# BRNO UNIVERSITY OF TECHNOLOGY

VYSOKÉ ÚČENÍ TECHNICKÉ V BRNĚ

## FACULTY OF BUSINESS AND MANAGEMENT

FAKULTA PODNIKATELSKÁ

# UTILIZING ACCOUNTING AND MACROECONOMIC VARIABLES IN THE PREDICTION OF SMES DEFAULT

VYUŽITÍ ÚČETNÍCH A MAKROEKONOMICKÝCH UKAZATELŮ V PREDIKCI ÚPADKU MALÝCH A STŘEDNÍCH PODNIKŮ

HABILITATION THESIS

HABILITAČNÍ PRÁCE

FIELD: ECONOMY AND MANAGEMENT

OBOR: EKONOMIKA A MANAGEMENT

Ing. Michal Karas, Ph.D.

BRNO 2021

**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.

**Klíčová slova**

Malé a střední podniky; predikce úpadku; Coxův model; makroekonomické ukazatele; finanční ukazatele

**Abstract**

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

The aim of this work is to verify whether the accuracy of the default prediction model could be significantly enhanced by incorporating macroeconomic variables to the otherwise same 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 DeLong test. The research is conducted on a comprehensive sample of 202,209 European (EU-28) small and medium enterprises (SMEs). The conducted analysis respects the multiperiod nature of the default process by employing the Cox regression method for deriving the model, furthermore there is also control for the effects resulting from the heterogeneity of SME segment or industry effect.

The results show that utilizing both macroeconomic and firm-specific variables has led to out-of-sample accuracy higher by 5.46 pp compared to the case, when only the same set of accounting ratios would be utilized, moreover the combination of macroeconomic and accounting ratios has led the accuracy higher by 2.15 pp comparing to when the set of accounting ratios would be reassessed.

**Key words**

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

**Bibliographic citation**

Karas, M. (2021). Utilizing Accounting and Macroeconomic Variables in the prediction of SMEs defaults, p. 149

**Sworn statements**

I declare under oath that I have prepared the thesis at hand independently and without the help of others and that I have not used any other sources and recourses than the ones stated. I declare that the citation of the sources used is complete, that I have not infringed the copyright in my work (in the sense of Act No. 121/2000 Coll., On Copyright and Rights Related to Copyright).

Brno …. 2021

Michal Karas

**Acknowledgement**

I would like to thank the following people who have helped me undertake this research: Prof. Mária Režňáková for all valuable advices, guidance, and motivation throughout all my research career. My wife Andrea for all love, support, and patience. My daughter Rozalie for all joy and hope. And to my parents, especially to my mother, gone, but not forgotten, for all love, support, and trust.

Contents

# 1 Introduction

The prediction of Small and medium enterprises (hereinafter referred as SMEs) default represents a research gap in the current state of the art in default prediction literature. This gap is a consequence of the opinion that the default prediction model could be effectively applied for prediction default in case of different business segments, time, and industry branches. Many studies show that this is not the case and motivate the effort for creating new models, while there are also studies claiming the opposite. Predicting default of SMEs is one of these issues. The need of adopting a special approach for SMEs segments in assessing the default risk is especially obvious from the limited access of SMEs to external funds, which as a consequence affect their capital structure, related working capital issues and investment decisions. This limited access can be considered as a market failure when the credit is not provided to an otherwise financial healthy business. The source of this, is application of the same metrics for SMEs and large business, which leads to inappropriate assessment of the related credit risk. A better understanding of the default risk factor of SMEs could help adopt policies that will alleviate this unfavourable situation.

Regarding the SME definition, authors usually SMEs definition coming from EU recommendation 2003/361, under which the business with less than 250 employees and with turnover lower or equal to 50 mil. EUR or total assets value lower or equal to 43 mil. EUR. For example this definition has been adopted in work on SMEs financing issues by De Moor et al (2016), Mocking et al (2016) and others. Studies focusing on non-EU SMEs, adopt slightly different definition of SMEs, e.g. Altman, Sabato (2007) adopt the definition of SMEs from Basel Capital Accord, under which a company will sales less than 65 mil. USD (approximately equal to 50 mil. EUR) are considered as SMEs. Eniola Entebang (2015) points out that the definition of SMEs significantly varies from country to country depending on factors such as the number of employees, the value of fixed assets, production capacity, basic characteristics of the inputs, level of technology used, capital employed, management characteristics, economic development, and the particular problems experienced by SMEs. In the course of this study, the business with operating revenue value less than or equal to 10 mil. EUR and total asset value less than or equal to 20 mil. EUR were included.

The aim of this work is to derive a default prediction model for SMEs, which would combine macroeconomic and firm-specific (mainly accounting) types of indicators 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 be adopting traditional approaches on a sample of SMEs. The traditional approach of utilizing accounting data as the main sources of information for assessing the default risk seems to be of limited potential, while this potential seems to be already exhausted by existing studies. A quest for potentially utilizable 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, utilizing information from the external environments seems to be of promising potential. There has been attempts to employ both macroeconomic indicators together with accounting information in the prediction of SMEs default, however the approach adopted in this work differs. The macroeconomic variables were adopted in different and more flexible manner. The current approach of addressing the issue is to adopt the macroeconomic indicator as a baseline hazard rate. Such an approach allows to utilize only one macroeconomic indicator at a time. In case of utilizing 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 specifics feature, that the estimation of model's coefficient is possible even if the baseline hazard rate is left unspecified. Under such approach, the macroeconomic indicators could enter the model as independent variables, while the baseline hazard rate is left unspecified. Such 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 panel not only allows the application, but also lead to obtaining sufficient variability of macroeconomic data and consequently benefits model robustness.

The work is organized as follows. At first, the review on 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.  Description of the methods employed in the presented research as follows. Afterwards, the results are presented, followed by hypothesis verification, discussion of the results, while the work ends with conclusion and notes about the contribution of the thesis.

# 2 Business default and its prediction

In general terms, when speaking about the risk, that the counterparty will not fulfil their obligation in full on the due date or at any time thereafter, we speak about **credit risk** (e.g. Andersson et al. 1999). While the issue of predicting credit risk in the context of banking business is often referred as credit **scoring**.

## 2.1 The term of default

The literature on credit scoring is quite rich and so is the terminology employed by the authors. There are several generic terms used in the literature for describing 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*. Berent et al. (2017) stressed that "each of these concepts can be defined/understood in many ways ", because of that, it will be described shortly:

**1)** The **financial distress** occurs when the business *is unable to meet its mature obligations* or in other words, when the "reservoir" of liquid assets drains out, while the cash flow from operations can be viewed as the net inflow of liquid assets to the "reservoir". The larger the inflows are, the lower the probability of failure is (see Beaver, 1966). It is worth to mentioned that different authors have adopted a different definition for the term of distress, the following studies defining financial distress as:

- Wruck (1990) - the situation where the cash flow of a firm it's not enough to cover is current obligations.
- Andrade and Kaplan (1998) - the first year that a firm's EBITDA is less than financial expenses, while the second condition has to be met, which is that the firm attempts to restructure its debt.
- Whitaker (1999) - the first year in which a firm's cash flow is less than the current maturity of long-term debt. Moreover, the market value is used to confirm the situation, while to confirm the distress the condition of either a negative rate or growth in market value or in industry-adjusted market value has to be simultaneously met.
- Tinoco and Wilson (2013) – this situation, which occurs whenever the firm's EBITDA is lower than its financial expenses for two consecutive years and whenever the firms suffer from a negative growth n market value for two consecutive years.

2) **Default** is a judicial decision declaring a company insolvent. In the US, it is often identified with the creditor's or management's filing for, e.g., Chapter 10 or Chapter 11[1] protection (see Berent et al., 2017). Moreover, there is a term of technical default, which "always involves the relationship between the debtor firm and a creditor class". Technical default occurs when the debtor violates a condition of an agreement with a creditor and can be grounds for legal action" see Altman and Hotchkiss (2006, p. 5).

3) **Failure** means, by economic criteria, "that the realized rate of return on invested capital, with allowances for risk consideration, is significantly and continually lower than prevailing rates on similar investments." (Altman and Hotchkiss, 2006).

4) **Business failure** was a term adopted by Dun & Bradstreet (D&B), according to whom these definitions covers business "businesses that cease operations following assignment or bankruptcy; those that cease with loss to creditors after such actions or execution, foreclosure, or attachment; those that voluntarily withdraw, leaving unpaid obligations, or those that have been involved in court actions such as receivership, bankruptcy reorganization, or arrangement; and those that voluntarily compromise with creditors." (Altman and Hotchkiss, 2006, p. 4).

5) **Insolvency** is another term describing negative firm performance and usually refers to so-called technical insolvency, which is a situation when a firm cannot meet its current obligations, signifying a lack of liquidity. Technical insolvency may be a temporary condition, although it often is the immediate cause of a formal bankruptcy declaration (see Altman Hotchkiss, 2006, p. 5).

**To sum up,** there are several terms describing the event, which is later subjected to prediction. In the following chapters, especially in those presenting the research results, the credit scoring term will be further adressed in terms of the company's financial condition, binomially defined into **default and nondefault status**. **The term of default will be addressed in the above defined manner, i.e., on the condition of business, which involves a judicial decision declaring a company insolvent**. The juridical decision is based on the given country legislative, in the example of Czech business it is Act No. 182/2006 Coll. on bankruptcy and settlement (Insolvency Law).

To be more specific, the past data were drawn on the defaulted businesses in the above stated meaning, while the focus is paid to their financial report anticipating the default by one year. The further predicted probabilities are to be interpreted such as the probability that the business is going

[1] Chapter 10 and 11 of the US Bankruptcy Code referring to possible solution of solving the bankruptcy. Chapter 10 becomes retired and in 1978 was revised an incorporated into chapter 11. Chapter 11 is often referred to as "reorganization" bankruptcy.

to meet the signs of a legally defaulted business within one year. As the default itself is decision of the insolvency court, as such the decision itself cannot be predicted.

## 2.2 Causes of corporate's default

The business default is a juridical decision made by the management in order to solve to unfavourable financial situation. For the management this is a last possible solution to deal with the situation. It should be stressed, that such a situation is not a sudden event, but was anticipated by a longer unfavourable development. Moreover, the decision is made when the future perspective about the environment are also poor, like the end of the economic crises.

The idea of predicting business default is not based on analysing the causes of the default, but rather its manifestation or traces, that could be spotted in the periods anticipating the defaults itself. However, several explanations on why the businesses default could be found in the literature.

A very often mentioned reason, is that the default is a consequence of "some type of managerial incompetence.", as mentioned by Altman and Hotchkiss (2006). Wu (2010) is more specific about the internal business causes of default, claiming that it could take a form of insufficient management skills, marketing, and inability to compete. The same author also suggests, that these factors manifestation is reflected in the company performance as recorded in the books. For this reason, accounting data or rather financial ratios are a frequent source of information for assessing the stability and viability of a business.

Chen and Hsiao (2008) suggest that the relationship between the cause of the default and its manifestation in company's book could be following:

- Companies lacking capital to manage the business and started to have problems meeting their short-term monetary obligations (Deakin, 1972; Gilson, 1989). Financially, this condition is detectable in the values of current liquidity, quick liquidity, accounts, receivable cash flow, total asset turnover, and other factors.
- Companies with a negative value of non-distributed profit for two consecutive periods or a negative growth for at least 1 year. The signs of financial problems appear in the following indicators (Altman, 1983): asset profitability, sales receipts, profits before and after taxes, and operating profit margin.
- Companies whose shares on a public stock market show an overall drop, are excluded from trading or withdrawn from the market.

The issue is more complicated by the fact, that the manifestation of default **might not share the same pattern among different types of business, industries or environments**. As summarized by Lin, Liang and Chen (2011), “Early studies tend to treat financial ratios measuring profitability, liquidity, and solvency as significant indicators for the detection of financial difficulties. However, reliance on these financial ratios can be problematic. The order of their importance, for example, remains unclear as different studies suggest different ratios as major indicators of potential financial problems.” The point of Lin, Liang and Chen (2011) was regarding the internal causes, while suggesting the need of analysing the whole context. This idea could be found already in earlier works, such as work of Mensah (1984) who came to conclusion that different sets of indicators were significant determinants of a firm's probability of distress for different periods of the business cycle, while other studies confirms this as well (e.g. Grice and Dugan, 2001). On one hand, there is no doubt that there is a relationship between business default and the business cycle, however, there is no agreement on the channels by which default and the business cycle interact, nor on how to measure the link between them Boratyńska (2016).

Altman and Hotchkiss (2006) add that the business is usually defaults due to multiple reasons, above the mentioned internal factors, there also many possible external factors, such as:

- Chronically sick industries (e.g., agriculture, textiles, department stores).
- Deregulation of key industries (i.e., airlines, financial services, health care, and energy).
- High real interest rates in certain periods.
- International competition.
- Overcapacity within an industry.
- Increased leveraging of corporate America.
- Relatively high new business formation rates in certain periods.

## 2.3 Current methodological approaches of predicting default

According to Berger and Frame (2007), credit scoring is "a statistical approach to predicting the probability that a credit applicant will default or become delinquent." The literature can differentiate the credit scoring models according to whether or not they recovery rate, what specific credit risk measures they employ, or what type of data they utilized. A brief description of the existing types of models follows.

Altman and Hotchkiss (2006) argues that there are three main components of the credit:

- The probability of default (PD),
- The "loss is given default" (LGD), which is equal to 1-recovery rate in the event of default (RR). Loss given default: Loss Given Default (LGD) is (as defined by Bellotti Crook, 2012): *the loss incurred by a financial institution when an obligor defaults on a loan, given as the fraction of exposure at default (EAD) unpaid after some period of time. It is usual for LGD to have a value between 0 and 1, where 0 means that the balance is fully recovered and 1 means the total loss of EAD.*
- Exposure to default (EAD).

While most of the attention in the literature is paid to the first (i.e. PD), while the rest of the factors are rather neglected (Altman and Hotchkiss, 2006).

### 2.3.1 A credit scoring model according to the consideration of recovery rate

Altman and Hotchkiss (2006) distinguish the following groups of credit risk model, with respect to the **treated recovery rate factor** and its relationship with the probability of default of an obligor.

**I.Credit pricing model**

- **First generation structural form model**– based on the framework of Merton (1974), under this framework the debt is represented by a zero-coupon bond. Other examples of this type of model are the work of Black and Cox (1976), Geske (1977), or Vasicek (1984). Black and Cox (1976) modified the model with respect to a more complex capital structures, represented by subordinated debt. Geske (1976) introduced interest-paying debt, Vasicek (1984) extended the model with the distinction between short and long-term debt.

- **Second generation structural form model** - the Merton approach is still used; however, these generations of models removed the unrealistic assumption that the default can occur only at the maturity of the debt when the firm's assets are no longer sufficient to cover the debt obligation. Instead of that, it is assumed that default may occur anytime between the issuance and maturity of the debt. Moreover, under these models, the recovery rate in the event of default is exogenous and independent from the firm's asset value (see Altman and Hotchkiss, 2006). Examples of such models are models of Kim and Ramaswamy, and Sundaresan (1993), Longstaff and Schwartz (1995), and others.
- **Reduced form model** – represents an attempt to overcome the shortcoming of structural models. Unlike structural model, the reduced form model does not condition default on the value of the firm, but they assume that an exogenous random variable drives the default, moreover these models treat defaults as an unpredictable Poisson event. Examples of such models could be found in the works of Litterman and Iben (1991), Jarrow and Turnbull (1995), and others.

**II.Portfolio credit value-at-risk (VaR)**

The basic ideas behind this model can enable banks to develop a credit risk models which would measure the potential loss, with a predetermined confidence level, that a portfolio of credit exposures could suffer within a specific time horizon (generally one year), see Altman and Hotchkiss (2006).

These models could be gathered into two categories:

- **Default mode model –** the default event is identified under the binomial approach, i.e., only two possible events are considered – default and survival.
- **Mark-to-market model –** the default event is identified under the multinomial approach, i.e., the default event arises in case of all possible changes of borrower creditworthiness occur (such changes are called "credit migration").

It is worth to mentioned that the Merton's framework (of structural model) **is applicable only to business listed on the stock market**, where the volatility of equity, when applied to the Black-Scholes option pricing model, is a key feature of determining the asset values, implied asset volatility, distance to distress and probability of default (see Mitchell, 2015).

### 2.3.2 A credit scoring model according to consideration of credit risk measures

The credit scoring model utilizes different proxies for measure the credit risk. Trujillo-Ponce et al. (2013) mention several alternative measures:

- **Company's financial condition –** the financial conditioned under this approach is binomial and it is either default or non-default. Such measure has been employed by studies like Altman (1968), Ohlson (1980), Hillegeist et al. (2004), Agarwal and Taffler (2008), and many others.
- **Credit rating assigned by agency rating** - such a measure has been employed by the studies of Ang and Patel (1975), Blume et al. (1998), or Demirovic and Thomas (2007).
- **Spread of bonds issued by the firm and listed on a secondary market** – this type of measure has been employed by the studies of Collin-Dufresne et al. (2001), Longstaff and Rajan (2008) or Wu and Zhang (2008).
- **Credit default swap (CDS) spread** – is considered as the most current approach (e.g., employed by Alexander and Kaeck, 2008, Das et al, 2009, or Trujillo-Ponce et al., 2013).

### 2.3.3 Credit scoring model according to treat the exposure to default

Carling et al. (2007) distinguish the credit scoring models into default risk model and portfolio risk models, based on the fact whether or not these models address the feature of exposure to default.

- **Default risk models** – perceiving the obligor as a rather isolated unit, for example, the models of Altman (1973, 1984) or Shumway (2001)
- **Portfolio risk models –** which additionally address the concertation risk resulting from a common relationship among the group of obligors, while there are groups of these models:
- **Structural models** – based on the work of Merton (1974), while an individual firm default when their asset values fall below the value of their liabilities.
- **Economic factor risk model** – where the default risk in homogenous subgroups is determined by a macroeconomic index and a number of idiosyncratic factors (Carling et al., 2007).
- **Actuarial models,** like Credit Suisse's Credit Risk+ that make no assumption regarding causality.
- **Non-parametric models** – such as the model of Carey (1998).

The need of a portfolio risk model is summarized by CreditMetrics technical document (RiskMetrics Group, 2007), according to which, the main reason is to be able to systematically address concentration risk, while the concentration risk refers to additional portfolio risk resulting from increased exposure to one obligor or groups of correlated obligors (e.g., by industry, by location, etc.).

### 2.3.4 Credit scoring model according the type of data utilized

When limiting the perspective to **the corporate** credit scoring model, we can distinguish two major groups, which are recognized in several studies (e.g., Altman, Sabato and Wilson, 2010; Trujillo-Ponce et al., 2013 or Mitchell, 2015). The **corporate** credit scoring model, applicable to **large listed** business into two major categories and their combination:

1) **Accounting based models**, which draw information from past financial statements of companies.
2) **Market based models**, which evaluate the credit risk employ the information from the financial market, especially on the volatility of the company's shares.
3) **Combination of accounting and market-based models**

It is worth to mentioned that none of these approaches are generally considered as superior over the others. First, both of these categories and their limitations will be presented, further, a more detail overview of the existing approaches will be provided.

**Ad 1) Accounting based model**

These types of models define the default in terms of financial conditions (in the meaning of Trujillo-Ponce et al., 2013), i.e., the business is either categorized as default or nondefault. While for evaluation of the credit risk is based on historical accounting data. This approach is being criticized for the historical nature of the information they take as an input and for not considering the volatility of a firm's assets in estimating the risk of default (e.g. Vassalou and Xing, 2004). On the other hand, Agarwal and Taffler (2008) provided three arguments for the favour of accounting-based approach, according to this study, corporate distress is not a sudden event, it is of low probability that due to a sudden change in economic environment a firm with good profit ability and strong balance sheet will fill for bankruptcy. Furthermore, corporate failure is the culmination of several years of adverse performance and, hence, will be largely recognized in the firm's accounting statements. Moreover, the double entry system of accounting ensures that window dressing the accounts or change in accounting policies will have minimal effect on a measure that combines different facets of accounting information in financial ratios.

The idea of predicting corporate distress based on the information contained in the financial statement was among the first analysed in the work of Beaver (1966) and Altman (1968). Altman (1968) was the first to create a multivariate prediction model. The approach, which Altman chose, became a paradigm in the upcoming years. The paradigm starts with compiling an initial set of variables (financial ratios). The absolutely fundamental idea of predicting bankruptcy by using financial ratios is that these ratios attain characteristically different values for a group of financially healthy companies and a group of companies that are approaching bankruptcy.

The following graph shows an example of such a situation, it illustrates a mean plot for return on asset ratio (ROA) for the group of non-defaulted business (referred as active) and for a distress business (referred as bankrupt). The existence of not overlapping values of financial ratios between the groups of distressed and not distress business is the vital assumption of accounting-based models.

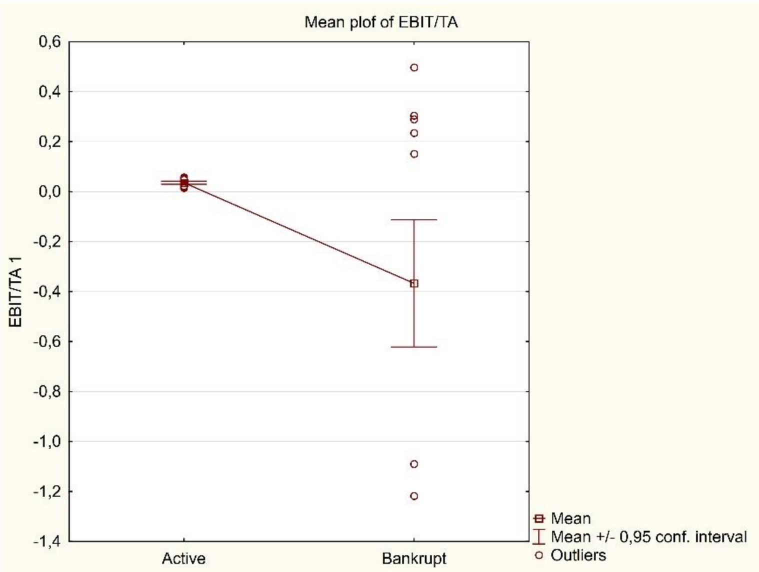



Figure 1, Example of mean plot, source: Own processing

This concept assumes that the mentioned statistically significant difference will occur in the future, which makes ex ante prediction possible. Usually this idea is utilized by employing a statistical method of classification analysis, most commonly (but not limited to) methods of linear discrimination analysis (LDA) and logistic regression (see Aziz and Dar 2006). The accounting-based model does not have to always take the form of an equation, but it takes a form of a classification tree (see e.g., Frydman et al. 1985).

To demonstrate this approach, the example of Altman (1968) model as a representative of LDA application, the example of Ohlson (1980) model as an example of logistic regression model, and the example of Frydman et al. (1985) model as an example of classification tree type model were chosen for a more detailed description.

- **The original model of Altman (1968)**

Altman's approach could be viewed as pioneering among the accounting-based models, while this approach is widespread among the accounting-based models. The Altman (1968) model is based on the method of linear discriminant analysis, while many other classification algorithms and their combinations have been employed successfully since.

Altman (1968) model takes following form:

$$Z = 0.012 \cdot \frac{WC}{TA} + 0.014 \cdot \frac{RE}{TA} + 0.033 \cdot \frac{EBIT}{TA} + 0.006 \cdot \frac{MVE}{TL} + 0.999 \cdot \frac{S}{TA}$$

Where: *Z-overall index, WC/TA – working capital over total assets, RE/TA – retained earnings over total assets, EBIT/TA - EBIT over total assets, MVE/TL – market value of equity over total liabilities, S/TA – sales over total assets*

The area of Z-score between 1.81 and 2.99 represents the zone of ignorance. For the value of Z-score over 2.99, a firm is considered as non-default, while under 1.81 is considered as threatened by default. The first version of the model was devoted to the listed firms only (because of the forth variable employing the market value of equity), however **Altman's approach is applicable to non-listed** firms as well (e.g. see Altman, 1983).

- **Revised Z-score Model**

The revised Z-score represents the original Z-score model (see Altman, 1968) which was adapted for non-listed companies (see Altman, 1983). The formula of the model is the following (see Altman and Sabato, 2013):

$$Z' = 0.717 \cdot \frac{WC}{TA} + 0.847 \cdot \frac{RE}{TA} + 3.107 \cdot \frac{EBIT}{TA} + 0.42 \cdot \frac{E}{TL} + 0.998 \cdot \frac{S}{TA} \qquad (1)$$

where: *Z' -revised Z-score, E – book value of equity*

The grey zone interval is (1.23; 2.9). For Z<1.23 the company is classified by the model as threatened by bankruptcy, for Z>2.9 is classified as not threatened by bankruptcy, i.e. financially healthy. Altman and Sabato (2007) tested a model on the sample of US SMEs over the period from

1994 to 2002. The resulted overall accuracy of the model was 68%, while type I error (a percentage of bankrupt firms classified as non-bankrupt) 25.81%.

- **The model of Ohlson (1980)**

Ohlson model was derived using the logistic regression analysis, so the output of the model is the probability of default (p). The interpretation of the model outcome is different to the one of LDA, while also the underlying assumptions of the model application differs. The probability of default is, under Ohlson's model is given by (Grice and Dugan, 2001) by:

$$p = \frac{1}{1+\exp[-(Y)]} \tag{2}$$

where:

$$Y = -1.3 - 0.4X_1 + 6.0X_2 - 1.4X_3 + 0.1X_4 - 2.4X_5 - 1.8X_6 + 0.3X_7 - 1.7X_8 - 0.5X_9\text{Y} \tag{3}$$

and $X_1$ *– log (total assets/GNP price-level index),* $X_2$ *– total liabilities/total assets,* $X_3$ *– net working capital/total assets,* $X_4$ *– current liabilities/current assets,* $X_5$ *– one if total liabilities > total assets, zero otherwise,* $X_6$ *– net income/total assets,* $X_7$ *– funds provided by operations (=EAT + depreciation)/total liabilities,* $X_8$ *– one if net income was negative for the last two years, zero otherwise,* $X_9$ *– measure of change in net income.*

For p>0.5, the business is evaluated as threatened by default, for p < 0,5 as non-threatened.

- **model of Frydman et al. (1985 )**

The model is interesting, as it was the first attempt to derive a model based on the methodology of classification trees. The advantage of this approach is its nonparametric nature, while the immunity to outlier's influence. The model takes the following form:

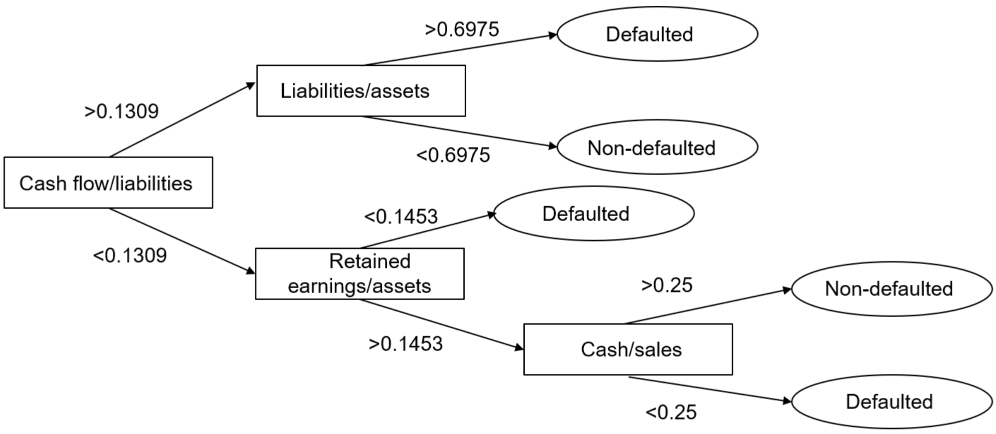



Figure 2, Model of Frydman et al (1985). Source: Own processing based on the literature Frydman et al. (1985)

**Ad 2) Market based model**

This is regarded as a more current approach, when comparing to accounting type of models. The market based model utilized the Merton approach (further referred as structural models). The advantage of market data employment or rather their potential superiority over accounting data is attributed to the theory that market price reflect investors' expectations about a firm's future performance. As a result, these prices contain forward-looking information, which is ideally suited for calculating the profitability that a firm will default in the future (see Trujillo-Ponce et al., 2013).

On the other hand, the mentioned study of Trujillo-Ponce et al. (2013) also stresses that the inefficiencies of capital markets may lead to prediction errors in market-based models.

Market based models employ the approach of Merton (1974) for modelling the default probability, under this approach, the equity of a firm is considered as a European-type call option on its assets, with the strike price being equal to the accounting value of the outstanding debt due for repayment in a defined time horizon (Trujillo-Ponce et al., 2013).

The original Merton approach is based on a very strict assumptions and its utilization is computationally rather difficult. For comparison purposes, we will describe this approach briefly.

- **The original Merton structural model approach**

Under the Merton approach, the total asset value (A) is equal to the market value of equity (E) and the value of the debt (D), while the debt is represented by a zero-coupon bond. The firm defaults at bond maturity (at time T) when the value of its assets (A) falls below the amount of debt is had to pay (D). The value of equity at time T is related to the value of the assets and debt by the following formula:

$$E_T = max(A_T - D, 0) \tag{4}$$

The model assumes that A follows a geometric Brownian motion (see Afik et al., 2016):

$$dA = \mu_A \cdot A \cdot dt + \sigma_A \cdot A \cdot dW \tag{5}$$

Where: $\mu_A$ *is the expected continuously compounded return on A,* $\sigma_A$ *is the volatility of asset return, and dW is the standard Wiener process*.

Agarwal and Taffler (2008) highlight that the parameters of volatility of asset and the assets value (i.e., $\sigma_A$ and A parameter) **are unobservable** and have to be estimated. According to the mentioned study, as a consequence of this procedure, the empirical evidence on the performance of this types of models are mixed.

The model applied the Black and Scholes (1973) formula to calculate the value of equity as a call option on its assets with expiration time (T) and strike price equal to amount of debt (D):

$$E = N(d)A - De^{-rT}N\left(d - \sigma_A\sqrt{T}\right) \tag{6}$$

Where

$$d = \frac{ln\left(\frac{A}{D}\right) + \left(r + 0.5 * \sigma_A^2\right) * T}{\sigma_A\sqrt{T}} \tag{7}$$

Where E is the value of the firm equity, r is the risk-free interest rate, N(•) us the cumulative standard **normal distribution function**. In fact, the assumption of normality of stock returns is often pointed out as an example of restrictive assumptions of Merton model (see Saunders and Allen, 2002, p. 58-61). The asset volatility and assets value parameter (i.e., $\sigma_A$ and A parameters) are estimated through simultaneous solving of equation (E) and the following equation (see Afik at, 2016):

$$\sigma_E = \frac{A}{E}N(d)\sigma_A \tag{8}$$

After the parameters $\sigma_A$ and A are estimated, the distance to default (DD) can be calculated:

$$DD = \frac{ln\left(\frac{A}{D}\right) + \left(\mu_A - 0.5 * \sigma_A^2\right) * T}{\sigma_A\sqrt{T}} \tag{9}$$

According to Afik et al. (2016) "DD may be regarded as the normalized distance between the firm asset value (A) and the face value of debt (D)". Furthermore, as the log asset value is normally distributed under the geometric Brownian motion, the probability of default (PD) is:

$$PD = N(-DD) \tag{10}$$

**Ad 3) Models combining accounting, market or even macro-based variables.**

Several studies aim at combining the above-mentioned types of variables, some of them even add macro-based variables. The result of these studies suggests that a synergic effect may occur in such a combination (e.g., Trujillo-Ponce et al., 2013, Agarwal and Taffler, 2008, Tinoco and Wilson, 2013). Trujillo-Ponce et al. (2013) used a reduced form model (in terms of Altman Hotchkiss, 2006), while the CDS swap were utilized as a proxy of credit risk. The study focuses on European markets. The accounting-based variables were drawn from the studies of Altman (1968) and Ohlson (1980). As a result, they find that there is little difference in the explanatory power of each approach and suggest that a comprehensive model, which includes both types of variables appear to be the best option.

