SMEs, showing that the external sources of finance in the form of debt are often not accessible for SMEs. Due to the coefficient sign, other common predictors of default were excluded, such as the retained earnings over assets (RE/TA) or net working capital over total assets (WC/TA).

Let's further focus on the significant variables with the expected sign. These results are the subject of the following table.

Table 7, The estimated coefficient of the first step model, firm's specific variables- significant variables with expected sing only

Abbreviation	B	SE	Wald	df	Sig.	Exp(B)
C/TA**	-0.1302	0.0191	46.5306	1	0.000000	0.8779
CA/S**	0.0029	0.0006	26.8848	1	0.000000	1.0029
CashR**	-0.0180	0.0041	18.9749	1	0.000013	0.9822
CE/TL**	-0.0114	0.0025	20.8588	1	0.000005	0.9887
CL/E**	0.0034	0.0003	116.2627	1	0.000000	1.0034
DCP**	0.0000	0.0000	34.2336	1	0.000000	1.0000
EBIT/S*	-0.0095	0.0040	5.5821	1	0.018145	0.9906
EBITDA/IE**	0.0000	0.0000	38.0962	1	0.000000	1.0000
EBITDA/TA**	-0.0643	0.0167	14.7426	1	0.000123	0.9377
FE/S**	0.7338	0.0551	177.6086	1	0.000000	2.0831
FE/TA**	1.0975	0.1169	88.1549	1	0.000000	2.9966
IA/TA**	0.4249	0.0519	67.1331	1	0.000000	1.5294
NI/S**	-0.0103	0.0036	8.2041	1	0.004180	0.9897
NI/TA*	-0.0162	0.0075	4.7218	1	0.029783	0.9839
QA/TA**	-0.0911	0.0161	31.8404	1	0.000000	0.9129
QR**	-0.0112	0.0019	31.6778	1	0.000000	0.9889
SHP**	0.0001	0.0000	59.1392	1	0.000000	1.0001
size**	-0.1361	0.0073	349.9673	1	0.000000	0.8727
ST/TA**	0.0678	0.0222	9.3156	1	0.002272	1.0702
T/TA**	-0.8435	0.1960	18.5268	1	0.000017	0.4302
TC/TD**	0.0017	0.0004	21.3250	1	0.000004	1.0017
TCPP**	0.0001	0.0000	17.1049	1	0.000035	1.0001
TL/NW**	0.0042	0.0006	48.1676	1	0.000000	1.0043
TL/TTA*	0.0001	0.0000	6.0562	1	0.013858	1.0001
Ln (age)**	-0.1634	0.0042	1493.2476	1	0.0000	0.8493

Note: *significant at 5% level, **significant at 1% Source: Own calculation based on Amadeus database. Note: B – estimated coefficient, SE – standard error, df – degrees of freedom.

A substantial part of the significant variables is represented by indicators dealing with working capital management, which is a part of financial health often mentioned as problematic in the case of SMEs. Some of these ratios describe the cash conversion cycle (SHP, DCP or TCPP) or describe the relationship between different items of net working capital, especially in terms of liquidity ratios (CashR or QR). Among the significant ratios, there is also the indicator of size, which defines the size of the business in terms of asset value. he significance of this indicator further confirms the heterogeneity of the SMEs sectors.

Another substantial group of significant ratios is dealing with business profitability, while both the operating profitability (EBITDA/TA or EBIT/S) and net income profitability (NI/S) play a significant role.

Moreover, the indicators describing the age of the business (i.e. ln(age)) proved to be significant. It should be mentioned that nevertheless the Cox regression model's aim is to model the time to default, while the time term which is the subject to this indicator is defined in different terms. The age of the business is the time since the business was established, and not the time since the business entered the study.

As this analysis was performed on a univariate basis and many significant ratios describe a similar area of financial health, the presence of a significant correlation between these ratios is expected.

A similar procedure was conducted for the macroeconomic factors under analysis.

Table 8, The estimated coefficient of the first step model, macro-economic variables- significant variables with expected sing only

Variable	Ex. sign	B	SE	Wald	df	Sig.	Exp(B)
Exchange rate**	(+)	-0.2110	0.0237	79.4277	1	0.0000	0.8098
Interest rate**	(+)	0.33598	0.0045	5657.954	1	0,0000	1.3993
GDP per capita**	(-)	0.00001	0.0000	166.1575	1	0.0000	1.0000
GDP annual growth rate	(-)	0.0091	0.0036	0.0609	1	0.8050	1.0091
GVA per employee **	(-)	0.0082	0.0004	362.1806	1	0.0000	1.0082
PC per employee **	(+)	0.0160	0.0009	314.6718	1	0.0000	1.0161
Inflation**	(+)	-0.7147	0.0073	9589.6140	1	0.0000	0.4894
Employment rate**	(-)	-0.0685	0.0019	1280.2992	1	0.0000	0.9338

Note: *significant at 5% level, **significant at 1%. Source: Own calculation based on Amadeus database. Note: B – estimated coefficient, SE – standard error, df – degrees of freedom.

All the analysed macroeconomic variables are significant at the 1% level, except for the annual GDP growth rate. A possible explanation might be that the analysed period was a relatively stable one for EU SMEs, with only the Greek economy in 2015 and 2016 going into recession and in 2015, also Croatia, Cyprus, Finland and Serbia, which experienced negative annual GDP growth. Speaking of country-year GDP data, 99% of observation values were positive (see the appendix), thus not significantly triggering business defaults. For example, Nouri and Soltani (2016) analysed the impact of GDP growth rate, interest rate and inflation on the bankruptcy of businesses listed on the Cyprus Stock Exchange and found that these variables have no significant impact. However, it should be noted that their results are based on different methodologies.

Regarding the expected sign of the **analysed variables**, only interest rates, PC per employee and employment rate variables have the expected sign, thus being kept for further analysis.

The next step was the correlation check. For this purpose, the Pearson's correlation coefficient was employed. The following table shows only highly correlated pairs of variables (the correlation coefficient was higher than 0.7 or lower than -0.7).

Table 9, High correlated pair of variables

Type of variables	Pairs of variables	Pearson correlation coefficient	Sig. (2-tailed)	N
Firm specific	NI/TA & EBITDA/TA	0.857	0.00000	102,457
	DCP & CA/S	0.751	0.00000	104,694
	NI/S & EBIT/S	0.948	0.00000	101,482
	QR & CashR	0.805	0.00000	119,102
Macro-economic	GVA & PC per employee	0.934	0.00000	140,075

Source: Own calculation based on Amadeus database

Regarding firm-specific variables, there were four highly correlated pairs of variables identified. The first pair deals with the return on assets. According to the Wald statistics, the EBITDA/TA represents a better measure than NI/TA. A possible cause of this is the different levels of corporate taxation among EU countries. The second correlated pair is composed of the Debtor Collection Period (DCP) indicator and the ratio of current assets over sales. These indicators have in common the features of sales and accounts receivable (debtors). The DCP represents a more significant measure, which is why the CA/S will be excluded from further analysis. The third correlated pair of variables measures the profit margin at different levels of profit (net profit or operating profit margin, i.e. NI/S or EBIT/S). The net profit margin (NI/S) achieved a more significant estimate, thus remaining for further analysis. The fourth correlated pair of variables deals with business liquidity. The pair consists of the quick ratio (QR) and the cash ratio (CashR). Moreover, the current ratio (CA/CL) was among the analysed ratios, but this ratio's estimated value was not significant at the 5% level. In the study presented, the Quick ratio (QR) represents a more significant measure, and that is why this ratio will be further analysed.

Furthermore, a multicollinearity check was also conducted. For this purpose, the Variance Inflation Factor (VIF) approach was adopted.

