Ádám Banai–Gyöngyi Körmendi–Péter Lang–Nikolett Vágó

Modelling the credit risk of the Hungarian SME sector

MNB Occasional Papers 123

2016

Ádám Banai–Gyöngyi Körmendi–Péter Lang–Nikolett Vágó

Modelling the credit risk of the Hungarian SME sector

MNB Occasional Papers 123

2016

The views expressed are those of the authors and do not necessarily reflect the official view of the central bank of Hungary (Magyar Nemzeti Bank).

MNB Occasional Papers 123

Modelling the credit risk of the Hungarian SME sector

(A magyar kis- és középvállalati szektor hitelkockázatának modellezése)

Written by Ádám Banai, Gyöngyi Körmendi, Péter Lang, Nikolett Vágó

Budapest, December 2016

Published by the Magyar Nemzeti Bank

Publisher in charge: Eszter Hergár

H-1054 Budapest, Szabadság tér 9.

www.mnb.hu

ISSN 1585-5678 (online)

Contents

Abstract

In banking practice, quantifying the probability of default is one of the most important elements of the lending decision, therefore it is also vital from a financial stability perspective. The aim of our research was to model the probability of default as precisely as possible in the case of micro, small and medium-sized enterprises. By linking the data from the Central Credit Information System (KHR) and companies' financial statements, a database was created that covers all the SMEs with loan contract, thus we were able to examine credit risk based on a uniquely large group of enterprises. In our research, we created models that enabled us to produce estimates based on certain corporate features about the probability of default of micro, small and medium-sized enterprises. Our analysis revealed that modelling these size categories separately and managing non-linear effects in the case of several variables are especially important. In addition, the impact of the macroeconomic environment on credit risk also proved to be important in the fitting of our estimates.

Keywords: SME, credit risk, credit register, logit model, probability of default
JEL codes: C25, G20, G21

We are indebted to Zsolt Oláh, Mihály Szoboszlai, Gábor Csiszár and Marianna Endrész, and to all our colleagues in the Magyar Nemzeti Bank the central bank of Hungary who contributed to the preparation of the study with their helpful comments. Any remaining mistakes are the sole responsibility of the authors.

1 Introduction

Modelling the probability of default (PD) has primary importance both in banking practice and in central bank roles. In banking practice, quantifying the probability of default is one of the most important elements of the scoring systems supporting lending decisions. Scoring is used by banks to decide whether to issue a loan to a client or not, and the price at which the funds can be acquired. Therefore, ultimately, scoring systems profoundly influence banks' lending activity, and they may exert a substantial impact on profitability as well. A well-functioning scoring system may considerably increase the profits and competitiveness of the individual institutions (Oravecz, 2007). The models enabling more precise pricing are also beneficial for the real economy, since they foster the appropriate allocation of funds. From a financial stability perspective, modelling PDs is also key, since it enables loans to be priced appropriately, which reduces the negative effect of the shocks to the financial system, and may boost lending activity. For the above reasons, developing scoring systems has become a priority in banking. Banks employ several different methodologies. These vary widely from the simpler linear regression models through those employing binary dependent variables (logit, probit) to the use of neural networks. The aim of our present study is to provide credit risk information that may support banks' lending activity.

In the literature, there are numerous studies that attempted to estimate credit default risk. One of the seminal works is the study by Altman (1968), in which the author modelled bankruptcy risk with the help of indicators capturing companies' financial position. Despite performing decidedly well, the model was based on a calculation involving a very small group of companies, which reduced its practical utility. Especially because the sample only contained large, listed enterprises. But from a banking perspective, the really important challenge is to model the SME sector, as in the case of large firms, scoring models play a much more limited role in lending decisions. Characteristics of the SME sector were detailed in Altman and Sabato's (2007) study. They demonstrated that due to the special features of the SME sector, the models applied to large enterprises cannot be used in the case of SMEs, and that separate models should be prepared for this segment. This finding tallies with the practice of the past 20 years, since the research into the probability of corporate default has been based on ever larger databases. Lykke et al. (2004) prepared a bankruptcy estimation on a corporate sample containing 300,000 annual accounts, and they showed that the differences in size, sector, liquidity position, age and equity all have a significant impact on the probability of default. Sjovoll (1999), who made a bankruptcy estimation based on a Norwegian corporate sample containing 500,000 observations, presented similar results. In his model, he also confirmed that qualitative information, such as transparency, can also significantly influence the probability of a bankruptcy for a company. From the Hungarian literature, it is important to mention the study by Bauer and Endrész (2016), which examines an issue similar to the focus of our present study. The greatest difference between the two analyses is that Bauer and Endrész (2016) did not use the data from the Central Credit Information System, therefore the dependent variable of their model is based on the legal definition of bankruptcy. An important lesson from the models created by them is that in the different corporate size categories, the relation between corporate characteristics and bankruptcy risk may vary, and that the fitting of the model is materially improved by the inclusion of macroeconomic variables. They also showed that in the case of certain variables, the use of non-linear functional forms improves the fit of the model.

The aim of our present study is to establish a model that enables us to estimate a company's probability of default. Based on the above, this information is relevant both from a central banking and a banking perspective. This may be a materially important, additional information especially for smaller banks with less developed risk management practices. For several reasons, we focus on micro, small and medium-sized enterprises during modelling. First, the literature attests that they behave differently from large enterprises. Second, we know that in banking practice, the role of scoring models is much greater in lending decisions in the SME sector than in the case of large corporations. And finally, information on credit risk can support SME lending which is a priority of the central bank.

Our research is based on the fact that the adoption of the proposal initiated by the MNB to amend the relevant legislation allowed the MNB to link the Central Credit Information System (KHR) containing information on corporate loans' performance and the corporate database of the Hungarian National Tax and Customs Administration (NTCA).[1] The linking of the data paved the way for the simultaneous examination of corporate credit data and economic activity. This gave rise to a database that is, in certain respects, more limited than several banks' information base about their clients, but it includes all the companies with bank loans. In view of the fact that the new database can be used exclusively by the MNB, the models that have just been built and that provide an overview about the credit risk of the whole SME sector may serve as useful information to commercial banks, too.

During our research, we created models that enabled us to produce estimates on the probability of default of micro, small and medium-sized enterprises based on certain corporate features. In the course of modelling, we paid attention to prepare the models in a form best suited to support the credit evaluation process. Our estimated models fit especially well in the case of small and medium-sized enterprises, and provide a relatively accurate picture about the probability of default. In line with our expectations, explanatory variables that influence corporate performance the most, such as equity, profitability, the available fixed assets, the ratio of FX debt or the volume of sales revenue, were significant. Certain explanatory variables have different effects on credit risk in the case of the different SME categories, which confirms that different models should be estimated for the various company sizes. Nevertheless, in addition to the results that are in line with the literature and intuition, it should also be noted that based on our study, managing non-linearities is especially important in the case of some variables. This tallies with the findings of Bauer and Endrész (2016).

The main difference between our present research and Bauer and Endrész (2016) is in the definition of the dependent variable, since they used a legal bankruptcy definition, while we used a default definition which is commonly used in banking practice. In the first case, the initiation of liquidation proceedings means the end of a company's operation, whereas in the case of studying the default risk of loans, a loan in arrears is not necessarily a final state, as it can also indicate a temporary problem. Thus the default may be the result of less major reasons, which undermines modelability. With respect to the population under review, the bankruptcy risk analysis provides a broader sample, since it also includes companies without outstanding loans, so the conclusions that can be drawn from that may be applied to the whole corporate sector. In contrast, our credit risk analysis characterises the population of companies with loans at credit institutions, which is only a subset of the whole population, but for credit institutions, the behaviour of this corporate population is relevant.

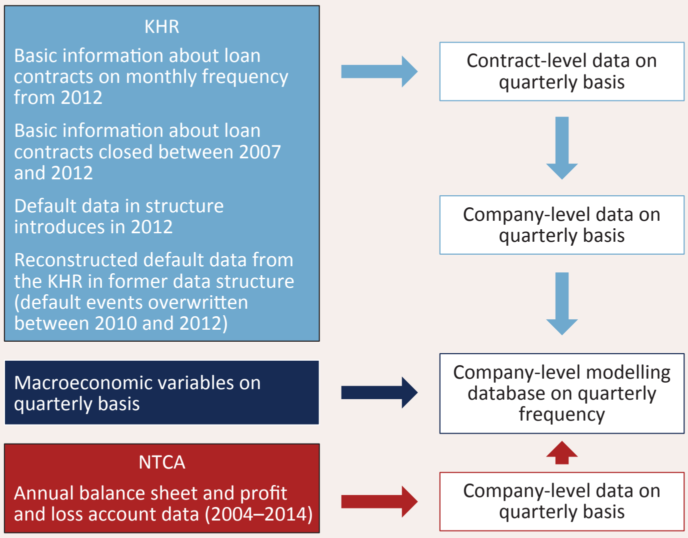
In Chapter 2 of the study, we will present the characteristics of the database we used. In Chapter 3, we will discuss the methodology we employed. After that, we will give a detailed overview about the features and interpretation of the models we arrived at as well as their potential use: Chapter 4 will show our baseline model, Chapter 5 will deal with our models for the different company sizes, specified for scoring, and Chapter 6 will present our models estimated for the different sectors in the national economy. Lastly, in Chapter 7, we will summarise our results.

[1] The amendment to Act CXXII of 2011 on the Central Credit Information System, effective from 1 September 2015.

2 The structure of the database

The database used in our credit risk estimate was based on the corporate contracts and the corresponding default events stored in the corporate subsystem of the Central Credit Information System (KHR). This was complemented by companies' financial indicators, which were calculated from the financial statements submitted to the Hungarian National Tax and Customs Administration (NTCA) in their annual tax returns. We also added macroeconomic variables to the thus created company-level database, because there was much information at the individual level that was not available to us, and that materially influences the credit risk of companies. However, the development of a part of that information in time can be approximated with macroeconomic variables (Chart 1).