The model of Trujillo-Ponce et al. (2013) is an example of a reduced form models which employs both accounting and market-based variables takes the following form:

$$Y_{i,t} = \beta_1 \cdot \frac{CL}{CA}_{i,t} + \beta_2 \cdot \frac{WC}{TA}_{i,t} + \beta_3 \cdot \frac{RE}{TA}_{i,t} + \beta_4 \cdot \frac{TL}{E}_{i,t} + \beta_5 \cdot coverage_{i,t} + \beta_6 \cdot \frac{EBITDA}{TL}_{i,t} + \beta_7 \cdot turnover_{i,t} + \beta_8 \cdot ROA_{i,t} + \beta_9 \cdot \frac{NI}{TA}_{i,t} + \beta_{11} \cdot DD_{i,t} + \beta_{12} \cdot \sigma_{Ei,t} + \beta_{13} \cdot \frac{P}{E}_{i,t} + \beta_{14} \cdot \frac{P}{C}_{i,t} + \beta_{15} \cdot \frac{P}{B}_{i,t} + \beta_{16} \cdot maturity_{i,t} + \beta_{16} \cdot industry\ (dummy)_{i,t} + + \beta_{16} \cdot country\ (dummy)_{i,t} + \beta_{16} \cdot year\ (dummy)_{i,t} + \varepsilon_{i,t} \quad (11)$$

Where $Y_{i,t}$ – *is the natural log of credit default swap (CDS) spread for the i^th^ firm at the end of year t, CA/CL – current assets over current liabilities, WC/TA – working capital over total assets, RE/TA – retained earnings over total assets, TL/E – debt to equity ratio, coverage – EBIT over interest payments, EBITDA/TL – EBITDA over total liabilities, turnover – asset turnover ratio, ROA – return on asset (EBIT over total assets), NI/TA – net income over total assets, DD – distance to default,* $\sigma_E$ *- annualized standard deviation of equity returns, P/E- price to earnings ratio, P/C – price to cash flow ratio, P/B – price to book value ratio, maturity – maturity of CDS contract, ε-disturbance (unobservable firms specific level and idiosyncratic error).*

The Trujillo-Ponce et al. (2013) models represents a reduced the form model, which takes form of **a general linear model,** while the applied **definition** of default (the depend variable is **continuous**), is alternative to that of accounting-based models (where the depend variable is dichotomous). The approach of Trujillo-Ponce et al. (2013) model is applicable to **listed firms** only, which might be viewed as a limitation.

Other studies aim at the comparison of accounting-based and structural models. Agarwal Taffler (2008) compared the predictive power of accounting-based Taffler (1982) models with the structural model of Hillegeist et al. (2004) and Bharath and Shumway (2004) on a sample of UK-listed business. They conclude that despite the criticism of the traditional accounting-based models, this approach is robust and not dominated by empirically by KMV type option-based models. Agarwal and Taffler (2008) also add that “the accounting-based approach produces significant economic benefit over the market-based approach.”

Mitchell (2015) in her study mentions that the structural models are mentioned as superior to accounting-based models (Default risk model, Altman’s Z-score type model), but also stressed that, as shown by Shumway (2001), the accounting model are often weakened due to the multicollinearity problem.

Tinoco and Wilson (2013) explored the possibilities of predicting financial distress (as a dichotomically defined condition) by combining accounting and market-based data together with macroeconomic data. They used the panel logit approach, which was priory suggested by Shumway (2001) to build the model, i.e., they employ market-based variables, but without adopting the Merton approach.

Their results showed that the financial data contain information which are not carried by accounting data and thus these two types of variables act as complements in the default prediction model. On the other hand, the macroeconomic variables contribute to the model only marginally.

### 2.3.5 Commonly used financial default variables (accounting based)

The aim of the work is strongly related with the topic of variables utilizable for default prediction. A following chapter provides a systematic overview of variables commonly employed for this purpose.

Default prediction models are usually employing accounting-based variables and market-based variables as mentioned by Mai et al. (2019), whereas multiple examples in literature can be found (e.g., Altman, 1968, Deakin, 1972; Martin, 1977; Altman, Haldeman and Narayanan, 1977; Altman, 2000; Ohlson, 1980; Taffler, 1982; Zmijewski, 1984; Tam and Kiang, 1992; Shumway, 2001; Sánchez-Lasheras et al., 2012 and many others). Above that, in recent years, researches have addressed the importance of governance indicators (see Liang, Lu, Tsai and Shih, 2016), country characteristics (Duompos et al., 2017). Mai et al. (2019) introduced a model employing the combination of textual data, accounting-based and market-based variables. Li and Faff (2019) created a hybrid model combining accounting-based variables with market-based variables, while the loading on both types of variables are non-monotonic, i.e., the hybrid model employs a regime-switching approach.

As the work is dealing with accounting type of ratios a further focus will be paid to this area. The macroeconomic indicators utilizable for default prediction will be discussed throughout chapter 6 Research methods and sample used.

The accounting type ratios could be divided into several groups, describing the business profitability, cash flow generating ability, liquidity, asset management and solvency. Also, other indicators, used in previous default studies could be found as well (referred as other less typical indicators). The indicators involving market type of data (especially market value of equity) is not presented, as such type of indicators are practically of no use for SMEs.

**a) Profitability ratio**

In the case of accounting-based ratios, the profit-based ratios play a significant role. Altman (1968) summarised the importance of asset profitability (EBIT/total assets) in the following way: *„Since a firm's ultimate existence is based on the earning power of its assets, this ratio appears to be particularly appropriate for studies dealing with corporate failure. "* Among Altman's variables, the return on assets (EBIT/TA) is regarded as the strongest predictor (see Shumway, 2001).

The EBIT/TA ratios represent the most common profitability ratios among the studies on bankruptcy prediction, it is the most significant indicator of the majority of Altman's variables (see Altman, 1968, 1973, 1977, 1993), it has been successfully applied by other studies as well, for example Li and Sun (2009), Mileris and Boguslaukas (2011), Psillaki, Tsolas and Margaritis (2009).

Alternatively, the profitability factor is represented by the ratio of EBITDA (i.e., the sum of EBIT and depreciation and amortization) over total assets (EBITDA/TA), which has been part of several studies, such as Perry et al (1984), Altman and Sabato (2007), Carling et al. (2007). The reason of adding the depreciation value to the EBIT may lay in fact, that the resultant indicator is a proxy of cash flow or as noted by Welc (2017), the advantage is also in making the variable less sensitive to the change of depreciation policy.

Own study (Karas, Režňáková, 2013a) showed, with respect to Czech manufacturing business, that the risk of defaulted could be spotted by analysis of EBITDA/TA value up to three years prior the bankruptcy.

A common feature of the EBIT/TA or EBITDA/TA indicators is that they compare the level of profit with the total financial sources (both equity and debt) invested in the assets. In other words, they compare the potentially available cash flow for all stakeholders, which the value of invested capital and for this reason the financial charges and taxes are not considered. However, it could be argued that the creditors perceive the worsening financial situation of the distressed business and they provide further capital only on the condition of higher interest rates, which results in an additional lowering of the net income. From this perspective, the return of assets based on net income (or rather Earnings After Taxes - EAT) may provide valuable insight in the situation of distress business. Such return on assets is given by the ratio of EAT over total assets and has been also subject to many studies on business distress (see Beaver, 1966, Deakin, 1972, Ohlson, 1980, Zmijewski, 1984, Cheng, Chen and Fu, 2006, Grunert et al., 2004, Lin, 2009, Wang and Lee, 2008).

All mentioned profitability ratios have in common that they assess the actual (or rather of the given year) results and do not consider the past results (in the value of the profit). Altman (1968) comes with idea that also the past profitability should be evaluated, while he suggests the ratio of Retained Earnings over total assets). This ratio contains the information of past profitability in terms of cumulating profit, the drawback is that such action effects simultaneously also the value of total assets, which may lower the value of added information. The past profitability ratio is also been utilized by more recent studies (e.g., Ding et al., 2008)

To sum up, the profitability is most frequently evaluated in relationship to total assets, but also some exceptions could be found, as an example, one could name the ratio of net income over operating revenue (NI/OR), which was utilized by the study of Wang and Lee (2008). It is worth to note that also in this case the numerator and denominator are being simultaneously affect by a common factor, such as the drop of sales, which may cause a drop of both operating revenue and the net profit. In case of the net profit, it further depends on the cost structure (the proportion of fixed cost).

**b) Cash flow ratio**

Jones (2016) pointed out that the return on assets (based on EBIT) is similar to cash flow returns (operating cash flow over total assets) and that none of Altman's studies test the cash flow indicators (see Altman, 2002). Above that, Jones (2016) stressed that earnings are often subject to systematic managed by companies, while the operating cash flows are relatively more difficult to manipulate as they do not involve accruals or deferrals of any kind (see Jones and Belkaoui, 2010). In the context of predicting bankruptcy, the situation is even complicated, as "*the distressed companies have a high propensity to engage in earnings management*" (see Jones, 2016). Suhaily, Rashidah and Mahenthiran (2013) specifically for Malaysian environment add that financial distress is significantly and positively related to fraudulent financial reporting.

Another issue discussed in the literature is the degree to which the financial data of distressed companies can be trusted. Berent el et al. (2017) pointed out that there has been no attempt yet in the literature dealing with bankruptcy prediction to accommodate for potential profit management. Pustylnick et al. (2017) showed that a reasonably good indication as to whether the financial statements of a company including the results of earnings management could be obtained by examination of liquidity-based financial variables and the indicators used in the DuPont Formula.

Etemadi and Tariverdi (2006) concluded that the final result of company operation is not the profit but the cash flow and added that "*while the profit is as an artificial concept, cash flow is objective and* real" (Etemadi and Tariverdi, 2006: in Kordestani et al., 2011).

Further arguments highlighting the importance of the cash flow-based indicators lie in the following facts:

1) The financial distress occurs when the business is unable to meet its mature obligations or in other words, when the “reservoir” of liquid assets drains out, while the cash flow from operations can be viewed as the net inflow of liquid assets to the “reservoir”. The larger the inflows are, the lower the probability of failure is (see Beaver, 1966). This applies especially for **operating cash flow-based** indicators.

2) Another definition of distress used the fair value of the business asset to describe the situation. In line with this definition, distress arises “*when the total liabilities exceed a fair valuation of the firm's assets with the value determined by the earning power of the assets*” (see Altman, 1968). The business value is often described in terms of the discounted cash flow methods, under which the business value is given by the present value of **free cash flow** (see, e.g., Pohl, 2017).

Cash flow-based indicators are often mentioned as strong predictors, especially in relation to total debt. Beaver (1996) was among the first to explore the potential cash flow over the total debt ratio. However, the cash flow was defined only as the sum of **net income and depreciation and amortization**. Ong et al. (2011) also conclude that cash flow over total debt is a powerful predictor of bankruptcy in the case of Malaysian companies; in his work the cash flow is defined in terms of EBITDA.

EBITDA is often applied as a simplified surrogate of operating cash flows (see Mulford and Comsikey, 2002). Study of Welc (2017) provides comparison of the power of EBITDA versus cash flow in bankruptcy prediction. Welc’s study mentioned several drawbacks of both types of measures. For example, as a pitfall of EBITDA, the omission of working capital changes is often mentioned (Fridson and Alvarez, 2002). On the other hand, the cash flow has also drawbacks (among others), such as sales of receivables, accounts in factoring transactions, or liquidation of inventories in “fire sales” (for more details see Welc, 2017).

Work of Karas, Režňáková (2020) showed, that different cash flow components can be utilized efficiently in default predicting using the hybrid model approach.

**c) Liquidity ratios**

The lack of capital resulting in the business inability to meet its short-term mature obligations represents one of the typical manifestations of imminent default (see Deakin, 1972, Gilson, 1989). For this reason, the liquidity ratios are often employed in financial distress prediction models.

The net working capital total assets ratio (WC/TA) and the current ratio (CR) represent perhaps the most common measures of liquidity in the mentioned models. Usually the WC/TA ratio are more favourable in the literature. The comparison of these two measures of liquidity could be found in Beaver (1966). According to Beaver, the CR failed in predicting the distress, as the mean value of CR for the group of distress business a year prior bankruptcy was slightly above 2, while WC/TA reached much better prediction results. Beaver's conclusion was also confirmed by many other researchers (e.g. Altman, 1968, Perry et al., 1984, Ding et al, 2008, Psillaki, Tsolas and Margaritis, 2009, Wu, Gaunt and Grey, 2010). Nevertheless, we can say that the WC/TA is dominating over CR in the financial distress studies. Many studies are employing CR as the liquidity measure (e.g., Zmijewski, 1984, Martens et al., 2008, Grunert et al, 2004, Wang and Ma, 2011).

Another example of a net working capital (NWC) based ratio is the ratio of NWC over sales (NWC/S), which has been utilized also by many studies (e.g., Beaver, 1966, Deakin, 1972, Ohlson, 1980, Martens et al., 2008, Lin, 2009 Shin and Han, 1999, 2001).

**d) Asset management**

Other frequently mentioned factor preceding bankruptcy is the lack of capital for business management (see Deakin, 1972, Gilson, 1989). Such a lack is identifiable from the asset turnover ratio, i.e., the ratio of sales over total assets (S/TA), such a factor was part of several previous studies (e.g., Altman (1968, 1977), Altman and Sabato (2007), Li and Sun (2009), Perry et al. (1984) or Ding et al. (2008). Altman (1968) highlights the usefulness of asset turnover ratios as a measure of management's ability to succeed in a competitive environment.

Other studies address the area of asset management in terms of asset composition or rather the proportion of fixed assets, which may serve as collateral (see Li and Sun, 2009, Psillaki, Tsolas and Margaritis (2009).

**e) Solvency (indebtedness) ratios**

In is general, it is often assumed that a high proportion of debt in present in the capital structure of distress business (see Zavgren, 1985, Stiglitz, 1972). For example, in Czech bankruptcy law, high indebtedness is directly stated as the legal condition of declaring bankruptcy.

The meaning of indebtedness ratios in the distress prediction model is summarized by Psillaki, Tsolas and Margaritis (2009), who claim the indebtedness feature *„is regularly used as an indicator of a company's ability to meet its long-term debt obligations and remain solvent. "*

The level of debt is often compared to the value of total asset (TL/TA), which describes the extent to which the business has issued debt for financing its assets. The TL/TA is a measure of total debt level and is very frequently employed in the distress studies (see Beaver, 1966, Deakin, 1972, Ohlson, 1980, Martens et al., 2008, Ding et al., 2008, Mileris and Boguslaukas, 2011, Psillaki, Tsolas and Margaritis, 2009, Shin and Han, 1999, 2001, Altman, 1983, Zavgren, 1985, Wang and Ma, 2011, Altman and Sabato, 2007, Carling et al., 2007).

The work of the pioneering work of Beaver (1966) showed even more value information detecting the upcoming business distress may be gained by analysing the proportion of cash flow over total debt (CF/TL). Or in other words, the level of debt may not necessarily show whether the business is clearly insolvent, but its ability of generating cash flow should be considered at the same time.

**f) Other less typical indicators**

The further development of the bankruptcy prediction models, to find more significant predictors, has led to the employment of other types of indicators, especially to those of non-ratio types.

- **Business size indicator**

All above-mentioned predictors have in common that they are a ratio indicator. The very basic idea of creation ratios indicators, which date back about 100 years back, was to make the results of differently large business comparable, i.e., to exclude the size of the business from the comparison.

The thing is that even the ratio value may systematically differ among the business of different sizes. Small businesses are considered as financially constrained, which means that the financial sources available to large business are not available to small business. Such facts may affect the indebtedness ratios and thus make them not ideally comparable between small and large business. Moreover, studies showed that small and medium companies are more vulnerable in the case of an economic recession than large companies or multinational companies (see Jin et al., 2018). The factor itself represents a significant bankruptcy predictor (see Karas, Režňáková, 2013). From the information perspective, the incorporation of the size factor in the prediction models brings the aspect of business market position (see Altman, 1977, Ding et al., 2008, Niemann et al, 2008, Psillaki, Tsolas and Margaritis, 2009). Moreover, Shumway (2001) mentions the size factors, in terms of market value of equity, as significant predictors of bankruptcy. Wu, Gaunt and Gray (2010) add that large businesses are more capable to survive a harsh period, while being less prone to bankruptcy. From the above-mentioned perspective, there is a connection between the business size and the a priori risk of bankruptcy, while there is also a potential interaction between the size of the business and value of the financial ratios. The question is whether or how this fact should be reflected in the prediction model. Historically, the efforts were to exclude the factor outside the model, usually the studies used a relatively small sample of companies, and thus the variability of the size factors was limited. Such an approach becomes to be known as match pair samples (see Altman, 1968) and the basic idea behind that lies in comparing only enterprises of identical size. Later this has been criticised due to two facts. First, the business size as such may itself be a significant bankruptcy indicator in the first place (see Ohlson, 1980; Peel and Peel, 1987). Second, as bankruptcy is a rare occurrence, this matching may influence the sample size and, thus, the number of degrees of freedom (Taffler, 1982). Nevertheless, some studies tend to incorporate the size factor into the prediction models. For example, Ohlson (1980) employed a size factors in the form of the logarithm of total assets divided by GNP price index, to ensure the size factors remains actual in the case of later applications.

- **Dichotomous type indicators**

The usage of logit and other probability-based procedures allow to incorporate dichotomous indicators into the models. For comparison, the LDA method, which was applied by Altman, allows to use only continuous variables (i.e., financial ratios or size factors).

1. Total liability exceeds total assets (1 if TL>TA, 0 otherwise); employed in the study of Ohlson (1980).
2. Negative income for two consequent periods (1 if, NI for two periods < 0, 0 otherwise); employed in the study of Ohlson (1980).
3. Ever negative the cash flow minus CAPEX (1 if cash flow – CAPEX is negative for the past 5 years, 0 otherwise), CAPEX – capital expenditure; employed in the study of Niemann et al. (2008).

- **Other less frequent indicator types**

Other indicators aim to cover more periods preceding the distress (or rather bankruptcy) in a single indicator, e.g.:

1. Ohlson's change of net income, which is given by following formula: $(NI_t - NI_{t-1})/(|NI_t| + |NI_{t-1}|)$, where: *NI – net income, t-time prior bankruptcy*.
2. Volatility indicators – Niemann et al. (2008) utilized the multiyear transformation of financial indicators, they suggested to include the 5-year volatility of EBIT or profit margin (EBIT/sales) into the potential predictor set. The 5-year volatility stands for the standard deviation of the given indicator, based on 5 years data.

# 3 SMEs specifics and related default predicting issues

The Small and Medium Enterprises (hereinafter referred as SMEs), which are under investigation in the research presented, can be viewed as a rather specific segments of businesses. On the one hand, their play a very significant role in the economy. Gupta et al. (2015), among others, pointed out that the SMEs are considered as the backbone of the global economy, furthermore, they are viewed as an important route to recovery in the aftermath of the global financial crisis of 2008–2009. De Moor et al. (2016) add that SMEs are regarded as an economy's engine for sustainable growth and stable employment. According to Eniola and Entebang (2015) "the importance of SMEs is in the evolution of economic reduction in poverty, increase in employment, output, innovation in technology, and lifting up in social position and standards is globally proven and acknowledged in emerging as well as in developed economies."

## 3.1 SMEs and financial constraints issues

A commonly discussed feature of SME is their **vulnerability to economic environment changes** and the **constraints** they must face in the course of their growth, especially in the term of availability of external financing, having consequences on capital structure, working capital management and investment decision. Due to these features, several studies focusing on predicting credit risk related to SMEs highlight the need of treating the SMEs business segment separately, as the model and metrics suitable for large businesses are unsuitable for SMEs. From the perspective of credit risk modelling, their specifics pose a challenge to researchers.

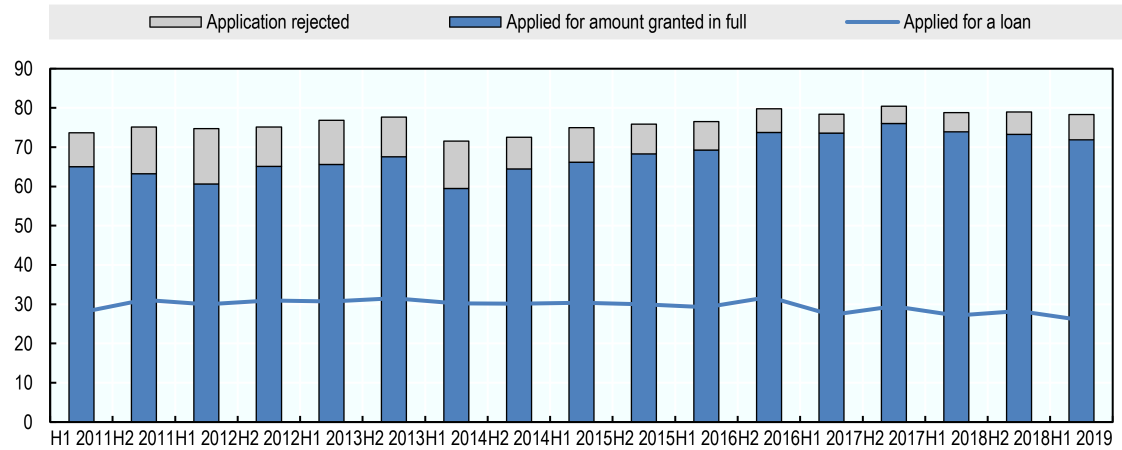
There, are several studies dealing with the topic of **financial constraints** that SMEs have to challenge (e.g., Ullah, 2019 or North et al., 2010). The result of the study differs according to the development of financial market of the SMEs' country.

Financial constraints are defined in the following manner (see Beck et al., 2006): "A firm is defined to be financially constrained if a windfall increase in the supply of internal funds results in a higher level of investment spending." The study of Beck et al. (2006) investigated the determinants of financial obstacles of the business, while the obstacles were perceived by the business themselves. They concluded that younger, smaller, and domestic firms report higher obstacles. Furthermore, they found that businesses in countries with higher level of financial intermediary development, more liquid stock markets, more efficient legal system, 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." Ullah's study focus was on Eastern European (Czech Republic included), Central Asian countries. The conclusion of the study is that in most transitions' economies where money and capital markets are underdeveloped, the growth of SMEs are severely constrained, because of their limited access to finance. North et al. (2010) pointed out that the main source of external finance to SMEs are commercial banks.

The ECB and European Commission (2019) presented a result of a survey on SMEs access to finance. Reporting that in 2019 25.96 % of SMEs applied for a bank loan, while the successful loan application reached 71.88 %, whereas the maximum was reached in H2 2017, peaking at 76%.

The study also suggests that the majority of SMES do not seek external financing, while this situation holds for the last four years.

Figure 3. Figure 4, ECB Survey on SME access to finance



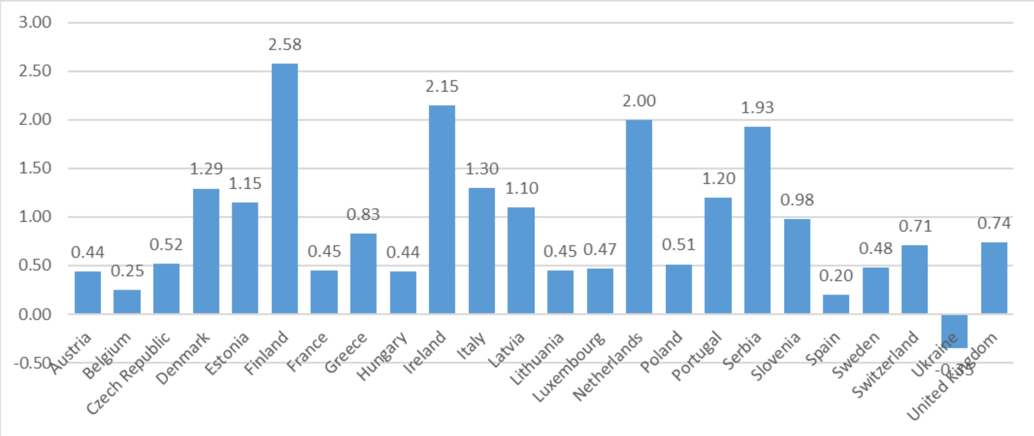

Source: ECB (2019)

The rejection rate is on the decline, where the maximum of 14.11 % was reached in 2012, while in 2019 the rejection rate was of 6.49 %. This can be interpreted as the constraints in credit supply is lowering and suggesting that a larger part of the demand for credit is being met.

From the perspective of commercial banks, SMEs are perceived as a riskier client to banks than large corporations are (e.g., see Dietsch and Petey, 2004; Saurina and Trucharte, 2004), while these studies suggest that banks should treat the SMEs client segment separately in terms of credit risk modelling.

This issue is reflected by the interest rate spread between loans to SMEs and to large firms, providing additional insight regarding SME credit conditions. In 2018, the average spread between loans to SMEs and to large firms was 0.91 pp (when focusing on EU SMEs), whereas the largest spread was recorded in Finland, reaching 2.58 %, the minimum positive value of 0.2pp was identified in Spain, while only in the case of Ukraine the spread was negative (of -0.35pp).

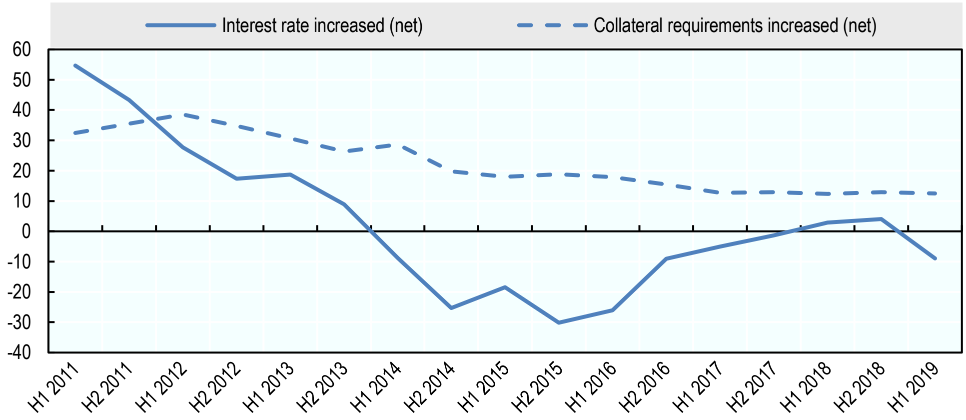
Figure 5, Interest rate spread between loans to SMEs and to large firms




Source: Own processing based on ECB (2019)

Addressing the non-EU countries, which were also subjected to the report of ECB, a much higher spread could be found, for example, in the case of Brazil the mentioned spread was of 12.9pp, while in the case of Peru it was 13.72pp. These ideas become even more important after releasing the Basel II regulatory framework, which allowed the banks to use their own measures as input for their minimum regulatory capital calculations (BCBS, 2005). North et al. (2010) were concerned about the existence of debt financing debt in the case of Scottish SMEs. They result showed those that new and early stage SMEs were not likely to seek finance than more established SMEs and more likely to encounter problems in doing so. Furthermore, over 25% of SMEs under investigation, which were experiencing difficulties in accessing external finance had been established in less than five years. The main obstacles of these younger SMEs when applying for bank loans were their lack of a trading record and insufficient collateral.  The data of ECB study shows that the increasing collateral requirement was experienced by 12.49 % of SMEs in 2019, whereas this number is relatively stable since 2017. On the other hand, a declining proportion of SMEs was experiencing increasing interest rates in the recent period.

Figure 6, ECB Survey on SME access to finance - interest rates and collateral requirements increased




Source: ECB (2019)

Altman and Sabato (2007) investigate the impact of separately modelling of credit risk for SMEs in the context of Advanced Internal Rating-Based approach of Basel II, they conclude that modelling credit risk specifically for SMEs also results in slightly lower capital requirements (around 0.5 per cent) for banks, than if one were applying a generic corporate model. Adoption of such an approach may lead to lower interest cost for SMEs customers.

Regarding the vulnerability, which is significant particularly in the case of an economic recession than large companies. Jin et al. (2018) study was concerned with businesses' recovery after the 2009 financial crisis, while their focus was on the relationship between financial recovery and financial constraints.  The results of the study showed that the recovery of SMEs is much slower than in the case of large businesses, while the recovery rate further depends on the level of external financing constrains they have to face (see Jin et al., 2018). Filipe et al. (2016) analysed the SMEs vulnerability in the context of systemic (macroeconomic) factors and came to conclusion that the larger the SMEs are, the less vulnerable to the macroeconomic situation they become, which is contrary to what Basel regulations assume.

Altman, Sabato and Wilson (2010) mentioned that the literature suggests that smaller firms both extend more credit to customers and take extended credit from suppliers when facing decline and financial stress. And further points out the importance of working capital for the survival of small business. Hudson (1987) argues that a large proportion of a firm's liability, especially for small firms, is represented by trade credit.

## 3.2 Default predictors specific for SMEs

First attempt to derive a default prediction model especially for the SMEs was the work of Edmister (1972), however the study did not explain the need of differentiating between small and large businesses. Altman and Sabato (2007) argue that the credit scoring techniques used for SMEs should be treated separately from other business segments. They derive a default prediction model especially for SMEs and compare its performance with a generic model (specifically the Z''-score) and come to conclusion that the model devoted for SMEs can reach by 30 per cent a higher accuracy level. Altman, Sabato and Wilson (2010) on a sample of UK SMEs further address the credit scoring issues specifically for SMEs by exploring the usefulness of adding nonfinancial information to enhance the model predictive ability. The mentioned study employed nonfinancial information on financial reporting, compliance, internal audit, and trade credit relationships and concluded, that this inclusion may lead to an increase in accuracy level by 13%. The study of Cressy (1992) was one of the fts dealing with the default predictors of small business. On a sample UK's small business found out that profitability and liability variables, especially the profit-to-debt ratio, are important in predicting bankruptcy. Pompe and Bilderbeek (2005) also addressed this topic with a sample of Belgian business and came to conclusion that the cash flow over total debt ratio is most useful for predicting the default of SMEs. Gupta et al. (2014) analysed whether the predictors of financial default are different in the case of domestic and international SMEs. They concluded that the same set of variables is significant in both cases, however, there is a statistical difference in their weights. This study highlights the role of ratios of cash over total assets, capital employed over total liabilities, tax over total assets, and trade creditors over total liabilities as significant predictors of SMEs default. Furthermore, they conclude that these financial ratios perform better for domestic SMEs than for international SMEs. Study of Campa et al. (2015) was concerned with prebankruptcy situation of SMEs and its impact on the earnings management tools they SMEs managers use. The main conclusion of their study is that the level of financial distress does affect the way the earnings are manipulated. The more serious the prebankruptcy distress is, the more the managers have the intent to manipulate earnings through real transactions and the less although discretionary accruals. The tools of real activity earnings managements are based on the firm's choices that effect the real conduct of the business and cash flows, such as reducing research and development expenses, cutting advertisement expenditures, altering production output, increasing revenues by offering large discounts or usually favourable credit terms to customers, and selling noncurrent assets. Thus, investors should be more cautious on earning information, especially when the firm are in financial distress (see Campa et al., 2015).

# 4 The nexus between macroeconomic development and business default

This chapter deals with the results of a literature review on direct and indirect relationships between macroeconomic conditions and various company default predicting issues. Especially with 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. Whereas, attention is also paid to a stream of literature suggesting that the model effectiveness is bounded by the prevailing macroeconomic conditions, while under alternative conditions the effectiveness is degraded. The rest of the chapter is dedicated to the application of macroeconomic factors in the course of predicting business default, with special focus to SMEs applications.

## 4.1 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 was addressed by several studies (e.g., Fama, 1986, Wilson, 1997, Carey, 1998), whereas the authors of these studies concluded that default rates **increase when the economy turns down**.

Koopman and Lucas (2005) have pointed out that the Altman-type models emphasize 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 the default rates tend to be higher in the time of recession states. Focusing on a very long data sample, the authors, among others, concluded that cyclical co-movements between GDP and business failures mainly arise at the longer frequency. Carling et al. (2007) was concerned with the survival time to default for borrowers in the business loan portfolio of a major Swedish bank. The major result of their study is that the macroeconomic variables have significant **explanatory power for firm default risk in addition to a number of common financial ratios**. The authors have 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 a little explored topic in the empirical literature. However, some improvement has been made in the recent years, the Carling's statement still partly holds.

## 4.2 The relationship between firm-level determinants of default and macroeconomic conditions

There can be found interesting studies 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) is of high attention among these studies. The importance of leverage factors as a determinant of corporate default is highlighted by recently Traczynski (2017) who showed that there are 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 business financial leverage might be the most important predictor of financial distress. This is in line with the general expectation noted by Zavgren (1985) or Stiglitz (1972) according to a high proportion of debt in present in the capital structure of distress business. The influence of macroeconomic variables of monetary policy on corporate leverage was analysed by Azofra et al. (2020), who focused on the question, whether this influence 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 less 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.

## 4.3 Accuracy of the models under alternative macroeconomic conditions

Many other authors were interested in applying the model in periods, areas, and industries different from those for which it was designed (see, for instance, Platt and Platt, 1990, Grice and Dugan, 2001; Carling et al., 2007; Wu, Gaunt and Gray, 2010; Režňáková and Karas, 2015 or Karas, Režňáková, 2017).

Grice and Dugan (2001) investigated Ohlson's (Ohlson, 1980) and Zmijewski's model (Zmijewski, 1984), concluding that the precision of both models was degraded significantly when they were applied to different data samples.

They postulate that the relationship between the financial figures and bankruptcy may change over time. This conclusion corresponds with Deakin's view (1972). Among them first, these doubts were raised by Scott (1981) who pointed out that there is a risk connected with every reduction of the potential predictors based on their significance for a given case, condition or environment. According to this study, such reduction could result in a lower robustness of the created model or in the group of predictors found (created), being ineffective when applied to different companies, a different period or a different economic environment, generally under conditions different to those that were used for deriving the model. Most of the models previously created (Grice and Dugan, 2001) were derived from data on manufacturing companies. According to some authors, these models are ineffective when used by companies in other branches. For example, Thomas, Wong and Zhang (2011) point out the need for creating models for branches such as construction, as the existing models are inappropriate for this branch. The specifics of the construction industry can be found in more detail in Heo Yang (2014), Sun Liao Li (2013), and Tserng et al. (2014). The effectiveness of the models in the branch of agriculture was addressed by Vavřina, Hampel and Janová (2013). In our previous research, I aimed to formulate a bankruptcy prediction model for Czech manufacturing companies by using accounting data (see Karas and Režňáková, 2013) and to analyse the significance of the same predictors in different industries (see Karas and Režňáková, 2015) or even in different economic environments (see Režňáková and Karas, 2015). However, our model also suffers from the same shortcomings as the models created by other authors.