Table 101, Collinearity Statistics

Variable	**Tolerance**	**VIF**	**Variable**	**Tolerance**	**VIF**
Interest rate	0.752	1.329	IA/TA	0.891	1.122
PC per employee	0.840	1.191	NI/S	0.484	2.066
Employment rate	0.682	1.467	QA/TA	0.516	1.939
C/TA	0.699	1.431	QR	0.422	2.371
CE/TL	0.411	2.435	SHP	0.697	1.435
CL/E	0.671	1.490	size	0.762	1.313
DCP	0.435	2.299	T/TA	0.506	1.976
EBITDA/IE	0.946	1.057	TCPP	0.418	2.390
EBITDA/TA	0.479	2.089	TL/NW	0.671	1.491
FE/S	0.403	2.482	TL/TTA	0.917	1.090
FE/TA	0.622	1.607	TC/TD	0.916	1.091
ln(age)	0.796	1.257			

Source: Own calculation based on Amadeus database

Based on the VIF results, no variable VIF score exceeds a value of 4, which represents a commonly used cut-off. Thus, the multicollinearity presence is not significant. Otherwise, such a feature would bias the coefficient estimates.

6.4 ESTIMATING THE MODELS' COEFFICIENTS

The results of estimating the models are presented in the following manner. At first, the overall model statistics are given, followed by variables excluded from the model during the stepwise selection procedure and finally the model coefficients are presented. Subsequently, the benchmark is presented – the re-estimated Altman model and Altman Sabato model.

The final step is testing the model and comparing outcomes. All three models are tested using the ROC curves, while the AUC values are later compared by using the procedure suggested by DeLong et al. (1988).

6.4.1 Details of model 1 estimates

Model 1 was estimated in a stepwise manner by employing a backward elimination procedure using conditional likelihood ratio (LR) statistics as a criterion, which are considered as least prone to error. As a result, the model is significant at the 1% level. In the case of model 1, the stepwise procedure can lead to the exclusion of eight variables out of the final model, while the residual chi-square is 9.910 (with df = 8), sig. = 0.271, which is not significant, thus forcing these variables into the model would not make a significant contribution to its predictive power.

Table 11, Variables not included in model 1

Variable	**Score**	**df**	**Sig.**
CE/TL	1.483	1	0.223
DCP	2.424	1	0.119
EBITDA/TA	1.288	1	0.256
FE/TA	0.399	1	0.528
IA/TA	1.405	1	0.236
TCPP	0.913	1	0.339
TL/TTA	0.355	1	0.551
TC/TD	1.691	1	0.194

Source: Own calculation based on Amadeus database

The details of variables, which enter the model are listed below. The final version of model 1 contains three macroeconomic indicators, twelve firm-specific indicators and categorical variables describing the industry and category of companies. Furthermore, two interaction terms enter the model and reached a significant estimate.

Table 12, Variables in model 1

Variables		B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
								Lower	Upper
Macroe.	Interest rate**	1.067	0.021	2631.976	1	0.000	2.906	2.790	3.027
	PC per employee **	0.010	0.002	38.962	1	0.000	1.010	1.007	1.013
	Employment rate**	-0.036	0.005	61.928	1	0.000	0.965	0.956	0.973
Firm specific	C/TA**	-1.953	0.164	141.968	1	0.000	0.142	0.103	0.196
	CL/E**	0.006	0.001	46.443	1	0.000	1.006	1.004	1.008
	EBITDA/IE**	0.000	0.000	8.797	1	0.003	1.000	1.000	1.000
	FE/S**	1.617	0.171	89.381	1	0.000	5.040	3.604	7.048
	ln(age) **	-0.069	0.017	16.674	1	0.000	0.934	0.904	0.965
	NI/S**	-0.207	0.019	116.354	1	0.000	0.813	0.783	0.844
	QA/TA**	-0.172	0.064	7.240	1	0.007	0.842	0.743	0.954
	QR**	-0.131	0.031	17.411	1	0.000	0.877	0.825	0.933
	SHP**	0.000	0.000	19.310	1	0.000	1.000	1.000	1.000
	size**	-1.147	0.042	738.019	1	0.000	0.317	0.292	0.345
	T/TA**	-2.749	0.699	15.470	1	0.000	0.064	0.016	0.252
	TL/NW**	0.006	0.002	7.520	1	0.006	1.006	1.002	1.010
Categorical (dummy)	SB**	1.236	0.064	371.831	1	0.000	3.441	3.034	3.901
	IND**			2095.415	4	0.000			
	IND (N/A) **	3.068	0.113	740.366	1	0.000	21.499	17.236	26.815
	IND (IND 1)	-0.416	0.087	23.132	1	0.000	0.660	0.557	0.781
	IND (IND 2) **	-0.356	0.091	15.261	1	0.000	0.701	0.586	0.837
	IND (IND 3) **	-0.517	0.106	23.568	1	0.000	0.597	0.484	0.735
	OENEG**	-0.170	0.048	12.865	1	0.000	0.843	0.768	0.926
Interaction	SB x NI/S**	0.306	0.022	199.008	1	0.000	1.358	1.302	1.417
	SB x QR**	0.187	0.032	34.116	1	0.000	1.205	1.132	1.283

Note: **Significant at 1% level. **significant at 5% level. Category of companies: 1 – medium business, 0 – small business. Source: Own calculation based on Amadeus database. Note: B – estimated coefficient, SE – standard error, df – degrees of freedom.

The industry effects and category of company effect are significant variables in the final model, which is in line with expectations (see Chava and Jarrow, 2004 or Gupta et al., 2015). The industry effect influences only the model intercept and not the slope. On the other hand, the category of company not only influences the model intercept but also the slope in the case of two variables – net income over sales (net profit margin - NI/S) and quick ratio (QR).

Further regarding the industry effect. As a default industry, the IND 4 industry, i.e. financial and real estate activities were chosen according to the estimated parameter sign, while other industries are less risky, except for non-specified industries, which means that the industry information is essential in default risk prediction, in line with expectations (see Grice and Dugan, 2001). Usually studies on the hazard models (e.g. Shumway, 2001) tend to exclude financial firms from the sample. Chava and Jarrow (2004) derived a model for nonfinancial businesses and for all businesses (including financials). They conclude that after such inclusion, the overall prediction accuracy of the model drops, and the authors of that study indicate that predicting bankruptcy for financial businesses is a more complicated exercise. It is worth mentioning that the study of Chava and Jarrow (2001) did not focus on SMEs.

There are three macroeconomic variables included in the model, i.e. the interest rate, personnel cost (PC) per employee and employment rate, while they had the expected signs. Regarding the coefficient value, the largest influence of the unit change of indicators on the default probability is related to the change in interest rate, where a unit change of interest rate (by 1pp) increases the default probability by 1.076 pp. The effect is further supported in the case of a business which has issued a loan, as the ratio of financial expenses to sales is also part of the model. On the other hand, a unit change of personnel cost per employee will lead to an increase in default probability by 0.010 pp. The firm-specific financial ratios included in the final model describe the working capital management level (SHP – stock holding period) or its structure (C/TA – cash over total assets, QA/TA) – quick assets over total assets). Other measures describing this area of financial health were not included in the model. However, on a univariate base they proved to be significant. This applies for the ratios of debtor collecting period (DCP) and trade creditors payment period (TCPP).

Further significant indicators are measures of business solvency (EBITDA/IE, CL/E, TL/NW, or CE/TL) or measure the relative size of financial expenses (FE/S) and net profit margin (NI/S). El Kalak and Hudson (2016) found that the net profit margin (NI/S) is a significant profitability measure for SMEs. However, when focusing solely on small businesses, this measure was insignificant. Gupta et al. (2018) report varying (insignificant) explanatory power across different periods, while the same applies for EBITDA/IE and CL/E indicators.