Chart 1
The structure of the database



The frequency of the three main data sources was different: the credit register provided monthly data, the balance sheet and profit and loss account figures from the financial statements showed annual data, while a portion of the macroeconomic variables was available on a quarterly basis. As the use of annual data would have substantially shortened the already relatively short time series of the database, we decided to use quarterly frequency, and we increased the frequency of the NTCA data with interpolation as described later. The degree of aggregation in the different databases also varied: while the KHR database contained contract-level data, the NTCA database comprised company-level data. As we wished to estimate company-level credit risk, we aggregated the KHR data to the company level.[2]

[2] This means a simple sum or weighted arithmetic mean, depending on the given variable.

2.1 THE CREDIT REGISTER DATA

Banks, specialised and co-operative credit institutions and financial enterprises have to record the information about all loan and loan-type contracts[3] that are signed with companies or sole traders[4] in the corporate subsystem of the Central Credit Information System. Since 2010, the MNB has received a snapshot of this constantly changing database in each month, therefore the static images can be linked to form a time series. However, there is no possibility for retroactive correction or the submission of missing data. In addition, in 2012 Q2, the data structure was changed, which also has a profound impact on the range of usable variables.

In the case of the default data, the introduction of the new data structure brought about significant changes. While in the new system each default event is stored sequentially, in the former system a concluded default event of a given loan was overwritten by a subsequent default event. However, this problem could be addressed to some extent: the overwritten defaults between 2010 and 2012 could be partly reconstructed. Yet as we move back in time from 2010, the data about the default is ever more likely to have been overwritten by 2010, therefore we wrongly believe a contract to be non-defaulting when in fact it is in default. This problem, and the complete lack of recording defaults can be especially acute in the case of the data from before 2007, since there are very few default observations for this period. This is not warranted by banks' portfolio quality statistics, therefore we excluded the period before 2007 Q2 from our estimation. The less serious part of the problem was tackled within the model: on account of the overwritten defaults for 2007–2010, a trend was introduced for this period as an explanatory variable. Over time, the value of the trend decreases steadily from a positive range right until 2010 Q3, and from then on it assumes a value of zero for all periods.

2.2 COMPANIES' FINANCIAL INDICATORS

Companies' financial indicators were calculated from the balance sheet and profit and loss account figures submitted with their annual tax returns. The NTCA database contains the data for all companies with tax liability and double-entry bookkeeping. From this group of companies, the government- or local government-owned companies (above a 25 per cent ownership ratio), non-profit institutions serving households and financial corporations were filtered out. The non-financial corporation sample we arrived at in this manner was linked to the KHR database, thereby narrowing down the population of non-financial corporations with loans. This corporate group was whittled down further based on the main goal of our analysis: large enterprises[5] and firms that operate in a concentrated or special-activity sector (linked to utilities, in the quasi-fiscal or financial sectors, or producing for own, household consumption) were excluded.[6] As we have already mentioned, in the case of the modelling database, a quarterly frequency was chosen, and the NTCA database's originally annual figures were interpolated to generate quarterly data.[7]

There are numerous companies that had ongoing loan contracts in a given period but did not report taxable sales revenue in the given fiscal year, therefore their financial statements for the year in question are not available in the NTCA database. As we did not want to exclude such a large number of observations from the

[3] The register includes credit and loan transactions, loan-type framework agreements, financial leases, bill of exchange discounting, business factoring transactions, guarantees, surety agreements, securities lending and unsecured letters of credit. The overwhelming majority of the contracts are credit and loan transactions, and loan-type framework agreements. The analysis includes all the transactions in the previously listed product types, provided they are on-balance sheet items in the given quarter.

[4] The present study only examines companies' credit risk, therefore the data for sole traders is not used. Based on their behaviour and characteristics, sole traders are more like households than companies. Their data are also not suitable for a similar analysis: on the one hand, their balance sheet figures are not available from an administrative source, and on the other hand, pursuant to the KHR Act, the information about them in the credit register can only be accessed by the MNB in an anonymised form, therefore it cannot be linked to other databases.

[5] Companies were considered large enterprises if they had a sales revenue in excess of EUR 50 million or at least 250 employees in the majority of the observed periods between 2004 and 2014. This definition is in line with those of the other company size categories that will be detailed later.

[6] The excluded NACE categories: mining and quarrying; electricity, gas, steam and air conditioning supply; water supply, sewerage, waste management and remediation activities; financial and insurance activities; public administration and defence, compulsory social security; activities of households as employers, undifferentiated goods- and services-producing activities of households for own use.

[7] In the case of the profit and loss account, interpolation is tantamount to the assumption that the profits of a company are generated evenly during the year, and that a given point in time records the sum of the profits in the four quarters ending with the quarter concerned. In line with this, we assumed that the balance sheet structure of a company changed gradually over the year.

analysis without assessing its impact on the estimation, we filled the missing data along technical assumptions.[8] The observations created in this manner, and those where in the four quarters ending in the given quarter the company did not have sales revenue and did not report a positive headcount, are from now on referred to as non-functioning companies. The whole database, i.e. the one containing the data points categorised as non-functioning companies as well, was only used for robustness checks (see Chapter 9.2), while the credit risk models also suitable for scoring, i.e. the main goal of this research, were estimated without these. In the case of companies that generated no taxable income in the year of the loan application, future credit risk can be better assessed on the basis of the business plan, therefore in such cases a different evaluation system needs to be set up based on completely different types of data.

Companies' size categories are critical from the perspective of our analysis. It is a general experience in both the literature – e.g. Altman and Sabato (2007) – and in banking practice that differentiating companies by size is essential for the examination of default. The decision-making process, a company's bargaining power on the market, its ability to adapt and its vulnerability, and therefore not only the degree of its credit risk but also credit risk's connection to corporate characteristics varies widely across the size categories. In banking practice, sales revenue plays a central role in the definition of the size categories, therefore we also regarded it as a starting point, and it was complemented with the headcount information. As individual banks segment their portfolio using different limits, when defining our own size categories, we did not rely on these, but on the limits of the common European categories. In line with this, companies with up to 10 employees and a sales revenue of up to EUR 2 million were considered microenterprises, those with up to 50 employees and a sales revenue of up to EUR 10 million were considered small enterprises, and those with up to 250 employees and a sales revenue of up to EUR 50 million were considered medium-sized enterprises.[9]

Classification by size was performed for each period of each company individually. Based on this classification, the size category of a considerable portion of the companies changes over time. As we did not wish to break up the credit histories of the individual companies, when differentiating models by size, companies were always classified in line with the most frequent (mode) size classification over the 2004–2014 horizon, disregarding the time spent in the category referred to as non-functioning. We were forced to diverge from the banking practice in managing corporate groups, since no information on the interlocking ownership structures could be included in our databases, therefore certain companies may have been classified into smaller size categories than warranted.

2.3 OUR MACROECONOMIC VARIABLES

There are several unique features of a company that cannot be observed and that still heavily influence whether the given company defaults. The dynamics of a portion of the unobserved variables over time can be approximated with macroeconomic variables. That is why in addition to unique features, macroeconomic indicators may also have explanatory power in credit risk and bankruptcy models, for example in the estimates by Carling et al. (2007) and Bauer and Endrész (2016). In our case, the effect of interest expenses' development over time was represented in the model not only through the profitability indicators of the given company, but also partly through the inclusion of the BUBOR as the general benchmark of floating interest rate forint loans. We deemed it important to examine the GDP and the GDP deflator (as producer prices) from the macroeconomic environment, and we also tested the indicator of external demand as well as the disposable income of households. In the case of the models specified for the companies in the individual economic sectors, not only the general macroeconomic variables were examined but also sectoral indicators such as the number

[8] All profit items and headcount figures were considered zero for these periods, and the balance sheet items were fixed at their last known value. In the cases where the figures from the first few years of the company's operation were missing, balance sheet items were supplied based on the first filed tax returns.

[9] The headcount figure is heavily influenced by the issue with the database that missing and zero values cannot be distinguished. Since providing this information is not compulsory, the headcount is zero in 15 per cent of the tax returns filed by companies also in the KHR in 2004–2014. Therefore we did not want to include the headcount as a continuous variable, but in the case of the size categories this does not cause a great problem, as these observations are classified based on sales revenue.

of workers in the sector, the production volume and implicit price index of the industry and its confidence indicator (subindex of the Economic Sentiment Indicator).

2.4 OUR DEPENDENT VARIABLE

The KHR records data on loans past due by over 30 days: date of the 30-day default, the amount of the item past due by 30 days and, if the default no longer exists, the date when the default ended. However, from a modelling perspective, the defaults of over 90 days are interesting, since this way technical defaults can be by and large excluded, therefore this indicator is more relevant from the perspective of credit risk.[10] Therefore, in our study we analyse 30-day defaults ongoing for at least 60 days. In a certain period, the given contract was considered to be in default if it was in default at some time during the period. This relatively strict approach was employed because in this manner, the assessment of a contract's default does not change according to frequency.[11] During the aggregation to the company level, the contract-level defaults were weighted with the contract amount, and companies were considered to be in default if they were in default with at least 10 per cent of their contracts during the given period.[12] By determining the minimum threshold, we wished to offset the strictness of the contract-level definition to some extent. However, all in all, we excluded relatively few defaults due to this: merely 3.7 per cent of the default periods. In more than half of their default periods, companies were in default in all their contracts in force.

Our model examines the probability of a company defaulting within a year. By using this annual definition at quarterly frequency we identified a default in each quarter when in the preceding quarter the company was in a non-default status, but during the given quarter or in the subsequent three quarters it assumed to be in a default status. This means that each default that was preceded by at least four quarters of non-default status, was observed four times. Therefore the scale of the PD also becomes annual instead of quarterly. The companies that returned from default were not excluded from our sample, since the database itself, due to its time horizon and its issues with data recording, is unsuitable for identifying the first default in the history of a company. Thus a company reappeared in our estimation sample if after the ending of its default it was a non-defaulter for at least a quarter of a calendar year.