Deakin (1972) found that the ranking of predictor significance changes with receding time, while Mensah (1984) found that different sets of predictors were significant determinants of a firm's probability of failure for different periods of the business cycle. These conclusions were confirmed by the work of Grice and Dugan (2001). Shumway's approach was highlighted by Berent et al. (2017) as follows: "The real economy as well as firms are driven by multiperiod processes. The models of Altman (1968), Ohlson (1980), and Zmijewski (1984), which have gained wide acceptance in academia and industry, do not follow the underlying nature of the modelled process. Shumway (2001), focusing on survival analysis, was the first to note this." We have also contributed to this topic by analysing the possibilities of incorporating the change form of indicators into the model with respect the time factor (see Režňáková and Karas, 2014; Karas and Režňáková, 2017). Most of the previously created models (Grice and Dugan, 2001) were derived from the data of manufacturing companies. Given that the values of financial ratios are industry-influenced, there is a need to construct bankruptcy models directly for individual fields of activities.

This problem is noted, for example, by Thomas, Wong and Zhang (2011), who point out the need of creating models for branches such as construction, as the existing models are inappropriate for this branch. According to Heo and Yang (2014), the specifics of construction companies show high values of liquidity ratio, high debt, and the fact that the positive cash flow generated from contracts is concentrated only in their later stages. Sun, Liao and Li (2013) add some more specifics of this sector: The construction industry is a capital-intensive industry that requires long-term project periods, large investment, and takes a long time to receive returns from the investment. Therefore, it has a different capital structure from other industries, and the same criteria used for other industries cannot be applied to effectively evaluate its financial risk (Sun, Liao and Li, 2013 in: Heo, Yang, 2014). The said opinion is also confirmed by another study (Barrie and Paulson, 1992 in: Tserng et al., 2014), as follows: "due to the distinctive operational behaviour of the construction industry, its financial characteristics also differ from other industries". Prediction of bankruptcy specifically for construction companies in the Czech Republic is dealt with, for example, by Kuběnka and Králová (2013) and Špička (2013). Špička (2013) states that the typical manifestation of bankruptcy of construction companies in the Czech Republic is high indebtedness, especially in the short term, as well as low labour productivity and negative return on assets.

Lee and Choi (2013) compared the accuracy of their model based exclusively on the data from construction companies with a similar model, but based on the data of companies from different industries. The model especially designed for construction companies reached a 6–12 % higher classification accuracy compared to the model created on the data of companies from different industries. The authors believe that the accuracy of the model would be even higher if predictor's specific for the construction industry were used.

Especially designated for agriculture companies. Bieliková, Bányiová and Piterková (2014) examined the potential of three different classification techniques for developing a new bankruptcy prediction model for agriculture companies. The mentioned study, namely, applies the methods of discriminant analysis, logistic regression, and decision tree. The best results were obtained by using the method of decision trees. Vavřina, Hampel and Janová (2013) suggest the use of production function in predicting the bankruptcy of agriculture companies. Furthermore, their study contains also the comparison with other methods, namely, Data Envelopment Analysis (DEA), logistic regression, and Z-score. The best results were obtained by the use of the logistic function, however, they mentioned that under some circumstances the method of DEA and production function could be superior to the logistic regression.

In the course of their research, they analyse the current prediction ability of Altman's Z-score and come to conclusion that the accuracy of this model in the case of agriculture companies is only 62%. The mentioned studies showed on the one hand that the agriculture business in specific in many features and on the other that for an effective bankruptcy prediction it is not possible to rely solely on the traditional bankruptcy prediction models. In a wider context, the line between corporate distress (or success) and changing macroeconomic conditions has been addressed in the studies by Kaminsky and Reinhart (1999), Edison (2003), and Knedlik (2014). These studies showed that successful companies often analyse foreign and domestic macroeconomic.

## 4.4 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 the default probability with respect to the time factor. 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 age, defined as a logarithm of the number of days the business is listed on 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), Campbell et al. (2008).

Chava and Jarrow (2004) focused on US companies traded on AMEX or NYSE or NASDAQ from 1962 to 1999, while they employed variables of Shumway (2001) and Altman (1968), Zmijewski (1984). The models were re-estimated to cover the period 1962-1991 and to test 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 industry effect for the 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 **the listed business**, from the period from 1980 to 2000, compared the information content of the discrete hazard model, re-estimated Altman (1968) and Ohlson (1980) models using the hazard model approach, with the 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 does either of the popular accounting-based measures (Altman's or Ohlson's accounting variables). A 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. Nam et al. (2008) study was conducted on a sample of Korean listed business over a period from 1991 to 2000. They applied the change in interest rates suggested by Hillegeist et al. (2001) and the volatility of 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 utilization of recent default rates, as suggest by Hillegeist et al. (2001) as the baseline hazard rate, arguing that "*can be interpreted as an actual realization of the unconditional baseline hazard rate in the previous period. Moreover, this autoregressive specification would have no forecasting power given unexpected macroeconomic shocks…*"

## 4.5 Application of macroeconomic variables to the SME business segment in course of default prediction

It is worth to mentioned that most of the applications were done on a sample of listed and thus large business and as noted by Filipe et al. (2016), most of the European SMEs are small and not satisfying 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), or Gupta et al. (2018), dealt with the application of the hazard model for the SMEs default modelling.

Gupta et al. (2015) argued that 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 is neglected by empirical studies on SME credit risk. The authors applied the discrete-time duration-depending 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 business, while by comparing the model's version, their results suggests that the segment of micro business should be separately treated 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 a period 1980-2013, while the SBA (Small Business Administration) was adopted. El Kalak and Hudson (2016) confirmed Gupta's et al. (2015) conclusion about the necessity to treat micro business separately from the rest of SMEs segments, due to different (i.e. lower) survival probabilities. On the other hand, El Kalak and Hudson (2016) points out that Gupta's approach of utilizing the insolvency rate variables as the baseline hazard rate, distorts the baseline hazard idea.

# 5 Aim of the work and methodology adopted

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

## 5.1 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 business and the attention paid to SMEs, while the specifics of SMEs segments pose a lot of **challenges to the modelling process**. In the following paragraphs, I will try to summarize the main issues on this.

- The research on business default prediction begins with accounting-based models of Altman type (see Altman, 1968), while in recent years most of the research attention gravitates towards the structural model approaches of Merton type (see e.g. Trujillo-Ponce et al., 2013), with the prevailing consensus regarding the superiority of structural approach models. It is worth to mention that the structural approach suffers from two main drawbacks, **first,** the default probability under this approach is based on the volatility of asset and the assets 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 of the European SMEs are not satisfying the entry requirements of stock exchanges, thus such an approach is for the SME segment practically inapplicable (see Filipe et al., 2016).
- The accounting type models of Altman type are often criticized due to their static nature or rather 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 to large business), for example, the financial constraints issue, 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 the **default prediction issues especially in SMEs segment.** *In other words, there is a consensus about the specificity of the SME segment, however with week reflection is the default prediction literature*. Nevertheless, several studies on default prediction issues of SMEs could be found (e.g. Edminster, 1972, Altman, Sabato, 2007, Altman, Sabato, 2010, Holmes et al.., 2010, Gupta et al, 2015, El Kalak, Hudson, 2016 or Gupta et al., 2018).

- The firm-level information utilizable for default prediction is in the case of SMEs are rather limited in comparison to large and listed business, due to the lack of financial market data, causing that the accounting data represents a main source of utilizable information. On the other hand, there is not 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.

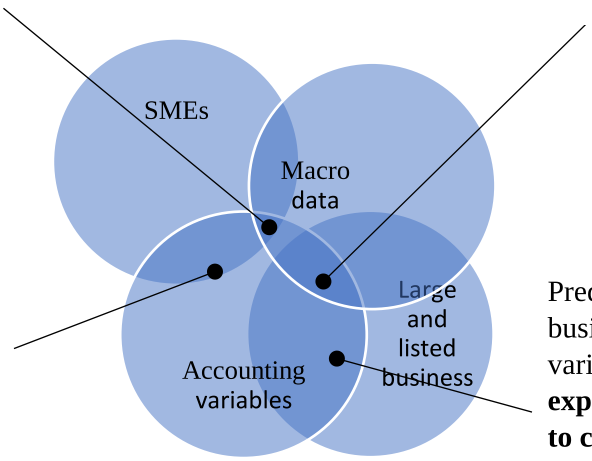
- There are studies show that systematic risk factors are correlated with macroeconomic conditions (e.g., Koopman, Lucas, 2005) or that the default rates tend to be higher when under distressed macroeconomic conditions (Fama, 1986, Wilson, 1997, Carey, 1998). **Based on this, it can be argued that the macroeconomic factors have an impact on the default probability of businesses or rather could represent an important distress predictor.**

- Several studies dealing with the accounting type of models demonstrated that the application of the model under alternative conditions (different countries, industries, period) is connected with the drop of model accuracy (see, for instance, Platt, Platt, 1990, Grice, Dugan, 2001, Wu, Gaunt, Gray, 2010). **These studies indirectly showed that environment 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 summarized in the following Venn diagram. The size of the overlap of the sets is representing rather the number of papers devoted to the given area, not the influence of one feature on another.

**Aim of the work**. Predicting default especially for SME, while respecting the specifics of this segment of business. Utilizing mainly firm-specific data and macroeconomic data – **very limited reseach sofar, increasing attention of researchers.**

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 SME, while respecting the specifics of this segment of business. Utilizing mainly firm-specific data – **several papers, however currently attaining incresing attention of researchers.**

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

***Note:*** *The SME and large businesses are presenting as disjoints set as the definition of SME clearly distinguishes 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 utilizing 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 (difference between small and medium business), industry specifics, and the multiperiod nature of the default process in the modelling phase.

To my best knowledge, this issue has not been addressed so far to this extent in the literature.

## 5.2 The Research hypothesis and their verification

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

> **Null hypothesis:** *The model combining a set of macroeconomic and* firm-specific *variables will reach a significantly* higher discrimination power**, in terms of AUC, than model utilizing only a set of* firm-specific *variables, while not employing the set of macroeconomic variables.*
>
> **Alternative hypothesis:** *The model combining a set of macroeconomic and firm-specific variables will not reach significantly* higher discrimination power**, in terms of AUC, than a model utilizing only a set of firm-specific variables, while not employing the 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 hypothesis is complicated by the very nature of researched phenomenon, which is the accuracy of the model. A model accuracy is a feature of model as a whole, whereas the accuracy (among others reason) depends on the choice of model's variables. Selecting optimal set of model variables are often done in terms of stepwise procedures, which aim is to ensure, that model contains only significant variables. **The core of the problem regarding the verification of the research hypothesis lies in fact, that it cannot be ensured that after changing the set of potential variables (i.e. adding macroeconomic variables to set of firm-specific variables), the original set of firm-specific variables will remain unchanged in the newly formed model.** Thus, potential difference between the original model and newly formed model will be simultaneously affected by two effects:

1) The effect of changing the set of variables.

2) The effect of adding 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 following:

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** – model was derived only on a set of firm-specific variables, which were included in model 1, while the variables were forced to entry into the model, i.e., the stepwise procedure was not applied.
3) **Model 3** – 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 purpose – to analyse the extent to which the macroeconomy variables improve the accuracy of the model solely based on firm-specific variables.

From a statistical point of view, all created model represents nested models. Demler et al. (2012) stressed that application of DeLong's test in case on nested models and in-sample testing will result in week 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 effects:

1. Effect of adding macroeconomic variables to the otherwise the same set of firm-specific variables (by comparing model 1 and model 2).
2. Effected gain by the combination of macroeconomic variables and accounting variables and solely firm-specific variables (by comparing model 1 and model 3).
3. Effect of changing the set of accounting variables (by comparing model 2 and model 3).

**The research hypothesis will be verified (i.e. accepted or rejected) under following schema:**

The **hypothesis** will be accepted (or rather not rejected) when three following conditions will be simultaneously met:

1) The AUC value reached in out-of-sample testing (on the test sample) of the model 1 will be higher than AUC value reached on the same sample by model 2.
2) The AUC value reached in out-of-sample testing (on the test sample) of the model 1 will be higher than AUC value reached on the same sample by model 3.
3) All the mentioned difference proved to be statistically significant in term of DeLong et al (1988) at least at 5% significance level.

Not meeting a single one of the mentioned condition would results in not accepting the presented hypothesis.

### 5.2.1 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:

$$ln[h(t)] = ln[h_0(t)] + \gamma_1 IND_2 + \cdots + \gamma_3 IND_4 + \gamma_4 SB + \beta X_{i,t} \quad (12)$$

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:

$$ln[h(t)] = ln[h_0(t)] + \gamma_1 \cdot IND_2 + \cdots + \gamma_3 \cdot IND_4 + \gamma_4 \cdot SB + \beta \cdot X_{i,t} + \delta_1 \cdot SB * X_{i,t} + \delta_2 \cdot IND * X_{i,t} + \delta_3 \cdot OENEG * X_{i,t} \quad (13)$$

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, $\gamma$;$\beta$;$\delta$ – regression coefficients.*

### 5.2.2 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 is 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 (12).

### 5.2.3 Gaining further insight into the relationship between macroeconomic variables and firm-specific ratios

Comparison of variables of model 2 and model 3 might show that an exclusion of the macroeconomic variables from a set of potential variables will lead to different choice of the variables. In such case, 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 situation.**

For further addressing this issue, a general linear model with fixed effect will be estimated.

$$Y_{j,t} = \alpha_0 + \alpha_1 \cdot status + \alpha_2 \cdot IND + \alpha_3 \cdot SB + \alpha_4 \cdot status \cdot IND + \alpha_5 \cdot status \cdot SB + \beta \cdot Z_{i,t} + \varepsilon_{i,j,t} \qquad (14)$$

Where: $Y_{j,t}$ *–a given macroeconomic indicator under investigation, Z- vector of investigated firm-specific indicators, t – given year of observation, i - given business,* α, *β – regression coefficients, ε – error term.*

The variables of *status*, *industry*, *category of company*, and *time* represent a fixed effect in the above presented model, as a different level of the rest of independent variables is to be expected for different groups of business (i.e. defaulted versus non-defaulted). Or in other words, a different level of effect of the given state of macroeconomic variables is expected to manifest in different levels of firm specific variables (accounting ratios). It is needed to mention that the idea behind formulating the above-described model is not driven by the effort to inferring the **cause-consequence relationship** between macroeconomic variables and firm specific level information (variables), but rather to analyse the relationship between the macroeconomic variables (conditions of the environment) and firm-specific information (represented by accounting variables) to put more light on composition of created model 1,2 and 3.

# 6 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.

## 6.1 Research sample

The importance of SMEs to economy is underlined by their number. By 2015, regarding the EU-28's nonfinancial businesses, the vast majority of 92.8% business were businesses with less than 10 persons employed. On the other hand, just 0.2 % of all enterprises had 250 or more persons employed and thus classified as large enterprises. In terms of value added and provided work, the large business may have a greater weight, as large businesses provided work to more than one-third (33.7 %) of the EU-28's nonfinancial business economy workforce and they generated 43.5 % of its value added (Key figures for Europe, 2018). In 2018, there were 25.079 million SMEs around Europe, while 93% of them were micro businesses, followed by small business (5.87%), whereas medium-sized business represented less (by 0.94%).

Table 1, Number of small and medium-sized enterprises (SMEs), within the European Union in 2018

| **Type of SME** | **Number (in 1000s)** | **Share (in %)** |
|---|---|---|
| Micro | 23 323,9 | 93.00 |
| Small | 1 472,4 | 5.87 |
| Medium-sized | 235,67 | 0.94 |
| Total SMEs | 25 079,31 | 100.00 |

Source: European Commission (2019)

Although the majority of businesses are of SME type, the share of SME number in the population differs significantly around European countries. The following table shows the share of SMEs in European Union countries in 2018.

Table 2, Share of SMEs in European Union countries in 2018

| **Country** | **Share (in %)** | **Country** | **Share (in %)** |
|---|---|---|---|
| Austria | 87.1 | Italy | 94.9 |
| Belgium | 94.6 | Latvia | 91.6 |
| Croatia | 90.9 | Lithuania | 93.1 |
| Cyprus | 92.9 | Luxembourg | 87.5 |
| Czech Republic | 96 | Malta | 93.1 |
| Denmark | 88.2 | Netherlands | 95.6 |
| Estonia | 91.3 | Poland | 96.1 |
| EU 28 | 93 | Portugal | 95.4 |
| Finland | 90.9 | Romania | 88.4 |
| France | 95.5 | Slovakia | 97.2 |
| Germany | 82 | Slovenia | 94.7 |
| Greece | 97.4 | Spain | 94.7 |
| Hungary | 94.1 | Sweden | 94.6 |
| Ireland | 91.9 | United Kingdom | 90 |

Source: European Commission (2019)

The highest share of SMEs in the total number of business could be spotted in Greece (97.4 %), while the lower share is present in Germany (82%).

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 bankruptcy) were analysed. As the focus is on SME, its definition needs to be specified. Usually, the authors adopt the SMEs definition of EU (EU recommendation 2003/361), under which the business with less than 250 employees and with turnover lower than or equal to 50 mil. EUR or total asset value lower than or equal to 43 mil. EUR. For example, this definition has been adopted in work on SMEs financing issues by De Moor et al. (2016), Mocking et al (2016) and others. Studies focusing on non-EU SMEs adopt a slightly different definitions of SMEs, e.g. Altman and Sabato (2007) adopt the definition of SMEs from Basel Capital Accord, under which a company will sales less than 65 mil. USD (approximately equal to 50 mil. EUR) are considered as SMEs. Eniola and Entebang (2015) points out "the definition of SMEs significantly varies from country to country depending on factors such as the number of employees, the value of fixed assets, production capacity, basic characteristics of the inputs, level of technology used, capital employed, management characteristics, economic development, and the particular problems experienced by SMEs ".

In the course of this study, the business is considered as 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 as a medium company, if it's not 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 3, 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 achieved to publish their financial statements of 2019, the number of observations per this period is significantly lower. The sample was randomly divided into learning part (70% of all observations) and testing part (30%) and later adopted the ROC curve method, this approach in relation to the hazard model was employed also e.g. by Gupta et al. (2015).

Table 4 Number of defaults per EU-28 countries

| **Country** | **Status** | | **Total** | **Country** | **Status** | | **Total** |
|---|---|---|---|---|---|---|---|
| | **Non-default** | **Default** | | | **Non-default** | **Default** | |
| AT | 2097 | 0 | 2097 | HU | 3151 | 6475 | 9626 |
| BE | 2757 | 7806 | 10563 | IE | 702 | 0 | 702 |
| BG | 1603 | 735 | 2338 | IT | 28249 | 7822 | 36071 |
| CY | 143 | 0 | 143 | LT | 877 | 115 | 992 |
| CZ | 4303 | 142 | 4445 | LU | 220 | 381 | 601 |
| DE | 12514 | 0 | 12514 | LV | 811 | 100 | 911 |
| DK | 387 | 4897 | 5284 | MT | 342 | 0 | 342 |
| EE | 794 | 1 | 795 | NL | 251 | 3403 | 3654 |
| ES | 17697 | 948 | 18645 | PL | 9061 | 1148 | 10209 |
| FI | 3502 | 1345 | 4847 | PT | 4080 | 0 | 4080 |
| FR | 22048 | 2410 | 24458 | RO | 3129 | 11025 | 14154 |
| GB | 10170 | 0 | 10170 | SE | 7763 | 9038 | 16801 |
| GR | 1674 | 0 | 1674 | SI | 1167 | 9 | 1176 |
| HR | 1036 | 1811 | 2847 | SK | 1972 | 98 | 2070 |

Source: Own calculation based on Amadeus database

The majority of the observations of the SMEs are from companies situated in Italy (17.8% of the observations), France (12.1% of the observations) and Spain (9.2% of the observations). On the other hand, the lowest number of observations are from companies situated in Cyprus (0.1% of the observations), Malta (0.2% of the observations) and Luxemburg (0.3% of the observations). The research sample covers about 0.81 % of SMEs population is EU.

### 6.1.1 Default definition adopted

Studies on credit scoring often employ several generic terms used for describing 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 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 business, as the SMEs are not a homogenous segment, to control for that, a dummy variable (called "category of company") was added.

Table 5, Share of small and medium business in the research sample

| Status/category | Small | Medium | Total |
|---|---|---|---|
| Non-default | 38 | 142,462 | 142,500 |
| default | 50,400 | 9,309 | 59,709 |
| Total | 50,438 | 151,771 | 202,209 |

Source: Own calculation based on Amadeus database

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

Table 6, Industry groups under analysis

| NACE rev. 2 Main section | Industry Dummy | Abbrev. |
|---|---|---|
| A - agriculture, forestry and fishing | Miscellaneous industries | IND 1 |
| F - construction | Miscellaneous industries | IND 1 |
| G - wholesale and retail trade; repair of motor vehicles and motorcycles | Miscellaneous industries | IND 1 |
| I - accommodation and food service activities | Miscellaneous industries | IND 1 |
| M - professional, scientific, and technical activities | Miscellaneous industries | IND 1 |
| N - administrative and support service activities | Miscellaneous industries | IND 1 |
| O - public administration and defence; compulsory social security | Miscellaneous industries | IND 1 |
| P - education | Miscellaneous industries | IND 1 |
| Q - human health and social work activities | Miscellaneous industries | IND 1 |
| R - arts, entertainment and recreation | Miscellaneous industries | IND 1 |
| S - other service activities | Miscellaneous industries | IND 1 |
| T - activity of households as employers; u0ndifferentiated goods- and service-producing activities of households for own use | Miscellaneous industries | IND 1 |
| U - activities of extraterritorial organisations and bodies | Miscellaneous industries | IND 1 |
| J - information and communication | Miscellaneous industries | IND 1 |
| B - mining and quarrying | Manufacturing and mineral industries | IND 2 |
| C - manufacturing | Manufacturing and mineral industries | IND 2 |
| D - electricity, gas, steam, and air conditioning supply | Transportation, communications and utilities | IND 3 |
| E - water supply; sewerage, waste management and remediation activities | Transportation, communications and utilities | IND 3 |
| H - transportation and storage | Transportation, communications and utilities | IND 3 |
| K - financial and insurance activities | Finance, insurance and real estate | IND 4 |
| L - real estate activities | Finance, insurance and real estate | IND 4 |

Source: Own processing based on Chava and Jarrow (2004).

The preliminary results of data analysis showed that several variables clearly exhibit extreme outlier values, to ensure that the results or the estimated parameters will not be negatively influenced by this feature, the variables under analysis were winsorized at the 1 or rather 99 percentile level. Usually the literature on credit risk or rather hazard models (e.g., Shumway, 2001, Altman, Sabato, Wilson, 2010, or Gupta et al., 2015) tends to exclude financial business from the sample, on the other hand, there are papers which aim to derive a model, which includes financial businesses, however, with industry dummy (e.g. Chava, Jarrow, 2004). As noted in Table 4, there was missing information about the industry group in case of a significant number of analysed business, thus such observations were treated as a separate group (referred as N/A).

Table 7, Number of observations per industry

| Count | | Industry group | | | | | **Total** |
|---|---|---|---|---|---|---|---|
| | | N/A | IND 1 | IND 2 | IND 3 | IND 4 | |
| Status | Non-default | 632 | 93,184 | 33,024 | 11,195 | 4,465 | 142,500 |
| | default | 8,947 | 37,218 | 6,593 | 3260 | 3,691 | 59,709 |
| Total | | 9,579 | 130,402 | 39,617 | 14,455 | 8,156 | 202,209 |

Source: Own calculation based on Amadeus database

In the case presented, the most populated industry group is group IND 1, i.e., miscellaneous industries.

### 6.1.2 Firm-specific potential variables

Reviewing previous studies with the static models might not be useful as the hazard approach analyses the significance of the variables over more than just one specific period. For these reasons, the empirical studies dealing with hazard approach and SMEs were reviewed. The information on the expected variable signs was drawn from these studies as well, in several cases the authors stated the expected sign explicitly, in other cases 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, the specific procedure will be described in the methodology section of this work.

Table 8, 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)

## 6.2 Macro-economic potential variables

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

Table 9, 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 SMEs survival as they are more likely to "*face competition from abroad and to be involved in exports and imports*". The changes of exchange rate are expected to have an adverse effect on firm survival, as change "*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 the data were drawn from Amadeus database, which quotes the exchange rate based on the data from International Monetary Fund (IMF) website and the exchange rates are referred to 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 case the interest rates are high. On the other hand, high interest causes rising costs on debt capital, 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 failure probability. 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, on contrary, the wage increase (personal cost) means a rise of the cost, thus is expected to have a rising probability of failure (see Holmes at all, 2010).

Inflation is expected to affect the probability 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 the inflation increases the risk-taking capacity of the investors and by that lower default probability (Tinoco and Wilson, 2013 or Qu, 2008). However, as acknowledged further by Qu (2008), the direction of the inflation effect on default probability has not been unequivocally established due to the 'complexity of inflation's effect on the economy. Mare (2012) noted that the high inflation rates are sing of weak macroeconomy condition 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 above-mentioned arguments, it is expected that an increase in the inflation rate is related to an increase of the firm's hazard probability.

The employment rate is expected with lower the probability of failure, employment is a proxy for demand, the higher the employment is, the higher the demand can be expected (see Holmes et al., 2010). Employment rate was drawn from EUROSTAT database and refers to the percentage of employed people between the age of 15 to 64 years 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 relation between GDP growth and firm default rate.

Beck et al. (2006) found that businesses in countries with higher level of financial intermediary development, more liquid stock markets, more efficient legal system, 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 Gupta et al. (2015) suggest treating small business separately from the media business was in their lower survival probability. The GDP per capita might serve as a proxy of financial obstacles the business has to face in their country, on the other hand, the growth obstacles seem to affect indirectly the survival probability. And for these reasons, a negative relation between GDP per capita and firm survival might be expected.

## 6.3 Specification of the adopted macroeconomic measures

The data on macroeconomic variables have been drawn mainly from EUROSTAT database, while there are more similar indicators available.

1) The **exchange rate from local currency to EUR** was employed in this study, while the data was drawn from the Amadeus database, which quotes the exchange rate based on data from the International Monetary Fund (IMF) website; the exchange rates are referred to the closing date of the statement.

2) **GVA per employee**, where GVA stands for Gross Value Added, which is defined *as the output value at basic prices, less intermediate consumption valued at purchasers' prices. GVA is calculated before the consumption of fixed capital.* (see Eurostat, 2020a), the definition per employee was adopted, as considered more suitable for comparing different countries at different stages of economic development.

3) **Personnel costs per employee** (PC) is defined by Eurostat, as *the total remuneration, in cash or in kind, payable by an employer to an employee for the work carried out. This is divided by the number of employees (paid workers), which includes part-time workers, seasonal workers etc, but excludes persons on long-term leave* (see Eurostat, 2020b).

4) The **inflation** or rather inflation rates, analysed in this work, takes the form of Harmonised index of consumer prices (HICP), defined as an annual average index, where the value of 2015 represents value of 100.

5) The **employment rate** was drawn from the EUROSTAT database and refers to the percentage of employed persons between the ages of 15 to 64 years expressed as a proportion of the total population.

6) The **GDP growth** – took the form of a percentage change on the previous years of the volume of real Gross Domestic Product (GDP).

7) **GDP per capita** – is measured by Gross domestic product at current market prices expressed in euro per capita.

8) **Interest rate** - in this work, the yield on government bonds with a maturity of ten years was adopted as the interest rate variable; such interest rates are used to define the Maastricht criterion on long-term interest rates.

The specific values of the analysed macroeconomic variables are listed in the appendix section of this work.

## 6.4 Classification methods used in default prediction model

According to the study of Aziz and Dar (2006), linear discriminant analysis and logistic regression represent the most commonly employed classification methods. The models based on these methods, however, are not capable of addressing the time factor, which is a point at which these methods are being criticized (e.g., Shumway, 2001), however, the logistic regression also exists in the multiperiod version is equivalent to Cox proportional model with discrete time, however the Cox model will be presented separately.

The point of describing the methods of linear discriminant analysis and logistic regression is given by the fact that there were selected two models for benchmark purposes – the model of Altman (1968) and Altman and Sabato (2007) which employs these two methods and above that, the original versions of the model were, in course of the research, re-estimated.

Above mentioned methods, also the methods assessing the data features, which were further utilized, are presented, specifically the tools of multicollinearity check (such as correlation coefficients or Variance Inflation Factor).

Finally, the methods of assessing model accuracy or rather the accuracy measures (total accuracy, ROC curves, and AUC values) are presented, together with DeLong test, which investigates whether two AUC values significantly differ, meaning a significant difference in model accuracy, which is the main question of the research presented.

### 6.4.1 Linear discriminant analysis

This approach was discovered by Fisher (1936) and first utilised for distress prediction models by Altman (1968). The objective of the LDA method is, according to Hebák et al. (2004), to find a linear combination of p monitored predictors, i.e., Y = bTx, where bT = [$b_1$, $b_2$, … , $b_p$] is a vector of parameters that would segregate better than any other linear combination of the groups under consideration, so that its variability within the groups would be minimal and its variability between the groups maximal.

The intragroup variability was denoted as E and the intergroup variability as B. Then the requirement for the highest intergroup variability and, at the same time, the lowest intragroup variability of variable Y could be simultaneously met by the maximisation of the F-ratio, known as Fisher's discrimination criterion (Hebák et al., 2004), which can be described in the following manner:

$$F = \frac{b^T B b}{b^T E b} \tag{15}$$

where:

$$E = \sum\sum(x_{ih} - \overline{x_h})\,(x_{ih} - \overline{x_h})^T \text{and } B = \sum\sum(\bar{\mathrm{x}}_\mathrm{h} - \bar{\mathrm{x}})\,(\bar{\mathrm{x}}_\mathrm{h} - \bar{\mathrm{x}})^T \qquad (16, 17)$$

Discriminant analysis works on the assumption of the multivariate normal distribution of data. The density of probability of a multivariate normal distribution of a variable x can be written as follows (Hastie et al., 2009, p. 108):

$$f_k(x) = \frac{1}{(2\pi)^{p/2}|\Sigma_k|^{1/2}} exp\left[-\frac{1}{2}(x - \mu_k)^T {\Sigma_k}^{-1}(x - \mu_k)\right] \tag{18}$$

where: *x is the vector of independent predictors, where x = ($x_1$, $x_2$, …, $x_p$), $\mu_k$ is the vector of mean values, $\Sigma_k$ is the covariance matrix of the $k^{th}$ group.*

Linear discriminant analysis (LDA) is a special kind of discriminant analysis that adds the assumption of identical covariance matrices ($\Sigma_k$). Under these assumptions, the discriminant rule, based on the Mahalanobis distance, can be written as follows (Hebák et al., 2004):

$$x^T\Sigma^{-1}(\mu_1 - \mu_2) > \frac{1}{2}(\mu_1 + \mu_2)^T\Sigma^{-1}(\mu_1 - \mu_2) \rightarrow \text{group 1 (e.g. non-default)} \tag{19}$$

$$x^T\Sigma^{-1}(\mu_1 - \mu_2) < \frac{1}{2}(\mu_1 + \mu_2)^T\Sigma^{-1}(\mu_1 - \mu_2) \rightarrow \text{group 2 (e.g. default)} \tag{20}$$

Where: *$\pi_1$ or $\pi_2$ is a priori probability of units belonging to the group corresponding to the group 1 or 2.*

Where the assumptions are not fulfilled for the identical covariant matrix (Σk) represents a quadratic form of discriminant analysis (QDA), a more suitable discriminant rule. However, the disadvantage of QDA is its significant sensitivity to deviations from normality, for which reason LDA is more frequently applied. The factors beneficial to the accuracy of the LDA method are: at least a roughly normal distribution of data (McLeay and Omar, 2000), negatively correlated indicators (Altman, 1968; Cochran, 1964) and the absence of extreme values (Zimmerman, 1994, 1995, 1998).