After the first deriving the model, the variables of QR (quick ratio) and net profit margin (NI/S) change sign to positive, which was contrary to prior expectation. According to Kennedy (2005), such a phenomenon could be (among others) explained by the presence of multicollinearity, outlier presence or missing interaction terms. As the data were winsorised and multicollinearity checked, the only explanation which have been left was a missing interaction term, especially resulting from data aggregation. As a potential missing interaction, that between industry groups, category of company and OENEG (dummy) variables was analysed. Only the interaction between QR (or rather NI/S) variables and the category of company indicators enter the model and lead to a change in the main effect estimate sign. The situation means that the ratio of net profit margin (NI/S) and quick ratio (QR) changes its behaviour depending on whether the business is of medium or small type. The main effect coefficient has to be interpreted together with the interaction coefficient. The expected sign is encountered only in the case of medium businesses (as the category of the company dummy is equal to zero), while in the case of small businesses, the positive sign of the interaction term coefficient prevails over the negative value of the main effect coefficient, which makes the overall effect positive. Thus, the higher value of NI/S and QR indicators represent a lower default probability only in the case of medium businesses, while in the case of small businesses the default probability is on the contrary, increased.

Moreover, the age of the company was subjected to analysis (i.e. the Ln(age) indicator) which refers to the natural logarithm of the number of days from establishment of the business to the day the business declared bankruptcy or to the end of the observed period. The Cox model requires a time term to establish its parameters, while in this case the time from the start of the observed period was used to the moment of bankruptcy, and thus these two terms are not interchangeable. The model also contains a size factor in terms of the natural logarithm of the asset size divided by the inflation rate. The size factors refer to the market position of the business (see Ding et al., 2008, Niemann et al, 2008, Psillaki, Tsolas and Margaritis, 2009). Shumway (2001) considered the company size factor to be a significant predictor of bankruptcy, but he derives that indicator from market data. Wu, Gaunt and Gray (2010) added that bigger firms are considered more capable of surviving tough economic times and less prone to bankruptcy. Although the model also contains the variable of the category of company (differentiating between small and medium businesses in the sample), the significance of the size factor, especially in a hazard model, may refer to the diminishing asset value of the defaulting business.

6.4.2 Details of model 2 estimates

The aim of the work is also to analyse the significance of macroeconomic variables in predicting the default of European SMEs and for this reason a second version of the model (referred to as model 2) was derived. This version of the model contains only firm-specific variables and industry and category of company dummy variables. All re-estimated coefficients have the expected sign, apart from the relative size of quick assets (QA/TA), which changes sign to positive, which might be a result of missing interaction, caused by the change of variable set. Furthermore, the indicators of TL/NW and EBITDA/IE become insignificant in the model.

Table 132, Variables in model 2

Variables		B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
								Lower	Upper
Firm specific	C/TA**	-1.590	0.110	210.389	1	0.000	0.204	0.165	0.253
	CL/E**	0.009	0.001	166.900	1	0.000	1.009	1.007	1.010
	EBITDA/IE	0.000	0.000	1.640	1	0.200	1.000	1.000	1.000
	FE/S**	1.012	0.122	69.065	1	0.000	2.750	2.166	3.491
	ln(age) **	-0.174	0.010	323.096	1	0.000	0.840	0.824	0.856
	NI/S**	-0.289	0.015	353.087	1	0.000	0.749	0.727	0.772
	QA/TA**	0.227	0.047	23.444	1	0.000	1.254	1.144	1.375
	QR**	-0.288	0.032	80.097	1	0.000	0.750	0.704	0.799
	SHP**	0.000	0.000	76.799	1	0.000	1.000	1.000	1.000
	size**	-0.694	0.028	619.631	1	0.000	0.500	0.473	0.528
	T/TA**	-2.930	0.512	32.800	1	0.000	0.053	0.020	0.146
	TL/NW	0.001	0.002	0.448	1	0.503	1.001	0.998	1.004
Categorical (dummy)	SB**	2.382	0.051	2158.309	1	0.000	10.827	9.792	11.971
	IND**			34.301	4	0.000			
	IND (N/A)	-0.020	0.077	0.067	1	0.796	0.980	0.843	1.140
	IND (IND 1)**	-0.213	0.065	10.849	1	0.001	0.808	0.712	0.917
	IND (IND 2) *	-0.161	0.069	5.457	1	0.019	0.851	0.744	0.974
	IND (IND 3) **	-0.322	0.080	16.413	1	0.000	0.725	0.620	0.847
	OENEG**	-0.293	0.036	67.867	1	0.000	0.746	0.696	0.800
Interac	SB x NI/S**	0.386	0.016	563.580	1	0.000	1.471	1.424	1.518
	SB x QR**	0.314	0.032	94.335	1	0.000	1.368	1.285	1.458

Note: **Significant at 1% level. **significant at 5% level. Source: Own calculation based on Amadeus database.
Note: B – estimated coefficient, SE – standard error, df – degrees of freedom.

Model 2 was derived in a forced entry manner, which is the opposite to the stepwise procedure. The reason for that is that applying a stepwise procedure to a set of variables, after excluding the macroeconomic variables, would also mean a change in firm-specific variables. Comparing such a model will not explain the extent to which macroeconomic variables influence the model's accuracy.

6.4.3 Details of model 3 estimates

In the case of model 3, the stepwise procedure led to the exclusion of six variables from the final model, while the residual chi-square is 6.191 (with df = 6), sig. = 0.402, which is not significant. The details of the variables, which enter the model are listed below.

Table 3, Variables not in model 3

Variables	Score	df	Sig.
FE/TA	2.385	1	0.122
T/TA	1.328	1	0.249
TCPP	2.29	1	0.13
TL/NW	0.456	1	0.5
TL/TTA	0.24	1	0.624
TC/TD	0.653	1	0.419

Source: Own calculation based on Amadeus database

The final version of model 3 contains three macroeconomic indicators, fifteen firm-specific indicators and categorical variables describing the industry and category of companies. Furthermore, two interaction terms significant in model 1 also entered the model.
By comparing the firm-specific variables of model 2 (or rather of model 1) and the variables of model 3, it can be concluded that there are four variables which enter model 3, while not entering model 2 (or

rather model 1). These variables are – capital employed over total liabilities (CE/TL) – a measure of capital structure; debtor collecting period (DCP) – a measure dealing with net working capital or rather cash collecting cycle; EBITDA over total assets – EBITDA/TA – a asset profitability ration and ratio of intangible assets and total assets (IA/TA) – assessing the asset structure. On the other hand, instead of these four variables, a set of two different variables enters model 3, while not being included in model 2 (or rather model 1). These ratios are – tax over total assets (T/TA) showing the relative size of the paid taxes, and the total liabilities over net working capital (TL/NW). The details of model 3 coefficient estimates are listed in the table below.

Table 15, Variables in model 3

Variables		**B**	**SE**	**Wald**	**df**	**Sig.**
Firm specific	C/TA**	-2.505	0.167	226.041	1	0.0000
	CE/TL**	-0.079	0.027	8.278	1	0.0040
	CL/E**	0.009	0.001	137.843	1	0.0000
	DCP**	0.000112	0.000032	12.184	1	0.0000
	EBITDA/IE**	-0.000043	0.000014	9.798	1	0.0020
	EBITDA/TA**	-0.221	0.066	11.159	1	0.0010
	FE/S**	1.666	0.177	88.812	1	0.0000
	ln(age) **	-0.189	0.016	142.286	1	0.0000
	IA/TA**	0.804	0.137	34.562	1	0.0000
	NI/S**	-0.241	0.021	135.859	1	0.0000
	QA/TA*	0.159	0.07	5.115	1	0.0240
	QR**	-0.16	0.041	15.287	1	0.0000
	SHP**	0	0	44.345	1	0.0000
	size**	-0.993	0.041	590.169	1	0.0000
Categorical (dummy)	OENEG**	-0.308	0.054	32.595	1	0.0000
	IND**			37.599	4	0.0000
	IND (N/A)	-0.164	0.101	2.646	1	0.1040
	IND (IND 1)**	-0.342	0.086	15.662	1	0.0000
	IND (IND 2)**	-0.24	0.091	6.919	1	0.0090
	IND (IND 3)**	-0.514	0.106	23.434	1	0.0000
Inter. t.	SB**	2.178	0.066	1087.503	1	0.0000
	SB x QR**	0.204	0.04	26.485	1	0.0000
	SB x NI/S**	0.409	0.022	347.483	1	0.0000

Note: **Significant at 1% level. **significant at 5% level. Source: Own calculation based on Amadeus database.
Note: B – estimated coefficient, SE – standard error, df – degrees of freedom.