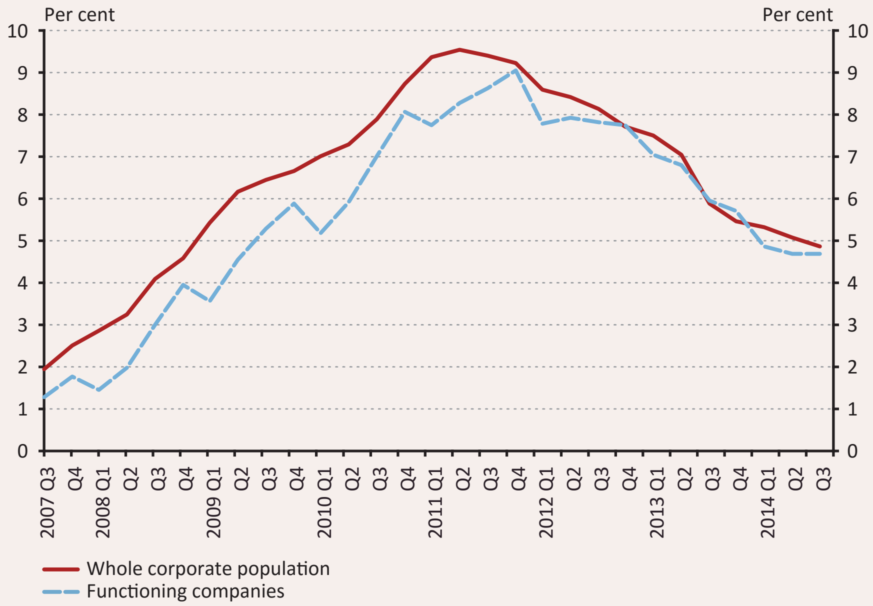
The time profile of our dependent variable can be examined for the period between 2007 Q3 and 2014 Q3 (Chart 2), both with respect to the whole population of the companies with outstanding loans, and to those companies – from now on referred to as functioning companies – that filed a tax return in the given period and, based on that, had positive sales revenue in the four quarters ending with the given point in time. In the case of both corporate populations, the previously mentioned effect of the overwriting of defaults can be clearly seen in the first years of the time horizon: the proportion of companies just defaulting is markedly lower than what we would expect based on banks' aggregate portfolio quality reports.

[10] About half of the 30-day defaults recorded in the database end in 60 days.

[11] Had we examined the default status at period end, we would have lost the defaults that started and ended within the period, and the number of the events thus excluded would have depended on whether we had used monthly or quarterly frequency. If we had stipulated a minimum multi-day period within the interval, we would have moved away from the definition of the 90-day default.

[12] This definition differed from banking practice in many aspects. First, banks take into account current exposures, but we only know the contract amount. Second, banks rate the debtor based on the contracts signed by themselves, while we considered all the company's contracts signed with credit institutions and financial enterprises as one. Compared to these basic differences, the difference in percentage is secondary.

Chart 2
The development of the proportion of defaulting companies over time



Note: The pattern discernible at the turn of the calendar year for the sample of functioning companies is caused by the companies shifting from the functioning to the non-functioning category, since the proportion of defaulters is higher among them than among those remaining in the category of functioning companies.

2.5 OUR EXPLANATORY VARIABLES

The most important explanatory variables in our model are the unique features of the companies, which are complemented by the previously mentioned macroeconomic variables and the trend introduced to manage data issues, which help model fitting. We avoided variables which, based on statistical correlation, fit well but the cause–effect relationships behind their fit are not suitable for preparing scoring.

From a technical perspective, corporate features are divided into two groups: financial indicators interpreted on a continuous scale and the features used as categorical variables. The continuous indicators include those that capture the profitability, liquidity, leverage, investments and exports of a company, the potential coverage, FX ratio and original matirity of its loans. The definitions of these are presented in Table 2 of the Appendix, and their descriptive statistics are shown in Table 4 of the Appendix. Since certain variables contain outliers, which cause distortions not only in the estimates but even in the average of the descriptive statistics, we used winsorising in their case. This primarily technical-driven modification that affects a relatively minor proportion of the observations was shown in the table containing the definition of the variable as “first winsorising”. After this, further winsorising and truncation[13] were necessary in the case of certain variables, but these were determined within the framework of examining non-linearities and were based on estimates. Obviously, the ratio of the population concerned was examined in these cases as well, and we endeavoured to minimise it as much as possible. The winsorisations and truncations performed during modelling are summarised in Table 3 of the Appendix.

With respect to our categorical variables (Appendix, Table 5), in the case of the ownership structure and equity the two defined states (negative vs non-negative equity, and majority foreign ownership vs majority domestic ownership) are qualitatively different, therefore they should be represented as binary variables instead of continuous variables. In the case of company age, identifying the differences and similarities between the

[13] Truncation means the process of replacing an originally continuous variable with a dummy variable on a section of its domain.

individual life stages is easier in the form of a categorical variable, therefore our analysis was performed in this form, and we did not return to a continuous form with an appropriate functional form. Within categorical variables, companies' classification by size is crucial. Two types of variables were defined for this purpose: first, the size of the company for the given quarter, and second, the most typical (mode) size between 2004 and 2014. In the case of the models estimated separately by size, the typical size was used, since that is constant in time, therefore it does not break up the credit history of the company. In the case of the models without size differentiation, this problem did not arise, therefore the current company size was used, which showed a more accurate picture about the situation of the company in the given period. As the distribution of the individual explanatory variables differs by company size, descriptive statistics are shown both in the case of continuous and categorical variables, by the most typical size categories.

In terms of the number of observations, the populations used for estimating the regressions are dominated by microenterprises (Table 1), but since the whole sample size is quite large, despite its modest proportion even the population of the medium-sized enterprises can be used for model building on its own.

Table 1
Number of companies and observations in the whole model and the model for functioning companies

	Number of companies		**Number of observations**	
Typical corporate size	**Functioning companies**	**Whole corporate population**	**Functioning companies**	**Whole corporate population**
Microenterprises	132,216	146,256	1,693,544	2,105,466
Small enterprises	21,404	22,705	385,162	428,750
Medium-sized enterprises	4,403	4,703	86,460	95,252
Total	**158,023**	**173,664**	**2,165,166**	**2,629,468**

:

3 Methodology

Research of corporations' credit risk is one of the primary issues regarding the functioning of credit markets, therefore in the past almost half a century, several statistical, econometric and simulation methods were used for measuring and modelling it. These methodologies are detailed by Oravecz (2007).[14] When choosing the applied methodology, a major criterion was that the model should be able to reproduce the observed default ratios with high accuracy by managing many variables simultaneously, and that the estimated probabilities of default should be adequately segmented. Another criterion was that the functioning of the model and the effect of the variables within it should be clear-cut and relatively easy to calculate. Within the framework of such a model, results can be presented in a much simpler fashion than in the case of a black box-type model. Of course, in addition to the above, a basic requirement was that the method should fit the available database well, and that it should be able to use it to the fullest possible extent. The three criteria together were met by the probit and logit models that are otherwise widely known and used, and due to its easier application, we decided to use the latter.

In the case of a logistic model, the probability of a company's default is stated as follows:

$$\hat{p} = \frac{1}{1 + \exp(-\beta' X)}$$

where $\hat{p}$ is the vector of the estimated probabilities of default, β is the vector of the regression parameters and X is the matrix of the explanatory variables. This model form ensures that the estimated probabilities fall between 0 and 1, in line with the theoretical limits. In order to facilitate the interpretation of our results, it is also important to point out that due to the model form, the magnitude of the partial effects of the individual explanatory variables depends on the values of the other explanatory variables.

In line with the recommendation of Shumway (2001), we used the observations about the companies for all periods when companies were under the risk of default. However, in our case this does not result in a survival model (hazard) structure, as a company may default several times during its history, therefore its age may not necessarily equal the period spent in non-default status.

The examination of non-linearities played a key role in our modelling strategy. On the one hand, we explored the functional forms that could represent the effects of the explanatory variables (the methodology for this is presented in Chapter 9.1 of the Appendix), and we also analysed the joint effect (cross effect) of several explanatory variables. The latter meant the examination of whether a given explanatory variable should be divided by size categories,[15] and the exploration of non-trivial cross effects with the help of decision trees. As the size of our database did not enable us to create a decision tree for the whole database, the calculation was performed for 10 per cent of the companies chosen at random. Based on the results, the variable simultaneously capturing the original maturities of the company's current loans and its capital level (whether it had positive equity) was included in the equation as an explanatory variable.

With respect to the estimated model, just like in the case of scoring models in general, the problem of selection bias arises, since we can only observe the default or non-default of the companies that received loans. However,

[14] The article presents all relevant methods with respect to measuring the risks of corporate credit extended by banks. In the literature, we may also come across the quantification of Merton models, for example Bauer and Agarwal (2014) test the performance of this model type. But this approach can be construed as the credit risk of corporate bonds or as companies' bankruptcy risk and not so much as the default on bank loans.

[15] As we have already mentioned, the population under review is quite heterogeneous by company size, and the operation of the companies in the different size categories varies extensively. Thus it is no coincidence that there are several financial indicators that exert a different degree of impact on the company's probability of default in the case of the different company sizes.

whereas a bank can take into account the features of the rejected companies when constructing its credit rating scheme,[16] we do not have such data. Nevertheless, we should note that in our case the extent of the bias is probably smaller than in the estimates by the individual banks, since we see all the banks together, thus a company is included in the sample even if at least one credit institution or financial enterprise is willing to enter into a contract with it. Therefore in our case, out of the credit institutions and financial enterprises available to the company at a given point in time, the entry threshold is determined by those with the most lenient requirements. It should also be noted that a company is included in the database even if it has a modest credit line, which is much easier to obtain than a large-value, long-term investment loan.

[16] For this, the sample selection model of Heckman (1979) can be used, for example.

4 Our estimates for functioning companies

In this chapter, we will present our main model specification explaining the annual probability of default, built on the KHR and NTCA databases linked in the manner described above. First, we will interpret the results of our model by variables. Then we will examine the goodness of fit of our model, and identify the main variables, groups of variables influencing the extent of the fit. The robustness analyses of the model presented in this chapter are detailed in Chapters 6 and 9.2.

4.1 DETAILED RESULTS

In our estimation, we used company-specific variables, categorical variables capturing a part of the remaining heterogeneity, macroeconomic variables capturing the unexplained heterogeneity in time and a trend in order to adjust for the overwritten default events. Our estimates can be found in Table 6 and 7 of the Appendix. In order to facilitate the interpretation of the coefficients, we also provided a chart for all complex, i.e. multiple-term company-specific variables (see the Appendix), in which we show the effect of the variable on the log-odds[17] ceteris paribus, i.e. with all other explanatory variables unchanged. During modelling, we strove to ensure that the signs of the coefficients are in line with economic intuition. We used a quarterly lag for all our company-specific variables. The reason behind this is that as we defined default as being in arrears for at least 90 days, the default sought to be explained actually happens a quarter of a year before its observation.