### 6.4.2 Logistic regression model

Logistic regression is a regression model for dichotomic data and is appropriate when the response ($y_i$) takes only two possible values representing success or failure. Yi is a realisation of the random variable $Y_i$, which can take a value of only one or zero with $\pi_i$ and 1- $\pi_i$. Under this assumption, Yi follows the Bernoulli distribution, which can be described as follows (see Rodriguez, 2007):

$$Pr\{Y_i = y_i\} = \pi_i^{y_i}(1 - \pi_i)^{1-y_i} \tag{21}$$

If we consider that we have *k* independent observations $y_1$, $y_2$, …, $y_k$ as a realisation of a random variable Yi, then Yi follows the binomial distribution given by:

$$Pr\{Y_i = y_i\} = \binom{n_i}{k_i}\pi_i^{y_i}(1 - \pi_i)^{n_i-y_i} \tag{22}$$

For modelling the πi, it is necessary to ensure the values are bound to an interval of <0,1>; this is done by a logit transformation, while the abbreviation **logit** stands for log-odds, i.e.

$$\eta_i = logit(\pi_i) = log\frac{\pi_i}{1-\pi_i} \tag{23}$$

By solving the equation for $\pi_i$, one will obtain:

$$\pi_i = \mathrm{logit}^{-1}(\eta_i) = \frac{e^{\eta_i}}{1+e^{\eta_i}} \tag{24}$$

Let’s further consider that the logit of the underlying probability $\pi_i$ is a linear function of the **predictors** (see Rodríguez, 2007):

$$logit(\pi_i) = x_1'\beta \tag{25}$$

where: *$x'_i$ is a vector of covariates and β is a vector of regression coefficients.*

By solving the equation for $\pi_i$, we get a nonlinear function of the predictors:

$$\pi_i = \mathrm{logit}^{-1}(\eta_i) = \frac{e^{x_1'\beta}}{1+e^{x_1'\beta}} = \frac{1}{1+e^{-x_1'\beta}} \tag{26}$$

The **logistic regression model** can be viewed as a generalised linear model with binomial errors and link logit. Interpreting the logistic regression model coefficient is rather complicated, as the derivate of the above-mentioned function with respect to $x_j$ (considering that $x_j$ is a continuous predictor) is:

$$\frac{d\pi_i}{dx_{ij}} = \beta_j \pi_i (1 - \pi_i) \quad (27)$$

From this perspective, the effect of the $j^{th}$ predictor on the probability $\pi_i$ depends on the coefficient $\beta_j$ and the value of the probability (see Rodriguez, 2007). The multiplicative form of the model tends to be employed to interpret the model parameters:

$$\frac{\pi_i}{1-\pi_i} = exp\{x_1'\beta\} \quad (28)$$

In this form, the unit change of the $j^{th}$ predictor (x) would multiply the odds by $exp(\beta_j)$.

The parameters β are estimated by maximising the following log-likelihood function (l) for (n) independent binomial observations:

$$l(\beta) = logL(\beta) = \sum\{y_i log(\pi_i) + (n_i - y_i) log(1 - \pi_i)\} \quad (29)$$

The log-likelihood function is maximised either by the Fisher scoring procedure or the Newton-Raphson method. Both methods are iterative in nature, i.e., they apply several steps repeated continuously until the convergent value is obtained.

The Newton-Raphson algorithm description is (Purba et al., 2018):

$$\beta^{(k+1)} = \beta^{(k)} - H\left(\beta^{(k)}\right)^{-1} l'^{(\beta^{(k)})} \quad (30)$$

where: *$H(\beta^{(k)})$ is Hessian matrix, l' is predicted by an algorithm using an extension of the Taylor series of $\beta^{(k)}$. $\beta^{(k)}$ is the initial estimate of the parameter.*

The Fisher scoring procedure is a rather similar procedure, although it is not based on a Hessian matrix but on an information matrix, i.e., (Purba et al., 2018):

$$\beta^{(k+1)} = \beta^{(k)} + I\left(\beta^{(k)}\right)^{-1} l'^{(\beta^{(k)})} \quad (31)$$

where: *I (β(k)) is the matrix of information*, which takes the form (Purba et al., 2018):

$$I = -E\left(\frac{\partial^2 L(\beta)}{\partial\beta\partial\beta}\right) \quad (32)$$

As noted by Altman and Sabato (2007), the benefits of logistic regression are that this methodology does not require restricting assumptions as in the case of linear discriminant analysis and, furthermore, the methodology is capable of working with disproportional samples.**Wald statistics** are a measure of parameter significance that is used in logistic regression models. To test the null hypothesis that a single parameter estimates equals 0, the Wald statistic is:

$$W = \left(\frac{\widehat{\beta_i}}{\hat{\sigma}_{\widehat{\beta_i}}}\right) \tag{33}$$

Where: $\widehat{\beta_i}$ *is the* $i^{th}$ *parameter estimate;* $\hat{\sigma}_{\widehat{\beta_i}}$ *is the estimated standard error of the* $i^{th}$ *coefficient estimate.* The Wald statistic is asymptotically $\chi^2$ distributed with 1 degree of freedom.

## 6.5 Methods of survival analysis used in default prediction models

There are several studies stress that the static approach to default probability modelling is resulting in biased estimates of the model (e.g., Shumway, 2001, Berent et al., 2017), while highlighting the need of treating the default as a multiperiod process. The methods suitable for this purpose are known and methods of survival analysis, while they are commonly used in the medicine sciences. In the course of the research, the following methods of survival analyses were adopted – the Cox semiparametric model with discrete time, the Kaplan-Maier estimator, and the log-rank test.

The application of Cox model usually begins with the estimation of survival curves and in the case that there are more groups in the research sample, by comparing the survival function of each group, while the task is whether the survival curves are not crossing, as this would harm the model assumption. This analysis is done by using Kaplan-Meier estimator and log-rank test.

### 6.5.1 Cox semiparametric proportional model

For deriving the model, Cox semiparametric proportional model approach was employed, and this approach was first adopted by Lando (1998) who was the first to model default with 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 also by other authors, e.g., Chava and Jarrow (2004) or Berent et al. (2017). Study of Berent et al. (2017) highlights the need of treating default as a multiperiod process, as "the real economy as well as firms are driven by multi-period processes", which advocates 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, also other opinions could be found, e.g., 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 the criticism, the Cox approach seem to be flexible and appropriate the multiperiod nature of the data, which is the main reason of employing the approach in course of this work.

The model was originally developed by Cox (1972), whereas the general formula of Cox model is:

$$\lambda(t; z) = exp(z\beta)\lambda_0(t) \tag{34}$$

The main problem behind Cox model is the relationship between the distribution of failure time (t) and variable z. Β 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 the logged form (see Landau and Everitt, 2004):

$$ln[h(t)] = ln[h_0(t)] + \beta_1 X_1 \ldots + \beta_q X_q \tag{35}$$

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 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 the discrete-time hazard model is equivalent to a multiperiod logit model, while the authors tend to specify the baseline hazard rate, e.g. Shumway (2001) specified the applied hazard model in the following way:

$$\phi(t; x; \theta_1; \theta_2) = \frac{1}{1+exp\left(g(t)'\theta_1 + x'\theta_2\right)} \tag{36}$$

where: *$\phi$ is the hazard function, g(t) is the natural logarithm of the number of days the business was listed on NYSE, $\theta_1$, $\theta_2$ – regression parameters, x is explanatory variable.*

Generally, there are two main approaches to the specification of the baseline hazard rate. The first is a 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 the firms has 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 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 this approach is different from the other mentioned studies (e.g., Nam et al., 2008). The main difference is that under this approach the macroeconomy variables influence the hazard rate through a shift of baseline hazard (as other explanatory variables), which seems to be useful as the analysis deals with panel data.

### 6.5.2 Kaplan-Maier estimator and log-rank test

The Kaplan-Maier estimator of survival function S(t) was developed by Kaplan Maier (1958) and it is given by:

$$\hat{S}(t) = \prod_{i=1}^{j}\left(1 - \frac{d_i}{n_i}\right) \tag{37}$$

Where $d_i$ - denotes the actual number of deaths (in this case defaulted businesses) at each of the times $t_i$ and $n_i$ denote the actual number of individuals remaining (in this case non-defaulted businesses) at each of the times $t_{i.}$

The **log-rank test** investigates the hypothesis that there is no difference in survival time between the groups studied. The log-rank test compares the observed and expected number of events for each group using the same test statistic as the chi-square test.

Test Statistic of log-rank test for comparing two groups A (in this case non-defaulted) and B (in this case defaulted) is following:

$$\chi^2_{log-rank} = \frac{(O_A - E_A)^2}{E_A} + \frac{(O_B - E_B)^2}{E_B} \tag{38}$$

Where: $E_A$ *is the expected number of events,* while it is given by*:*

$$E_{Aj=} \sum \frac{d_j n_{Aj}}{n_j} \tag{39}$$

Where: *$d_j$ is the number of events at time $t_j$; $n_A$ is the number of subjects at risk at time j in group A and $n_j$ is the total number of subjects at risk.*

### 6.5.3 Selecting model variables

For selecting the variables, I employed the same two test procedures, as was employed by El Kalak, Hudson (2016), however, with a modification which seems to be essential due to the heterogeneity of the analysed sample. The starting point of using El Kalak and Hudson (2016) procedure is deriving a univariate model for each of the analysed variables, while the variables which exhibit significant estimates and enjoy the expected sign are kept for further analysis. Such an approach has been widely adopted also by other authors (see Altman, Sabato and Wilson, 2010, Gupta et al., 2015, or Nam et al, 2008). These papers have in common that they focus on a relatively homogenous, in terms of country specifics, dataset (UK businesses or Korean businesses). On the contrary to these studies, this presented study dealt with a dataset from EU-28 country businesses with different business environments. To control for this heterogeneity, the following categorical or rather dummy variables were formed - "*category of companies*" and "*industry". The role of these variables seems to be essential as was expected, in line with the literature (see Grice and Dugan, 2001 or Gupta et al., 2015) that the financial ratios do not perform in all business' environments with the same efficiency or credit risk metrices for small and medium businesses might differ. Before adding these variables into the otherwise univariate model or rather first-step model. The Kaplan-Maier procedure was run together with the log-rank test to test the equality of survival functions to get insight into survival functions for all these categorical variables. In the case of unequal survival functions, the common approach of deriving a univariate model needs to be adjusted for the control of the difference between the groups. In such preselection, I control for the difference between groups in terms of the model constant, not in terms of the slope. Chava and Jarrow (2004) for the industry group might affect the slope. In the case of this paper, a simpler model design is chosen due to the number of groups and variables under analysis. The following steps which were taken are the same as in El Kalak and Hudson (2016) paper, i.e., running the correlation analysis or rather the multicollinearity check. In case of identification of a high correlated pair of variables, the covariate with lower chi-square value was excluded from the final multivariate model.

## 6.6 The issue of multicollinearity and methods for its detection

According to Tinoco and Wilson (2013), multicollinearity is present when there is a linear dependency between two or more independent variables in a multivariate model. The source of this issue often lies in similarly defined variables (e.g., EAT/TA and EBIT/TA). The consequences for the model are **unstable parameter estimates**, while the standard error is inflated. Freund and Littell (2000) showed how the instability of coefficient estimates is increased by the existence of multicollinearity. The study by Balcean and Ooghe (2006) stressed the fact that the method of logistic regression (logit), which is commonly applied in credit scoring areas, is extremely sensitive to multicollinearity. There are several approaches to testing the presence of multicollinearity and its significance.

### 6.6.1 Correlation coefficients

Correlation is "a connection or relationship between two or more facts, numbers, etc." (Cambridge Dictionary, 2019). There are several correlation coefficients used for measuring the strength of the mentioned relationship, e.g. Pearson's correlation coefficient, Spearman's rank correlation coefficient, or Kendall's Tau. I will briefly describe them.

**Pearson's correlation coefficient** (r) in the following form (as there are various ways of expressing it) represents the cantered and standardised sum of the cross-product of two variables (see Rogers and Nicewander, 1988) :

$$r = \frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{[\sum(X_i-\bar{X})^2 \sum(Y_i-\bar{Y})^2]^{\frac{1}{2}}} \tag{40}$$

where: *r – sample correlation coefficient,* $X_i$*;* $Y_i$ *is the* $i^{th}$ *value of random variable X or Y;* $\bar{X}, \bar{Y}$ *are mean values of X or Y.*

The following test statistics are used when the relationship realised in the sample is strong enough to model the relationship in the population. The null hypothesis of the test is that the population correlation coefficient $\rho = 0$. The test statistics have a t-distribution with n-2 degrees of freedom:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{41}$$

where: *n is the number of observations.*

**Spearman's rank correlation coefficient**, unlike Pearson's, represents a non-parametric approach to measuring correlation. The coefficient, when there are no tied ranks, is given by:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{42}$$

where: $d_i^2$ *is the difference between two ranks of each observation, n is the number of observations*.

The significance of this correlation coefficient can also be tested, while there are two approximations:

a) for values of **n** > **20,** the test statistic is

$$\frac{r_s\sqrt{n-2}}{\sqrt{1-r_s^2}} \tag{43}$$

where: $r_s$ *is Spearman's sample correlation coefficient*. The statistics have approximately a t-distribution with n-2 degrees of freedom.

b) for values of **n** > **40,** the test statistic is $r_s\sqrt{n-1}$ (44)

The statistics have approximately an N(0,1) distribution. One of the limitations of Spearman's approach is the assumption of the presence of no tied ranks in the sample. If tied ranks are present (i.e., two observations columns of the same rank), a possible solution lies in the use of a different correlation coefficient, e.g., Kendal's Tau.

### 6.6.2 The variance inflation factor (VIF)

VIF is a tool for assessing the degree of multicollinearity. The VIF is part of the General Linear Model (GLM) diagnostic method. Its aim is to evaluate how much of the independent variable variance could be explained by the combination of other independent variables (see Craney and Surles, 2002). The GLM could be rewritten in the form:

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \ldots + \beta_k \cdot X_k + \varepsilon \tag{45}$$

Where: *Y – dependent variable,* $X_1$, $X_2$, *…,* $X_k$ *– independent variables, ε – error term.*

The VIF value for a given variable (e.g., $X_1$) is estimated in two steps. In the first step, the following regression model, with $X_1$ as the independent variable and with $X_2$, X3,… $X_k$ as dependent variables, is estimated:

$$X_1 = \alpha_0 + \alpha_1 \cdot X_2 + \alpha_2 \cdot X_3 + \ldots + \alpha_k \cdot X_k + \varepsilon_2 \tag{46}$$

In the second step, the determination index ($R^2$) of the $i^{th}$ model is estimated. The VIF is then given by the following $R^2$ transformation:

$$VIF = \frac{1}{1 - R_i^2} \tag{47}$$

For VIF values lower than 10 or 4, the presence of multicollinearity is considered nonsignificant (see Kim and Kang, 2010). On the other hand, VIF values higher than 10 or 4 mean a situation in which it is possible to explain the variance of a given independent variable (e.g., $X_1$) by the remaining independent variables ($X_2$, $X_3$, …, $X_k$), while 90 % or 75 % of the $X_1$ variance would be explained. In other words, it is possible in such cases to exclude X1 from the original model while retaining 90 % or 75 % of its variance.

## 6.7 Accuracy measures

There are several ways how to measure the classification performance of the distress prediction model or in general the performance of the produced classifier. We will introduce several most commonly used and mentioned limitations of their application. These measures are based on the confusion matrix. Bradley (1997) describes the confusion matrix as a "*matrix which is a form of a contingency table showing the differences between the true and predicted classes for a set of labelled examples*". The confusion matrix could take the following form (see Sokolova et al., 2006):

Table 10, Confusion matrix

| **Class/Recognized** | **As Positive** | **As Negative** |
|---|---|---|
| Positive | tp (true positive) | fn (false negative) |
| Negative | fp (false positive) | tn (true negative) |

Source: Own processing based on Sokolova et al. (2006)

Based on the confusion matrix, several other important terms are defined – the type I and II error, specificity and sensitivity and total accuracy.

### 6.7.1 Type I and type II errors

The *false positive (fp)* is also called type I error, while the *false negative (fn)* is entitled as type II error. The type I error case of distress prediction may materialise in a situation when a bankruptcy-prone company is assessed as financially stable, while the type II error situation would be the opposite, that is, evaluating a financially stable company as facing bankruptcy.

In the first situation, the creditor who would grant the credit lost all issued capital, while in the second situation the creditor would lose only the potentially earned interest. By Zhou and Elhag (2007), the type I error is 2 to 20 times more serious (thus costly) than the type II error.

### 6.7.2 Sensitivity and specificity

The *true positive* **rate** is often called sensitivity, while the *true negative* **rate** is marked as specificity. These rates are given by (see Sokolova et al., 2006):

$$sensitivity = \frac{tp}{tp+fn} \tag{48}$$

or rather

$$specificity = \frac{tn}{fp+tn} \tag{49}$$

The sensitivity and specificity are both varying with the variation of the decision threshold (See Bradley, 1997). The decision threshold in the case of distress prediction models is sometimes referred as the cut-off score.

### 6.7.3 Total accuracy

The (total) accuracy represents the most commonly applied measurement of prediction accuracy.

$$accuracy = \frac{tp+tn}{tp+fn+tn+fp} \tag{50}$$

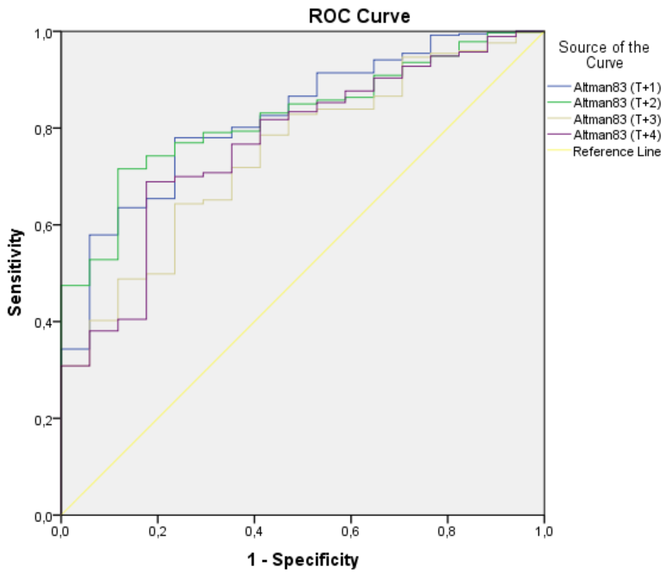
The disadvantage of the accuracy measure is in fact, that the number of correct classifications of different classes is not distinguished (see Sokolova et al., 2006).

Nevertheless, this measure is very popular, as noted by Ling, Huang and Zhang (2003), the accuracy "has been used as the main and often only evaluation criterion for the predictive performance of classification learning algorithms ". On the other hand, it is worth to mention that the explanatory power of the accuracy measure is often criticized from the perspective of the sample proportion. Berent et al. (2017) summarized that: "an accuracy rate, defined as the percentage of correctly designated ratings, of 95%, may indicate a very poor model performance in the case of a big, representative sample of thousands of firms with, say, 3% of bankrupt companies, as well as quite an achievement for a model with matched pairs."

### 6.7.4 Receiver Operating Characteristics (ROC) and Area Under Curve (AUC)

Due to the drawbacks of total accuracy, the ROC approach and the corresponding AUC values as the performance measures of learning algorithm (and of that by distress prediction model) are becoming preferable measures of learning algorithm performance. Bradley (1997) summarizes the need, which was behind the creation of ROC, as follows: "*there is a need for a single measure of classifier performance that is invariant to the decision criterion selected, prior probabilities, and is easily extended to include cost/benefit analysis.*" And added that such features where not met by the total accuracy measure or specificity and sensitivity.

Figure 7, An example of ROC curve




Source: Own processing using SPSS

### 6.7.5 Estimating the AUC value

There are several ways to estimate the AUC values, Bradley (1997), among others mentioned that it is possible to calculate AUC by assuming that the underlying probabilities of predicting negative or positive are Gaussian and the AUC can be fitted by Maximum Likelihood Estimation. Another way to calculate the AUC is to use a trapezoidal approach, which does not pose any assumptions on the underlying probability distribution.

The trapezoidal integration can be applied by the following formula (see Bradley, 1997):

$$AUC = \sum_i \left\{(1 - \beta_i \cdot \Delta\alpha) + \frac{1}{2}[\Delta(1 - \beta) \cdot \Delta\alpha]\right\} \quad (51)$$

where

$$\Delta(1 - \beta) = (1 - \beta_i) - (1 - \beta_{i-1}) \quad (52)$$

$$\Delta\alpha = \alpha_i - \alpha_{i-1} \tag{53}$$

where: *1-β – is the sensitivity, 1-α – is the specificity*

As noted by Hanley McNeil (1982), trapezoidal approach systematically underestimates the AUC, which is because all points of ROC are connected with a straight line instead of smooth concave curves. However, in case of a reasonable number of points, the underestimation should not be too severe. Hand Till (2001) build on the work of Bradley (1997) and present a different approach to calculate AUC, which is equivalent to Wilcoxon statistic rank test. Under this approach, the AUC for a G classifier is given by:

$$A = \frac{S_0 - \frac{n_0(n_0+1)}{2}}{n_0 n_1} \tag{54}$$

where: *$n_0$ and $n_1$ are the numbers of positive and negative examples, respectively*,

and

$$S_0 = \sum r_i \tag{55}$$

Where: *$r_i$ is the rank of the ith positive example in the ranked list*.

Ling and Zhang (2002) showed that if we build a classifier, which maximizes AUC, instead of accuracy, such a classifier would not only produce higher AUC but also higher accuracy than would be achieved in the opposite case (i.e. building a classifier which maximizes accuracy at the first place).

### 6.7.6 Comparing two ROC curves

For comparing two or more ROC curves, a DeLong test is a commonly adopted approach (see DeLong et al., 1988). The aim of comparing ROC curves is often motivated by measuring the increase of model discrimination ability. The null hypothesis of DeLong test, under the assumption of comparing two ROC curves (i.e., two AUC values) can be formulated as follows:

$$H_0: \theta_1 = \theta_2 \tag{56}$$

Where: *θ is the true AUC, to derive the empirical AUC, it is needed to determine its probability distribution.*

The test statistic under the null hypothesis has a standard normal distribution N(0,1) while taking the following form:

$$z = \frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{L_T\left(\frac{1}{m}S_{10} + \frac{1}{n}S_{01}\right)L}} \tag{57}$$

Where: $\hat{\theta}$ *is the empirical AUC; L = (1 -1)* $^T$*;* $S_{10}$ *and* $S_{01}$ *are K x K matrices (in this case of two AUCS it is 2x2 matrices), with the (r, s)th element* defined as follows:

$$(S_{10})_{r,s} = \frac{1}{m-1}\sum_{i=1}^{m}\left[V_{10}^{r}(X_i) - \hat{\theta}^r\right]\left[V_{10}^{s}(X_i) - \hat{\theta}^s\right] \quad (58)$$

and

$$(S_{01})_{r,s} = \frac{1}{m-1}\sum_{i=1}^{m}\left[V_{01}^{r}(Y_j) - \hat{\theta}^r\right]\left[V_{01}^{s}(Y_j) - \hat{\theta}^s\right] \quad (59)$$

where

$$V_{10}^{r}(X_i) = \frac{1}{n}\sum_{j=1}^{n}\psi\left(X_i^r, Y_j^r\right) \quad (60)$$

and

$$V_{01}^{r}(Y_j) = \frac{1}{n}\sum_{j=1}^{m}\psi\left(X_i^r, Y_j^r\right) \quad (61)$$

The empirical AUC (i.e. $\hat{\theta}$) given by the trapezoidal rule is given by:

$$\hat{\theta} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\psi(X_i, Y_j) \quad (62)$$

Where

$$\psi(X_i, Y_j) = \begin{cases} 1 & Y < X \\ \frac{1}{2} & Y = X \\ 0 & Y > X \end{cases} \quad (63)$$

Where: *m – is the number of individuals truly belonging to group 1 (e.g., default business), n – is the number of individuals truly belonging to group 2 (e.g. non-default business),* $X_i$ *or rather* $Y_i$ *– probability of i-the individual belonging to group 1 or rather group 2, where the probability is given by a binary classifier.*

# 7 Results

In this chapter, the results of deriving the new default prediction model will be presented. The results will be compared to results 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 either with the original setting and with re-estimated coefficients. The model's re-estimation was done on the learning samples.

## 7.1 Surviving times of small and medium companies

The underlying idea of Cox regression is to analyse the time to an event (in case presented the default of a company). From this perspective, a closer look at the surviving time of the companies under analysis is useful. The surviving time was analysed separately for small and medium companies, as a different serving time is expected, especially from the perspective mentioned by Gupta et al. (2015).

Table 11, Life table of the analysed companies

| **Category of company (SB)** | **Interval Start Time** | **Cumulative Proportion Surviving at End of Interval** |
|---|---|---|
| Small | 0 | 0.66 |
| | 5 | 0.47 |
| | 10 | 0.38 |
| | 15 | 0.32 |
| | 20 | 0.29 |
| | 25 | 0.26 |
| | 30 | 0.00 |
| Medium | 0 | 0.99 |
| | 5 | 0.97 |
| | 10 | 0.96 |
| | 15 | 0.95 |
| | 20 | 0.95 |
| | 25 | 0.94 |
| | 30 | 0.85 |

Source: Own calculation based on Amadeus database

In the case of small business, the highest default rate could be observed in the first five years of their existence, while in the end of this period only 66% still survive, while after the next five years, only 38% of the business survive (i.e. till 10 years after establishing the business).

The median surviving time for small business is 9.31 years. In the case of media business, the situation is significantly different, where 99% of the business survives till the end of the first five-year period, while till the end of the second five-years period (i.e. till 10 year after the business was established), still 97% of the business are still active. The median surviving time of media business is 30.00 years. A more detailed view is provided by the following chart, which shows the survival function of the companies under analysis.

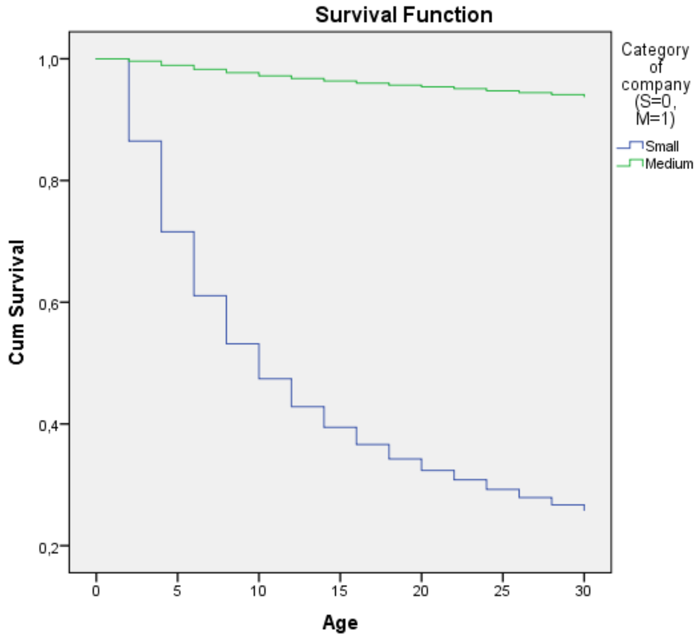


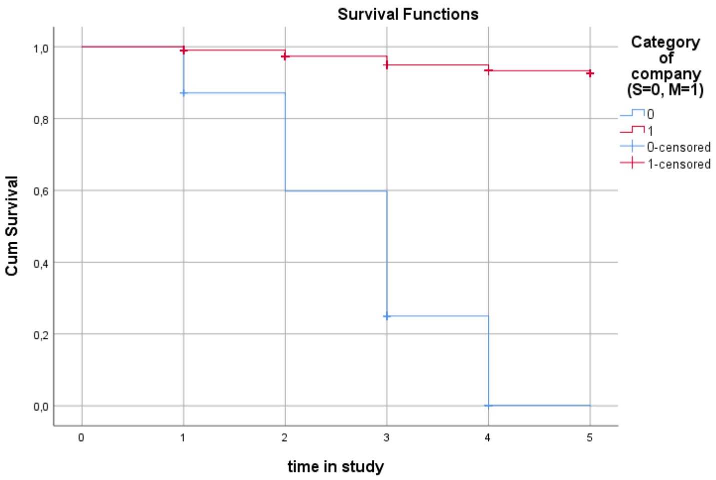

Figure 8, Survival functions of analysed companies

The data in studies are left censored, which means that the data do 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, the further analysis will deal with the time of study.

## 7.2 Initial step of deriving the hazard model - survival function comparison

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

Figure 9, Survival functions of SMEs




Source: Own calculation based on Amadeus database

According to the results of log-rank test, there are significant differences between the small and medium business, while there are significant differences between businesses operating in different industries and regions of Europe as well. The procedure was applied for the OENEG indicator, which differentiated between the business with negative net profit and those with positive net profit.

Table 12, 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

## 7.3 Initial discrimination analysis

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

Table 13, 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 (-) in the case 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 (+) sign. Among the excluded variables, a significant part is represented by profitability ratio (EBIT/CE, EBIT/TA, and NI/E) and liquidity ratio (CA/CL, log(CA/CL)) and asset management ratio (S/TA, S/TTA).

Two of the profitability ratios EBIT/CE (i.e., EBIT over capital employed) and NI/E (net income over equity) over that exhibit an unexpected sign. As both indicators dealt implicitly with capital structure, an analysis of the capital structure differences between default and nondefault groups of business could provide some explanation. For this purpose, the value of TL/TA (total liabilities over total assets) indicator was put under analysis, the mean value of the indicator on a sample of nondefault businesses was 0,66 while the mean value of the same indicator on the sample of default business was 2,02 (see for details appendix), which means, that an average default 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), the expectation is that a higher value of such ratio is related to a higher probability of default. Based on the descriptive statistics of the sample, the mean of the mentioned ratio for nondefault business is 0.2801, while in the case of default business it is 0.2577. A potential explanation for this could be drawn from the results of McGuinness et al. (2018), who found that the SMEs survival can depend on the extension of additional trade credit and/or relax payment terms by their unconstrained creditors.

It could suggest that the nondefault business are able to relax the payment terms for their customers, however based on the available date it cannot be distinguished from the overdue credit. Moreover, the table shows the ratios at which the estimated variable sign was not meeting the expectation and due to this reason, the given ratio was excluded from the sample.

Table 14, 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) as one of the most significant indicators of bankruptcy, e.g., Cathcart et al. (2020), according to whom financial leverage might be the most important predictor of financial distress of unlisted business, furthermore Zavgren (1985) or Stiglitz (1972) suggest that a high proportion of debt in present in the capital structure of distress business. In the case presented, the TL/TA reached a coefficient estimated with a negative sign, whereas the magnitude of the coefficient is relatively low. A possible explanation could be seen is the financial constraints aspect, often mentioned in relation to SMEs, showing that the external sources of finances in the form of debt are often not accessible for the SMEs.

Due to the coefficient sign reason, other common predictors of default were excluded, such as return retained earnings over assets (RE/TA) or net working capital over total assets (WC/TA).

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

Table 15, 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 substa

ntial 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 relation among 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, the 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 aim is to model the time to default, the time which is the subject to this indicator, is defined in different terms. The age of the business is time since the business was established, not time since the business enters the study.

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

A similar procedure was conducted for macroeconomic factors under analysis.

Table 16, 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 analysed macroeconomic variables are significant at the 1% level, except for the GDP annual growth rate. The possible explanation might be that the analysed period was a relatively stable period of time for EU SMEs, within only Greek economy in 2015 and 2016 turning into recession and in 2015, so did Croatia, Cyprus, Finland and Serbia, which experienced negative annual GDP growth. Speaking about 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 business listed in 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 are enjoying the expected sign, thus they will be 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 a high correlated pair of variables (the correlation coefficient was higher than 0.7 or lower than -0.7).

Table 17, 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 the firm specific variables, there were four highly correlated pairs of variables identified. The first pair is dealing with return on assets, according to the Wald statistics, the EBITDA/TA represents a preferable measure than NI/TA. A possible cause for this is the different level of corporate taxation among EU countries.

The second correlated pair is composed of Debtor collection period (DCP) indicator and the ratio of current assets over sales, these indicators have in common the features of sales and accounts receivables (debtors). The DCP represents a more significant measure, that is why the CA/S will be excluded from further analysis.

Third, correlated pair of variables measure 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) reached a more significant estimate, thus staying for further analysis.

Fourth, correlated pair of variables deals with business liquidity, the pair consists of quick ratio (QR) and the cash ratio (CashR). Moreover, the current ratio (CA/CL) was among the analysed ratios, however this ratio's estimated was not significant at the 5% level. In the study presented, the Quick ratio (QR) represents a more significant measure, that is why this ratio will be further analysed.

Furthermore, the multicollinearity check was also conducted, as the given variable may be explained not only but the other variables, but a combination of several variables. For this purpose, the Variance Inflation Factor (VIF) approach was adopted.

Table 18, 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 the 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.

## 7.4 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.

Final step is in testing the model and comparing the model 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).

### 7.4.1 Models' overall statistics

Model 1 was estimated in a stepwise manner by employing a backward elimination procedure using conditional likelihood ratio (LR) statistics as a criterion, which is considered as least prone to error measurement. As a result, the model is significant at 1% level.

Model 2 was derived by using the same variables (with the exception of macroeconomic variables) as model 1, the model is significant as 1% level as well.