Comparison of the variables of model 2 and model 3 clearly showed that excluding macroeconomic variables from a set of potential variables led to different sets of firm-specific variables. **This could mean that some of the information content carried by macroeconomic variables was supplemented by other firm-specific variables, which might represent a manifestation of the macroeconomic variables in the firm's situation.** To further explore this issue and by that gain a deeper insight into the researched issue a regression model was estimated. The reason for that is to assess the extent to which the information carried by macroeconomic variables is unique and irreplaceable within the context of analysed firm-specific variables.

6.5 MODELS FOR BENCHMARK PURPOSES

To assess the performance of the derived model properly, two models were selected – the model of Altman (1983) and that of Altman and Sabato (2007). The model of Altman (1983) is a **generic** type model (of use for unspecified types of business). The reason for selecting this model is its worldwide popularity, because of which this model is often selected as a benchmark and from this perspective would make the results more comparable.

6.6 MODELS' TESTING RESULTS

Models 1, 2 and 3 were tested in terms of Area Under Curve while the survival probability as model outcome was subjected to testing. For this purpose, the following formula for survival probability was employed (see Bharat et al. 2018):

$$S(t) = exp[-H_0(t)exp(PI)] \quad (8)$$

Where: *S(t) is the survival probability at time t. $H_0(t)$ is the baseline cumulative hazard function. PI is the prognostic index,* which is given by:

$$PI = \sum_k^q \beta_k x_k \quad (9)$$

Where: *β are the regression parameters. x are the model variables.*

To estimate the survival function values at a given time (t), the following estimates of baseline cumulative hazard function values were utilised. As the form of baseline hazard function was unspecified, the specific values cannot be interpreted, however they are needed for estimating the predicted probability of default for a given observation. The survival probabilities shown in the table reflects the situation of the media business, while there is a penalisation for small businesses (represented as a dummy variable "category of company").

Table 16. Baseline cumulative hazard function values

Time			**1**	**2**	**3**	**4**	**5**
Model 1	Baseline Cumulative		0.298	1.362	3.767	5.816	
	At mean of covariates	Survival	0.998	0.992	0.979	0.967	
		SE	0.000	0.000	0.001	0.001	
		Cum Hazard	0.002	0.008	0.021	0.033	
Model 2	Baseline Cumulative		0.707	1.993	4.391	7.070	9.150
	At mean of covariates	Survival	0.994	0.983	0.962	0.940	0.923
		SE	0.000	0.000	0.001	0.001	0.005
		Cum Hazard	0.006	0.017	0.039	0.062	0.080
Model 3	Baseline Cum		2.357	6.833	14.926	20.550	27.729
	At mean of covariates	Survival	0.995	0.986	0.970	0.959	0.945
		SE	0.000	0.000	0.001	0.001	0.005
		Cum Hazard	0.005	0.014	0.030	0.042	0.057

Source: Own calculation based on Amadeus database

To evaluate performance of the estimated hazard model, receiver operating characteristics (ROC) curves were employed, while the comparison between the model's AUCs was subjected to a nonparametric Delong test. The Area Under Curve was conducted under the assumption of a binomial distribution, while there is also the possibility of estimating the AUC under nonparametric assumptions using the trapezoidal approach, whereas the results might slightly differ.

6.7 THE AUC VALUES OF THE TESTED MODELS

Both the set of derived models and the models used for benchmark purposes were tested on the same sample (i.e. learn and test sample). In the case of the original version of the models used for benchmark purposes, it is assumed that such splitting of the same would have an insignificant impact on the estimated AUC value, as both samples present out-of-sample testing for these models. However, in the case of the derived and re-estimated models, the test sample results are preferable for drawing the implications, from this perspective, in further text, the test sample results would be preferred.

Table 4, Models testing results.

Model	Learn sample			Test sample		
	AUC	SE	95% CI	AUC	SE	95% CI
Model 1	0.880	0.00305	0.877 to 0.882	0.884	0.00443	0.880 to 0.888
Model 2	0.821	0.00392	0.818 to 0.824	0.829	0.00571	0.825 to 0.834
Model 3	0.854	0.00348	0.852 to 0.857	0.862	0.00498	0.858 to 0.866
Z' score	0.754	0.00379	0.751 to 0.758	0.758	0.00567	0.753 to 0.763
AS original	0.746	0.00345	0.742 to 0.749	0.747	0.00524	0.741 to 0.752
Z' score re-est.	0.761	0.00378	0.758 to 0.765	0.766	0.00569	0.761 to 0.771
AS re-est.	0.781	0.00327	0.778 to 0.784	0.790	0.00480	0.785 to 0.795

Source: Own calculation based on Amadeus database. Note: CI – confidence interval, SE – standard error.

All tested models reached AUC higher than 0.5, which is the value representing a model without discrimination power, which means that all models have a significant discrimination power, whereas none of the models reached AUC lower than 0.7. Or more specifically said, the AUC values of the models selected for the benchmark range from 0.746 to 0.754 in its original setting, while its re-estimated versions' AUCs range from 0.761 to 0.781. The AUC values of the derived model range from 0.821 to 0.880. Comparing the results of the models selected for the benchmarks lead to conclusions that contradict the assumptions, as the Z'-score as a general model was assumed to reached a lower accuracy for AS model, as the model specific for SMEs segment (in line with Altman and Sabato, 2007).

The results of testing the original version of the model did not approve, whereas the Z¨ - score in its original setting outperforms AS model in its original setting. However, after both models were re-estimated on learning sample, the model in its re-estimated version outperforms re-estimated Z'-score. According to the results gained on the test sample, the highest AUC value was reached by Model 1, followed by Model 3 and Model 2, followed by Z score model and Z'-score-re-est. The estimated AUC values were subjected to DeLong's test to assess whether the mentioned between the two models AUC is significantly different, which would mean a significant difference in model accuracy, i.e. its quality. The results of DeLong's test will be presented in the following manner, first the derived models will be compared with the benchmarks (in order Model 3, followed by Model 2 and finally Model 1) and then the derived model will be compared between themselves).

6.8 COMPARING MODEL 3 WITH THE BENCHMARKS

Model 3 represents that version of the model derived using the Cox regression methodology, while it was derived from a full set of analysed firm-specific variables. The difference is AUC values of the model and the re-estimated models used as benchmarks can be assigned to differing methodology of model estimation (considering the time factor) and to effect gain by other factors (industry specifics, SME segment heterogeneity and different set of firm-specific variables) which applies for the AS model, while in the case of the Z-score model, the effect of the non-specific focus of the model (not distinguishing between large, medium and small businesses) is included above that.