Companies' equity and indebtedness are shown by our leverage variable that is taken into account continuously for companies with positive equity. Yet our database also contains firms with negative equity, i.e. companies that are in theory insolvent but are still functioning. These are controlled for with a dummy variable, that, in line with our expectations, has a substantially significant, positive coefficient, i.e. these companies are materially riskier. On the chart about leverage (Appendix, Chart 2), it can be seen that higher leverage results in higher default risk all other features being equal.

In the case of the sales revenue's growth rate, the effect of the explanatory variable on default risk varies by company size, therefore we capture the effect of sales revenue changes for companies in different size categories with different variables. Chart 3 of the Appendix shows that a drop in sales revenue increases default risk considerably and at a diminishing pace for all three size categories.

In addition to sales revenue, the probability of default is also heavily influenced by profitability. Our results attest that in line with our expectations, a higher return on assets (ROA) reduces a company's default risk, in fact, the drop in risk happens at an increasing pace (Appendix, Chart 4).

The change in the profitability of companies is represented in the model with the change by percentage points in the after-tax ROA (Appendix, Chart 5). Positive change in the ROA reduces risk. The larger size category the company is in, the more pronounced this effect is.

The model also includes the original maturities of the company's outstanding loans weighted with the contractual loan amounts as an explanatory variable. Longer original maturity decreases risk, since ceteris

[17] The odds is the probability of default divided by the complementary probability. Due to the shape of the logistic function, the linear combination of the variables with the coefficients gives the logarithm of the odds, therefore we can draw conclusions about the functional form to be tested in the case of the individual variables through the examination of the log-odds.

paribus it means a later due date in the case of a lump-sum repayment, or lower repayment instalments, i.e. a lower repayment burden, in the case of regular payments. Based on the decision tree analysis, the variable was taken into account in interaction with the positive or negative value of equity. The variable gives a negative coefficient in the case of both positive and negative equity. For companies with negative equity, this risk-reducing effect is stronger ceteris paribus, which means that in the case of extreme indebtedness, the repayment burden becomes more important.

The liquidity position has a strong influence on the credit risk of a company, which is why our analysis includes a liquidity indicator, the proportion of a company's liquid assets less its short-term liabilities relative to its balance sheet total. The variable was divided up across size categories, because had we used it as one, it would have obscured the clear-cut heterogeneity across SME categories seen in our results of the models differentiated by company size. Our expectation that higher liquidity may reduce default risk was confirmed for the size categories of small and medium-sized enterprises (Appendix, Chart 6). In fact, the greater the size category of the company, the stronger the effect. However, in the case of microenterprises, a higher liquidity indicator significantly increases default risk.

The proportion of a company's stock of fixed assets relative to the contractual amounts of the company's outstanding loans is an explanatory variable in our model. It attempts to capture the potential coverage behind the loans. This is important because greater coverage substantially reduces moral hazard. In line with our expectations, we arrived at a negative coefficient in the case of all three size categories (Appendix, Chart 7).[18]

In addition to the company-specific variables listed so far, we regard the export activities of the companies an important explanatory variable from the perspective of their credit risk, therefore we also show the proportion of exports relative to sales revenue in our model. Usually, the variable is expected to produce a negative coefficient, i.e. a risk-reducing effect, as controlling for the foreign ownership of companies, the firms that can be deemed more competitive are in a better position to engage in export activities. Nevertheless, the mechanisms in the variable do not necessarily point in one direction, since if a company produces almost exclusively for exports, the potential instability of its external business relations may arise as an additional risk, and this effect is stronger as the bargaining power of a company weakens. In addition, foreign exchange risk (in the case of potentially inadequate hedging) may also be greater in the case of companies that obtain a larger portion of their sales revenue from abroad. According to our results (Appendix, Chart 8), the best solution for small and medium-sized enterprises may be to diversify their market demand through export activities, thereby generating a stable demand for their products. The export ratio, which reduces risk the most, is higher among medium-sized enterprises, the bargaining power of which is probably greater. Within microenterprises, we distinguished between firms with an export ratio above 15 per cent and those below that. We found that a low export ratio significantly reduces a company's default risk, but export activities of over 15 per cent increase it.

FX indebtedness was represented in our model with the ratio of the company's FX loans relative to the contractual loan amount. As the natural hedge for FX liabilities was already partly controlled for with our export ratio variable, this variable was expected to increase default risk. Our results (Appendix, Chart 9) confirm this relation. Furthermore, in the case of a mixed-denomination financing structure, which is a characteristic of relatively few companies, we estimated higher credit risk than with typically exclusively forint or FX financing.

The model controlled for majority foreign ownership with a dummy variable. A company was considered majority foreign-owned if the proportion of foreigners among its owners was greater than 50 per cent. As foreign owner typically means a corporate owner, i.e. a foreign parent company, ceteris paribus we expect this to reduce default risk, since we assume that the parent company can assist its subsidiary in the case of solvency problems through lending or capital injection. And the variable does indeed have a negative coefficient (Appendix, Table 6).

[18] Although in some cases the linear or the quadratic term of the variable did not prove to be significant, we have also shown these coefficients in Table 6 of the Appendix to facilitate the comparison with the model prepared for the whole database and to present the behaviour of the indicator in detail.

We used various categorical variables to capture the further heterogeneity among companies. In addition to dividing up company-specific variables across size categories in the estimated equation in order to more precisely identify their heterogeneous effect, we used a categorical variable to control for the differences between the SME categories that were not indirectly captured by the above-mentioned variables. Compared to microenterprises used as a reference point, the greater the size category the given company belongs to, the lower its default risk is ceteris paribus (Appendix, Table 7).[19] We also used a categorical variable to control for the place of operation of a company, or more precisely the region where the company is registered. In addition, we used a categorical variable to take into account the NACE category of the company's main activity, divided up by economic sectors.[20] We also took into consideration the age of the company, i.e. the time since its founding in full years.[21] Ceteris paribus, companies with outstanding loans have the highest default risk when they are two years old, and from then on this risk decreases monotonously until the company is 18 years old (Appendix, Table 7). The intuitively expected monotonous reduction does indeed happen in almost the whole life cycle.

We should manage temporal heterogeneity in our model not explained by our other explanatory variables, first and foremost because of the above-mentioned data problems. We can capture all unexplained temporal heterogeneity if we use a separate dummy variable for each quarter. In such a scenario, however, the model would only be suitable for credit risk evaluation based on recent data with substantial additional uncertainty, therefore we used macroeconomic variables and a trend instead of quarter dummies. In addition to the trend, the final model also included the one-quarter lag of the annual growth rate of the GDP deflator and the 3-month BUBOR. The rise in the GDP deflator may reduce the probability of default, since it may increase the nominal value of the company's profit provided that the repayment instalments are constant. In line with our expectations, the sign is indeed negative (Appendix, Table 7). The opposite sign is expected in the case of a change in the BUBOR, since in view of the fact that corporate loans typically have a floating interest rate, the rise in the BUBOR may translate into increased financing costs for companies, which clearly heightens the default risk. The estimated coefficient confirms this expectation.

4.2 THE GOODNESS OF FIT AND THE ACCURACY OF CLASSIFICATION

The goodness of fit is analysed with the method usually employed in case of a logistic regression, i.e. the so-called (McFadden's or) pseudo-R^2, which compares the likelihood function of the estimated model to the likelihood function of a model in which the coefficients of all the variables are zero. Although the indicator is between 0 and 1, and a higher value means a better fit, its concrete value is usually not considered to be overly important. The indicators capturing the accuracy of the classification are more important success indicators, and out of these, we will present sensitivity and classification accuracy. Sensitivity shows what portion of defaulters are classified as such, and classification accuracy expresses the proportion of the correct classifications relative to all the observations. It is important to note that these two indicators depend on the so-called cutoff value, i.e. a threshold PD above which the given observation is classified as a default. Moreover, the two indicators change in opposite direction to the cutoff value. To illustrate this, we charted the indicators together with the proportion of the correctly accepted as a function of the cutoff (Appendix, Chart 10). [22]

When determining the optimal cutoff, one should consider, for example, the two types of mistakes one can make during the lending decision: incorrectly classifying future non-defaulters as defaulters, i.e. the rejection of their loan application, which results in unrealised profits, and classifying bad clients as good ones, i.e. lending to them, which entails loan losses. Choosing the ratio of these, i.e. the cutoff used by the bank, depends on

[19] Although we excluded the companies from the sample that were most often grouped into the large enterprise category on the time horizon for the estimation, we still have a large enterprise variable. This includes the companies that, for the most part, are not classified as large enterprises but currently they are.

[20] As this variable is only a control variable in the equation, this part of the results is not presented.

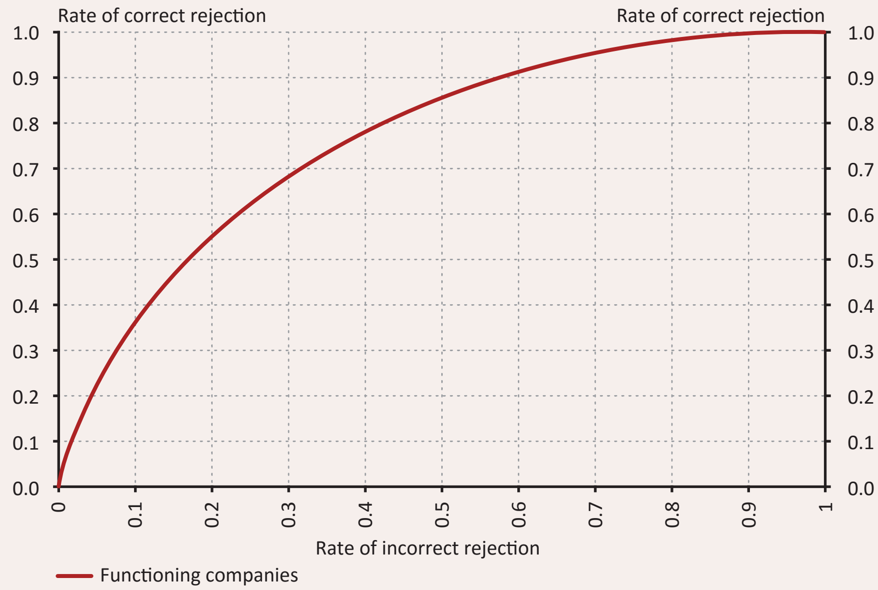
[21] Our results do not include a zero-year age category because in the case of the variables that represent change, we used an annual difference, therefore the companies that had not been functioning for a full year dropped out of our estimation database.