Table 19, Models' overall statistics

| Model version | -2 Log Likelihood | Overall (score) | | | Change from Previous Block | | |
|---|---|---|---|---|---|---|---|
| | | Chi-square | df | Sig. | Chi-square | df | Sig. |
| 1 | 72958.984 | 50541.546 | 23 | 0.000 | 17242.308 | 23 | 0.000 |
| 2 | 130834.524 | 59904.249 | 20 | 0.000 | 21488.126 | 20 | 0.000 |
| 3 | 76800.56 | 48328.41 | 22 | 0.000 | 13587.01 | 22 | 0.000 |

Source: Own calculation based on Amadeus database

### 7.4.2 Details of model 1 estimates

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 the predictive power of the model.

Table 20, 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 21, 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 |
| Interactio n terms | 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 of the final model, which is in line with expectation (see Chava and Jarrow, 2004 or Gupta et al., 2015). The industry effect influences only the model intercept, not the slope. On the other hand, the category of company is not only influencing 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, other industries are less risky, except for non-specified industries, which means that the industry information is essential in default risk prediction, which is in line with the expectation (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 business and for all business (including financials), they conclude that after such inclusion, the overall prediction accuracy of the model drops, the authors of that study indicate, that predicting bankruptcy for financial business is a more complicated exercise. It is worth to mention 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, personal cost (PC) per employee, and employment rate, while they enjoy 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 of 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, the unit change of personal cost per employee will lead to an increase of 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 business only, this measure was insignificant. Gupta et al. (2018) report varying (insignificant) explanatory power across different time 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 of positive, which was contrary to the prior expectation. According to Kennedy (2005), such phenomenon could be (among others) explained by the presence of multicollinearity, outlier's presence, or missing interaction terms. As the data were winsorized and multicollinearity checked, thus the only explanation which have been left was a missing interaction term, especially resulting from data aggregation. As a potential missing interaction, the interaction 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 enters the model and lead to the change of 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 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 met 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 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 business, 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 since establishment of the business till the day the business declared bankruptcy or till the day to the end of the observed period. The Cox model requires a time term to establish its parameters, while in this case the time since the start of the observed period was used till the moment of bankruptcy, thus these two terms are not interchangeable.

The model also contains a size factors 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) considers the company size factor as a significant predictor of bankruptcy, but he derives that indicator from market data. Wu, Gaunt and Gray (2010) add 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 companies (differentiating between small and medium business 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.

### 7.4.3 Details of model 2 estimates

The aim of the work is also to analyse the significance of the macroeconomic variables in predicting the default of European SMEs and for this reason a second version of the model (referred 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 an expected sign, with the exception of the relative size of quick assets (QA/TA), which change sign for 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 not significant in the model.

Table 22, 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. terms | 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 force entry manner, which is the opposite procedure 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 on firm-specific variables. Comparing such a model will not explain the extent to which the macroeconomic variables influence the model accuracy.

Regarding the change of sign of the QA/TA indicator. Once again, the effect could be caused by missing the interaction terms or in other words, it is possible that there is an interaction between the macroeconomic variables (which were excluded from this version of model) and the mentioned QA/TA indicator. For this purpose, the correlation between the macroeconomic indicators (included in model 1) and the QA/TA ration was analysed separately for small and for medium businesses.

Table 23,Correlation between macroeconomic indicators and QA/TA indicators

| **Macro-economic indicator** | **Statistic** | **Category of company** | |
|---|---|---|---|
| | | **Small** | **Medium** |
| Interest rate | Pearson Correlation | -0.052** | -0.026** |
| | Sig. (2-tailed) | 0.000 | 0.000 |
| | N | 39694 | 144758 |
| PC per empl. | Pearson Correlation | 0.053** | 0.131** |
| | Sig. (2-tailed) | 0.000 | 0.000 |
| | N | 39202 | 145922 |
| Employment rate | Pearson Correlation | 0.032** | -0.019** |
| | Sig. (2-tailed) | 0.000 | 0.000 |
| | N | 39695 | 146828 |

Note: **Significant at 1% level. **significant at 5% level. Source: Own calculation based on Amadeus database.
Note: N – number of observations.

The correlation between employment rate and QA/TA reaches positive and exhibits positive correlation in case of small business, while negative in case of medium business. The different signs of correlation might suggest that these two variables interact, while missing this interaction in model 2 has led to a change of the sign of QA/TA coefficient estimate.

### 7.4.4 Details of model 3 estimates

In the case of model 3, the stepwise procedure lead to the exclusion of six variables out of 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 24, 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 enter 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 25, 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 variables of model 2 and model 3 clearly showed that excluding the macroeconomic variables form a set of potential variables lead 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 situation.** To further explore this issue and by that gain a deeper insight into research problematic 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 unreplaceable within the context of analysed firm-specific variables.

## 7.5 Estimating the general linear model assessing the dependence between firm-specific and macroeconomic indicators

The first and most important information about the estimated general linear models, from the perspective of solving the above-mentioned issue, is given by analysing the R-squares measures, which give information about the proportion of the dependent variables (in the macroeconomic variables) explained by the set of independent variables (in this case, the set of firm-specific variables entering model 3, above the set of variables entering model 1).

Table 26, R-squared of general linear model's

| Dependent variable | R-squared | R-adjusted |
|---|---|---|
| Interest rate | 0.295 | 0. 294 |
| PC per empl. | 0.154 | 0.154 |
| Employment rate | 0.231 | 0.230 |

Source: Own calculation based on Amadeus database

The R-squared measures showed that the above added firm-specific variables (namely, CE/TL, DCF, EBIT/DA, and IA/TA) were able to explain partially the information otherwise carried by the macroeconomic variables included in model 1. The full or rather significant replacement of the missing information was not expected, however, the amount explained is not a marginal one. In other words, the information carried by **interest rate** variability was substituted from 29.5% by a set of added firm-specific variables, while **70.5% remains missing.** The information carried by personal cost per employee (**PC per empl.)** variability was substituted from 15.4% by a set of added firm-specific variables, while **84.6% remains missing** in case of **employment rate** variability the information was substituted from 23.9%, leaving **76.1%** of the information missing. The lack-of-fit test was conducted to analyse whether there is no missing interaction in the model, or the form (linear) is appropriate for modelling the dependent variable. The null hypothesis of the test is that the model fits the data well.

Table 27, Results of lack of fit test for GLM model

| Dependent Variable | Source | Sum of Squares | df | Mean Square | F-stat. | Sig. |
|---|---|---|---|---|---|---|
| Interest rate | Lack of Fit | 92,627.643 | 83607 | 1.108 | 1.365 | 0.305 |
| | Pure Error | 8.114 | 10 | 0.811 | | |
| PC per empl. | Lack of Fit | 14,751,615.880 | 83357 | 176.969 | 2.031 | 0.104 |
| | Pure Error | 871.227 | 10 | 87.123 | | |
| Employment rate | Lack of Fit | 1,584,076.696 | 83811 | 18.901 | 2.572 | 0.048 |
| | Pure Error | 73.487 | 10 | 7.349 | | |

Source: Own calculation based on Amadeus database. Note: df – degrees of freedom.

In the case of the GLM the model were the dependent variables of *interest rate* and *PC per empl.* The test statistic is insignificant, meaning the model fits the data well. On the other hand, the results of the model for *employment rate* are significant, however, p-value is very near $\alpha = 5\%$ level.

**To conclude this part, the analysis of the selected macroeconomic variables has led to deduction that the information content carried by these factors is relevant for default prediction, cannot be replaced by changing the set of firm-level firm-specific variables within the set of analysed potential variables.** For gaining a more detailed insight into the issued test of between-subject effects results for all three versions of the estimated general linear model was conducted. First, the details of model describe the variability of **interest rate**.

Table 28, GLM for the interest rate variable

| **Source** | **Type III Sum of Squares** | **df** | **Mean Square** | **F-stat.** | **Sig.** |
|---|---|---|---|---|---|
| Corrected Model** | 38744.024 | 73 | 530.740 | 479.069 | 0.000 |
| Intercept** | 1280.034 | 1 | 1280.034 | 1155.41 | 0.000 |
| CETL** | 335.166 | 1 | 335.166 | 302.535 | 0.000 |
| DCP** | 90.426 | 1 | 90.426 | 81.623 | 0.000 |
| EBITDATA** | 24.757 | 1 | 24.757 | 22.347 | 0.000 |
| IATA** | 337.726 | 1 | 337.726 | 304.846 | 0.000 |
| Status | 0.002 | 1 | 0.002 | 0.002 | 0.968 |
| IND** | 337.520 | 4 | 84.380 | 76.165 | 0.000 |
| Time ** | 2724.162 | 4 | 681.040 | 614.736 | 0.000 |
| SB | 0.576 | 1 | 0.576 | 0.520 | 0.471 |
| Status x IND** | 150.054 | 4 | 37.513 | 33.861 | 0.000 |
| Status x time** | 17.047 | 4 | 4.262 | 3.847 | 0.004 |
| Status x SB | 0.861 | 1 | 0.861 | 0.777 | 0.378 |
| IND x time ** | 57.273 | 16 | 3.580 | 3.231 | 0.000 |
| IND x SB | 5.083 | 4 | 1.271 | 1.147 | 0.332 |
| time x SB* | 12.323 | 4 | 3.081 | 2.781 | 0.025 |
| Status x IND x time** | 91.979 | 12 | 7.665 | 6.919 | 0.000 |
| Status x time x SB | 1.041 | 1 | 1.041 | 0.939 | 0.332 |
| IND x time x SB* | 25.107 | 12 | 2.092 | 1.889 | 0.031 |
| Error | 92635.757 | 83617 | 1.108 | | |
| Total | 686304.113 | 83691 | | | |
| Corrected Total | 131379.781 | 83690 | | | |

Note: **Significant at 1% level. **significant at 5% level. Source: Own calculation based on Amadeus database. Note: df – degrees of freedom.

The significance of the corrected model means that the model including all terms except for the intercept term is significant, implying that there is a significant relationship between the dependent and independent variables. while the significant value of the intercept term of the model can be interpreted as a significant amount of information about the interest rate variation is not captured by the model's variables, which is line with the expectation and correspond with the R-squared value. All analysed firm-specific variables are significant, meaning that there is a significant relationship between these factors and the interest rate. Moreover, the *IND* and *time* fixed effects are significant, meaning that the relationship is not constant over time, while not the same over different industries. On the other hand, the *status* and *category of companies* mean that the relationship does not differ for default and nondefault businesses or do not differ for small and medium businesses **in general**. However, the significant value of the interaction terms between *Status x and time* suggest, that the relationship between the firm-specific ratios and interest rate differs between default and non-default business. Moreover, this depends on time, same applies for the category of company, as the interaction *between time x Category of company* (i.e. SB) is significant, meaning that this relation also is highly time dependent.

Second, the details of model describing the variability of **PC per empl**. variable.

Table 29, GLM for the personal cost per empl.

| **Source** | **Type III Sum of Squares** | **df** | **Mean Square** | **F-stat.** | **Sig.** |
|---|---|---|---|---|---|
| Corrected Model** | 2693934.044 | 65 | 41445.139 | 234.208 | 0.000 |
| Intercept** | 178803.747 | 1 | 178803.747 | 1010.428 | 0.000 |
| CETL** | 211366.081 | 1 | 211366.081 | 1194.440 | 0.000 |
| DCP** | 130133.019 | 1 | 130133.019 | 735.388 | 0.000 |
| EBITDATA | 483.969 | 1 | 483.969 | 2.735 | 0.098 |
| IATA** | 338161.390 | 1 | 338161.390 | 1910.966 | 0.000 |
| Status | 63.544 | 1 | 63.544 | 0.359 | 0.549 |
| IND** | 104868.986 | 4 | 26217.247 | 148.155 | 0.000 |
| time** | 9249.215 | 3 | 3083.072 | 17.423 | 0.000 |
| SB** | 1181.846 | 1 | 1181.846 | 6.679 | 0.010 |
| Status x IND** | 18840.186 | 4 | 4710.047 | 26.617 | 0.000 |
| Status x time** | 4625.514 | 3 | 1541.838 | 8.713 | 0.000 |
| Status x SB | 438.449 | 1 | 438.449 | 2.478 | 0.115 |
| IND x time** | 28967.433 | 12 | 2413.953 | 13.641 | 0.000 |
| IND x SB** | 2535.430 | 4 | 633.857 | 3.582 | 0.006 |
| time x SB | 919.331 | 3 | 306.444 | 1.732 | 0.158 |
| Status x IND x time** | 5917.758 | 12 | 493.146 | 2.787 | 0.001 |
| Status x time x SB | 22.795 | 1 | 22.795 | 0.129 | 0.720 |
| IND x time x SB | 1818.922 | 12 | 151.577 | 0.857 | 0.592 |
| Error | 14752487.107 | 83367 | 176.958 | | |
| Total | 114514748.04 | 83433 | | | |
| Corrected Total | 17446421.151 | 83432 | | | |

Note: **Significant at 1% level. **significant at 5% level. Source: Own calculation based on Amadeus database. Note: df – degrees of freedom.

In the case of the general linear model describing the variability of personal cost per employee, it is also valid that both the corrected model and the interaction term are significant. Regarding the significant of the analysed firm-specific variables, it was found that except for the EBITDA/TA, all firm-specific variables under analysis are significant, implying there is relationship between the level of personal cost per employee (on a national level) and the profitability of the assets. This result is rather surprising, as a relationship between the level of cost (in this case personal cost) and the profitability measures could be anticipated. On the other hand, this analysis (in terms of GLM) in addressing the macroeconomic variables in a univariate level (i.e. separately), thus by this analysis only the cost perspective is affected, while the revenue part is not. Another possible expiation could lie in the wage rigidity issue. The revenue part is rather addressed under the analysis of employment rate, as employment rates are regarded as a proxy of demand.

The insignificance of the *status* variables on the one hand and the significance of the interaction terms of *status x time* or *status x IND* lead to conclusion that the personal costs per employee are not related, and do not differ in the case of defaulting and non-defaulting business in general, but it further depends on the industry and time. It is worth to say; the results also suggest that the situation of small significantly differs from the situation of medium businesses.

Third, the details of models describing the variability of **the employment rate** variables.

Table 30, GLM for the employment rate

| **Source** | **Type III Sum of Squares** | **df** | **Mean Square** | **F-stat.** | **Sig.** |
|---|---|---|---|---|---|
| Corrected Model** | 474610.703 | 73 | 6501.516 | 344.010 | 0.000 |
| Intercept** | 731207.150 | 1 | 731207.150 | 38689.838 | 0.000 |
| CETL** | 15164.327 | 1 | 15164.327 | 802.379 | 0.000 |
| DCP** | 7080.362 | 1 | 7080.362 | 374.638 | 0.000 |
| EBITDATA** | 1766.783 | 1 | 1766.783 | 93.485 | 0.000 |
| IATA** | 1180.442 | 1 | 1180.442 | 62.460 | 0.000 |
| Status** | 202.695 | 1 | 202.695 | 10.725 | 0.001 |
| IND** | 9945.423 | 4 | 2486.356 | 131.559 | 0.000 |
| time** | 2270.064 | 4 | 567.516 | 30.029 | 0.000 |
| SB | 49.947 | 1 | 49.947 | 2.643 | 0.104 |
| Status x IND** | 4122.834 | 4 | 1030.709 | 54.537 | 0.000 |
| Status x time** | 261.824 | 4 | 65.456 | 3.463 | 0.008 |
| Status x SB | 53.241 | 1 | 53.241 | 2.817 | 0.093 |
| IND x time** | 1506.810 | 16 | 94.176 | 4.983 | 0.000 |
| IND x SB | 88.595 | 4 | 22.149 | 1.172 | 0.321 |
| time x SB | 170.627 | 4 | 42.657 | 2.257 | 0.060 |
| Status x IND x time** | 1615.886 | 12 | 134.657 | 7.125 | 0.000 |
| Status x time x SB | 4.406 | 1 | 4.406 | 0.233 | 0.629 |
| IND x time x SB | 258.451 | 12 | 21.538 | 1.140 | 0.322 |
| Error | 1584150.182 | 83821 | 18.899 | | |
| Total | 422542086.500 | 83895 | | | |
| Corrected Total | 2058760.886 | 83894 | | | |

Note: **Significant at 1% level. **significant at 5% level. Source: Own calculation based on Amadeus database

The significance of the corrected model and intercept term is also true when regarding the model describing the relationship between the employment rate and the given set of firm-specific variables. The relationship, unlike in the case of the rest of the macroeconomic variables, is however significantly different in the case of default and nondefault businesses, in general, which also holds for business from different industries. This, however, is not the case for small and medium businesses, while such a distinction does not change this relationship (between the employment rate and macroeconomic indicators), not even under the dependency of time or industry effects.

## 7.6 Models for benchmark purposes

To assess the performance of the derived model properly, two models were selected – the model of Altman (1983) and model of Altman and Sabato (2007). The model of Altman (1983) is a representant of **a generic** type model (of use for not specified types of businesses). The reason behind selecting this model is its worldwide popularity, because of which this model is often selected as a benchmark and from this perspective, such comparation would make the results more comparable.

On the other hand, the model of Altman and Sabato (2007) was derived especially for application on SME segment of business, whereas the authors stated that the effectiveness of the model on a sample of SME is about 30% higher (in terms of accuracy), than the effectiveness of a generic model. This version of the model was chosen as it is based on firm-specific data only, which allows the model application on the analysed sample. The Altman and Sabato (2007) model was developed on a sample of 2,010 U.S. SMEs out of which 120 was defaulted. Data were drawn from COMPUSTAT database over the period of 1994-2002. This a model is the representant of model derived especially in application for SME segment of businesses. It is worth to mentioned that the same authors developed a more recent model of SMEs, however, the availability of the data does not allow the comparison with a more recent version of the model, i.e. the Altman, Sabato, Wilson (2010) model, mainly due to the fact, that the model incorporate also non-financial data.

Both models were applied in its original setting and with re-estimated coefficients to ensure the coefficients setting is **not adversely affected by the influence of different period or business environment conditions.** For re-estimating the models, the original methodology was applied, i.e., the linear discriminant analysis in the case of Altman (1983) model and the logistic regression method in the case of Altman and Sabato (2007) model.

### 7.6.1 Introduction model selected for benchmark.

The selected models take the following form in their original settings.

**Model of Altman (1983)**

This model is a version of the Altman (1968) model devoted to unlisted business. The revised Z-score represents the original Z-score model (see Altman, 1968), which was adapted for non-listed companies (see Altman, 1983). The formula of the model is following:

$$Z' = 0.717 \cdot \frac{WC}{TA} + 0.847 \cdot \frac{RE}{TA} + 3.107 \cdot \frac{EBIT}{TA} + 0.42 \cdot \frac{E}{TL} + 0.998 \cdot \frac{S}{TA} \quad (64)$$

where: *NWC – net working capital (=current assets-current liabilities), TA – total assets, RE – retained earnings, EBIT – earnings before interest and taxes, E – book value of equity, S – sales*

The grey zone interval is (1.23; 2.9). For Z<1.23,Z<1.23, the company is classified by the model as threatened by bankruptcy, for Z>2.9 is classified as not threatened by bankruptcy, i.e. financially healthy. Altman and Sabato (2007) tested a model on the sample of US SMEs over the period from 1994 to 2002. The resulted overall accuracy of the model was 68%, while type I error (a percentage of bankrupt firms classified as non-bankrupt) was 25.81%.was 25.81%.

**Model of Altman, Sabato (2007)**

The model version with unlogged predictors was employed, the model takes the following form:

$$log \frac{PD}{1-PD} = 4.28 + 0.18 \cdot \frac{EBITDA}{TA} - 0.01 \cdot \frac{CL}{E} + 0.08 \cdot \frac{RE}{TA} + 0.02 \cdot \frac{C}{TA} + 0.19 \cdot \frac{EBITDA}{IE} \quad (65)$$

where: *PD – probability of default, while the modelled the probability is probability that a business will default within one year, EBITDA – Earnings before interest taxes, depreciation and amortization, TA – total assets, CL – short-term debt, E – book value of equity, RE – retained earnings, C – cash, IE- interest expenses.*

### 7.6.2 Details of Altman (1983) model re-estimation on the learning sample

The model was re-estimated on the same learned sample as was the developed hazard model. First, the results of re-estimation of Altman (1983) model will be presented, afterwards the results of re-estimating the Altman and Sabato (2007) model will be shown. First, model overall statistics assessed by Wilk's lambda.

Table 31, the overall discrimination ability of the re-estimated Altman model

| Model/statistics | Wilks' Lambda | Chi-square | df | Sig. |
|---|---|---|---|---|
| Altman (1983) re-est.** | 0.927 | 3729.673 | 5 | 0.0000 |

Note: **Significant at 1% level. **significant at 5% level. Source: Own calculation based on Amadeus database. Note: df – degrees of freedom.

The re-estimated Altman (1983) model is significant at the 1% level; however, its discrimination power is rather weak (as the Wilk's lambda is near 1, which is the value representing no discrimination power at all). The details of the contribution of the model's variables to the discrimination power of the model can be assessed by the partial Wilk's lambda, which is equal to Wilk's lambda of the whole model, in case the given variable would be excluded from the model.

Table 32, Partial Wilk's lambda of the re-estimated model

| Variable | Wilks' Lambda | F-stat. | df1 | df2 | Sig. |
|---|---|---|---|---|---|
| WC/TA** | 0.952 | 2504.492 | 1 | 49137 | 0.0000 |
| RE/TA** | 0.946 | 2783.636 | 1 | 49137 | 0.0000 |
| EBIT/TA** | 0.95 | 2599.51 | 1 | 49137 | 0.0000 |
| E/TL** | 0.996 | 185.955 | 1 | 49137 | 0.0000 |
| S/TA** | 0.999 | 35.677 | 1 | 49137 | 0.0000 |

Note: **Significant at 1% level. **significant at 5% level. Source: Own calculation based on Amadeus database.
Note: df – degrees of freedom.

All variables of re-estimated model are statistically significant at the 1% level, which means that none of the variables could be excluded from the model, without significant loss of model discrimination ability. The contribution of the model variables to model discrimination ability could be also assessed by the analysis of standardized canonical discriminant function coefficients.

Table 33, Standardized canonical discriminant function coefficients

| Variable | Coefficient |
|---|---|
| WC/TA | 0.089 |
| RE/TA | 0.525 |
| EBIT/TA | 0.547 |
| E/TL | 0.042 |
| S/TA | 0.268 |

Source: Own calculation based on Amadeus database

The EBIT/TA and RE/TA represent the variables with the highest contribution to model discrimination ability, while the E/TL variable contributes the mentioned ability less. The re-estimation led clearly to the change of the variable importance rank of the E/TL indicator, as the author of the model (see Altman, 2000) evaluated the E/TL as the third most important model variable.

Finally, the Fisher's discrimination function coefficients of the re-estimated are shown below.

Table 34, Fisher's discrimination function coefficients

| Variable/status | Non-default | Default |
|---|---|---|
| WC/TA | -0.097 | -0.130 |
| RE/TA | 0.153 | -0.001 |
| EBIT/TA | 0.353 | -0.318 |
| E/TL | 0.105 | 0.098 |
| S/TA | 0.393 | 0.330 |
| (Constant) | -1.203 | -1.118 |

Source: Own calculation based on Amadeus database

For comfort purpose, the function could be rewritten in the following form:

$$Z'(re-est).=0.033\cdot\frac{WC}{TA}+0.154\cdot\frac{RE}{TA}+0.671\cdot\frac{EBIT}{TA}+0.007\cdot\frac{E}{TL}+0.063\cdot\frac{S}{TA} \quad (66)$$

Where the business is evaluated as threatened with default if Z'(re-est.) < 0.085, otherwise, the business is evaluated as not threatened with default.

### 7.6.3 Details of Altman and Sabato (2007) model re-estimation on the learning sample

The results of re-estimation of Altman and Sabato (2007) model will be presented in a similar manner, i.e., starting with the model overall characteristics, followed by analysis of the contribution of the given model's variables. The result of testing the model will be presented together with the results of testing the derived models.

Table 35, Model fitting criteria

| | Model Fitting Criteria | Likelihood Ratio Tests | | |
|---|---|---|---|---|
| **Model** | **-2 Log Likelihood** | **Chi-Square** | **df** | **Sig.** |
| Intercept Only | 42352.180 | | | |
| Final** | 37422.752 | 4929.429 | 5 | 0.000 |

Note: **Significant at 1% level. **significant at 5% level. Source: Own calculation based on Amadeus database

The improvement of the fitted model over a model with intercept only is statistically significant at the 1% level. For assessing the extent to which the fitted model provides better discrimination ability than an intercept model pseudo R-squares measures are commonly employed.

Table 36. Pseudo R-squares of the re-estimated model

| Measure | Value |
|---|---|
| Cox and Snell | 0.143 |
| Nagelkerke | 0.195 |
| McFadden | 0.116 |

Source: Own calculation based on Amadeus database

The pseudo R-squares measure interpretation should not be confused with the interpretation of the R-square measures of the general linear model, i.e., they do not provide information on how many percent of the data variability is explained by a set of explanatory variables. However, they share a common feature of reaching a maximum value of 1. According to the Nagelkerke R square measure, the improvement of the fitted model over the intercept model is rather low, while not considering the number of model parameters. Whereas the Cox and Snell pseudo R square accounts also for a number of model parameters.

The details of parameter estimation are listed below, while the reference category is that the business has defaulted, thus the modelled probabilities are the probabilities of default.

Table 37 Re-estimated Altman Sabato model - estimation details

| Variable | B | SE | Wald | df | Sig. | Exp(B) | 95% CI for Exp(B) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Lower bound | Upper bound |
| Intercept** | 0.08189 | 0.018 | 21.160 | 1 | 0.000 | | | |
| C/TA** | 1.01764 | 0.091 | 124.510 | 1 | 0.000 | 2.767 | 2.314 | 3.308 |
| EBITDA/TA** | 0.93441 | 0.071 | 172.752 | 1 | 0.000 | 2.546 | 2.215 | 2.926 |
| EBITDA/IE** | 0.00004 | 0.000 | 24.937 | 1 | 0.000 | 1.000 | 1.000 | 1.000 |
| RE/TA** | 1.62248 | 0.041 | 1533.032 | 1 | 0.000 | 5.066 | 4.670 | 5.494 |
| CL/E** | -0.00819 | 0.001 | 134.443 | 1 | 0.000 | 0.992 | 0.990 | 0.993 |

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

All estimated coefficients are significant at 1% level. The re-estimated coefficients have kept their sign; however, the value of the coefficients change in many cases significantly, especially in the case of RE/TA (retained earnings over total assets) and EBITDA/IE (EBITDA over interest expenses).

## 7.7 Models' testing results

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

$$S(t) = exp[-H_0(t)exp(PI)] \tag{67}$$

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:

$$PI = \sum_k^q \beta_k x_k \tag{68}$$

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

For estimating the survival function values of given time (t). the following estimates of baseline cumulative hazard function values were utilized. 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 given observation. The survival probabilities shown in table reflects the situation of media business, while there is a penalization for small business (represented as a dummy variable "the category of companies").

Table 38. 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, the 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 binomial distribution, while there is also a possibility to estimating the AUC under nonparametric assumption using the trapezoidal approach, whereas the results might slightly differ.

### 7.7.1 The AUC values of the tested models

Both the set of derived models and the models used for benchmark purposes where 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 model and re-estimated model, the test sample testing results represent the key results, from which the implication could be drawn. From the above-mentioned perspective, in further text, the test sample would be preferred.

Table 39, 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).

### 7.7.2 Comparing model 3 with the benchmarks

Model 3 represents that version of the model, which was 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 different 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 case of Z-score model, the effect of non-specific focus of the model (not distinguishing between large and medium and small business) is included above that.

Table 40, 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 mentioned differences are statistically significant by 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) by 10.5 pp, while the re-estimated version of Z-score is still being outperformed by 9.59 pp. The original version of 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 hardly to be achieved in practice, thus the space for model improvement is limited.

### 7.7.3 Comparing model 2 with the benchmark

Comparison of model 2 and the models selected for the benchmark will give a similar answer to the one mentioned in the case of model 3. The set of variables in model 2 is only a subset of 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 cause a potential lower efficiency of this model, in comparison with model 3.

Table 41, 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 Z-score by (on the test samples) by 7.14 pp, while the re-estimated version of Z-score is still being outperformed by 6.28 pp. The original version of 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 by 1% level.

### 7.7.4 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 SME, while respecting the SME segment heterogeneity and industry effects, while utilizing both firm-specific type and macroeconomic type of variables) provides better classification accuracy then the models selected for benchmark (not utilizing the above-mentioned features).

Table 42, 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 Z-score by (on the test samples) by 12.6 pp, while the re-estimated version of Z-score is still being outperformed by 11.7 pp. The original version of 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 by 1% level.

### 7.7.5 Comparing the derived models

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

Table 43, 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 1%. 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 reedifying the set of firm-specific ratios led to an increase of AUC by 3.31 pp, whereas the difference is significant at the 1% level.

# 8 Discussion

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

The issue with assessing a potential contribution of adding explanatory variables to a model, is that the accuracy of a model could be viewed as an outcome of testing an estimated multivariate model. The accuracy of the model is strongly affected by significance of model variables. Under the multivariate context, the significance of a given variable is assessed in the relationship to other variables. In course of the research, it was needed to **tend to isolate the effect, that change of variable set would have on the accuracy and influence the result as an unobservable factor.** To control for this factor, two possible ways were considered. First way is to estimate and compare two full models – first with full set of firm-specific variables and second with a full set of firm-specific variables accompanied by a full set of macroeconomic variables. Second way is to estimate the models in a reduced form, which allows to analyse the most significant subset of variables of each group of variables. The drawback of first way of solving the problem is, that for estimating a full model complete observation would be needed, which would significantly lower the number of observation utilizable for estimating the model. The drawback of the second approach is, that there is always a risk when reducing a set of potential set of variables, as it might harm model robustness. On the other hand, reducing the set of variables in a stepwise manner would prove, whether macroeconomic variable could enter the model, which mean that such variables can significantly increase model accuracy.

In course of the research a following hypothesis was formulated: *The model combining a set of macroeconomic and firm-specific variables will reach a significantly higher discrimination power, in terms of AUC, than model utilizing a set of firm-specific variables, while not employing the set of macroeconomic variables.* This hypothesis was assessed by comparison of the AUC values of model 1 and model 2 or rather by testing whether the difference between these two AUC values is significant, while testing was conducted using the methodology of DeLong et al. (1988).

The AUC value of model 1 (combining the firm-specific variables and the macroeconomic variables) reached AUC value (on the test sample) of 0.884, whereas model 2 (containing the same set of firm-specific variables) reached AUC value 0.829. Due to the large sample, it was possible to reach relatively narrow confidence intervals of the mentioned AUC values – in the case of model 1, the 95% confidence interval is from 0.880 to 0.888, whereas in the case of model 2 it is, from 0.825 to 0.834. These two intervals are not overlapping thus, these values are clearly different. The

DeLong et al. (1988) test result p-value is lower than 0.001, confirming the statistical difference between these two AUC values.

Furthermore, a comparison of the AUC values of model 1 and model 3 results are, that the AUC value of model 3 was (on the test sample) of 0.862, whereas the 95% confidence interval was from 0.858 to 0.866, i.e. even in this case the confidence intervals of AUC values of model 1 and model 3 are not overlapping. The results of DeLong et al. (1988) test confirms that this difference in AUC values is significant at the 1% level (p<0.0001).

To sum up, the incorporation of macroeconomic variables to a set of firm-specific variables lead to a model version with a significantly higher discrimination power, than was the case in model version utilizing only firm-specific variables. The results holds, even when the set of firm-specific variables was changed (optimized).

**Based on this, it can be concluded that the evidence of the results is in favour of hypothesis, which based on the results cannot be rejected.**

The focus on SME segment, paid throughout this work, is given by several factors. For instance, despite the number of papers, which have been published on predicting business default, there are still some areas or topics which deserve scientific attention or rather there are still problems that could be regarded as a research gap in the current knowledge. Predicting the default of small and medium business is one of these areas, especially when addressing the role of environment factors. The need of focusing on modelling SME defaults apart from default large and listed business is given by several reasons. The first reason is that there has been a clear disparity between the scientific efforts devoted to the default of a large and listed business (numerous papers since the 1960's), where a lot of research has been done to verify appropriate approaches to model the default employing various methods (both of parametric and nonparametric nature) and the number of papers devoted especially to SME's.