Table 18, DeLong's test results – model 3 vs. benchmark

Model/Sample		Difference between areas	Standard Error	95% Conf. Interval	z statistic	Sign. level
Learn	Z' score	0.101	0.00480	0.0920 to 0.111	21.136	P < 0.0001
	Z' score re-est.	0.0950	0.00479	0.0856 to 0.104	19.832	P < 0.0001
	AS original	0.109	0.00464	0.0996 to 0.118	23.399	P < 0.0001
	AS re-est.	0.0752	0.00429	0.0668 to 0.0836	17.548	P < 0.0001
Test	Z' score	0.105	0.00713	0.0905 to 0.118	14.660	P < 0.0001
	Z' score re-est.	0.0959	0.00709	0.0820 to 0.110	13.529	P < 0.0001
	AS original	0.115	0.00698	0.101 to 0.128	16.403	P < 0.0001
	AS re-est.	0.0722	0.00634	0.0598 to 0.0846	11.393	P < 0.0001

Source: Own calculation based on Amadeus database

All the mentioned differences are statistically significant at the 1% level, which means the mentioned results will with high probability hold, even with an alternative sample of data.
The AUC of model 3 outperforms the AUC of Z-score by (on the test samples) 10.5 pp, while the re-estimated version of the Z-score is still being outperformed by 9.59 pp. The original version of the AS model is outperformed by model 3 by 11.5 pp, while after re-estimation of the model, the difference is 7.22 pp.
The differences might be considered marginal, but one has to keep in mind that the AUC has a maximum value of 1, which is hard to achieve in practice, and thus the space for model improvement is limited.

6.9 COMPARING MODEL 2 WITH THE BENCHMARK

Comparison of model 2 and the models selected for the benchmark will give a similar answer to that mentioned in the case of model 3. The set of variables in model 2 is only a subset of the variables of model 1 (only the firm-specific type variables), which means that model 2 is based on a suboptimal set of firm-specific variables, which causes the potential lower efficiency of this model, in comparison with model 3.

Table 19, DeLong's test results – model 2 vs. benchmark

Model/Sample		Difference between areas	Standard Error	95% Conf. Interval	z statistic	Sign. level
Learn	Z' score	0.0671	0.00510	0.0571 to 0.0771	13.163	P < 0.0001
	Z' score re-est.	0.0599	0.00509	0.0499 to 0.0699	11.780	P < 0.0001
	AS original	0.0757	0.00492	0.0661 to 0.0853	15.389	P < 0.0001
	AS re-est.	0.0422	0.00466	0.0330 to 0.0513	9.053	P < 0.0001
Test	Z' score	0.0714	0.00749	0.0568 to 0.0861	9.540	P < 0.0001
	Z' score re-est.	0.0628	0.00746	0.0482 to 0.0774	8.423	P < 0.0001
	AS original	0.0812	0.00743	0.0667 to 0.0958	10.927	P < 0.0001
	AS re-est.	0.0391	0.00692	0.0255 to 0.0526	5.651	P < 0.0001

Source: Own calculation based on Amadeus database

The AUC of model 2 outperforms the AUC of the Z-score by (on the test samples) 7.14 pp, while the re-estimated version of the Z-score is still being outperformed by 6.28 pp. The original version of the AS model is outperformed by model 2 by 8.12 pp, while after re-estimation of the model, the difference is 3.91 pp. All the mentioned differences are statistically significant at the 1% level.

6.10 COMPARING MODEL 1 WITH THE BENCHMARK

Comparison of model 1 with the models selected for the benchmark showed the extent to which model 3 (i.e. the duration model, especially derived for SMEs, while respecting SME segment heterogeneity and industry effects, and utilising both firm-specific type and macroeconomic type variables) provides better classification accuracy then the models selected for the benchmark (not utilising the above-mentioned features).

Table 20, DeLong's test results – model 1 vs. benchmark

Model/Sample		Difference between areas	Standard Error	95% Conf. Interval	z statistic	Sign. level
Learn	Z'_score	0.126	0.00409	0.118 to 0.134	30.678	P < 0.0001
	Z'_score_re_est.	0.118	0.00409	0.110 to 0.126	28.949	P < 0.0001
	AS_original	0.134	0.00431	0.126 to 0.143	31.129	P < 0.0001
	AS_re_est.	0.101	0.00371	0.0934 to 0.108	27.118	P < 0.0001
Test	Z'_score	0.126	0.00615	0.114 to 0.138	20.499	P < 0.0001
	Z'_score_re_est.	0.117	0.00612	0.105 to 0.129	19.183	P < 0.0001
	AS_original	0.136	0.00642	0.123 to 0.148	21.133	P < 0.0001
	AS_re_est.	0.0937	0.00545	0.0830 to 0.104	17.188	P < 0.0001

Source: Own calculation based on Amadeus database

The AUC of model 1 outperforms the AUC of the Z-score by (on the test samples) 12.6 pp, while the re-estimated version of the Z-score is still being outperformed by 11.7 pp. The original version of the AS model is outperformed by model 1 by 11.6 pp, while after re-estimation of the model, the difference is 9.37 pp.
All mentioned differences are statistically significant at the 1% level.

6.11 COMPARING THE DERIVED MODELS

Finally, the comparison of the derived models is subjected to the following table, while such a comparison enables the assessing of the extent to which the difference in accuracy could be attributed to the addition of macroeconomic variables to what is otherwise the same set of firm-specific ratios.

Table 21, DeLong's test results – derived models

Model/Sample		Difference between areas	Standard Error	95% Conf. Interval	z statistic	Sign. level
Learn	Model 1 ~ Model 2	0.0584	0.00420	0.0502 to 0.0667	13.903	P < 0.0001
	Model 1 ~ Model 3	0.0255	0.00388	0.0178 to 0.0331	6.557	P < 0.0001
	Model 2 ~ Model 3	0.0330	0.00107	0.0309 to 0.0351	30.798	P < 0.0001
Test	Model 1 ~ Model 2	0.0546	0.00624	0.0424 to 0.0669	8.753	P < 0.0001
	Model 1 ~ Model 3	0.0215	0.00572	0.0103 to 0.0327	3.767	P = 0.0002
	Model 2 ~ Model 3	0.0331	0.00169	0.0298 to 0.0364	19.563	P < 0.0001

Source: Own calculation based on Amadeus database

The model combining firm-specific and macroeconomic variables (i.e. Model 1) outperforms models containing the same set of firm-specific variables (i.e. Model 2) by 5.46 pp, while this difference is statistically significant at the 1% level. The potential of the firm-specific model could be increased by redefining the set of potential variables, while the previous conclusion that the addition of macroeconomic variables leads to significantly higher accuracy still holds, as Model 1 outperforms Model 3 by 2.15 pp, while this difference is statistically significant at the 1% level. By comparison of Model 2 and Model 3, it can be concluded that redefining the set of firm-specific ratios led to an increase of AUC by 3.31 pp, where the difference is significant at the 1% level.

7 CONTRIBUTION OF THE THESIS FROM A SCIENTIFIC, PRACTICAL, AND PEDAGOGICAL PERSPECTIVE

The presented work contributes to the **current scientific state of art, while there is also a possible overlap to praxis,** especially in the following manner:

- Presenting a new hazard model on a comprehensive sample of EU-28 data, whereas previous approaches usually address country-specific datasets. Whereas the presented model significantly outperforms the competing models, which were specially derived for the SME segment, while this holds even after the re-estimation of the competing models on a same data sets (i.e. the learning sample under analysis).
- Showing that macroeconomic factors could be effectively incorporated in the hazard model in the form of explanatory variables, which allows the use of several factors at once, instead of addressing the macroeconomic factor univariately, by adopting them into the baseline hazard function.
- Presenting that macroeconomic indicators carry a unique type of information that can be only partly supplemented (and thus not satisfactorily) by a different set of firm-specific variables.
- Proving that adding the combination of macroeconomic and firm-specific indicators is results in significantly higher accuracy than could be achieved by employing solely firm-specific indicators.
- Presenting a prediction model especially derived for SME segment of business, showing the specific risk factors of SMEs, which should be examined during the credit application process by banks or other credit providers.
- Presenting a tool for estimating the probability of default, which is one of the most important parts of credit risk, which also reflects in the provided interest rates for the credit applicant.
- Showing the extent to which, the probability of SME default is magnified by current macroeconomic conditions, apart from firm-specific measures (category of a company and given level of financial health represented by the specific values of financial ratios).