[22] Let us choose, for example, 50 per cent as the cutoff. As the default rate is low in the database (approximately 5.5 per cent for functioning companies), with the 50 per cent cutoff, we classify almost every observation as a non-defaulter. In such a case, the classification accuracy is close to 100 per cent, while sensitivity is around 0 per cent. If we reduce the cutoff, classification accuracy will drop, while sensitivity will rise.

a business decision.[23] For this reason, in our table we use a cutoff value, the proportion of defaults in the population, that can be considered a reference point.

The dependence of the previously mentioned indicators on the cutoff is eliminated by the AUROC indicator, which is independent from the threshold and which gives us the area under the so-called ROC curve. The ROC curve (Chart 3) describes the classification feature of the scoring model in case of various cutoff values. The greater the area under the curve, the better the scoring model. Completely random classification, i.e. the model with basically zero separation power, corresponds to the ROC curve on the 45-degree line. The AUROC indicator, i.e. the area under the curve, is used accordingly. It provides help in evaluating the classification capacity of the model irrespective of the cutoff.

Chart 3
ROC curve – Model for functioning companies



Note: The vertical axis shows the previously defined sensitivity, while the horizontal axis shows the proportion of the good ones classified as bad with the given cutoff. This means that the curve shows for each proportion of those excluded from the good ones (i.e. the incorrectly rejected) – and, equivalent to this, for each corresponding cutoff value – the corresponding proportion of those excluded from the bad ones (i.e. the correctly rejected). In line with this, the best model, i.e. the one which can best separate future non-defaulters from defaulters corresponds to the ROC curve with the greatest positive slope, since in such a scenario at first we only exclude exclusively bad ones as the cutoff increases.

With the help of our classification indicators, we examined the temporal fit of our model fitted onto functioning companies – which from now on is referred to as the baseline model – and the importance of some key variable groups with respect to the explanatory power of the model. By calculating the AUROC indicator for each quarter, we find that the model exhibits a slightly better fit at the beginning of the sample (Chart 4). This is probably partly caused by the fact that at the beginning of the sample period, the default events that proved to be final in the history of the companies persisted in greater proportion. The model performs well over the whole time horizon, since if we use quarter dummies instead of macroeconomic variables and the trend – and this specification gives a perfect temporal fit, in line with the logit regression's feature – the pseudo-R^2 and the AUROC for the whole time horizon barely improve (Table 2). The macroeconomic variables play a substantial part in the temporal fit, and the fit particularly declines from 2011 without them, but the difference can be seen in the indicators for the whole sample as well. However, the explanatory power of the baseline model measured over the whole time horizon is not determined by the macroeconomic variables: the influence of the trend and the included quadratic terms is much more pronounced. By excluding the quadratic terms from the baseline model, goodness of fit indicators deteriorate considerably (both the AUROC and the

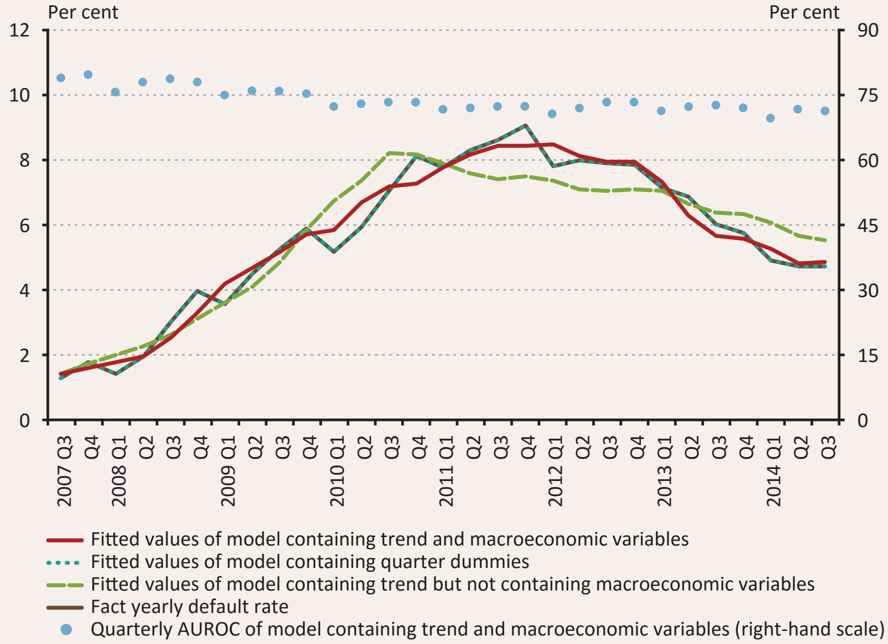
[23] The method for choosing the optimal cutoff is detailed in Oravecz (2007).

pseudo-R^2 decrease), which confirms the importance of non-linear effects. The re-estimation of the baseline model without a trend entails a quite considerable drop in the fit of the model, which proves that data issues should be addressed.

Table 2
The goodness of fit and the classification accuracy indicators for functioning companies in the case of different specifications

	Baseline model	With quarter dummies	Without macroeconomic variables	Without trend variable	Without quadratic terms
Pseudo-R^2	11.2%	11.3%	11.0%	9.3%	10.4%
AUROC	76.2%	76.2%	76.0%	74.0%	75.5%
Cutoff	5.5%	5.5%	5.5%	5.5%	5.5%
Sensitivity	71.4%	71.4%	71.3%	68.6%	71.2%
Classification accuracy	67.5%	67.6%	67.4%	66.8%	66.7%

Chart 4
Fit of the models for functioning companies to the factual time series of the default rate



Note: The pattern discernible at the turn of the calendar year for the sample of functioning companies is caused by the companies shifting from the functioning to the non-functioning category, since the proportion of defaulters is higher among them than among those remaining in the category of functioning companies.

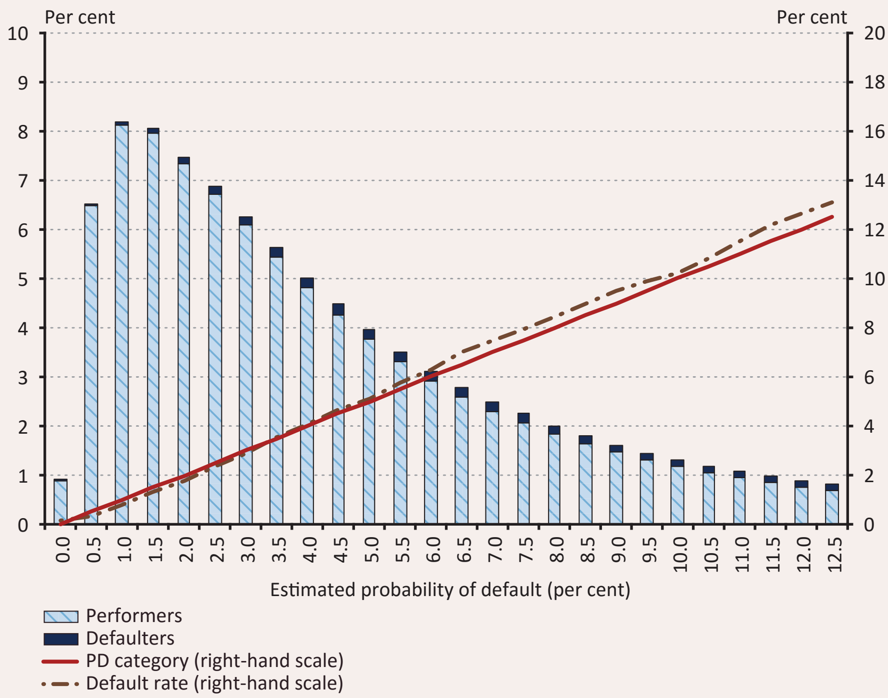
Chart 5, which can also be used for establishing the accuracy of the baseline model, shows the distribution of the PDs estimated by the model and the deviation of the estimated PDs from the actual default rate at the same time. The bars on the chart show the relative frequency of the observations with the given estimated PDs (or with the PDs rounded to those).[24] Furthermore, each bar shows the portion of defaulters and non-defaulters

[24] The estimated PDs of the observations were rounded to 0.5 per cent. The 10 per cent of the whole population with the highest estimated PD (or, more precisely, the PD categories above the one containing the 90th percentile) is not included in the chart, since, as it can be seen, the distribution stretches far to the right, therefore there are very few observations (less than 1 per cent of the sample) in the higher estimated PD categories.

within the observations with the given estimated PDs. Based on the bar chart, we can say that the fact that the distribution stretches far to the right, with a large portion of the observations having a low probability of default of a couple of percentage points, tallies with our expectations. The separation of the bars to defaulters and non-defaulters shows that as the estimated PD grows, the proportion of defaulters increases, and it is also clear that although their proportion rises, defaulters do not become dominant in the greater estimated PD categories, therefore the model does not help in choosing a well-separating cutoff.

The continuous lines on the chart show the actual default rates of the companies in the given estimated PD categories, and, as a reference point, the 45-degree line of the PD estimated for the category. The actual default rate fitting onto this line means that out of the subpopulation with the given estimated PD, the proportion of defaults equalled our probability-of-default estimate. Based on the chart, the actual default rate is close to the line, which confirms the goodness of our PD estimate.