The effort of predicting default especially for SMEs started in 1978 with the work of Edminster (1978), but the paper did not clarify the need of focusing on SMEs apart from large business when modelling the default risk. This has been done much later by Altman and Sabato (2007), who showed that a model especially derived for SME segment would reach up to 30% higher accuracy than the compared generic model. Second reason, which cannot be neglected is the role that SMEs play in the economy, whereas the SMEs are regarded as the backbone of the global economy (among others, Gupta et al, 2015) or as an economy's engine for sustainable growth and stable employment (De Moor et al., 2016). Moreover, their role in reduction in poverty, increase in employment, output, innovation in technology, and lifting up in social positions and standards is well recognized (see, e.g., Eniola and Entebang, 2015 or Altman and Sabato, 2007). The up-to-day scientific attention paid to SME default predictions could be viewed as being in sharp contrast to their role in the economy, however, the situation is currently getting better. Third reason is that the SMEs in order of their development have to face the financial constraints issues (see, e.g., Beck et al, 2006; Fauceglia, 2015; Ullah, 2019; Erdogan, 2018; McGuinness et al., 2018). As first, this issue was noted by Fazzari et al. (1988), who showed that the investment decisions of financially constrained firms are more sensitive to the availability of internal cash flows than it is in the case of unconstrained firms. The limited availability of external sources of finance are commonly mentioned in the context of SMEs. Altman Sabato (2007) showed that modelling default risk, especially for SMEs, would result in lower capital requirements for the banks (under BASEL III), than in the case of applying a generic model. Fourth reason is that the contemporary approaches employing Merton's structural approaches or CDS (Credit Default Swap) spreads or other approach utilizing direct capital market data to estimate the probability of default are simply unsuitable or even inapplicable for SMEs segments. The reasons were summarized by Filipe (2016) who highlights that SMEs usually do not meet the requirements to enter the capital market. The thing is that the market approach to default modelling is currently regarded as theoretically superior to the competing or rather historically antecedent firm-specific (accounting) ratio approaches, thus **research on SME default has to find its own methodological way**. Researchers either rely on accounting or nonfinancial business specific data (Altman and Sabato, 2007 or Altman, Sabato and Wilson, 2010) or they utilize data from the external environments by employing a hazard model approaches (e.g. Gupta et al., 2014, 2015). Incorporating the external environment (especially macroeconomic data) seems to be especially suitable for SME segments, as this segment of business is commonly viewed as more vulnerable to environment changes and financially constrained.

On the one hand, the researchers pose no doubt that there is a link between macroeconomic factors and business default risk (especially in the case of SMEs), however, the specific mechanism in which their factors interact is much less clear (for example highlighted among others by Boratyńska, 2016). Currently, there are a limited number of papers, which deal with utilizing macroeconomic factors in predicting SME default. The existing papers usually employ a hazard model approach, while the macroeconomic variables serve as the baseline hazard rates. **If so, most of them focus on US or UK data samples only, while there is a clear gap in the knowledge regarding the SME from EU countries. And this is the reason behind choosing the topic of this work**.

The next issue was the choice of an appropriate method for deriving the new model. The majority of the created models over the several past decades employed either discrimination analysis or logit models (see Aziz and Dar, 2006). Since the criticism was raised by Shumway (2001) regarding the dynamic nature of the default process, which was neglected by the majority of the models, suggest the need of employing the hazard model approach or rather the Cox regression model.

The Cox model approach is a rather flexible method, especially in terms of defining the baseline hazard rate, while the model's parameters could be estimated even if the baseline hazard rate is left unspecified. Most of the studies on hazard models specified the hazard rate in terms of time dummies or macroeconomic variables. **In this work, the macroeconomics variables were rather utilized as explanatory variables, which could be viewed as a more flexible approach to the utilization of the environment approach, thus the methodology employed in this the work is rather different to work of other authors.**

When suggest the model form, several issues had to be respected. The first issue was the heterogeneity of SME segments. Previously, it has been shown that the default risk is different for large business and for SME highlighting the need of treating SME segments especially (see Altman and Sabato, 2007), however further studies showed that a similar difference can be spotted between small and medium business, causing the SME segment being heterogenous (see Gupta et al., 2015). As a reflection of this, a small business dummy (SB) variable was added into the model, whereas the initial application of log-rank test confirms the need of adopting such procedure. The same applies for the industry specific, while the industry effect was treated by adding an industry categorial variables into the model. The industry effect was addressed only in terms of model constant, not in terms of the slope (as an interaction between the industry variables and continuous variables), the main reason was an increase of the number of potential variables, which would, in relation with the number of observations, cause significantly higher computational demand.

In some cases, the preliminary results showed that there is a need for adding an interaction term into the model, as there was a change in the estimate coefficient gain by the initial discrimination and the final model. Such a situation could be explained by several factors, as noted by Kennedy (2005), such as outliers' existence or missing interaction terms. The outliers have been winsorized, thus the missing interaction term was the only explanation which has left, further results confirm this preliminary conclusion.

**The results proved, that incorporating the macroeconomic indicators to a set of firm-specific indicators have a potential of significant increase in model accuracy and provide an unique and uncorrelated (to other explanatory variables) information.** A general linear model was formulated in order to analyse the extent to which the information carried by macroeconomic indicator provide a unique contribution to the information content of the model. In this model the macroeconomy factor served as response variable, while the firm-specific indicators served as explanatory variables. The results showed, that only part of the information content, about 15-30 %, carried by macroeconomic indicators, can be supplemented by firm-specific variables, while the remaining 70-85 % represent a unique contribution. The aim of the model was not to infer the cause-relationship, but rather to explore the extent to which the firm-specific ratios reflects the macroeconomic conditions. Such conclusion on one hand clearly advocates the need of incorporating macroeconomic indicator in the default prediction models and on the other suggest that the there is a limit of increasing accuracy of the default prediction models solely based on firm-specific ratios. Above this, it shed light into often observed phenomenon of drop of accuracy related with the application of default prediction model under alternative condition, where this has been showed by many researchers (among others by Grice and Dugan, 2001).

The limited availability of external financing to SMEs is often perceived as a burden to their further growth, but as one of the results of this study, it can be shown that the interest rate has a direct effect on SMEs survival probability. The effect is especially significant in case of the change of long-term interest rate, whereas a change of long-term interest rate by one percentage point will multiply the relative risk of SME default by 2.9 times. This result is further confirmed by the presence of firm-specific variables, which entered the model. Three out of twelve firm-specific variables of model 1 are directly related to the cost of debt – the ratio of financial expenses over sales (FE/S), the interest cover ratio (EBITDA/IE) and consequently the net profit margin (NI/S), whereas the operating profit margin (EBIT/S) also showed high significance, however was removed from the sample due to high correlation with net profit margin, which above that exhibits a slighter higher significance during the initial discrimination analysis. The profitability ratios play in general a significant role in the default probability studies, whereas there that is importance is frequently being highlighted (e.g., Altman, 1968, Li and Sun, 2009, Altman and Sabato, 2007, Psillaki, Tsolas and Margaritis, 2009) moreover their role also withstand the criticism raised by Shumway (2001). All these studies have in common that the profitability term was employed in relation to assets (in the form of EBIT/TA or EBITDA/TA), but not in relation to sales (as NI/S or EBIT/S). The profitability of assets (EBIT/TA or EBITDA/TA) was also analysed, however did not enter model 1, whereas the net profit margin did. When searching for an explanation of such phenomena, one has to keep in mind two issues. First is that the profitability of assets (in terms of DuPont formula) can be divided into operating profit margin (EBIT/S) and asset turnover (S/TA), whereas the asset profitability could be viewed as an interaction between these two terms. Second, in the case of drop of sales, the profitability of assets and operating profit margin ratios are affected in a different manner. Drop of sales is affecting negatively the profit margin, however it depends on the cost structure (the proportion of fixed cost over the total cost), while in the case of return on assets this further depends on the asset turnover factor. Analysing the results of initial discrimination, it was found that the asset turnover term was not significant at the univariate level, thus not further analysed, which is the reason model 1 or model 2 cannot contain this variable. However, the return on assets is present in model 3, but not in model 1. Such situation also applies to other ratios, namely – capital employed over total liabilities (CE/TL), debtor collection period (DCP), and intangible assets over total assets (IA/TA).

**This has led to a presumption that some of the information carried by macroeconomic indicators could be partly reflected in the values of other firm-specific variables**. To address this issue further, the macroeconomic variables present in model 1 were regressed with firm-specific variables which enter model 3 extra over the set which enter model 1. **The motivation behind this is not to address the cause-effect relationship**, but rather can be viewed as an analysis of the extent to which the information carried by macroeconomic variables and useful in default prediction could be reflected by a set of firm-specific indicators. In this specific case, it was found that there is a significant relationship between the macroeconomic indicators and the firm-specific ratios, however the absence of macroeconomic indicators cannot be sufficiently supplemented by the presence of other firm-specific variables, as there is a large and unique contribution of macroeconomic indicators in explaining the default of SME. To be more specific. Model 3 (i.e., the model with solely firm-specific variables) reached an AUC value of 0.831, whereas model 1 (combining firm-specific and macroeconomic variables) reached an AUC of 0.881, while this difference between the mentioned AUC values is statistically significant. The variance of macroeconomic indicators values, which could be supplemented by the variance of firm-specific variables (extra added to model) could be explained from 15.4 to 29.5 %, whereas the rest of the macroeconomic indicator variance (i.e. its unique contribution), which ranges from 70.5 to 84.6 % will remained unexplained. **These results explain the role of macroeconomic indicators from an inner perspective, while the AUC values explain the role from an outer perspective, while both contribute in fulfilling the aim of the work.**

To make the result more comparable to other author’s work or rather to assess the contribution to the current state of art in the field. The created models were compared to well-known models of Altman (1983) – a representant of a generic type the model and compared to model of Altman and

Sabato (2007) – a representant of the model especially derived for SMEs. An interesting comparison of the mentioned models and model, which were created in the course of this research could be done, based on the perspective suggested by Shumway (2001), according to whom not incorporating of time factor results in biased and unstable predictor estimates. Above that, accepting the specificity of SMEs default prediction, the comparison could be done among the Altman and Sabato (2007) model predictors and the predictors of the created models. The model of Altman and Sabato (2007) contains five firm-specific variables, while four of three of them are also incorporated in all created models (model 1, model 2, and model 3).

These ratios are – short-term debt over equity (CL/E), cash over total assets (C/TA) and interest cover (EBITDA/IE) – these ratios prove their importance in SMEs default prediction, even when the time factors is incorporated, moreover even when the macroeconomic factors are added. The model of Altman Sabato (2007) also contains indicators measuring the profitability of assets, both current (EBITDA/TA) and the past (RE/TA). The RE/TA indicator (retained earning overs total assets) was excluded from the analysed sample of indicators, as in the phase of initial discrimination analysis the sign of the estimate was in contrary to the theoretical expectations. The EBITDA/TA entered only model 3, not model 1, which could be interpreted in the following manner, the significance of the indicators holds even when the time factors in included, when above that the macroeconomic factors are added, the information carried by the EBITDA/TA indicator could be supplemented, causing insignificance of the indicator or in other words, it can be suggested that the EBITDA/TA value are strongly affected by macroeconomic factors, such as interest rates, employment rate and personal cost per employee. However, further research would be needed to confirm this presumption.

To compare the accuracy of the created models, and a comparison with other approaches or rather other models, applicable for SMEs was carried out. To avoid the risk mentioned by (among others) by Grice and Dugan (2001) dealt with a drop of accuracy when the model is applied under alternative conditions (comparing to the conditions of the learning sample, on which the model was derived), both the models' coefficients were re-estimated on the learning sample (i.e. on the sample on which the created model 1, model 2 and model 3 were estimated).

Further, regarding the variables of the created models, it should be mentioned that also several non-ratios indicators were adopted or even nonfinancial ones. The first mentioned is the business age indicator ("ln(age)"), which is defined as a natural logarithm of the number of days since business establishment till the last closing day of the last financial statement available. This is to differentiate the business that are new to those which have already stabilized the market position, while still remaining in the SME segment. This is a rather an indirect way how to address the time factor, whereas the adopted approach of Cox regression addresses the time factor directly. Nevertheless, the time under this analysis is left censored, meaning that the data under analysis are not since the business was established, but since the business enters the study and from this perspective, these two ways of addressing the time factors are not competing and not harming the model assumptions. The results showed that this factor is significant predictor of SME default. This results further contribute the issue of heterogeneity of SME segments, which has been previously noted by Gupta et al. (2014), however, from a different point of view.

The conducted research also has some limitations, which can be addressed in future research. The results on parameter estimation showed, that the SME heterogeneity poses challenges to the modelling process. The control for this feature was in employing small business (SB) dummy variable, however first results of estimation of multivariate model suggest, that there is a missing interaction terms and further analysis exhibit significant interaction term between quick ratio (QR) and net profit margin (NI/S). Addition of these interaction terms was a necessity, highlighted by a change of estimated sign between the univariate model analysis (initial discrimination) and the estimated multivariate model. There is no proof to believe, that there might be other interaction between the factors – especially between the firm-specific ratios and macroeconomy factors, suggesting that the influence of given firm-specific ratio (e.g. indebtedness) could be enhance in high interest rate, and vice versa.

# 9 Contribution of the thesis from a scientific, practical, and pedagogical perspective

The presented work contributed 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 address usually country-specific datasets. Whereas the presented model outperforms significantly the competing models, which were specially derived for SMEs 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 the 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 the macroeconomic indicators carry a unique type of information that can be just partly supplemented (and thus not satisfyingly) by a different set of firm-specific variables.
- Proving that adding the combination of macroeconomic and firm-specific indicators is resulting in significantly higher accuracy that could be reached by employing solely firm-specific indicators.
- Presenting especially derived for SME segment of business, showing the specific risk factors of SMEs, which should be examined during the credit application process by the banks or other credit providers.
- Representing a tool for estimating the probability of default, which is one of the most important part 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 following:

- For the teaching of "**Financial Management of Small Company**" the subject – presenting the student 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 to highlight the factor, which should be paid a specific attention when preparing a business plan, which task for the student to complete the subject. Furthermore, showing then students that the when analysing the statements of financial analysis in case of small business, attention has to be paid to more specific factors, 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 a business, the students are required to conduct a strategic analysis, in which they need to assess the macroeconomic factors as well and assess whether the business is meeting a going concern assumption – which otherwise mean, that the business has a perspective. Providing a key result of this work might give the students guidance or rather an insight into the mechanism in which the environment factors and firm level factors interact with the business perspective or rather its survival probability.
- Some of the methodological aspects of this work could be inspiring for the student, who decided for a master thesis the topic of "**Default predicting**". The topic is very challenging for the student on the one hand, on the other, there have been several such being successfully defended at the faculty till nowadays, while some of them also have gained recognition by practitioners during the held competitions.
- The research topic has not been fully explained yet, thus there is a space for a Ph.D. student research to address several existing gaps.

# 10 Conclusion

The presented research focuses on EU-28 countries, the reason behind this is twofold. Firstly, as previously mentioned, a complex study on EU level countries was missing in the current literature. Secondly, it was necessary to obtain enough data variability, whereas 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 were captured by added macroeconomic variables. There was also a try to add categorial variables describing geographic regions of Europe (west, east, north, south), however such procedure would split the model into four smaller models and limit the variability of the macroeconomic indicator 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 8 different macroeconomic factors. To verify the research hypotheses, at 3 different models were created and tested (both in and out of the sample), the results were compared with two models of the other authors, both in their original setting and in their re-estimated form, whereas the re-estimation was done for the research presented.

The main ideas behind the adopted methodological approach could describe 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 (preferable increase) of model accuracy. At first, the appropriate measure of model accuracy had to be selected and the appropriate was to test 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 a strong criticism on employing the accuracy measure, as this measure is highly sensitive to the current model settings (the set of cut-off scores), moreover, the proportion of the tested sample (the proportion of nondefault and default businesses) is also highly affecting 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 were 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 above limited to a value of 1. In the case of 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 set to the rest of the model's variables, i.e., the variable is treated as significant when a removal of such a variable would result in a significant drop of model overall significance. In other words, the significant 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 an insignificance of a previously significant variables (or rather significant in another model). From this perspective, it was a need to derive three different models. For selecting model variables, the stepwise procedure was employed. The first question was whether a macroeconomic indicator would enter the model, otherwise fitted with firm-specific variables, as such entering will mean that there is unique information carried by the given macroeconomic indicators, which increase the overall significance of the model. This idea results in creating 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 the otherwise same model. At this phase, the aim of the research could not fully met, as there was still a chance that the reassessment of the firm-specific variables set would result in a more significant model, than it 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 such added 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 personal cost per employee seems to play a significant role in SMEs survival probability.

# References


1. Act No. 182/2006 Coll. on bankruptcy and settlement (Insolvency Law).

2. AFIK, Z., ARAD, O. and K. GALIL. Using Merton model for default prediction: An empirical assessment of selected alternatives. Journal of Empirical Finance. 2016, 35. 43-67. https://doi.org/10.1016/j.jempfin.2015.09.004

3. 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

4. ALEXANDER, C. and A. KAECK. Regime dependent determinants of credit default swap spread. Journal of Banking and Finance. 2008, 32(6): 1008-1021. https://doi.org/10.1016/j.jbankfin.2007.08.002

5. 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

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

7. 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

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

9. 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.

10. 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

11. 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

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

13. 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

14. 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

15. ANDERSSON, F., MAUSSER, H., ROSEN, D. and S. URYASEV. Credit Risk Optimization with Conditional Value-at Risk Criterion. Mathematical Programming. 2001, 89: 273-291. https://doi.org/10.1007/PL00011399

16. ANDRADE, G. and S. N. KAPLAN. How costly is financial, not economic distress? Evidence from highly leveraged transactions that became distressed. Journal of Finance. 1998, 53(5):1443-1493. https://doi.org/10.1111/0022-1082.00062

17. ANG, J. and K. PATEL. Bond rating methods: comparison and validation, Journal of Finance. 1975, 30(2): 631–640. https://doi.org/10.1111/j.1540-6261.1975.tb01836.x

18. AZIZ, M. and H. DAR. Predicting corporate bankruptcy: Where we stand? Corporate Governance. 2006, 6(1):18-33. https://doi.org/10.1108/14720700610649436

19. 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

20. BALCEAN, S. and H. OOGHE. 35 years of studies on business failure: an overview of the classical statistical methodologies and their related problems. The British Accounting Review. 2006, 38(1): 63-93. https://doi.org/10.1016/j.bar.2005.09.001

21. BARRIE, D. S. and B. C. PAULSON. 1992. *Professional construction management: including CM, design-construct, and general contracting*. New York: McGraw-Hill.

22. BCBS. International convergence of capital measurement and capital standards: A revised framework. Bank for International Settlements. 2005, 25: 1–251.

23. BEAVER, W. H. Financial Ratios as predictors of Failure. Journal of Accounting Research. 1966, 4: 71-111. https://doi.org/10.2307/2490171

24. BEAVER, W. H., MCNICHOLS, M. F. and J. W. RHIE. Have Financial Statements Become Less Informative? Evidence of the Ability of Financial Ratios to Predict Bankruptcy. Review of Accounting Studies. 2005, 10: 93-122. https://doi.org/10.1007/s11142-004-6341-9

25. 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

26. BELLOTTI, T. and J. CROOK. Loss default model incorporating macroeconomic variables for credit cards. International Journal of Forecasting. 2012, 28(1): 171-182. https://doi.org/ 10.1016/j.ijforecast.2010.08.005

27. 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

28. BERGER, A. N. and W. S. FRAME. Small Business Credit Scoring and Credit Availability. Journal of Small Business Management. 2007, 45(1): 5-22. https://doi.org/10.1111/j.1540-627X.2007.00195.x

29. BHARATH, S. AND T. SHUMWAY. 2004. *Forecasting default with the KMV-Merton model*. Working paper, University of Michigan.

30. BIELIKOVÁ, T., BÁNYIOVÁ, T. and A. PITERKOVÁ. Prediction Techniques of Agriculture Enterprises Failure. Procedia Economics and Finance. 2014, 12: 48–56. https://doi.org/10.1016/S2212-5671(14)00319-0

31. BLACK, F. and C. C. COX. Valuing Corporate Securities: Some Effects of Bond Indenture Provisions. Journal of Finance. 1976, 31(2): 351-367. https://doi.org/10.1111/j.1540-6261.1976.tb01891.x

32. BLACK, F. and M. SCHOLES. The pricing of options and corporate liabilities. Journal of Political Economy. 1973, 81(3): 637-659. https://doi.org/10.1086/260062

33. BLUME, M., LIM, F. and A. C. MACKINLAY. The declining credit quality of U.S. corporate debt: Myth or reality? Journal of Finance. 1998, 53(4): 1389-1413. https://doi.org/10.1111/0022-1082.00057

34. BORATYŃSKA, K. Corporate bankruptcy and survival on the market: Lessons from evolutionary economics. Oeconomia Copernicana. 2016, 7(1): 107-129. https://doi.org/10.12775/OeC.2016.008

35. BRADLEY, A. P. The use of the area under the roc curve in the evaluation of machine learning algorithms. Pattern Recognition. 1997, 30(7): 1145-1159. https://doi.org/10.1016/S0031-3203(96)00142-2

36. CAMBRIDGE DICTIONARY. Correlation. 2019.Available at: https://dictionary.cambridge.org/dictionary/english/correlation

37. CAMPA, D. and M. CAMACHO-MIÑANO. The impact of SME's pre-bankruptcy financial distress on earnings management tools. International Review of Financial Analysis. 2015, 42: 222-234. https://doi.org/10.1016/j.irfa.2015.07.004

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

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

40. 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

41. 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

42. CHEN, L. H. and H. D. HSIAO. Feature selection to diagnose a business crisis by using a real GA-based support vector machine: An empirical study. Expert Systems with Applications. 2008, 35: 1145-1155. https://doi.org/10.1016/j.eswa.2007.08.010

43. CHENG, C. B., CHEN, C. L. and C. J. FU. Financial Distress Prediction by a Radial Basis Function Network with Logit Analysis Learning. Computers and Mathematics with Applications. 2006, 51(3-4): 579-588. https://doi.org/10.1016/j.camwa.2005.07.016

44. 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

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

46. COCHRAN, W. G. On the Performance of the Linear Discriminant Function. Technometrics. 1964, 6(2): 179-190. https://doi.org/10.2307/1266150

47. COLLIN-DUFRESNE, P. and R. GOLDSTEIN. Do credit spreads reflect stationary leverage ratios? Journal of Finance. 2001, 56(5): 1929-1957. https://doi.org/10.1111/0022-1082.00395

48. 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

49. CRANEY, T. A. and J. G. SURLES. Model-Dependent Variance Inflation Factor Cutoff Values. Quality Engineering. 2002, 14(3): 391-403. https://doi.org/10.1081/QEN-120001878

50. CREDITREFORM. Vývoj insolvencí v České republice v roce 2017 [Number of insolvencies in Czech Republic in 2017]. 2018. Available at: https://www.creditreform.cz/fileadmin/user_upload/CR-International/local_documents/cz/Presseartikel/Tiskova_zprava_vyvoj_insolvenci_v_CR_v_roce_2017.pdf

51. CREFOPORT. Vývoj insolvencí v České republice v roce 2018 [Number of insolvencies in Czech Republic in 2018]. 2019. Available at: http://crefoport.cz/wp-content/uploads/2019/01/TZ_Insolvence_2018_fin.pdf

52. CRESSY, R. C. U.K. small firm bankruptcy prediction: A logit analysis of financial trend-, industry-, and macro-effects. Journal of Small Business Finance. 1992. 1(3): 233-253.

53. DAS, S., HANOUNA, P., and A. SARIN. Accounting-based versus market-based cross-sectional models of CDS spreads. Journal of Banking and Finance. 2009, 33(4): 719-730. https://doi.org/10.1016/j.jbankfin.2008.11.003

54. DE MOOR, L., WIECZOREK-KOSMALA, M. and J. BLACH. SME Debt Financing Gap: The Case of Poland. Transformations in Business & Economics, 2016, 15(39): 274-291.

55. DEAKIN, E. B. A Discriminant Analysis of Predictors of Business Failure. Journal of Accounting Research. 1972, 10(1): 167-179. https://doi.org/10.2307/2490225

56. 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.

57. DEMIROVIC, A. and D. C. THOMAS. The relevance of accounting data in the measurement of credit risk. The European Journal of Finance. 2007, 13(3): 253-268. https://doi.org/10.1080/13518470601025177

58. 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.

59. DIETSCH, M. and J. PETEY. Should SME exposure be treated as retail or as corporate exposures? A comparative analysis of default probabilities and asset correlation in French and German SMEs. Journal of Banking and Finance. 2004, 28(4): 773-788. https://doi.org/10.1016/j.jbankfin.2003.10.006

60. DING, Y., SONG, X. and Y. ZEN. Forecasting financial condition of Chinese listed companies based on support vector machine. Expert Systems with Applications. 2008, 34(4): 3081-3089. https://doi.org/10.1016/j.eswa.2007.06.037

61. DOUMPOS, M., ANDRIOSOPOULOS, K., GALARIOTIS, E., MAKRIDOU, G. and C. ZOPOUNIDIS. Corporate failure prediction in the European energy sector: a multicriteria approach and the effect of country characteristics. European Journal of Operational Research, 2017, 262(1): 347–360. https://doi.org/10.1016/j.ejor.2017.04.024

62. EDISON, H. J. Do indicators of financial crisis work? An evaluation of an early warning system. International Journal of Finance and Economics. 2013, 8(1): 11-53. https://doi.org/10.1002/ijfe.197

63. EDMISTER, R. An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction. Journal of Financial and Quantitative Analysis. 1972, 7(2): 1477-1493. https://doi.org/10.2307/2329929

64. 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

65. ENIOLA, A. A. and H. ENTEBANG. SME Firm Performance-Financial Innovation and Challenges. Procedia - Social and Behavioural Sciences. 2015, 195: 334-342. https://doi.org/10.1016/j.sbspro.2015.06.361

66. 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

67. ETEMADI, H. and Y. TARIVERDI. The effect of cash flow statement on the predictions of professional investors. Iranian Accounting and Auditing Review. 2006, 45: 69-86.

68. EUROPEAN COMMISSION. Annual Report on European SMEs 2018/2019. 2019. Available at: file:///C:/Users/karas/Downloads/SME%20Annual%20report%202018-2019%20(1).pdf

69. EUROSTAT. Key figures in Europe. 2018. Available at: https://ec.europa.eu/eurostat/web/products-statistical-books/-/KS-EI-18-001

70. 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

71. FAMA, E. F. Term premiums and default premiums in money markets, Journal of Financial Economics. 1986, 17(1): 175-196.

72. FAUCEGLIA D. Credit constraints, firm exports, and financial development: Evidence from developing countries. The Quarterly Review of Economics and Finance. 2015, 55: 53–66. https://doi.org/10.1016/j.qref.2014.08.003

73. FAZZARI, S., HUBBARD, R. and B. PETERSEN. 1988. *Financing Constraints and Corporate Investment*. National Bureau of Economic Research, Inc, NBER Working Papers. 1.

74. 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

75. FISHER, R. A. The use of multiple measurements in taxonomic problems. Annals of Human Genetics 1936, 7(2): 179-188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x

76. FREUND, R. J. and R. C. LITTELL 2000. *SAS system for regression*. New York: Wiley.

77. FRIDSON, M. and F. ALVAREZ. 2002. *Financial Statement Analysis: A Practitioner's Guide*. New York: John Wiley & Sons.

78. FRYDMAN, H., ALTMAN, E. I. and D. L. KAO. Introduction Recursive Partitioning for Financial Classification: The Case of Financial Distress. Journal of Finance. 1985, 40(1): 269-291. https://doi.org/10.1111/j.1540-6261.1985.tb04949.x

79. GESKE, R. The Valuation of Corporate Liabilities as Compound Options. Journal of Financial and Quantitative Analysis. 1977, 12(4):541-552. https://doi.org/10.2307/2330330

80. GILSON, S. C. Management turnover and financial distress. Journal of Financial Economics. 1989, 25(2): 241-262. https://doi.org/10.1016/0304-405X(89)90083-4

81. 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

82. GRUNERT, J., NORDEN, L. and M. WEBER. The role of nonfinancial factors in internal credit ratings. Journal of Banking & Finance. 2005, 29(2): 509–531. https://doi.org/10.1016/j.jbankfin.2004.05.017

83. 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.

84. 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.

85. 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

86. HAND, D. J. and R. J. TILL. A simple generalisation of the area under the ROC curve for multiple class classification problems. Machine Learning. 2001, 45: 171-186. https://doi.org/10.1023/A:1010920819831

87. HANLEY, J. A. and B. J. MCNEIL. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982, 143(1): 29-36. https://doi.org/10.1148/radiology.143.1.7063747

88. HASTIE. T., TIBSHIRANI, R. and J. FRIEDMAN. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer.

89. HEBÁK, P., HUSTOPECKÝ, J., JAROŠOVÁ, E. and I. PECÁKOVÁ. 2004. *Vícerozměrné statistické metody (1) [Multivariate statistical methods]*. Praha: Informatorium.

90. HEO, J. and J. Y. YANG. AdaBoost based bankruptcy forecasting of Korean construction companies. Applied Soft Computing. 2014, 24(1): 494-499. https://doi.org/10.1016/j.asoc.2014.08.009

91. 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

92. 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

93. 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

94. JARROW, R. A. and S. M. TURNBULL. Pricing Derivatives on Financial Securities Subject to Credit Risk. Journal of Finance. 1995, 50: 53-86. https://doi.org/10.1111/j.1540-6261.1995.tb05167.x

95. 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

96. JONES, J. Earnings Management Following Import Relief Investigations. Journal of Accounting Research. 1991, 29(2): 193-228. https://doi.org/10.2307/2491047

97. JONES S. A cash flow-based model of corporate bankruptcy in Australia. Journal of Applied Management Accounting Research. 2016, 3(2): 21-35.

98. JONES, S. and R. H. BELKAOUI. 2010. *Financial Accounting Theory* (3rd Edition). Sydney: Cengage.

99. JONES, S. Does the capitalization of intangible assets increase the predictability of corporate failure? Accounting Horizons. 2011, 25(1): 41-70. https://doi.org/10.2308/acch.2011.25.1.41

100. KAMINSKY, G. and C. REINHART. The Twin Crises: The Causes of Banking and Balance-of-Payments Problems. American Economic Review. 1999, 89(3): 473-500. https://doi.org/10.1257/aer.89.3.473

101. KAPLAN, E. L. and P. MEIER. Nonparametric Estimation from Incomplete Observations, Journal of the American Statistical Association. 1958, 53(282): 457-481.

102. 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.

103. KARAS, M. and M. REŽŇÁKOVÁ. Bankruptcy Prediction Model of Industrial Enterprises in the Czech Republic. International Journal of Mathematical Models and Methods in Applied Sciences, 2013, 7: 519-531.

104. KARAS, M. and M. REŽŇÁKOVÁ. Predicting bankruptcy under alternative conditions: the effect of a change in industry and period on the accuracy of the model. Procedia Social and Behavioural Sciences. 2015, 213(1): 397-403. https://doi.org/10.1016/j.sbspro.2015.11.557

105. KARAS, M. and M. REŽŇÁKOVÁ. The potential of dynamic indicators in the development of bankruptcy prediction models: The case of construction companies. Acta Universitatis Agriculturae Et Silviculturae Mendelianeae Brunensis. 2017, 65(2): 641-652. https://doi.org/10.11118/actaun201765020641

106. KARAS, M. and REŽŇÁKOVÁ, M. The stability of bankruptcy predictors in the construction and manufacturing industries at various times before bankruptcy. E a M: Ekonomie a Management. 2017, 20(2): 116-133. https://dx.doi.org/10.15240/tul/001/2017-2-009

107. KARAS, M. and REŽŇÁKOVÁ, M. Cash Flows Indicators in the Prediction of Financial Distress. Inzinerine Ekonomika-Engineering Economics. 2020, 31(5): 525-535. https://doi.org/10.5755/j01.ee.31.5.25202

108. KENNEDY, P. Oh No! I Got the Wrong Sign! What Should I Do? The Journal of Economic Education. 2005, 36 (1): 77-92.

109. KIM, I. J., RAMASWAMY, K. and S. SUNDARESAN. Does Default Risk in Coupons Affect the Valuation of Corporate Bonds? A Contingent Claims Model. Financial Management. 1993, 22 (3):117-131. https://doi.org/10.2307/3665932

110. KIM, M. J. and D. K. KANG. Ensemble of neural networks for bankruptcy prediction. Expert Systems with Applications. 2010, 37: 3373–3379. https://doi.org/10.1016/j.eswa.2009.10.012

111. KNEDLIK T. The impact of preferences on early warning systems - The case of the European Commission's Scoreboard. European Journal of Political Economy. 2014, 34: 157-166. https://doi.org/10.1016/j.ejpoleco.2014.01.008

112. 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

113. KORDESTANI, G. BIGLARI, S. V. and M. BAKHTIARI. Ability of combinations of cash flow components to predict financial distress. Business: Theory and Practice. 2011, 12(3): 277-285. https://doi.org/10.3846/btp.2011.28

114. LANDAU, S., EVERITT, B. S. 2004. *A handbook of statistical analysis using SPSS*. Chapman & Hall/CRC Press LLC.

115. LANDO, D. On Cox Processes and Credit Risky Securities. Review of Derivatives Research. 1998, 2(2): 99-120.

116. LEE, S. and W. S. CHOI. A multiindustry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis. Expert Systems with Applications. 2013, 40(8):2941-2946. https://doi.org/10.1016/j.eswa.2012.12.009

117. LI, H. and J. SUN. Predicting business failure using multiple case-based reasoning combines with support vector machine. Expert Systems with Applications. 2009, 36(6): 10085-10096. https://doi.org/10.1016/j.eswa.2009.01.013

118. LI, H., SUN, J. and J. WU. Predicting business failure using classification and regression tree: An empirical comparison with popular classical statistical methods and top classification mining methods. Expert Systems with Applications. 2010, 37(8): 5895-5904. https://doi.org/10.1016/j.eswa.2010.02.016

119. LI, L. and R. FAFF. Predicting corporate bankruptcy: What matters? International Review of Economic and Finance. 2019, 62: 1-19. https://doi.org/10.1016/j.iref.2019.02.016

120. LIANG, D., LU, C., TSAI, F. C. and G. SHIH. Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. European Journal of Operational Research. 2016, 252(2): 561-572. https://doi.org/10.1016/j.ejor.2016.01.012

121. LIN, F., LIANG. D. and E. CHEN. Financial ratio selection for business crisis prediction. Expert Systems with Applications. 2011, 38(12): 15094-15102. https://doi.org/10.1016/j.eswa.2011.05.035

122. LIN, S. L. A new two-stage hybrid approach of credit risk in the banking industry. Expert Systems with Applications. 2009, 36(4): 8333-8341. https://doi.org/10.1016/j.eswa.2008.10.015

123. LING, C. X. and H. ZHANG. On Bayesian classifiers with accurate probabilities. In: Proceedings of the Sixth Pacific-Asia Conference on KDD. 2002.