The results have an overlap to **pedagogical practice** as well, especially:

- For the teaching of "**Financial Management of Small Companies**" the subject – presenting students with further issues regarding the financial specifics of small businesses and the link between macroeconomic factors and the financial condition of the business. The results of the thesis enable highlighting the factor, which should be paid specific attention when preparing a business plan, which is a task for the student to complete the subject. Furthermore, showing then students that when analysing the financial statements in the case of small business, attention has to be paid to more specific factors, and not only the generic ones.
- During "**Rating and valuation of a business**", the students are usually provided with some insight into the distress prediction model, while when evaluating business students are required to conduct a strategic analysis, in which they need to assess macroeconomic factors as well and assess whether the business is meeting a going concern assumption – which otherwise means that the business has prospects. Providing a key result of this work might give the students guidance or rather an insight into the mechanism in which environmental and firm level factors interact with the business' prospects or rather its survival probability.
- Some of the methodological aspects of this work could be inspiring for students who decide on a master's thesis on the topic of "**Default prediction**". The topic is very challenging for the student on the one hand, while on the other, there have been several such successfully defended at the faculty up till now, while some of them also have gained recognition by practitioners during competitions held.
- The research topic has not been fully explained yet, thus there is a space for PhD student research to address several existing gaps.

8 CONCLUSION

The presented research focuses on the EU-28 countries. The reason for this is twofold. Firstly, as previously mentioned, a complex study of EU countries was missing in the current literature. Secondly, it was necessary to obtain enough data variability, where the specific value of the macroeconomic indicators differs for each of the countries, even at the same time, thus focusing on such panels would result in higher data variance and consequently will favourably affect the robustness of the created model(s). The difference between the economic development of the countries was captured by added macroeconomic variables. There was also an attempt to try to add category variables describing geographical regions of Europe (west, east, north and south), however such procedure would split the model into four smaller models and limit the variability of the macroeconomic indicators and that might result in biased coefficient estimates. To meet the aim of the work, a sample of 202,209 EU SMEs was collected and a set of 42 firm-specific ratios, accompanied by a set of 8 different macroeconomic factors. To verify the research hypotheses, three different models were created and tested (both in and out of the sample), the results were compared with two models by other authors, both in their original setting and in their re-estimated form, where re-estimation was carried out for the research presented.

The main ideas behind the adopted methodological approach could be described in the following manner. The main aim of the work was to verify the extent to which adding macroeconomic variables to a set of firm-specific ratios would result in a change (preferably an increase) in model accuracy. At first, the appropriate measure of model accuracy had to be selected, along with an appropriate test to assess the differences between two specific values (results) of testing model accuracy. In the current literature there are two main accuracy concepts – total accuracy and area under the curve (AUC) measures. There is strong criticism of employing the accuracy measure, as this is highly sensitive to the current model settings (the set of cut-off scores), moreover, the proportion of the tested sample (the proportion of non-defaulting and defaulting businesses) also highly affects the total accuracy results. As these negative features did not influence the AUC values, the AUC measure was rather adopted. The difference between two specific AUCs was tested using the DeLong et al. (1988) methodology. The difference between AUC values in some cases might be viewed as marginal, but one may have to keep in mind that the AUC values are limited to a value of 1. In the case of the AUC value related to a credit risk model, it can be also argued that a relatively small increase in model accuracy might result in significant saving when the model is applied to a large portfolio (for the example of credit applicants).

Analysing the effect which will result from adding a set of variables (in this case macroeconomic variables) to the other sets of variables (in this case the firm-specific variables) is complicated by the fact, that is has to be done in terms of the regression model's variables, thus before such a comparison could be made a model containing such types of variables has to be derived. At this point, another issue has to be considered – the significant value of a specific model's variable is given by its contribution to the rest of the model's variables, i.e. the variable is treated as significant when removal of such a variable would result in a significant drop in the model's overall significance. In other words, the significance of the model variables is in the context of the other variables' presence. And this is the nature of the problem, when two sets of otherwise significant variables are mixed in a new model, the significance of a given variable might change, even causing the insignificance of a previously significant variable (or rather significant in another model). From this perspective, it was necessary to derive three different models. For selecting model variables, a stepwise procedure was employed. The first question was whether a macroeconomic indicator would enter the model, otherwise fitted with firm-specific variables, as such entry will mean that there is unique information carried by the given macroeconomic indicators, which increase the overall significance of the model. This idea resulted in the creation of model 1. To assess the importance of the added macroeconomic indicators, as model 2 was formulated, this model contains only firm-specific variables which were present in model 1. Such a procedure was chosen to assess the extent to which the included macroeconomic indicators increase the accuracy of what is otherwise the same model. At this phase, the aim of the research could not fully be met, as there was still a chance that the reassessment of the firm-specific variables set would result in a more significant model than was achieved in the case of model 2, whereas the given set of firm-specific variables represents a subset of model 1 variables, while it was of high probability that deriving a model from a full set of firm-specific variables could lead to slightly different set of variables. This idea was behind the estimation of model 3.

Estimation of model 1 showed that the macroeconomic variables are able to provide a significant contribution to the set of firm-specific ratios to predict the default of SMEs of EU countries. Moreover, further analysis showed that information content of the added variables is of unique character and cannot be supplemented by firm-specific variables, within the set of analysed variables.

Regarding the specific macroeconomic variables, which enter the model - the employment rate, together with long-term interest rates and the personnel cost per employee seems to play a significant role in SMEs survival probability.

REFERENCES USED IN HABILITATION THESIS SUMMARY

AGARWAL, V. and R. TAFFLER. Comparing the performance of market-based and accounting-based bankruptcy prediction models. Journal of Banking and Finance. 2008, 32(8): 1541-1551. https://doi.org/10.1016/j.jbankfin.2007.07.014

ALLEN, L. and A. SAUNDERS. Incorporating systemic influences into risk measurements: A survey of the literature. Journal of Financial Services Research. 2004, 26 (2): 161-191. https://doi.org/0.1023/B:FINA.0000037545.38154.8a

ALTMAN, E. I. 1983. *Corporate financial distress: A complete guide to predicting, avoiding, and dealing with bankruptcy*. New York: John Wiley and Sons.

ALTMAN, E. I. 2000. Predicting financial distress of companies: Revisiting the Z-score and Zeta® models [online]. 2000 Available at: http://pages.stern.nyu.edu/~ealtman/PredFnclDistr.pdf

ALTMAN, E. I. 2002. *Bankruptcy, Credit risk, and High Yield Junk Bonds*. Malden, MA: Blackwell Publishers. https://doi.org/10.1111/1468-0300.00084

ALTMAN, E. I. and E. HOTCHKISS. 2006. *Corporate Financial Distress and Bankruptcy: Predict and Avoid Bankruptcy, Analyse and Invest in Distressed Debt*. New York: John Wiley & Sons, Inc.