Chart 5
The distribution of estimated PDs and their fit to the default rate – baseline model for functioning companies



5 Scoring models by company size

Among the SME categories, there may be considerable differences as regards the heterogeneity of the companies within the size category, and as a result, small and medium-sized enterprises can be better modelled than microenterprises that are typically substantially different from each other. Moreover, according to banks' experiences, the credit risk of microenterprises is in many cases substantially influenced by qualitative factors that are relatively difficult to measure (e.g. the financial awareness, financial position and motivation of the managing director), and these cannot be taken into account in the present study. Therefore, we can give a more reliable estimate about the credit risk of small and medium-sized enterprises by using the company-level financial indicators available in our database than for microenterprises. Furthermore, it can be seen from the estimates of the previously presented model fitted onto functioning companies, and also from the results of the estimates using the categorised forms of the company-specific variables, that the extent of influence of the individual variables on the credit risk of micro, small and medium-sized enterprises varies widely, in fact, in certain cases the influence can be described with different functional forms. In addition, among the undivided company-specific explanatory variables of the model fitted onto functioning companies, there may be some whose influence on credit risk differs in the case of the various SME categories. Due to the above-mentioned reasons, we believe it is important to specify different scoring models for the different company sizes. This separation considerably improves the fit of the models in case of the two larger corporate size categories, as we will see from the results presented in the chapter.

We based the differentiation by company size solely on the dataset containing the functioning SMEs, and from that, we created three subsamples by taking into account the company's headcount and sales revenue as described in Chapter 2.2. Companies were grouped into the subsamples containing micro, small and medium-sized enterprises, depending on which SME category they belonged to the most frequently between 2004 and 2014. Based on this, 78.2, 17.8 and 4 per cent of the companies were included in the microenterprise, small enterprise and medium-sized enterprise subsamples, respectively. As microenterprises make up most of the sample, out of the scoring models differentiated by company size, the microenterprise model's results are expected to mostly resemble the results of the estimation performed without size differentiation.

During the creation of the models for the different company sizes, we strove to come up with models that may directly help banks' credit evaluation process. This scoring approach means that, in contrast to earlier attempts, we only include significant variables in the final model specifications presented below. Furthermore, the fact that all variables in the final models should have an intuitive sign was an important aspect during the selection of variables (for example in the case of microenterprises, we excluded the liquidity indicator from the scoring model for this reason). We also took into account that not all indicators related to companies' credit risk are suitable for a scoring model.[25] The requirement that the scoring models for the different company sizes should be suitable for supplementing banks' credit scoring practices was taken into consideration during the decision on truncating and winsorising variables, too.

[25] For example we have information about whether a company obtained a new loan in the given period, and this feature would be represented in our equation with a risk-reducing sign. But actually this variable would only capture the fact that a company passed a bank's credit evaluation in that moment, and that the bank performed a materially better choice during the credit evaluation compared to a random selection. However, this in itself does not necessarily mean that the given company's credit risk is lower than that of a peer identical with respect to all its other characteristics, because it is possible that the other company did not obtain a loan in the given quarter simply because it did not apply for it.

5.1 THE RESULTS OF THE MODELS DIFFERENTIATED BY COMPANY SIZE

Now we will present the estimates of the models specified for the different company sizes (Table 3, Table 8 of the Appendix), emphasising the differences compared to the model for functioning companies. One of the reasons behind the differences may be that the whole model of the functioning companies and the scoring models differentiated by size have partly different variable sets. Furthermore, although in the case of certain company-specific variables we also employed differentiation by size in the model for functioning companies, their impact on credit risk may be different in the scoring models differentiated by company size presented below, simply because the coefficients of the undivided company-specific explanatory variables and the macroeconomic variables are dominated by the behaviour of microenterprises that make up most of the population of functioning companies. It is also worth pointing out that the three models not only differ in the populations under review, but also partly in their variable sets. Accordingly, the results gained from the models differentiated by company size and presented on the same chart cannot be compared directly, without taking into account these considerations.

Table 3
Estimates of the scoring models for the different SME categories – Coefficients of selected company-specific variables

	Micro	Small	Medium-sized
Leverage for companies with positive equity (t-1)	–0.29 ***	2.44 ***	5.10 ***
Square of leverage for companies with positive equity (t-1)	1.28 ***		–1.62 ***
Sales revenue growth rate (t-1) (micro truncated)	–1.07 ***	–1.64 ***	–1.67 ***
Square of sales revenue growth rate (t-1) (micro truncated)	–0.11 ***		
Sales revenue growth rate (t-1) (dummy above truncation point for micros)	0.04 ***		
After-tax ROA (t-1)	–0.62 ***	–2.55 ***	–5.72 ***
Square of after-tax ROA (t-1)	–0.53 ***	–2.47 ***	–7.04 ***
Change in after-tax ROA (t-1)	–0.37 ***	–0.55 ***	–1.71 ***
Square of change in after-tax ROA (t-1)	–0.22 ***	–0.40 ***	–1.29 ***
Majority foreign ownership (t-1)	–0.18 ***	–0.41 ***	–1.03 ***
Original maturity for companies with positive equity (t-1)	–0.03 ***		
Original maturity for companies with negative equity (t-1)	–0.07 ***	–0.06 ***	–0.09 ***
Liquidity indicator (t-1)			–0.57 ***
Potential coverage of loans (t-1)	–0.19 ***	–0.05 ***	–0.07 ***
Square of potential coverage of loans (t-1)	0.01 ***		
Negative equity (t-1)	1.20 ***	2.78 ***	3.76 ***
Export ratio (relative to sales revenue) (t-1) - dummy for values under 15 per cent	–0.17 ***		
Export ratio (relative to sales revenue) (t-1) - dummy for values of 15 per cent and above	0.05 ***		
Export ratio (relative to sales revenue) (t-1)			–1.03 ***
Square of export ratio (relative to sales revenue) (t-1)			1.08 ***
Proportion of FX loans (t-1)	3.39 ***	3.20 ***	3.57 ***
Square of proportion of FX loans (t-1)	–3.08 ***	–2.98 ***	–3.31 ***
Change in fixed assets (truncated)		–0.92 ***	
Square of change in fixed assets (truncated)		0.93 ***	
Change in fixed assets (dummy above truncation point)		–0.01	

*Note: The table contains the estimated coefficients of the model. The ***, **, and * next to the coefficients mean that the coefficients differ from zero at a confidence level of 99, 95 and 90 per cent, respectively.*

All in all, according to all three models, as leverage increases, the credit risk of a company with positive equity heightens, yet there is some difference in the actual functional forms (Appendix, Chart 13).

On the continuous domain, an increase in sales revenue reduces the company's credit risk (Appendix, Chart 14). Similar to the model fitted onto functioning companies, the probability of default of the various SME categories is influenced differently by the company's sales revenue rising to more than 1.5 times of the previous year's value: in the case of a microenterprise it slightly increases, while in the case of small and medium-sized enterprises it markedly decreases credit risk.

Increasing profitability (ROA) reduces credit risk in all three SME categories, and in the case of all three models, there is a significant non-linear relationship between the ROA level and the company's credit risk (Appendix, Chart 15). Nonetheless, there is a difference in magnitude between the coefficients across the different size categories: the level of profitability has a much more pronounced effect on the probability of default in the case of a medium-sized enterprise than in the case of micro and small enterprises, which also exhibit substantial differences.

Chart 16 of the Appendix shows the influence of the change in the ROA on credit risk. Similar to the model for functioning companies, if a firm's profitability diminishes, its credit risk rises in the case of the models for the different company sizes as well. The functional form is similar in the three cases, but the extent of the effect is considerably different in the various size categories. The difference is especially great in the case of medium-sized enterprises, whose probability of default is influenced much stronger by the change in profitability than in the case of smaller companies.

After the differentiation by size, we see slightly different results with respect to the liquidity indicator, as compared to the model fitted onto functioning companies (Table 3). When including the liquidity indicator in the microenterprise model, we arrive at significantly positive coefficients both for the linear and the quadratic terms. However, in our final scoring model, we did not include the liquidity position of microenterprises, since we are not convinced that greater liquidity would directly increase the credit risk of a company. In the small enterprise model, the liquidity indicator proved to be insignificant, while in the model fitted onto medium-sized enterprises, in line with economic intuition, a better liquidity position significantly reduces the company's probability of default.

Longer original maturity decreases credit risk for companies with either positive or negative equity, but in the case of companies with negative equity, the risk-reducing effect is typically much more pronounced. For firms with positive equity, original maturity only proved to be significant in the microenterprise model, therefore in the scoring models for small and medium-sized enterprises, we only take it into account in the case of companies with negative equity.

In the baseline model fitted onto functioning companies, the indicator capturing the potential coverage of loans had a widely varying effect on credit risk in the different SME categories. Similarly, there is also a difference regarding the functional forms in the models estimated for the separate size categories (Appendix, Chart 17): in the case of a microenterprise, the probability of default diminishes at a decelerating pace as the coverage increases, while in the case of small and medium-sized enterprises, due to the insignificance of the quadratic term, companies' credit risk drops in proportion to the rise in potential coverage. Based on the log-odds, the credit risk-reducing effect of the variable is much greater in the microenterprise model than in the case of small and medium-sized enterprises.

The change in companies' stock of fixed assets, with which we endeavour to capture firms' investment activity, is only included in the final scoring model for small enterprises. According to the estimates, if a small enterprise considerably expanded its stock of fixed assets in the previous year, this has a diminishing effect on the company's probability of default (Appendix, Chart 18).

The impact of the export variable on credit risk was examined with a functional form similar to the baseline model fitted onto functioning companies. The only change we detected in our results was that in the small enterprise category the variable lost its significance, therefore we ultimately left it out of that model (Appendix, Chart 19).

The proportion of FX contracts basically has the same impact on companies' probability of default in the scoring models fitted onto each size category (Appendix, Chart 20): both the functional forms and the values of the coefficients are very similar. Consequently, irrespective of the size of company, we can say that the probability of default only increases with the growth in the proportion of FX-denominated loans to a certain extent, and beyond a point, a higher FX ratio entails lower credit risk.

Majority foreign ownership reduces companies' credit risk, regardless of the size category (Table 3). The value of the coefficient, however, varies significantly: the risk-reducing impact of majority foreign ownership is the most dominant in medium-sized enterprises, and it may be the least relevant in the case of microenterprises.