124. LING, X. C., HUANG, J. and H. ZHANG. AUC: Statistically Consistent and more Discriminating Measure than Accuracy. In: Proceedings of the 18th International Joint Conference on Artificial Intelligence. 2003.

125. LITTERMAN, R. and T. IBEN. Corporate Bond Valuation and the Term Structure of Credit Spreads. Journal of Portfolio Management. 17, 52-64. https://doi.org/10.3905/jpm.1991.409331

126. LONGSTAFF, F. A. and E. S. SCHWARTZ. A Simple Approach to Valuing Risky Fixed and Floating Rate Debt. Journal of Finance. 1995, 50: 789-819. https://doi.org/10.1111/j.1540-6261.1995.tb04037.x

127. LONGSTAFF, F. and A. RAJAN. An empirical analysis of the pricing of collateralized debt obligations. Journal of Finance. 2008, 63(2): 529–563. https://doi.org/10.1111/j.1540-6261.2008.01330.x

128. MAI, F., TIAN, S., LEE, C. and L. MA. Deep learning model for bankruptcy prediction using textual search. European Journal of Operational Research. 2019, 274(2): 743-758. https://doi.org/10.1016/j.ejor.2018.10.024

129. MAKRIDAKIS, S. and N. TALEB N. Living in a world of low levels of predictability. International Journal of Forecasting. 2009, 25(4): 840-844. https://doi.org/10.1016/j.ijforecast.2009.05.008

130. MANN, H. B. and D. R. WHITNEY. On a test of whether one of the 2 random variables are stochastically larger than the other. Annals of Mathematical Statistics. 1947, 18: 50-60.

131. MARE, D. S. Contribution of macroeconomic factors to the prediction of small bank failures. In: Proceedings of the 4th International IFABS Conference. 2012.

132. MARTENS, D., BRUYNSEELS, L., BAESENS, B., WILLEKENS, M. and J. VANTHIENEN. Predicting going concern opinion with data mining. Decision Support Systems. 2008, 45(4): 765-777.https://doi.org/10.1016/j.dss.2008.01.003

133. MARTIN D. Early warning of bank failure: A logit regression approach. Journal of Banking & Finance. 1977, 1(3): 249-276. https://doi.org/10.1016/0378-4266(77)90022-X

134. MCGUINNESS G, HOGAN T and R. POWELL. European trade credit use and SME survival. Journal of Corporate Finance. 2018, 49: 81-103. https://doi.org/10.1016/j.jcorpfin.2017.12.005

135. MCLEAY, S. and A. OMAR. The sensitivity of prediction models to the non-normality of bounded and unbounded financial ratios. British Accounting Review. 2000, 32(2): 213-230. https://doi.org/10.1006/bare.1999.0120

136. MENSAH, Y. M. The Differential Bankruptcy Predictive Ability of Specific Price Level Adjustments: Some Empirical Evidence. The Accounting Review. 1983, 58(2): 228-246.

137. 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

138. MILERIS, R. and V. BOGUSLAUKAS. Credit Risk Estimation Model Development Process: Main Steps and Model Improvement. Inzinerine Ekonomika Engineering Economics 2011, 22(2): 126-133. http://dx.doi.org/10.5755/j01.ee.22.2.309

139. MITCHELL, T. Bank Default Prediction: A Comparative Model using Principal Component Analysis. Journal of Stock & Forex Trading. 2015, 4(2):149. http://dx.doi.org/10.4172/2168-9458.1000149

140. MOCKING, R., MÖHLMANN, J., PALALI, A. and S. VAN VELDHUIZEN. 2016. *Dependence on External Finance and SME Survival*. CPB Background Document. CPB Netherlands Bureau for Economic Policy Analysis.

141. 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

142. NIEMANN, M., SCHMIDT, J. H. and M. NEUKIRCHEN. Improving the performance of corporate rating prediction models by reducing financial ratio heterogeneity. Journal of Banking & Finance. 2008, 32(3): 434-446. https://doi.org/10.1016/j.jbankfin.2007.05.015

143. 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

144. 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

145. 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

146. ONG, S., YAP, V. and R. KHONG. Corporate failure prediction: a study of publicly listed companies in Malaysia, Managerial Finance. 2011, 37(6): 553-564. https://doi.org/10.1108/03074351111134745

147. PATEL, S. and C. D. BHAVSAR. Analysis of pharmacokinetic data by Wilk's lambda (An important tool of MANOVA). International Journal of Pharmaceutical Science Invention. 2013, 2(1): 36-44.

148. PEEL, M. J. and D. A. PEEL. Some further empirical evidence on predicting private company failure. Accounting and Business Research. 1987, 18: 57-66. https://doi.org/10.1080/00014788.1987.9729348

149. PERRY, L., HENDERSON, G. JR and T. CRONAN. Multivariate analysis of corporate bond ratings and industry classification. Journal of Financial Research. 1984, 7(1): 27-36. https://doi.org/10.1111/j.1475-6803.1984.tb00351.x

150. 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

151. POMPE, P. P. M. and J. BILDERBEEK. The prediction of bankruptcy of small- and medium-sized industrial firms. Journal of Business Venturing. 2005, 20(6): 847-868. https://doi.org/10.1016/j.jbusvent.2004.07.003

152. PSILLAKI, M., TSOLAS, I .T. and M. MARGARITIS. Evaluation of credit risk based on firm performance. European Journal of Operational Research. 2010, 201(3): 873-881. https://doi.org/10.1016/j.ejor.2009.03.032

153. PURBA, S., SUTARMAN and O. SEMBIRING. Maximum Likelihood Based on Newton Raphson, Fisher Scoring and Expectation Maximization Algorithm Application to Accident Data. International Journal of Advanced Research. 2018, 6: 965-969.

154. PUSTYLNICK, I., TEMCHENKO, O. and S. GUBARKOV. Estimating the Influence of Accounting Variables Change with Earnings Management Detection. Journal of International Studies. 2017, 10(1): 110-122. https://doi.org/10.14254/2071-8330.2017/10-1/7

155. QU, Y. Macroeconomic factors and probability of default. European Journal of Economics, Finance and Administrative Sciences. 2008, 192-215

156. REŽŇÁKOVÁ, M. and M. KARAS. Bankruptcy Prediction Models: Can the prediction power of the model be improved by using dynamic indicators? Procedia Economics and Finance. 2014, 12: 565-574. https://doi.org/10.1016/S2212-5671(14)00380-3

157. REŽŇÁKOVÁ, M. and M. KARAS. The Prediction Capabilities of Bankruptcy Models in a Different Environment: An example of the Altman Model under the Conditions in the Visegrad Group Countries. Ekonomicky časopis. 2015, 63(6): 617-633.

158. RISKMETRICS GROUP. CreditMetrics—Technical Document. RiskMetrics Group. 2007.

159. RODRÍGUEZ, G. Lecture Notes on Generalized Linear Models. 2007. Available at: http://data.princeton.edu/wws509/notes/

160. ROGERS, J. L. and W. A. NICEWANDER. Thirteen Ways to Look at the Correlation Coefficient. The American Statistician. 1988, 42(1): 59-66. https://doi.org/10.1080/00031305.1988.10475524

161. SAUNDERS, A. and L. ALLEN. 2002. Credit Risk Measurement: New Approaches to Value -at-Risk and Other Paradigms. second ed. New York: Wiley Finance.

162. SAURINA, J. and C. TRUCHARTE. The impact of BASEL II on lending to small- and medium-sized firms: A regulatory policy assessment based on Spanish credit register data. Journal of Finance Service Research. 2004, 26(3): 121-144. https://doi.org/10.1023/B:FINA.0000037542.16864.11

163. SCOTT J. The probability of bankruptcy: a comparison of empirical predictions and theoretical models. Journal of Banking & Finance, 1981, 5(3): 317-344. https://doi.org/10.1016/0378-4266(81)90029-7

164. SHIN, K. and I. HAN. A case-based approach using inductive indexing for corporate bond ratings. Decision Support Systems. 2001, 32(1): 41-52. https://doi.org/10.1016/S0167-9236(01)00099-9

165. SHIN, K. and I. HAN. Case-based reasoning supported by genetic algorithms for corporate bond rating. Expert Systems with Applications. 1999, 16(2): 85–95. https://doi.org/10.1016/S0957-4174(98)00063-3

166. SHUMWAY, T. Forecasting Bankruptcy More Accurately: A Simple Hazard Model. Journal of Business. 2001, 74(1): 101-24. https://doi.org/10.1086/209665

167. SOKOLOVA, M., JAPKOWICZ, N. and S. SZPAKOWICZ. Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. In: Australian Conference on Artificial Intelligence. 2006, 1015-1021. http://dx.doi.org/10.1007/11941439_114

168. ŠPIČKA J. The financial condition of the construction companies before bankruptcy. European Journal of Business and Management. 2013, 5: 160-166.

169. 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.

170. SUHAILY H., RASHIDAH, A. R. and S. MAHENTHIRAN. Management Motives, Weak Governance, Earnings Management, and Fraudulent Financial Reporting: Malaysian Evidence. Journal of International Accounting Research. 2013, 12(1): 1-27. https://doi.org/10.2308/jiar-50353

171. SUN, J., LIAO, B. and H. LI. AdaBoost and bagging ensemble approaches with neural network as a base learner for financial distress prediction of Chinese construction and real estate companies. Recent patents on computer science. 2013, 6(1): 47-59. 10.2174/2213275911306010007

172. TAFFLER, R. J. Forecasting company failure in the UK using discriminant analysis and financial ratio data. Journal of the Royal Statistical Society. 1982, 145(3): 342–358. https://doi.org/10.2307/2981867

173. TAM, K. and M. KIANG. Managerial applications of neural networks: the case of bank failure prediction. Management Science. 1992, 38(7): 926-947. https://doi.org/10.1287/mnsc.38.7.926

174. THOMAS NG, S. T., WONG, J. M. W. and J. ZHANG. Applying Z-score model to distinguish insolvent construction companies in China. Habitat International. 2011, 35(1): 599-607. https://doi.org/10.1016/j.habitatint.2011.03.008

175. TINOCO, M. H. and N. WILSON. Financial distress and bankruptcy prediction among listed companies using accounting, market, and macroeconomic variables. International Review of Financial Analysis. 2013, 30: 394-419. https://doi.org/10.1016/j.irfa.2013.02.013

176. 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

177. 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

178. TSAI C. Feature selection in bankruptcy prediction. Knowledge-Based Systems. 2009, 22: 120-127. https://doi.org/10.1016/j.knosys.2008.08.002

179. TSERNG, P. CHEN, P., HUANG, W., LEI, M. and Q. TRAN. Prediction of default probability for construction firms using the logit model, Journal of Civil Engineering and Management. 2014, 20(2): 247-255. https://doi.org/10.3846/13923730.2013.801886

180. 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

181. VASICEK, O. A. 1984. *Credit Valuation*, KMV Corporation, March.

182. VASSALOU, M. and Y. XING. Default risk in equity returns. Journal of Finance. 2004, 59(2): 831-868. https://doi.org/10.1111/j.1540-6261.2004.00650.x

183. VAVŘINA J., HAMPEL D. and J. JANOVÁ. New approaches for financial Distress classification in agribusiness. Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis. 2013, 61(4): 1177-1182. https://doi.org/10.11118/actaun201361041177

184. WANG, G. and J. MA. Study of corporate credit risk prediction based on integrating boosting and random subspace. Expert Systems with Applications. 2011, 38(11): 13871-13878. https://doi.org/10.1016/j.eswa.2011.04.191

185. WANG, Y. J. and H. S. LEE. A clustering method to identify representative financial ratios. Information Sciences. 2008, 178(4): 1087-1097. https://doi.org/10.1016/j.ins.2007.09.016

186. WELC, J. EBITDA vs. Cash Flows in Bankruptcy Prediction on the Polish Capital Market. European Financial and Accounting Journal. 2017, 12(2):91-103. https://doi.org/10.18267/j.efaj.183

187. WHITAKER, R. The early stages of financial distress. Journal of Economics and Finance. 1999, 23:123–132. https://doi.org/10.1007/BF02745946

188. WILCOXON F. Individual comparisons by ranking methods. Biometrics Bulletin. 1945, 1: 80-83. https://doi.org/10.1007/978-1-4612-4380-9_16

189. WILSON, T. C. Portfolio credit risk (I). Risk. 1997, 10(9): 111-17.

190. WRUCK, K. H. Financial distress, reorganization, and organizational efficiency. Journal of Financial Economics. 1990, 27(2): 419-444. https://doi.org/10.1016/0304-405X(90)90063-6

191. WU, L. and F. X. ZHANG. A non-arbitrage analysis of macroeconomic determinants of the credit spread term structure Management Science. 2008, 54(6): 1160-1175. https://doi.org/10.1287/mnsc.1070.0835

192. WU, W. Beyond business failure prediction. Expert Systems with Applications. 2010, 37(3): 2371-2376. https://doi.org/10.1016/j.eswa.2009.07.056

193. 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

194. 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

195. ZHOU, Y. and T. M. S. ELHAG. Apply Logit analysis in Bankruptcy Prediction. In: Proceedings of the 7th WSEAS International Conference on Simulation, Modelling and Optimization. 2007, 301-308.

196. ZIMMERMAN, D. W. A note on the influence of outliers on parametric and nonparametric tests. The Journal of General Psychology. 1994, 121(4): 391-401. https://doi.org/10.1080/00221309.1994.9921213

197. ZIMMERMAN, D. W. Increasing the power of nonparametric tests by detecting and down-weighting outliers. Journal of experimental education. 1995, 64(1): 71-79. https://doi.org/10.1080/00220973.1995.9943796

198. ZIMMERMAN, D. W. Invalidation of parametric and nonparametric statistical tests by concurrent violation of two assumptions. Journal of Experimental Education. 1998, 67(1): 55-69. https://doi.org/10.1080/00220979809598344

199. 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

## List of figures

# List of tables

# List of abbreviations

A – asset value

AMEX - American Stock Exchange

AS – Altman, Sabato model

AT - Austria

AUC – Area Under Curve

AUC – Area Under Curve

B – estimates of regression coefficient

BCBS - Basel Committee on Banking Supervision

BE - Belgium

BG - Bulgaria

C - cash

CA - current assets

CAPEX - capital expenditures

CashR - Cash Ratio

CDS - Credit default swap

CE - capital employed

CF – cash flow

CL - current liability

CR - current ratio (CR)

CY - Cyprus

CZ - Czechia

D - debt

D&B - Dun & Bradstreet

DCP - Debtor collection period

DD - distance with default

DE - Germany

DEA – Data Envelopment Analysis

df – degrees of freedom

DK - Denmark

EAD - exposure by default

EAT – Earnings Taxes

EBITDA – Earnings Before Interest, Taxes, Depreciation and Amortization

ECB – European Central Bank

EE - Estonia

EL - Greece

ES - Spain

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

FI - Finland

Fn - false negative

Fp - false positive

FR - France

GDP – gross domestic product

GLM – general linear model

GNP - Gross national income

GVA - Gross Value Added

H0(t) - baseline cumulative hazard function

HICP - Harmonised index of consumer prices

HR - Croatia

HU - Hungary

IE – interest expenses

IE - Ireland

IMF - International Monetary Fund

IND – industry (group)

IT - Italy

LDA - linear discrimination analysis

LGD - Loss Default

LT - Lithuania

LU - Luxembourg

LV - Latvia

MT - Malta

MVE – market value of equity

N – number of observations

N(0,1) - the normal distribution with mean 0 and variance 1

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

NL - Netherlands

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

PL - Poland

pp – percentage point

PT - Portugal

QA – quick assets

QDA - quadratic form of discriminant analysis

R – Pearson’s correlation coefficient

$R^2$ – determination index

RE – retained profit

RO - Romania

ROA - return on asset ratio

ROC – Receiver Operating Characteristics

S - sales

S(t) – survival probability at time t

SB – small businesses

SE – standard error

SE – Sweden

SHP - Stock holding period

SI - Slovenia

SIC - Standard Industrial Classification

sig. – significance (p-value)

SK - Slovakia

SME – small and medium enterprise

ST – stock (inventory)

T – time

TA –total assets,

TC – trade creditors

TCPP - Trade creditors payment period

TL – total liability

tn - true negative

tp - true positive,

TTA - tangible assets

UK – United Kingdom

US – United States

USD – United States Dollar

VIF – Variance Inflation Factor

WC - working capital

Z- Z-score

Z’ – revised Z-score

$\sigma_A$ - volatility of assets return

# List of appendices

Appendix 1 Inflation rates per analysed country and year

| GEO/TIME | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 |
|---|---|---|---|---|---|---|---|---|---|---|
| Belgium | 92.09 | 95.18 | 97.68 | 98.90 | 99.38 | 100.00 | 101.77 | 104.03 | 106.44 | 107.77 |
| Bulgaria | 96.66 | 99.94 | 102.33 | 102.72 | 101.08 | 100.00 | 98.68 | 99.85 | 102.48 | 104.99 |
| Czech Republic | 92.6 | 94.6 | 98.0 | 99.3 | 99.8 | 100.0 | 100.7 | 103.1 | 105.1 | 107.8 |
| Denmark | 94.1 | 96.6 | 98.9 | 99.4 | 99.8 | 100.0 | 100.0 | 101.1 | 101.8 | 102.5 |
| Germany | 92.7 | 95.0 | 97.0 | 98.6 | 99.3 | 100.0 | 100.4 | 102.1 | 104.0 | 105.5 |
| Estonia | 87.96 | 92.43 | 96.33 | 99.46 | 99.93 | 100.00 | 100.80 | 104.48 | 108.05 | 110.50 |
| Ireland | 96.2 | 97.4 | 99.2 | 99.7 | 100.0 | 100.0 | 99.8 | 100.1 | 100.8 | 101.7 |
| Greece | 99.27 | 102.36 | 103.42 | 102.54 | 101.11 | 100.00 | 100.02 | 101.15 | 101.94 | 102.46 |
| Spain | 94.08 | 96.94 | 99.31 | 100.83 | 100.63 | 100.00 | 99.66 | 101.69 | 103.46 | 104.26 |
| France | 94.05 | 96.20 | 98.33 | 99.31 | 99.91 | 100.00 | 100.31 | 101.47 | 103.60 | 104.95 |
| Croatia | 92.55 | 94.59 | 97.76 | 100.04 | 100.26 | 100.00 | 99.37 | 100.67 | 102.23 | 103.04 |
| Italy | 92.6 | 95.3 | 98.4 | 99.7 | 99.9 | 100.0 | 99.9 | 101.3 | 102.5 | 103.2 |
| Cyprus | 95.09 | 98.40 | 101.45 | 101.84 | 101.57 | 100.00 | 98.78 | 99.45 | 100.23 | 100.78 |
| Latvia | 92.96 | 96.88 | 99.09 | 99.11 | 99.79 | 100.00 | 100.10 | 103.00 | 105.63 | 108.53 |
| Lithuania | 92.43 | 96.24 | 99.28 | 100.44 | 100.68 | 100.00 | 100.68 | 104.42 | 107.07 | 109.47 |
| Luxembourg | 91.44 | 94.85 | 97.59 | 99.25 | 99.94 | 100.00 | 100.04 | 102.15 | 104.21 | 105.93 |
| Hungary | 89.47 | 92.98 | 98.24 | 99.92 | 99.94 | 100.00 | 100.45 | 102.84 | 105.84 | 109.46 |
| Malta | 91.79 | 94.10 | 97.13 | 98.08 | 98.84 | 100.00 | 100.90 | 102.18 | 103.95 | 105.54 |
| Netherlands | 92.05 | 94.32 | 96.99 | 99.47 | 99.79 | 100.00 | 100.11 | 101.40 | 103.02 | 105.78 |
| Austria | 90.14 | 93.35 | 95.75 | 97.77 | 99.20 | 100.00 | 100.97 | 103.22 | 105.41 | 106.98 |
| Poland | 92.7 | 96.3 | 99.8 | 100.6 | 100.7 | 100.0 | 99.8 | 101.4 | 102.6 | 104.8 |
| Portugal | 93.22 | 96.54 | 99.22 | 99.65 | 99.50 | 100.00 | 100.64 | 102.20 | 103.40 | 103.71 |
| Romania | 87.73 | 92.84 | 95.98 | 99.04 | 100.41 | 100.00 | 98.93 | 100.00 | 104.08 | 108.15 |
| Slovenia | 93.86 | 95.81 | 98.50 | 100.40 | 100.77 | 100.00 | 99.85 | 101.40 | 103.36 | 105.11 |
| Slovakia | 91.69 | 95.43 | 99.00 | 100.45 | 100.35 | 100.00 | 99.52 | 100.90 | 103.46 | 106.33 |
| Finland | 90.83 | 93.85 | 96.81 | 98.96 | 100.16 | 100.00 | 100.39 | 101.23 | 102.42 | 103.58 |
| Sweden | 96.43 | 97.75 | 98.66 | 99.10 | 99.30 | 100.00 | 101.14 | 103.02 | 105.12 | 106.93 |
| United Kingdom | 89.4 | 93.4 | 96.1 | 98.5 | 100.0 | 100.0 | 100.7 | 103.4 | 105.9 | 107.8 |
| Iceland | 85.85 | 89.46 | 94.84 | 98.76 | 99.74 | 100.00 | 100.79 | 99.13 | 99.86 | 101.83 |
| Norway | 92.8 | 94.0 | 94.3 | 96.2 | 98.0 | 100.0 | 103.9 | 105.8 | 109.0 | 111.5 |
| Switzerland | 101.39 | 101.49 | 100.76 | 100.83 | 100.84 | 100.00 | 99.47 | 100.11 | 101.03 | 101.41 |
| North Macedonia | 92.57 | 95.52 | 97.26 | 99.91 | 99.87 | 100.00 | 100.24 | 102.35 | 104.66 | 105.42 |
| Serbia | 74.9 | 83.3 | 89.4 | 96.3 | 98.5 | 100.0 | 101.3 | 104.7 | 106.8 | 108.8 |

Appendix 2 Personal cost per analysed country and year

| **GEO/TIME** | **2008** | **2009** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** |
|---|---|---|---|---|---|---|---|---|---|---|
| Belgium | 46.6 | : | 48.1 | 48.7 | 50.9 | 51.9 | 52.2 | 52.4 | 52.0 | 52.5 |
| Bulgaria | 4.0 | : | 4.5 | 4.9 | 5.3 | 5.6 | 6.0 | 6.5 | 7.0 | 7.7 |
| Czech Republic | 14.9 | 14.4 | 15.4 | 16.2 | 16.1 | 15.7 | 15.2 | 15.9 | 16.7 | 18.3 |
| Denmark | 49.7 | 49.8 | 49.2 | 47.1 | 48.0 | 48.7 | 49.1 | 50.2 | 51.5 | : |
| Germany | 35.7 | 34.6 | 34.9 | 35.4 | 36.3 | 37.2 | 37.9 | 38.8 | 39.1 | 40.3 |
| Estonia | 12.3 | 11.6 | 11.7 | 12.6 | 13.3 | 14.1 | 14.8 | 15.4 | 16.4 | 17.5 |
| Ireland | : | 41.4 | 38.9 | 38.7 | 38.7 | : | 44.8 | 45.0 | 46.4 | 43.2 |
| Greece | 22.9 | 24.4 | 24.0 | 23.7 | 22.5 | 20.2 | 18.6 | 18.5 | 17.4 | 17.5 |
| Spain | 29.4 | 30.2 | 30.2 | 30.8 | 30.5 | 30.5 | 30.8 | 30.8 | 29.9 | 30.2 |
| France | : | 44.4 | 44.5 | 45.7 | 46.9 | 48.1 | 47.7 | 50.4 | 48.3 | 49.6 |
| Croatia | 12.7 | 12.6 | 12.6 | 12.2 | 12.1 | 12.0 | 12.2 | 12.6 | 13.2 | 13.6 |
| Italy | 33.6 | 33.7 | 34.7 | 35.4 | 35.2 | 35.8 | 36.3 | 36.8 | 36.7 | 37.0 |
| Cyprus | 22.7 | 23.2 | 23.3 | 23.3 | 23.3 | 22.3 | 21.7 | 20.5 | 20.4 | 20.6 |
| Latvia | 8.4 | 7.5 | 7.0 | 7.7 | 8.3 | 8.6 | 8.8 | 9.5 | 10.3 | 10.9 |
| Lithuania | 8.7 | 7.7 | 7.4 | 7.9 | 8.2 | 8.8 | 9.2 | 9.8 | 10.6 | 11.8 |
| Luxembourg | 44.3 | 45.7 | 46.4 | 47.8 | 48.7 | 49.6 | 51.4 | 52.7 | 53.1 | 54.8 |
| Hungary | 11.9 | 10.9 | 11.2 | 11.7 | 11.7 | 11.8 | 11.8 | 12.1 | 12.8 | 13.9 |
| Malta | 14.9 | 15.0 | 15.7 | 16.6 | 17.3 | 17.4 | 18.2 | 18.6 | 18.8 | 20.3 |
| Netherlands | 33.5 | 47.7 | 49.5 | 37.1 | 37.7 | 38.1 | 38.4 | 38.9 | 39.1 | 39.2 |
| Austria | 38.9 | 39.5 | 40.0 | 41.1 | 42.5 | 43.7 | 44.3 | 45.2 | 46.4 | 46.5 |
| Poland | : | 9.8 | 11.1 | 11.5 | 11.9 | : | 12.4 | 12.7 | 12.5 | 13.7 |
| Portugal | 16.2 | 16.6 | 17.0 | 17.1 | 17.1 | 17.2 | 17.2 | 17.4 | 17.5 | 17.9 |
| Romania | 6.0 | 5.7 | 5.9 | 6.1 | 6.2 | 6.5 | 7.0 | 7.6 | 8.3 | 9.3 |
| Slovenia | 19.8 | 19.8 | 20.8 | 21.4 | 21.4 | 21.6 | 22.1 | 22.5 | 23.4 | 23.9 |
| Slovakia | 11.7 | 12.9 | 13.1 | 13.8 | 14.6 | 14.8 | 14.6 | 14.9 | 15.5 | 16.2 |
| Finland | 40.4 | 39.8 | 40.6 | 42.0 | 43.5 | 43.1 | 43.2 | 44.4 | 44.8 | 45.1 |
| Sweden | 46.8 | 42.8 | 47.9 | 52.6 | 55.6 | 56.7 | 55.4 | 55.6 | 56.2 | 56.9 |
| United Kingdom | 31.8 | 28.0 | 29.7 | 29.4 | 32.1 | 31.3 | 33.6 | 37.2 | 33.3 | 32.5 |
| Iceland | : | : | : | : | : | : | : | : | 48.8 | 58.8 |
| Norway | : | 54.3 | 60.1 | 64.5 | 70.0 | 70.5 | 66.6 | 62.0 | 59.3 | 60.0 |
| Switzerland | : | : | : | : | : | : | : | : | : | : |
| North Macedonia | : | : | : | 4.3 | 4.3 | 4.6 | 4.9 | 4.9 | 5.4 | 5.7 |
| Serbia | : | : | : | : | : | : | : | : | 8.1 | 8.6 |
| Turkey | : | : | : | : | : | : | : | : | : | : |
| Bosnia and Herzegovina | : | : | : | : | : | : | 7.8 | 8.0 | 7.8 | 7.8 |

Appendix 3, Gross Value added per analysed country and year

| GEO/TIME | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 |
|---|---|---|---|---|---|---|---|---|---|---|
| Belgium | 79.9 | : | 83.0 | 85.7 | 89.0 | 90.5 | 91.8 | 94.4 | 98.1 | 100.5 |
| Bulgaria | 9.7 | : | 9.7 | 10.5 | 11.1 | 11.5 | 12.0 | 13.7 | 14.5 | 15.8 |
| Czech Republic | 30.0 | 28.1 | 30.2 | 31.8 | 30.9 | 30.9 | 31.1 | 32.8 | 33.3 | 35.8 |
| Denmark | 85.1 | 75.6 | 81.5 | 78.9 | 79.5 | 82.1 | 82.4 | 86.4 | 88.3 | : |
| Germany | 59.9 | 55.2 | 57.0 | 57.6 | 57.3 | 58.8 | 60.6 | 60.5 | 62.3 | 63.6 |
| Estonia | 19.5 | 18.5 | 20.7 | 24.1 | 24.8 | 25.9 | 26.9 | 26.7 | 29.1 | 31.1 |
| Ireland | : | 82.9 | 84.4 | 90.2 | 89.9 | : | 107.0 | 153.7 | 151.3 | 159.0 |
| Greece | 44.6 | 44.3 | 43.5 | 41.0 | 39.4 | 36.2 | 31.2 | 32.7 | 27.2 | 30.1 |
| Spain | 50.0 | 48.2 | 48.9 | 49.7 | 48.9 | 48.9 | 50.0 | 50.4 | 49.4 | 49.8 |
| France | 70.9 | 59.7 | 61.9 | 63.1 | 63.0 | 64.6 | 63.8 | 68.5 | 66.0 | 67.9 |
| Croatia | 23.3 | 21.9 | 22.1 | 21.7 | 20.8 | 21.8 | 22.2 | 23.3 | 24.0 | 24.5 |
| Italy | 61.8 | 55.8 | 64.3 | 65.6 | 62.8 | 63.2 | 64.8 | 66.8 | 67.7 | 67.9 |
| Cyprus | 42.2 | 40.1 | 39.2 | 37.4 | 37.0 | 36.4 | 35.9 | 35.8 | 36.3 | 37.4 |
| Latvia | 16.1 | 13.7 | 14.7 | 15.9 | 17.0 | 17.0 | 16.9 | 17.9 | 19.0 | 19.8 |
| Lithuania | 14.4 | 11.5 | 12.9 | 15.2 | 15.6 | 15.6 | 17.4 | 18.5 | 19.4 | 21.2 |
| Luxembourg | 78.8 | 73.2 | 77.8 | 82.5 | 82.4 | 83.9 | 90.4 | 92.4 | 90.8 | 95.5 |
| Hungary | 22.3 | 20.3 | 22.0 | 23.3 | 22.1 | 23.0 | 23.7 | 24.6 | 24.4 | 27.0 |
| Malta | 30.9 | 29.7 | 33.9 | 34.0 | 35.7 | 37.6 | 39.9 | 44.8 | 44.7 | 50.9 |
| Netherlands | 59.9 | 82.7 | 88.8 | 65.9 | 66.3 | 67.1 | 67.8 | 70.3 | 71.7 | 72.7 |
| Austria | 66.0 | 63.0 | 66.1 | 68.8 | 69.0 | 69.7 | 70.0 | 72.2 | 75.1 | 76.4 |
| Poland | : | 22.2 | 25.0 | 26.6 | 26.0 | : | 27.1 | 27.3 | 27.1 | 29.5 |
| Portugal | 28.7 | 28.4 | 29.1 | 28.0 | 27.7 | 28.4 | 29.1 | 29.6 | 30.4 | 31.3 |
| Romania | 13.7 | 11.5 | 12.9 | 12.9 | 13.0 | 14.2 | 15.0 | 14.4 | 15.7 | 17.1 |
| Slovenia | 33.3 | 29.0 | 33.0 | 34.8 | 34.2 | 35.2 | 37.3 | 38.2 | 39.9 | 41.9 |
| Slovakia | 22.7 | 21.8 | 28.3 | 29.6 | 30.7 | 28.8 | 28.2 | 29.2 | 29.9 | 30.6 |
| Finland | 66.6 | 59.2 | 64.6 | 65.4 | 64.7 | 64.7 | 65.1 | 67.7 | 71.5 | 74.2 |
| Sweden | 69.9 | 62.7 | 75.1 | 80.5 | 82.3 | 83.6 | 83.1 | 84.8 | 84.6 | 84.6 |
| United Kingdom | 60.3 | 51.4 | 56.1 | 57.5 | 60.9 | 60.5 | 68.1 | 75.9 | 68.1 | 64.9 |
| Iceland | : | : | : | : | : | : | : | : | 79.6 | 89.4 |
| Norway | : | 118.4 | 133.8 | 148.2 | 162.2 | 152.4 | 138.5 | 125.5 | 115.6 | 123.1 |
| Switzerland | : | : | : | : | : | : | : | : | : | : |
| North Macedonia | : | : | : | 9.8 | 9.6 | 10.0 | 9.8 | 10.1 | 11.4 | 11.7 |
| Serbia | : | : | : | : | : | : | : | : | 15.1 | 15.9 |
| Turkey | : | : | : | : | : | : | : | : | : | : |
| Bosnia and Herzegovina | : | : | : | : | : | : | 15.6 | 16.4 | 16.5 | 15.9 |