ALTMAN, E. I. and G. SABATO. Modelling credit risk for SMEs: Evidence from US market. Abacus. 2007, 43(3): 332-357. https://doi.org/10.1111/j.1467-6281.2007.00234.x

ALTMAN, E. I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. The Journal of Finance. 1968, 23(4): 589-609. https://doi.org/10.2307/2978933

ALTMAN, E. I. Predicting Railroad Bankruptcies in America. Bell Journal of Economics. 1973, 4(1):184-211. https://doi.org/10.2307/3003144

ALTMAN, E. I., HALDEMAN, R. G. and P. NARAYANAN. A new model to identify the bankruptcy risk of corporations. Journal of Banking and Finance. 1977, 1(1):22-54. https://doi.org/10.1016/0378-4266(77)90017-6

ALTMAN, E. I., SABATO, G. and G. WILSON. The Value of Non-Financial Information in SME risk management. Journal of Credit Risk. 2010, 6(2): 1-33. https://doi.org/10.1007/PL00011399

AZOFRA, V., RODRÍGUEZ-SANZ, J. A. and P. VELASCO. The role of macroeconomic factors in the capital structure of European firms: How influential is bank debt? International Review of Economics and Finance. 2020, 69: 494-514. https://doi.org/10.1016/j.iref.2020.06.001

BECK, T., DEMIRGUC-KUNT, A., LAEVEN, L. and V. MAKSIMOVIC. The determinants of financing obstacles. Journal of International Money and Finance. 2006, 25(6): 932-952. https://doi.org/10.1016/j.jimonfin.2006.07.005

BERENT, T., BŁAWAT, B., DIET, M., KRZYK, P. and R. REJMAN. Firm's default - new methodological approach and preliminary evidence from Poland. Equilibrium. Quarterly Journal of Economics and Economic Policy. 2017, 12(4): 753-773. https://doi.org/10.24136/eq.v12i4.39

CAMPBELL, J. Y., HILSCHER, J. and J. SZILAGYI. In Search of Distress Risk. The Journal of Finance. 2008, 58(6): 2899-2939.

CAREY M. Credit risk in private debt portfolios. Journal of Finance. 1998, 53(4): 1363-1387. https://doi.org/10.1111/0022-1082.00056

CARLING, K., JACOBSON, T., LINDÉ, J. and K. ROZSBACH. Corporate credit risk modelling and macroeconomy. Journal of Banking & Finance. 2007, 31(3): 845-868. https://doi.org/10.1016/j.jbankfin.2006.06.012

CATHCART, L., DUFOUR, A., ROSSI, L. and S. VAROTTO. The differential impact of leverage on the default risk of small and large firms. Journal of Corporate Finance. 2020, 60: 101541. https://doi.org/10.1016/j.jcorpfin.2019.101541

CHRISTIDIS, A. and A. GREGORY. 2010. *Some New Models for Financial Distress Prediction in the UK*. Xfi - Centre for Finance and Investment Discussion Paper no: 10/04

CLEVES, M., GOULD, W. W., GUTIERREZ, R. G. and MARCHENKO, Y. 2008. *An Introduction to Survival Analysis Using Stata*. Stata Press.

COX, D. R. Regression Models and Life–Tables. Journal of The Royal Statistical Society. 1972,

34(2): 187-220. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x
DELONG E.R., DELONG, D.M. and D. L. CLARKE-PEARSON. Comparing the areas under two or more correlated Receiver Operating Characteristic curves: a nonparametric approach. Biometrics. 1988, 44(3): 837-45.
DEMLER, O.V., PENCINA, M. J R. and B. D'AGOSTINO SR. Misuse of DeLong test to compare AUCs for nested models. Stat Med. 2012, 31(23): 2577-87. https://doi.org/10.1002/sim.5328.
EL KALAK, I. and R. HUDSON. The effect of size on the failure probabilities of SMEs: An empirical study on the US market using discrete hazard model. International Review of Financial Analysis. 2016, 43: 135-145. https://doi.org/10.1016/j.irfa.2015.11.009
ERDOGAN, A. I. Factors affecting SME access to bank financing: An interview study with Turkish bankers. Small Enterprise Research. 2018, 25(1): 23-35. https://doi.org/10.1080/13215906.2018.1428911
FAMA, E. and K. FRENCH. Industry Costs of Equity. Journal of Financial Economics. 1997, 43(2): 153-193. https://doi.org/10.1016/S0304-405X(96)00896-3
FAMA, E. F. Term premiums and default premiums in money markets, Journal of Financial Economics. 1986, 17(1): 175-196.
FILIPE, S. F., GRAMMATIKOS, T. and D. MICHALA. Forecasting distress in European SME portfolios. Journal of Banking & Finance. 2016, 64: 112-135. https://doi.org/10.1016/j.jbankfin.2015.12.007
GRICE, J. S. and M. T. DUGAN. The limitations of bankruptcy prediction models: Some cautions for the researchers. Review of Quantitative Finance and Accounting. 2001, 17: 151-166. https://doi.org/10.1023/A:1017973604789
GUNGORAYDINOGLU, A. and Ö. ÖZTEKIN. Firm- and country-level determinants of corporate leverage: Some new international evidence. Journal of Corporate Finance. 2011, 17(5): 1457-1474.
GUPTA, J., GREGORIOU, A. and J. HEALY. Forecasting Bankruptcy for SMEs Using Hazard Function: To What Extent Does Size Matter? Review of Quantitative Finance and Accounting. 2015, 45(4): 845-869.
GUPTA, J., WILSON, N., GREGORIOU, A. and J. HEALY. The effect of internationalisation on modelling credit risk for SMEs: Evidence from UK market. Journal of International Financial Markets, Institutions & Money. 2014, 31: 397-413. https://doi.org/10.1007/s11156-014-0458-0
HILLEGEIST, S. A., KEATING, E. K., CRAM, D. P. and K. G. LUNDSTEDT. Assessing the probability of bankruptcy. Review of Accounting Studies. 2004, 9: 5-34. https://doi.org/10.1023/B:RAST.0000013627.90884.b7
HOLMES, P., HUNT, A. and I. STONE. An analysis of new firm survival using a hazard function. Applied Economics. 2010, 42(2): 185-195. https://doi.org/10.1080/00036840701579234
HUDSON J. age, regional and industrial structure of company liquidations. Journal of Business Finance and Accounting. 1987, 14(2): 199-213. https://doi.org/10.1111/j.1468-5957.1987.tb00539.x
JIN, Y., LUO, M. and C. WAN. Financial constraints, macro-financing environment, and post-crisis recovery of firms. International Review of Economics and Finance. 2018, 55: 54-67. https://doi.org/10.1016/j.iref.2018.01.007
KARAS, M. and REŽŇÁKOVÁ, M. Identification of financial signs of bankruptcy: A case of industrial enterprises in the Czech Republic. In Proceedings of the 6th International Scientific Conference: Finance and the performance of firms in science, education, and practise. Zlín: 2013a. s. 324-335. ISBN: 978-80-7454-246- 6.
KENNEDY, P. Oh No! I Got the Wrong Sign! What Should I Do? The Journal of Economic Education. 2005, 36 (1): 77-92.
KOOPMAN, S. J. and A. LUCAS. Business and Default Cycles for Credit Risk. Journal of Applied Econometrics. 2005, 20: 311–323. https://doi.org/10.1002/jae.833
LANDAU, S., EVERITT, B. S. 2004. *A handbook of statistical analysis using SPSS*. Chapman &