Similar to the baseline model for functioning companies, firms' further characteristics were also controlled for with categorical variables in the scoring models for the different company sizes. In the case of company age, the categorisation for micro and small enterprises was similar to the one employed in the baseline model, while in the case of medium-sized enterprises, due to the smaller population size, the use of larger categories proved to be appropriate. It can be seen in the case of all three SME categories that, in line with the baseline model for the functioning companies, firms that were founded earlier are in general less risky than their younger peers (Appendix, Table 8).

To adjust for the overwritten default events, all three scoring models included the previously described trend, and, in order to ensure better temporal fitting, the two macroeconomic variables already seen in the model fitted onto functioning companies (Appendix, Table 8). In the case of the GDP deflator, the impact is slightly stronger in the medium-sized enterprise model than in the other two size categories.

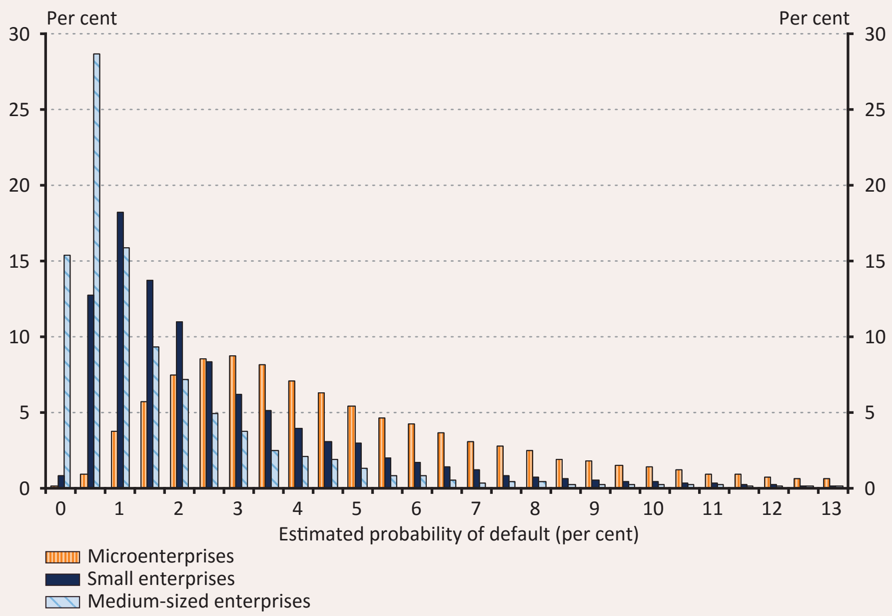
Table 4
The indicators of goodness of fit and classification accuracy in the scoring models for the different SME categories

	Microenterprises	Small enterprises	Medium-sized enterprises
Number of observations	1,693,544	385,752	87,245
Pseudo-R^2	10.1%	14.3%	20.6%
AUROC	74.7%	79.4%	84.4%
AUROC for the given corporate size in the baseline model of functioning companies	74.8%	78.4%	82.7%
Cutoff	7.0%	3.9%	2.5%
Sensitivity	63.0%	73.3%	77.7%
Classification accuracy	71.8%	71.4%	75.6%

Based on both the values of the pseudo-R^2 and the AUROC indicators (Table 4), the fit of the scoring models considerably improves with the growth in company size. Among other factors, this may be due to the fact that in the case of small enterprises, and especially medium-sized enterprises, the firms in the category are much more homogeneous than microenterprises, and their credit risk depends more on the financial indicators of the company. The AUROC indicators are very similar to the results of Antunes et al. (2016), both in terms of their values and their relations between company sizes. When comparing the AUROC indicator of the scoring models estimated separately for the different company sizes with the AUROC indicators restricted by the given company size that were based on our baseline model estimated for functioning companies, we find that the estimation for the separate size categories only yields improvement in categories with less data points. This is because in the case of the variables not divided up by size, the estimated coefficient is dominated by microenterprises, therefore eliminating these restrictions improves the performance of the small and medium-sized enterprise categories' models. In the case of microenterprises, the somewhat looser fit of the estimate differentiated by size is due to the exclusion of the liquidity indicator that is significant but does not have the appropriate sign from the perspective of scoring. When comparing sensitivity and classification accuracy in the different models, one should take into account that the cutoff values are different, since the proportion of defaulting companies varies across populations of the different SME categories.

We presented the distributions of the probabilities of default estimated by the scoring models as well as the divergence of the estimated PDs from the population's default rate on Charts 21–23 in the Appendix. In all three cases, 90 per cent of the companies in the given SME category are shown on the charts, as the other 10 per cent of firms are in an estimated PD category where there are relatively few observations. The actual default rate is close to the probability of default estimated in the model in all three cases. We can see that, similar to the model fitted onto functioning companies, the distribution of the estimated PDs assumes a shape that stretches to the right, but the extent of this varies widely across the size categories. The small enterprise model classifies much more companies into relatively lower default categories than the microenterprise version, but the medium-sized enterprise model does even more so. We can observe similar differences in distribution by company size if, unlike previously, we only consider the most recent PD distribution estimates, for 2014 Q3, and not for the whole modelling time horizon (Chart 6).

Chart 6
Distribution of estimated PDs by SME categories for 2014 Q3



Note: The chart does not show the companies with the greatest estimated PDs that approximately represent 10, 5 and 3 per cent of the sample in the case of microenterprises, small enterprises and medium-sized enterprises, respectively.

5.2 ROBUSTNESS CHECK ON THE LENGTH OF THE ESTIMATION SAMPLE

In addition to their goodness of fit, our scoring models for the different size categories were also analysed for stability, i.e. we examined how much the estimated relations change due to the changes in the estimation time horizon, and to what extent the defaulting and non-defaulting categories are modified when the default ratio in the population is chosen as the cutoff from our estimated probabilities of default. To this end, we shortened the time horizon serving as a basis for our estimation by two, four and six quarters as compared to the last quarter of our estimation time horizon, i.e. 2014 Q3. Then we estimated models in all three size categories, for all three shortened time horizons with the specifications detailed above.

The estimated coefficients of the models produced in this manner and the impact of the complex variables on the dependent variable were first compared to the coefficients of the model estimated for the whole time horizon spanning until 2014 Q3 and to the impacts on the dependent variable in that model. The only major difference could be observed in the medium-sized enterprise model, where the effect and significance of the indicator measuring the potential coverage of the loans drop sharply. Based on the AUROC indicator, the fit of the models estimated on the shortened time horizon almost equals the AUROC value calculated for the given period with the model estimated for the whole sample (Appendix, Table 9). The AUROC values of the two-, four- and six-quarter out-of-sample fittings produced based on the models with the shortened time horizons barely fall short of the AUROC values calculated for the same period with the model estimated for the whole sample. Thus, overall, the fit considering the whole time horizon deteriorates very little through the shortening of the horizon by a couple of quarters.

Nonetheless, in our model, not only the aggregate fitting statistics are interesting, but also the fluctuations of their implied classifications. Therefore we compared the defaulting and non-defaulting categories identified with the help of the cutoff and the PDs estimated from the models calculated for the shortened time horizons, with the categories derived in a similar fashion from the models estimated for the whole time horizon. This was examined for the estimation time horizon of the model with the shortened time horizon – this sample is part of the estimation samples of the models for both time horizons – and for the quarters beyond the estimation time horizon of the model with the shortened time horizon, for which an out-of-sample estimate is produced from the shorter time horizon model. In Table 5, we compare the model of all three size categories estimated for the whole time horizon that lasts until 2014 Q3 with the models estimated for the time horizons spanning until 2013 Q1, 2013 Q3 and 2014 Q1 with respect the periods detailed above. The table contains the ratio of the company-date pairs classified into different categories based on the two models (i.e. to non-defaulting instead of defaulting, or vice versa) in proportion to the company-date pairs for which a PD estimation was produced with both models.

Table 5
Assessment of the stability of the scoring models for the different size categories

Size categories	Time horizons for comparison	2014 Q1	2013 Q3	2013 Q1
Micro enterprises	Whole time horizon	0.50%	0.97%	1.34%
	Out-of-sample estimation of model with shortened time horizon	0.78%	1.59%	1.71%
	In-sample estimation of model with shortened time horizon	0.48%	0.89%	1.26%
Small enterprises	Whole time horizon	0.60%	1.34%	2.36%
	Out-of-sample estimation of model with shortened time horizon	0.92%	2.26%	4.11%
	In-sample estimation of model with shortened time horizon	0.58%	1.23%	1.99%
Medium-sized enterprises	Whole time horizon	0.66%	1.43%	2.41%
	Out-of-sample estimation of model with shortened time horizon	0.63%	1.61%	2.91%
	In-sample estimation of model with shortened time horizon	0.66%	1.41%	2.33%

Based on the table, the following can be stated. First, the more difference there is between the time horizons of the two estimates, the more the PDs estimated from the models and the thus created categories differ. Second, in line with our expectations, there are smaller differences on the shared section of the two estimates' time horizons than when comparing the categories fitted out-of-sample from the model estimated for the shortened time horizon with the results of the model estimated for the whole time horizon. And third, the greatest differences in the classification changes of the PDs estimated for the shortened time horizon and thus fit outside the time horizon can be found in the small enterprise model. This is probably explained by the fact that small enterprises' characteristics vary in time the strongest. Yet taking into account that in practice the change in the estimation time horizon of our models modifies less than 5 per cent of the observations' categories, our models can be considered stable, which underlines their utility, despite the fact that it has been more than a year since the closing date of the estimation sample. Nevertheless, we believe it is important to re-estimate our models regularly, for example annually or once every other year, since, as we have shown, the estimated coefficients and the significance levels of our models' variables and thus the relations sought to be captured may be modified more and more as we get further and further away from the estimation time horizon.

6 Models specified for different economic sectors

In addition to the differentiation by size categories, we specified separate models for some key economic sectors to examine how the relations derived for functioning companies are modified. Our aim was thus a sort of robustness analysis, and not the creation of models suitable for scoring. Among the economic sectors under review, trade, manufacturing, construction and agriculture represent 27.9, 13.4, 11.5 and 4.9 per cent of the sample of functioning companies, respectively. We tested the impact of the explanatory variables included in the baseline model for functioning companies and the sector-specific macroeconomic variables on credit risk for the samples differentiated by economic sectors. Most variables behave in these models similarly to the estimate for the whole population of functioning companies, however, there are shifts in emphasis reflecting certain sector-specific features (Appendix, Table 10).