Appendix 4, Interest rates per analysed country and year

| **GEO/TIME** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** | **2018** | **2019** |
|---|---|---|---|---|---|---|---|---|---|---|
| Belgium | 3.46 | 4.23 | 3.00 | 2.41 | 1.71 | 0.84 | 0.48 | 0.72 | 0.79 | 0.19 |
| Bulgaria | 6.01 | 5.36 | 4.50 | 3.47 | 3.35 | 2.49 | 2.27 | 1.60 | 0.89 | 0.43 |
| Czech Republic | 3.88 | 3.71 | 2.78 | 2.11 | 1.58 | 0.58 | 0.43 | 0.98 | 1.98 | 1.55 |
| Denmark | 2.93 | 2.73 | 1.40 | 1.75 | 1.32 | 0.69 | 0.32 | 0.48 | 0.45 | -0.18 |
| Germany | 2.74 | 2.61 | 1.50 | 1.57 | 1.16 | 0.50 | 0.09 | 0.32 | 0.40 | -0.25 |
| Estonia | : | : | : | : | : | : | : | : | : | : |
| Ireland | 5.74 | 9.60 | 6.17 | 3.79 | 2.37 | 1.18 | 0.74 | 0.80 | 0.95 | 0.33 |
| Greece | 9.09 | 15.75 | 22.50 | 10.05 | 6.93 | 9.67 | 8.36 | 5.98 | 4.19 | 2.59 |
| Spain | 4.25 | 5.44 | 5.85 | 4.56 | 2.72 | 1.73 | 1.39 | 1.56 | 1.42 | 0.66 |
| France | 3.12 | 3.32 | 2.54 | 2.20 | 1.67 | 0.84 | 0.47 | 0.81 | 0.78 | 0.13 |
| Croatia | 6.28 | 6.54 | 6.13 | 4.68 | 4.05 | 3.55 | 3.49 | 2.77 | 2.17 | 1.29 |
| Italy | 4.04 | 5.42 | 5.49 | 4.32 | 2.89 | 1.71 | 1.49 | 2.11 | 2.61 | 1.95 |
| Cyprus | 4.60 | 5.79 | 7.00 | 6.50 | 6.00 | 4.54 | 3.77 | 2.62 | 2.18 | 1.07 |
| Latvia | 10.34 | 5.91 | 4.57 | 3.34 | 2.51 | 0.96 | 0.53 | 0.83 | 0.90 | 0.34 |
| Lithuania | 5.57 | 5.16 | 4.83 | 3.83 | 2.79 | 1.38 | 0.90 | 0.31 | 0.31 | 0.31 |
| Luxembourg | 3.17 | 2.92 | 1.82 | 1.85 | 1.34 | 0.37 | 0.25 | 0.54 | 0.56 | -0.12 |
| Hungary | 7.28 | 7.63 | 7.89 | 5.92 | 4.81 | 3.43 | 3.14 | 2.96 | 3.06 | 2.47 |
| Malta | 4.19 | 4.49 | 4.13 | 3.36 | 2.61 | 1.49 | 0.89 | 1.28 | 1.39 | 0.67 |
| Netherlands | 2.99 | 2.99 | 1.93 | 1.96 | 1.45 | 0.69 | 0.29 | 0.52 | 0.58 | -0.07 |
| Austria | 3.23 | 3.32 | 2.37 | 2.01 | 1.49 | 0.75 | 0.38 | 0.58 | 0.69 | 0.06 |
| Poland | 5.78 | 5.96 | 5.00 | 4.03 | 3.52 | 2.70 | 3.04 | 3.42 | 3.20 | 2.35 |
| Portugal | 5.40 | 10.24 | 10.55 | 6.29 | 3.75 | 2.42 | 3.17 | 3.05 | 1.84 | 0.76 |
| Romania | 7.34 | 7.29 | 6.68 | 5.41 | 4.49 | 3.47 | 3.32 | 3.96 | 4.69 | 4.54 |
| Slovenia | 3.83 | 4.97 | 5.81 | 5.81 | 3.27 | 1.71 | 1.15 | 0.96 | 0.93 | 0.28 |
| Slovakia | 3.87 | 4.45 | 4.55 | 3.19 | 2.07 | 0.89 | 0.54 | 0.92 | 0.89 | 0.25 |
| Finland | 3.01 | 3.01 | 1.89 | 1.86 | 1.45 | 0.72 | 0.37 | 0.55 | 0.66 | 0.07 |
| Sweden | 2.89 | 2.61 | 1.59 | 2.12 | 1.72 | 0.72 | 0.54 | 0.65 | 0.65 | 0.04 |
| United Kingdom | 3.36 | 2.87 | 1.74 | 2.03 | 2.14 | 1.79 | 1.22 | 1.18 | 1.41 | 0.88 |
| North Macedonia | 92.57 | 95.52 | 97.26 | 99.91 | 99.87 | 100 | 100.24 | 102.35 | 104.66 | 105.42 |
| Serbia | 74.9 | 83.3 | 89.4 | 96.3 | 98.5 | 100 | 101.3 | 104.7 | 106.8 | 108.8 |

Appendix 5, GDP per capita (EUR) per analysed country and year

| GEO/TIME | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|---|---|---|---|---|
| Albania | 2 960 | 3 090 | 3 190 | 3 300 | 3 320 | 3 450 | 3 560 | 3 730 | 4 020 | : |
| Austria | 34 530 | 35 390 | 36 970 | 37 820 | 38 210 | 38 990 | 39 890 | 40 880 | 42 100 | 43 640 |
| Belgium | 32 090 | 33 330 | 34 200 | 34 890 | 35 320 | 36 050 | 37 080 | 38 100 | 39 330 | 40 320 |
| Bulgaria | 4 930 | 5 050 | 5 610 | 5 750 | 5 770 | 5 940 | 6 360 | 6 820 | 7 390 | 7 980 |
| Croatia | 10 460 | 10 480 | 10 440 | 10 280 | 10 250 | 10 220 | 10 560 | 11 120 | 11 820 | 12 560 |
| Cyprus | 23 110 | 23 400 | 23 270 | 22 500 | 20 880 | 20 420 | 21 030 | 22 160 | 23 320 | 24 290 |
| Czech Republic | 14 170 | 14 900 | 15 630 | 15 360 | 15 010 | 14 880 | 15 980 | 16 690 | 18 100 | 19 550 |
| Denmark | 41 880 | 43 840 | 44 500 | 45 530 | 46 100 | 47 090 | 48 050 | 49 420 | 50 700 | 52 010 |
| Estonia | 10 640 | 11 150 | 12 660 | 13 620 | 14 420 | 15 340 | 15 820 | 16 490 | 18 070 | 19 740 |
| Finland | 34 040 | 35 080 | 36 750 | 37 130 | 37 570 | 37 880 | 38 590 | 39 580 | 40 990 | 42 500 |
| France | 29 930 | 30 690 | 31 510 | 31 820 | 32 080 | 32 420 | 33 020 | 33 430 | 34 220 | 34 980 |
| Germany | 30 390 | 31 940 | 33 550 | 34 130 | 34 860 | 36 150 | 37 090 | 38 060 | 39 260 | 40 340 |
| Greece | 21 390 | 20 320 | 18 640 | 17 310 | 16 480 | 16 400 | 16 380 | 16 380 | 16 760 | 17 220 |
| Hungary | 9 420 | 9 900 | 10 180 | 10 050 | 10 310 | 10 730 | 11 400 | 11 740 | 12 830 | 13 690 |
| Iceland | 29 530 | 32 490 | 34 130 | 35 730 | 37 260 | 40 900 | 47 400 | 55 590 | 63 130 | 62 340 |
| Ireland | 37 470 | 36 790 | 37 310 | 38 090 | 38 890 | 41 870 | 55 970 | 57 210 | 61 870 | 66 670 |
| Italy | 26 470 | 26 930 | 27 450 | 26 920 | 26 590 | 26 770 | 27 260 | 27 970 | 28 690 | 29 220 |
| Kosovo | : | : | : | : | : | : | : | : | : | : |
| Latvia | 8 780 | 8 500 | 9 820 | 10 870 | 11 350 | 11 860 | 12 350 | 12 800 | 13 810 | 15 130 |
| Liechtenstein | : | : | : | : | 128 400 | 133 200 | 148 400 | 146 700 | 148 200 | : |
| Lithuania | 8 520 | 9 030 | 10 310 | 11 160 | 11 830 | 12 460 | 12 850 | 13 560 | 14 940 | 16 160 |
| Luxembourg | 74 220 | 79 160 | 83 100 | 83 000 | 85 270 | 89 240 | 91 440 | 93 930 | 95 170 | 98 640 |
| Malta | 14 880 | 15 920 | 16 410 | 17 060 | 17 940 | 19 560 | 21 690 | 22 710 | 24 130 | 25 560 |
| Montenegro | 4 840 | 5 050 | 5 270 | 5 130 | 5 410 | 5 560 | 5 870 | 6 350 | 6 910 | 7 490 |
| Netherlands | 37 800 | 38 470 | 38 960 | 38 970 | 39 300 | 39 820 | 40 730 | 41 590 | 43 090 | 44 920 |
| North Macedonia | 3 300 | 3 460 | 3 660 | 3 680 | 3 950 | 4 140 | 4 380 | 4 660 | : | : |
| Norway | 57 620 | 66 220 | 72 350 | 79 000 | 77 440 | 73 180 | 66 970 | 63 690 | 66 950 | 69 230 |
| Poland | 8 240 | 9 390 | 9 870 | 10 100 | 10 250 | 10 680 | 11 190 | 11 100 | 12 160 | 12 920 |
| Portugal | 16 600 | 16 990 | 16 680 | 16 010 | 16 300 | 16 640 | 17 350 | 18 060 | 19 020 | 19 830 |
| Romania | 6 150 | 6 190 | 6 550 | 6 640 | 7 190 | 7 550 | 8 090 | 8 650 | 9 580 | 10 420 |
| Serbia | 4 440 | 4 330 | 4 900 | 4 680 | 5 080 | 4 970 | 5 030 | 5 200 | 5 580 | 6 140 |
| Slovakia | 11 830 | 12 540 | 13 190 | 13 590 | 13 740 | 14 070 | 14 710 | 14 920 | 15 540 | 16 470 |
| Slovenia | 17 760 | 17 750 | 18 050 | 17 630 | 17 700 | 18 250 | 18 830 | 19 550 | 20 810 | 22 080 |
| Spain | 23 060 | 23 040 | 22 760 | 22 050 | 21 900 | 22 220 | 23 220 | 23 980 | 24 970 | 25 730 |
| Sweden | 33 730 | 39 920 | 43 590 | 45 050 | 45 850 | 45 130 | 46 350 | 47 000 | 47 690 | 46 310 |
| Switzerland | 50 190 | 56 150 | 63 700 | 64 990 | 64 080 | 65 320 | 73 970 | 72 460 | 71 260 | 70 120 |
| Turkey | 6 400 | 7 900 | 8 000 | 9 000 | 9 400 | 9 100 | 9 900 | 9 800 | 9 400 | 8 000 |
| United Kingdom | 27 900 | 29 750 | 30 220 | 33 150 | 32 730 | 35 760 | 40 560 | 37 090 | 35 780 | 36 410 |

Appendix 6, Employment rate per analysed country and year

| GEO/TIME | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|---|---|---|---|---|
| Belgium | 66.9 | 67.7 | 66.7 | 66.9 | 67.5 | 67.7 | 67.6 | 67.6 | 68.0 | 68.6 |
| Bulgaria | 67.2 | 66.7 | 65.9 | 67.1 | 68.4 | 69.0 | 69.3 | 68.7 | 71.3 | 71.5 |
| Czech Republic | 70.1 | 70.2 | 70.5 | 71.6 | 72.9 | 73.5 | 74.0 | 75.0 | 75.9 | 76.6 |
| Denmark | 78.7 | 78.0 | 77.8 | 77.2 | 76.6 | 76.6 | 76.9 | 77.5 | 77.9 | 78.2 |
| Germany | 76.3 | 76.7 | 77.3 | 77.2 | 77.6 | 77.7 | 77.6 | 77.9 | 78.2 | 78.6 |
| Estonia | 74.0 | 73.9 | 74.7 | 74.8 | 75.1 | 75.2 | 76.7 | 77.5 | 78.8 | 79.1 |
| Ireland | 73.0 | 71.6 | 71.2 | 71.1 | 71.8 | 71.8 | 72.1 | 72.7 | 72.7 | 72.9 |
| Greece | 67.4 | 67.8 | 67.3 | 67.5 | 67.5 | 67.4 | 67.8 | 68.2 | 68.3 | 68.2 |
| Spain | 73.1 | 73.5 | 73.9 | 74.3 | 74.3 | 74.2 | 74.3 | 74.2 | 73.9 | 73.7 |
| France | 69.9 | 70.0 | 69.9 | 70.4 | 70.9 | 71.0 | 71.3 | 71.4 | 71.5 | 71.9 |
| France | 70.3 | 70.3 | 70.1 | 70.7 | 71.1 | 71.2 | 71.5 | 71.7 | 71.8 | 72.2 |
| Croatia | 65.6 | 65.1 | 64.1 | 63.9 | 63.7 | 66.1 | 66.9 | 65.6 | 66.4 | 66.3 |
| Italy | 62.3 | 62.0 | 62.1 | 63.5 | 63.4 | 63.9 | 64.0 | 64.9 | 65.4 | 65.6 |
| Cyprus | 73.0 | 73.6 | 73.5 | 73.5 | 73.6 | 74.3 | 73.9 | 73.4 | 73.9 | 75.0 |
| Latvia | 73.5 | 73.0 | 72.8 | 74.4 | 74.0 | 74.6 | 75.7 | 76.3 | 77.0 | 77.7 |
| Lithuania | 69.6 | 70.2 | 71.4 | 71.8 | 72.4 | 73.7 | 74.1 | 75.5 | 75.9 | 77.3 |
| Luxembourg | 68.7 | 68.2 | 67.9 | 69.4 | 69.9 | 70.8 | 70.9 | 70.0 | 70.2 | 71.1 |
| Hungary | 61.2 | 61.9 | 62.4 | 63.7 | 64.7 | 67.0 | 68.6 | 70.1 | 71.2 | 71.9 |
| Malta | 59.4 | 60.4 | 61.8 | 63.9 | 66.3 | 67.8 | 68.8 | 70.6 | 72.2 | 74.7 |
| Netherlands | 78.1 | 77.9 | 78.1 | 79.0 | 79.4 | 79.0 | 79.6 | 79.7 | 79.7 | 80.3 |
| Austria | 74.3 | 74.4 | 74.6 | 75.1 | 75.5 | 75.4 | 75.5 | 76.2 | 76.4 | 76.8 |
| Poland | 64.7 | 65.3 | 65.7 | 66.5 | 67.0 | 67.9 | 68.1 | 68.8 | 69.6 | 70.1 |
| Portugal | 73.4 | 73.7 | 73.6 | 73.4 | 73.0 | 73.2 | 73.4 | 73.7 | 74.7 | 75.1 |
| Romania | 63.1 | 64.9 | 64.1 | 64.8 | 64.9 | 65.7 | 66.1 | 65.6 | 67.3 | 67.8 |
| Slovenia | 71.8 | 71.5 | 70.3 | 70.4 | 70.5 | 70.9 | 71.8 | 71.6 | 74.2 | 75.0 |
| Slovakia | 68.4 | 68.7 | 68.7 | 69.4 | 69.9 | 70.3 | 70.9 | 71.9 | 72.1 | 72.4 |
| Finland | 75.0 | 74.5 | 74.9 | 75.2 | 75.2 | 75.4 | 75.8 | 75.9 | 76.7 | 77.9 |
| Sweden | 78.9 | 79.1 | 79.9 | 80.3 | 81.1 | 81.5 | 81.7 | 82.1 | 82.5 | 82.7 |
| United Kingdom | 75.7 | 75.4 | 75.5 | 76.1 | 76.4 | 76.7 | 76.9 | 77.3 | 77.6 | 77.9 |
| Iceland | 84.6 | 84.7 | 84.5 | 84.9 | 85.8 | 87.4 | 88.4 | 89.3 | 88.7 | 87.5 |
| Norway | 78.9 | 78.1 | 77.8 | 78.2 | 78.2 | 78.0 | 78.2 | 78.0 | 77.3 | 77.9 |
| Switzerland | : | 81.3 | 82.0 | 82.3 | 82.4 | 82.9 | 83.3 | 83.9 | 84.0 | 84.2 |
| Montenegro | : | : | 57.2 | 58.8 | 58.6 | 61.6 | 62.6 | 63.4 | 63.5 | 64.7 |
| North Macedonia | 64.0 | 64.2 | 64.2 | 63.9 | 64.9 | 65.3 | 64.9 | 64.5 | 65.3 | 65.4 |
| Serbia | : | : | : | : | : | 63.4 | 63.7 | 65.6 | 66.7 | 67.8 |
| Turkey | 50.8 | 51.9 | 53.2 | 53.3 | 54.4 | 55.1 | 56.0 | 56.9 | 57.9 | 58.5 |

Appendix 7, GDP growth rate per analysed country and year

| **geo\time** | **2008** | **2009** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** | **2018** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Belgium | 0.4 | -2 | 2.9 | 1.7 | 0.7 | 0.5 | 1.6 | 2 | 1.5 | 2 | 1.5 |
| Bulgaria | 6.1 | -3.4 | 0.6 | 2.4 | 0.4 | 0.3 | 1.9 | 4 | 3.8 | 3.5 | 3.1 |
| Czech Republic | 2.7 | -4.8 | 2.3 | 1.8 | -0.8 | -0.5 | 2.7 | 5.3 | 2.5 | 4.4 | 2.8 |
| Denmark | -0.5 | -4.9 | 1.9 | 1.3 | 0.2 | 0.9 | 1.6 | 2.3 | 3.2 | 2 | 2.4 |
| Germany | 1 | -5.7 | 4.2 | 3.9 | 0.4 | 0.4 | 2.2 | 1.7 | 2.2 | 2.5 | 1.5 |
| Estonia | -5.1 | -14.4 | 2.7 | 7.4 | 3.1 | 1.3 | 3 | 1.8 | 2.6 | 5.7 | 4.8 |
| Ireland | -4.5 | -5.1 | 1.8 | 0.3 | 0.2 | 1.4 | 8.6 | 25.2 | 3.7 | 8.1 | 8.2 |
| Greece | -0.3 | -4.3 | -5.5 | -9.1 | -7.3 | -3.2 | 0.7 | -0.4 | -0.2 | 1.5 | 1.9 |
| Spain | 0.9 | -3.8 | 0.2 | -0.8 | -3 | -1.4 | 1.4 | 3.8 | 3 | 2.9 | 2.4 |
| France | 0.3 | -2.9 | 1.9 | 2.2 | 0.3 | 0.6 | 1 | 1.1 | 1.1 | 2.3 | 1.7 |
| Croatia | 1.8 | -7.4 | -1.5 | -0.3 | -2.2 | -0.5 | -0.1 | 2.4 | 3.5 | 3.1 | 2.7 |
| Italy | -1 | -5.3 | 1.7 | 0.7 | -3 | -1.8 | 0 | 0.8 | 1.3 | 1.7 | 0.8 |
| Cyprus | 3.6 | -2 | 2 | 0.4 | -3.4 | -6.6 | -1.9 | 3.4 | 6.7 | 4.4 | 4.1 |
| Latvia | -3.3 | -14.2 | -4.5 | 6.3 | 4.1 | 2.3 | 1.9 | 3.3 | 1.8 | 3.8 | 4.6 |
| Lithuania | 2.6 | -14.8 | 1.5 | 6 | 3.8 | 3.6 | 3.5 | 2 | 2.6 | 4.2 | 3.6 |
| Luxembourg | -1.3 | -4.4 | 4.9 | 2.5 | -0.4 | 3.7 | 4.3 | 4.3 | 4.6 | 1.8 | 3.1 |
| Hungary | 1.1 | -6.7 | 0.7 | 1.8 | -1.5 | 2 | 4.2 | 3.8 | 2.2 | 4.3 | 5.1 |
| Malta | 3.3 | -2.5 | 3.5 | 1.3 | 2.8 | 4.8 | 8.8 | 10.9 | 5.8 | 6.7 | 7 |
| Netherlands | 2.2 | -3.7 | 1.3 | 1.6 | -1 | -0.1 | 1.4 | 2 | 2.2 | 2.9 | 2.6 |
| Austria | 1.5 | -3.8 | 1.8 | 2.9 | 0.7 | 0 | 0.7 | 1 | 2.1 | 2.5 | 2.4 |
| Poland | 4.2 | 2.8 | 3.6 | 5 | 1.6 | 1.4 | 3.3 | 3.8 | 3.1 | 4.9 | 5.1 |
| Portugal | 0.3 | -3.1 | 1.7 | -1.7 | -4.1 | -0.9 | 0.8 | 1.8 | 2 | 3.5 | 2.4 |
| Romania | 9.3 | -5.5 | -3.9 | 2 | 2.1 | 3.5 | 3.4 | 3.9 | 4.8 | 7.1 | 4.4 |
| Slovenia | 3.5 | -7.5 | 1.3 | 0.9 | -2.6 | -1 | 2.8 | 2.2 | 3.1 | 4.8 | 4.1 |
| Slovakia | 5.6 | -5.5 | 5.7 | 2.9 | 1.9 | 0.7 | 2.8 | 4.8 | 2.1 | 3 | 4 |
| Finland | 0.8 | -8.1 | 3.2 | 2.5 | -1.4 | -0.9 | -0.4 | 0.5 | 2.7 | 3.1 | 1.7 |
| Sweden | -0.2 | -4.2 | 6.2 | 3.1 | -0.6 | 1.1 | 2.7 | 4.4 | 2.4 | 2.4 | 2.2 |
| United Kingdom | -0.3 | -4.2 | 1.9 | 1.5 | 1.5 | 2.1 | 2.6 | 2.4 | 1.9 | 1.9 | 1.3 |
| Iceland | 2 | -6.8 | -3.4 | 1.9 | 1.3 | 4.1 | 2.1 | 4.7 | 6.6 | 4.4 | 4.8 |
| Norway | 0.5 | -1.7 | 0.7 | 1 | 2.7 | 1 | 2 | 2 | 1.1 | 2.3 | 1.3 |
| Switzerland | 2.2 | -2.2 | 3 | 1.7 | 1 | 1.9 | 2.4 | 1.3 | 1.7 | 1.8 | 2.8 |
| Montenegro | 7.2 | -5.8 | 2.7 | 3.2 | -2.7 | 3.5 | 1.8 | 3.4 | 2.9 | 4.7 | 5.1 |
| North Macedonia | 5.5 | -0.4 | 3.4 | 2.3 | -0.5 | 2.9 | 3.6 | 3.9 | 2.8 | 1.1 | 2.7 |
| Albania | 7.5 | 3.4 | 3.7 | 2.5 | 1.4 | 1 | 1.8 | 2.2 | 3.3 | 3.8 | 4.1 |
| Serbia | 5.7 | -2.7 | 0.7 | 2 | -0.7 | 2.9 | -1.6 | 1.8 | 3.3 | 2 | 4.4 |
| Turkey | 0.8 | -4.7 | 8.5 | 11.1 | 4.8 | 8.5 | 5.2 | 6.1 | 3.2 | 7.5 | 2.8 |
| Bosnia and Herzegovina | 5.4 | -3 | 0.9 | 1 | -0.8 | 2.4 | 1.1 | 3.1 | 3.1 | 3.2 | 3.1 |
| Kosovo | : | 3.6 | 3.3 | 4.4 | 2.8 | 3.4 | 1.2 | 4.1 | 4.1 | 4.2 | 3.8 |

Appendix 8, Descriptive statistics of the learned sample (part 1)

| Abbr. | Default | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|---|
| C/TA | 0 | 95321 | 0.000029 | 0.976412 | 0.156326 | 0.177634 |
| | 1 | 33021 | 0.000029 | 0.976412 | 0.162557 | 0.255631 |
| CA/CL | 0 | 93196 | 0.005981 | 34.686543 | 2.462462 | 4.024672 |
| | 1 | 36498 | 0.005981 | 34.686543 | 1.899842 | 4.515398 |
| CA/S | 0 | 87514 | 0.038541 | 55.955425 | 0.739945 | 3.147657 |
| | 1 | 22288 | 0.038541 | 55.955425 | 4.424359 | 12.154233 |
| CashR | 0 | 90916 | 0.000034 | 10.867352 | 0.638563 | 1.421845 |
| | 1 | 32245 | 0.000034 | 10.867352 | 0.457830 | 1.475395 |
| CE/TL | 0 | 96035 | -0.916760 | 15.140040 | 1.470100 | 2.159381 |
| | 1 | 35791 | -0.916760 | 15.140040 | 0.707634 | 2.311561 |
| CL/E | 0 | 97582 | -26.859321 | 113.280196 | 4.222706 | 14.351762 |
| | 1 | 37525 | -26.859321 | 113.280196 | 3.294795 | 16.680780 |
| CL/TA | 0 | 97716 | 0.000000 | 12.357861 | 0.495006 | 0.441318 |
| | 1 | 37331 | 0.000000 | 12.357861 | 1.527586 | 2.545304 |
| DCP | 0 | 85987 | 0.000000 | 4158.307986 | 72.736610 | 168.853393 |
| | 1 | 18733 | 0.000000 | 4158.307986 | 392.949116 | 995.339491 |
| EBIT/CE | 0 | 85653 | -3.708332 | 4.460039 | 0.216272 | 0.605387 |
| | 1 | 33630 | -3.708332 | 4.460039 | 0.163282 | 1.174288 |
| EBIT/S | 0 | 79442 | -7.396304 | 0.692522 | 0.037610 | 0.330316 |
| | 1 | 22272 | -7.396304 | 0.692522 | -0.545192 | 1.684824 |
| EBIT/TA | 0 | 87309 | -2.846154 | 0.763695 | 0.079358 | 0.187749 |
| | 1 | 34749 | -2.846154 | 0.763695 | -0.189053 | 0.665123 |
| EBITDA/IE | 0 | 67053 | -447.940557 | 24330.858571 | 577.685919 | 2932.944371 |
| | 1 | 11287 | -447.940557 | 24330.858571 | 145.317091 | 1665.678692 |
| EBITDA/TA | 0 | 80153 | -1.306037 | 0.719408 | 0.116025 | 0.153397 |
| | 1 | 22564 | -1.306037 | 0.719408 | -0.048343 | 0.398562 |
| FE/S | 0 | 75554 | 0.000000 | 0.615849 | 0.011071 | 0.042423 |
| | 1 | 18988 | 0.000000 | 0.615849 | 0.049968 | 0.131377 |
| FE/TA | 0 | 79903 | 0.000000 | 0.239956 | 0.011122 | 0.021160 |
| | 1 | 29896 | 0.000000 | 0.239956 | 0.026117 | 0.048553 |
| IA/TA | 0 | 94707 | 0.000000 | 0.559565 | 0.025767 | 0.076679 |
| | 1 | 32312 | 0.000000 | 0.559565 | 0.033680 | 0.107505 |
| ln(age) | 0 | 99204 | 5.899897 | 10.670582 | 8.885609 | 1.030797 |
| | 1 | 39989 | 5.899897 | 10.670582 | 8.381462 | 1.524067 |
| log (CA/CL) | 0 | 93150 | -1.150817 | 1.732394 | -0.000009 | 0.009776 |
| | 1 | 35893 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| NI/E | 0 | 86951 | -5.169160 | 4.758995 | 0.198329 | 0.731512 |
| | 1 | 34955 | -5.169160 | 4.758995 | 0.144898 | 1.367618 |
| NI/S | 0 | 79279 | -8.366977 | 0.614884 | 0.019478 | 0.367177 |
| | 1 | 22251 | -8.366977 | 0.614884 | -0.628596 | 1.876628 |
| NI/TA | 0 | 87042 | -3.119521 | 0.654908 | 0.055450 | 0.179283 |
| | 1 | 34768 | -3.119521 | 0.654908 | -0.230030 | 0.710473 |

Appendix 9. Descriptive statistics of the learned sample (part 2)

| **Abbr.** | **Default** | **N** | **Minimum** | **Maximum** | **Mean** | **Std. Deviation** |
|---|---|---|---|---|---|---|
| OENEG | 0 | 97751 | 0,000000 | 1,000000 | 0,048020 | 0,213810 |
| | 1 | 37659 | 0,000000 | 1,000000 | 0,426644 | 0,494596 |
| QA/TA | 0 | 96700 | 0,008167 | 1,000000 | 0,568038 | 0,270587 |
| | 1 | 33603 | 0,008167 | 1,000000 | 0,554613 | 0,342310 |
| QR | 0 | 92208 | 0,003729 | 23,916683 | 1,862148 | 2,916805 |
| | 1 | 32923 | 0,003729 | 23,916683 | 1,305297 | 3,035906 |
| RE/TA | 0 | 97743 | -15,224966 | 0,969177 | 0,341431 | 0,515097 |
| | 1 | 37395 | -15,224966 | 0,969177 | -1,005320 | 3,312064 |
| S/TA | 0 | 87982 | 0,000000 | 13,913714 | 2,193113 | 1,947233 |
| | 1 | 26746 | 0,000000 | 13,913714 | 1,829883 | 2,706335 |
| S/TTA | 0 | 84158 | 0,000000 | 3946,744640 | 150,049857 | 525,342984 |
| | 1 | 19895 | 0,000000 | 3946,744640 | 75,980000 | 326,834366 |
| SHP | 0 | 86565 | 0,000000 | 2744,737021 | 56,667798 | 186,001485 |
| | 1 | 19858 | 0,000000 | 2744,737021 | 209,684685 | 582,879330 |
| size | 0 | 97794 | 0,278298 | 4,232678 | 3,524997 | 0,384141 |
| | 1 | 37362 | 0,278298 | 4,232678 | 2,143669 | 0,861419 |
| ST/TA | 0 | 96700 | 0,000000 | 0,897991 | 0,170485 | 0,198254 |
| | 1 | 33603 | 0,000000 | 0,897991 | 0,161774 | 0,250980 |
| St/WC | 0 | 96643 | -38276,98 | 8208267,96 | 109,8553 | 27102,5481 |
| | 1 | 33596 | -3460,129 | 14793,666 | 1,269807 | 108,190812 |
| T/TA | 0 | 79757 | -0,052749 | 0,156626 | 0,019748 | 0,028795 |
| | 1 | 26579 | -0,052749 | 0,156626 | 0,014070 | 0,033365 |
| TC/TA | 0 | 90537 | 0,000000 | 1,429992 | 0,200801 | 0,207758 |
| | 1 | 29515 | 0,000000 | 1,429992 | 0,240406 | 0,353175 |
| TC/TD | 0 | 78571 | 0,000000 | 118,165000 | 3,137605 | 12,677581 |
| | 1 | 18404 | 0,000000 | 118,165000 | 6,441405 | 19,522214 |
| TC/TL | 0 | 88914 | 0,000000 | 0,970178 | 0,318678 | 0,268004 |
| | 1 | 28340 | 0,000000 | 0,970178 | 0,205813 | 0,256202 |
| TCPP | 0 | 81205 | 0,000000 | 2543,341912 | 55,690369 | 136,926905 |
| | 1 | 18353 | 0,000000 | 2543,341912 | 209,894107 | 577,142466 |
| TD/TA | 0 | 96134 | 0,000000 | 0,970900 | 0,257778 | 0,234905 |
| | 1 | 32610 | 0,000000 | 0,970900 | 0,280164 | 0,306246 |
| TL/NW | 0 | 96102 | -17,260255 | 51,500642 | 2,146538 | 6,320823 |
| | 1 | 36232 | -17,260255 | 51,500642 | 1,701311 | 8,683340 |
| TL/QA | 0 | 95183 | 0,076491 | 135,655743 | 2,073669 | 5,965401 |
| | 1 | 32165 | 0,076491 | 135,655743 | 11,943901 | 28,655687 |
| TL/TA | 0 | 96141 | 0,035111 | 16,224842 | 0,661513 | 0,529951 |
| | 1 | 36085 | 0,035111 | 16,224842 | 2,021961 | 3,323952 |
| TL/TTA | 0 | 91053 | 0,205753 | 1852,764772 | 54,116469 | 216,646943 |
| | 1 | 26052 | 0,205753 | 1852,764772 | 76,547800 | 264,646168 |
| WC/S | 0 | 85855 | -84,380939 | 20,811625 | 0,041246 | 3,652998 |
| | 1 | 20287 | -84,380939 | 20,811625 | -5,701434 | 20,983521 |
| WC/TA | 0 | 97705 | -11,530992 | 0,984118 | 0,244101 | 0,456751 |
| | 1 | 37322 | -11,530992 | 0,984118 | -0,807172 | 2,538665 |