Hall/CRC Press LLC.
LANDO, D. On Cox Processes and Credit Risky Securities. Review of Derivatives Research. 1998, 2(2): 99-120.
MARE, D. S. Contribution of macroeconomic factors to the prediction of small bank failures. In: Proceedings of the 4th International IFABS Conference. 2012.
MERTON, R. On the pricing of corporate debt: The risk structure of interest rates. Journal of Finance. 1974, 29(2): 449-470. https://doi.org/10.2307/2978814
NAM, C. W., KIM, T. S., PARK, N. J. and H. K. LEE. Bankruptcy prediction using a discrete-time duration model incorporating temporal and macroeconomic dependencies. Journal of Forecast. 2008, 27: 493-506. https://doi.org/10.1002/for.985
NORTH, D., BALDOCK, R. and I. EKANEM. Is there a debt finance gap relating to Scottish SMEs? A demand-side perspective. Venture Capital. 2010, 12(3): 173-192. https://doi.org/10.1080/13691061003658670
NOURI, B. A. and M. SOLTANI. Designing a bankruptcy prediction model based on accounts, market, and macroeconomic variables (Case Study: Cyprus Stock Exchange). Iranian Journal of Management Studies. 2016, 9(1): 125-147. https://doi.org/10.22059/IJMS.2016.55038
OHLSON, J. A. Financial Ratios and the Probabilistic Prediction of Bankruptcy. Journal of Accounting Research. 1980, 18(1): 109-131. https://doi.org/10.2307/2490395
PLATT, D. H. and M. B. PLATT. Development of a Class of Stable Predictive Variables: The Case of Bankruptcy Prediction. Journal of Business Finance & Accounting. 1990, 17(1): 31-51. https://doi.org/10.1111/j.1468-5957.1990.tb00548.x
QU, Y. Macroeconomic factors and probability of default. European Journal of Economics, Finance and Administrative Sciences. 2008, 192-215
SAUNDERS, A. and L. ALLEN. 2002. Credit Risk Measurement: New Approaches to Value -at-Risk and Other Paradigms. second ed. New York: Wiley Finance.
SHUMWAY, T. Forecasting Bankruptcy More Accurately: A Simple Hazard Model. Journal of Business. 2001, 74(1): 101-24. https://doi.org/10.1086/209665
STIGLITZ, J. E. Some Aspects of the Pure Theory of Corporate Finance: Bankruptcies and Takeovers. Bell Journal of Economics and Management Science. 1972, 3(2): 458-482.
TRACZYNSKI, J. Firm Default Prediction: A Bayesian Model-Averaging Approach. Journal of Financial and Quantitative Analysis. 2017, 52(3): 1211-1245. doi:10.1017/S002210901700031X
TRUJILLO-PONCE A, MEDINA R, CARDONE RIPORTELLA, C. examined what best explains corporate credit risk: accounting-based versus market-based models. Journal of Business, Economics and Management. 2013, 15(2):1-24. https://doi.org/10.3846/16111699.2012.720598
ULLAH B. Financial constraints, corruption, and SME growth in transition economies. The Quarterly Review of Economics and Finance. 2019, 75: 120-132. https://doi.org/10.1016/j.qref.2019.05.009
WU, Y., GAUNT, C. and S. GRAY. A comparison of alternative bankruptcy prediction models. Journal of Contemporary Accounting & Economics. 2010, 6(1): 34-45. https://doi.org/10.1016/j.jcae.2010.04.002
ZAVGREN, C. V. Assessing the vulnerability to failure of American industrial firms: Logistic analysis. Journal of Business Finance and Accounting. 1985, 12(1): 19-45. https://doi.org/10.1111/j.1468-5957.1985.tb00077.x
ZMIJEWSKI, M. E. Methodological issues related to the estimation of financial distress prediction models. Journal of Accounting Research. 1984, 22: 59-82. https://doi.org/10.2307/2490859

ABSTRACT

The topic of predicting SME default is a research gap in the current literature, as there is a relatively limited number of research papers dealing with default prediction issues in SMEs, while an even smaller number of them deal with utilising macroeconomic variables for prediction purposes, whereas existing studies usually focus on US or UK specifics.

The aim of this work is to verify whether the accuracy of a default prediction model could be significantly enhanced by incorporating macroeconomic variables to an otherwise identical set of accounting ratios. The accuracy is assessed in terms of Area Under Curve as a metric of ROC curves, while the difference in accuracy is evaluated in terms of the DeLong test. The research is conducted on a comprehensive sample of 202,209 European (EU-28) small and medium enterprises (SMEs). The analysis conducted respects the multiperiod nature of the default process by employing the Cox regression method to derive the model, furthermore there is also a control for the effects resulting from the heterogeneity of the SME segment or industry effect.

The results show that utilising both macroeconomic and firm-specific variables led to an out-of-sample accuracy higher by 5.46 pp compared to the case when only the same set of accounting ratios would be utilised.

ABSTRAKT

Téma predikce úpadku malých a středních podniků představuje v současném literatuře mezeru v poznání, neboť existuje relativně omezený počet vědeckých článků, které pojednávají o specifikách úpadku MSP, přičemž ještě menší počet z nich se zabývá využitím makroekonomických proměnných k predikčním účelům. Existující studie se zabývají podniky z USA nebo Velké Británie.
Cílem práce je ověřit možnost, zdali přesnost modelu predikce úpadku lze statisticky významným způsobem zvýšit doplněním makroekonomických ukazatelů k jinak stejnému souboru účetních (finančních) ukazatelů. Přesnost v rámci této práce je hodnocena ukazatelem plochy pod křivkou (angl. Area Under Curve – AUC), přičemž, rozdíly v přesnosti mezi dvěma modely jsou posuzovány DeLongovým testem. Výzkum byl proveden na souboru 202.209 MSP. Provedená analýza respektuje multiperiodickou podstatu procesu úpadku, to s využitím Coxova regresního modelu k odvození modelu, přičemž jsou zohledněny i efekty plynoucí z heterogenity segmentu malých a středních podniků a oborové rozdíly.
Výsledky ukazují, že využití jak makroekonomických, tak i podnikově-specifických proměnných vedou k zvýšení přesnosti modelu mimo vzorek o 5.46 pb v porovnání s případem, kdy byly využity pouze účetní ukazatele. Kromě toho zmíněná kombinace ukazatelů vede k přesnosti vyšší o 2,15 pb, než byla dosažena v případě, že soubor účetních ukazatelů byl přehodnocen.

APPENDIX – LIST OF ABBREVIATIONS

A – asset value
AMEX - American Stock Exchange
AS – Altman, Sabato model
AUC – Area Under Curve
B – estimates of regression coefficient
C - cash
CA - current assets
CAPEX - capital expenditures
CashR - Cash Ratio
CE - capital employed
CF – cash flow
CL - current liability
CR - current ratio (CR)
D - debt
DCP - Debtor collection period
df – degrees of freedom
EAT – Earnings Taxes
EBITDA – Earnings Before Interest, Taxes, Depreciation and Amortization
EU – European Union
EU-28 – European Union Countries (28 members)
EUR – EURO
EUROSTAT - Statistical Office of the European Communities
Exp(B) – exponentiation of the B coefficient
F- test statistics of F-test
GDP – gross domestic product
GVA - Gross Value Added
H0(t) - baseline cumulative hazard function
HICP - Harmonised index of consumer prices
IE – interest expenses
IMF - International Monetary Fund
IND – industry (group)
N – number of observations
N/A - not available
NACE - Statistical Classification of Economic Activities in the European Community – From the French "Nomenclature Statistique des activités économiques dans la Communauté européenne"
NI – net income
NYSE – New York Stock Exchange
OENEG – dichotomic variable, equal to 1 if the net profit is negative for two consequent years, 0 otherwise
p – probability (of default)
P/E- price to earnings ratio, P/C – price to cash flow ratio, P/B – price to book value ratio
PD - probability of default
Personal Cost (PC)
PI - the prognostic index
pp – percentage point
QA – quick assets
R – Pearson's correlation coefficient
R^2 – determination index
RE – retained profit
ROA - return on asset ratio
ROC – Receiver Operating Characteristics
S - sales
S(t) – survival probability at time t
SB – small businesses
SE – standard error
SHP - Stock holding period
SIC - Standard Industrial Classification
sig. – significance (p-value)
SME – small and medium enterprise
ST – stock (inventory)
T – time
TA –total assets,
TC – trade creditors
TCPP - Trade creditors payment period
TL – total liability
TTA - tangible assets
USD – United States Dollar
VIF – Variance Inflation Factor
WC - working capital
Z- Z-score
Z' – revised Z-score
σ_A - volatility of assets return