The impact of a portion of the company-specific explanatory variables on companies' probability of default does not vary substantially in the economic sectors under review. Higher leverage typically impairs the probability of repayment at an increasing pace, and companies with negative equity honour their obligations towards credit institutions with a significantly lower probability. Growing sales revenue reduces credit risk in all the sectors under review, but, similar to the model for functioning companies, to a significantly varying extent in the different size categories. Also, a company's credit risk is considerably heightened in the case of all four sectors if it has FX-denominated loans. In all sectors, the curve representing this effect has its maximum at the companies indebted in forint and FX in roughly the same proportion. In the model for construction companies, the risk-increasing effect of FX loans may be significantly higher.

With respect to profitability indicators, there are some differences between certain sectors. Based on the estimated coefficients for the companies in the construction industry, the positive after-tax ROA may reduce credit risk much less than in the case of the other three economic sectors, and while in construction the risk-reducing effect is steady, it happens at an increasing pace in the other sectors under review. The improvement of a company's profitability reduces credit risk at an accelerating pace in all cases, and the magnitude of this effect becomes greater as firms' size category grows. The change in the after-tax ROA does not have a significant impact on credit risk for microenterprises in manufacturing, and for small enterprises in agriculture.

The liquidity indicator exerts a significant impact on the probability of default in the case of all the sectors under review, with different directions and extent for the different company sizes. The phenomenon observed in the case of functioning companies, namely that for microenterprises, in contrast to larger companies, a lower liquidity indicator indicates lower credit risk, emerged in the models of all economic sectors under review. Furthermore, in the case of medium-sized enterprises, the liquidity position typically has greater importance than in the case of small enterprises.

In these models as well, the impact of original maturity was examined separately for companies with positive and negative equity. In the models for construction, trade and manufacturing, longer maturity reduces the probability of default, and in the case of negative equity, it does so to a much larger extent. However, in the agriculture model, we arrive at a different result: for companies with positive equity, original maturity significantly increases credit risk. This statistical relation may uncover a previously unknown feature of lending to companies operating in agriculture. Due to the result supporting short-term lending, which runs contrary to our intuition described earlier, we believe that in its present form the model for agriculture is not suitable for scoring.

The company sizes in the case of which potential coverage exerts a significant impact on credit risk vary by economic sectors, but the sign of the significant coefficients is negative in all models, i.e. ceteris paribus the expansion of loans' potential coverage reduces a company's probability of default. In the models for trade and manufacturing, the risk-reducing effect is steady in all three SME categories, getting substantially greater with the growth of the size category, and the effect of potential coverage on credit risk is more substantial in the case of trade than in manufacturing. In the model for construction, the coefficients of micro and small enterprises are significant, and the estimated risk-reducing effect is steady in both cases, while in the case of agriculture, potential coverage reduces credit risk only for microenterprises, and at a diminishing pace.

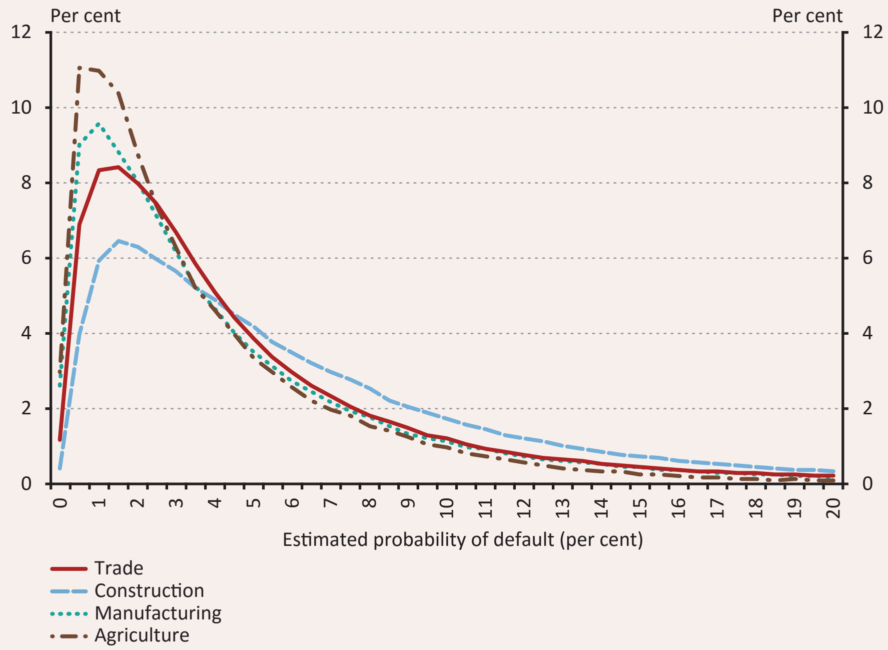
There are considerable differences with respect to the impact of export activities on credit risk across the economic sectors. In the model for construction, the export indicator has no significant effect on companies' credit risk, which tallies with the characteristics of the industry. In the model for trade, exports only play a role in the case of microenterprises: relatively limited export activity (of up to 15 per cent of the sales revenue) reduces, while a more intense activity increases companies' probability of default. In the model for agriculture, the opposite is the case, in that exports play a role with respect to the credit risk of small and medium-sized enterprises: to a certain extent, the expansion of the proportion of export revenues decreases credit risk, then the favourable effect on credit risk gradually diminishes, and in the case of companies producing almost solely for export, it ultimately even slightly increases companies' probability of default. Export activities have a significant impact on the credit risk of manufacturing companies in all three size categories, similar to the model fitted onto functioning companies. For microenterprises, export activities of less than 15 per cent relative to all sales revenues reduce, while more substantial export activities increase companies' estimated credit risk. By contrast, in the case of small enterprises, the impact of exports gets more and more favourable up to a certain point, then it diminishes, and for small enterprises barely producing for the domestic market it even boosts the probability of default. For medium-sized enterprises, an increase in the indicator mitigates credit risk at a diminishing pace, but over the whole domain.

In addition, several sector-specific features emerge in our categorical variables. For example in the companies that are majority foreign owned in construction, trade and manufacturing, credit risk is ceteris paribus lower, but in the case of companies operating in agriculture, majority foreign ownership has an insignificant impact. With respect to regions, one should point out agriculture's model because, consistent with its strong ties to agricultural land, credit risk is the lowest in the Great Plain regions and in Western Transdanubia.

Based on our models, we produced an estimate for the probability of default of the individual companies over the whole time horizon under review. According to the estimated PDs (Chart 7), among the companies with outstanding loans in the economic sectors under review, credit risk was the lowest for agricultural companies[26] (with the distribution being substantially peaked towards zero then flattening out relatively early on), and the highest in construction (with the distribution being less peaked in the case of lower PD categories, but having a heavier tail). Three effects emerge in these results simultaneously: first, the distribution of the sector's companies by risk, second, the risk-sensitivity of the lenders with respect to the given sector, and third, the demand of the given sector for bank loans.

[26] Our sample does not include small-scale agricultural producers, only companies operating as enterprises.

Chart 7
Distribution of the estimated probabilities of default across the economic sectors



7 Summary

In our study, we presented a new database suitable for estimating the credit risk of Hungarian companies, which is unique in terms of both its size and information content. On the one hand, it contains information on loan contracts and the contracts' performance for all non-financial corporations with loans or loan-type financial products for the past several years, and on the other hand, it links this information with companies' financial indicators from their annual accounts. As the current legislation only allows the Magyar Nemzeti Bank (the central bank of Hungary) to make such a connection between the aforementioned databases, the publication of the conclusions drawn from the linked database is especially important, since the results may be of interest to a more general public. Moreover, due to the exclusive access, all the issues going beyond the classic central banking areas are worth examining that have the potential to provide relevant information to some economic agent, thereby facilitating the better fulfilment of the economic agent's role in the economic system. Therefore we examined corporate credit risk in a form that enables its results and the lessons learnt from them to be used in the credit evaluation practice of lending institutions.

In our models presented in the previous chapters, the literature's recommendation that companies' credit risk analysis should be performed separately for the different size categories because one may observe different relations across the categories seems to have been validated. Companies' financial indicators increasingly influence their credit risk as the firms grow.

During the estimation of our models, we paid special attention to the analysis of the functional form of the relation between the individual explanatory variables and the dependent variable. We found a non-linear relation in the case of several variables, and we included a quadratic term in the equation of our estimation to represent this. Furthermore, in the case of some variables, the winsorising performed based on a statistical approach was overwritten by further winsorising or truncation. Thereby we managed to replace the linear approach with a form better resembling the actual relation between the variables in the case of the growth rate of the sales revenue, the export ratio and the FX loan ratio of a company, as well as profitability and its annual change.

Furthermore, we analysed non-linear effects through the interaction of multiple explanatory variables, and based on that we can say that the sign of equity proved to be a categorical variable that should be used – in addition to the separation by size categories mentioned above – to differentiate when examining the relation of other, continuous variables to credit risk. As the proportion of companies operating with negative equity for a longer period is high among Hungarian microenterprises and considerable among small and medium-sized enterprises, the special problems posed by this deserve particular attention.

In addition to the differentiation by company size, we also examined the distinction among economic sectors. Although in line with our expectations and conforming to the characteristics of the given sector, the impact of some corporate features – such as the export ratio, the potential coverage and the region of the company's registered office – exerted on credit risk and the magnitude of the impact differed from the estimate for the whole population, the importance of the differentiation by size did not vanish in these models either.

As we have already pointed out, the analyses of the new database presented here are only the first of the many possibilities. In addition to the fact that the models described here in the form best suited for credit evaluation can be estimated in a slightly different structure, with another aim – for example to examine the credit risk of the whole outstanding portfolio –, with another modelling approaches one can use the data to find answers to other questions, too. For example with hazard-type models, one could explore the relation between the duration of a company's default and its characteristics, or separately estimate the credit risk of previously defaulting companies turning once again back into the non-defaulting status. However, the length of the sample period is particularly important in these estimations, therefore the expansion of our database with new periods will materially improve the analysability of the issues.

8 References

Altman, E. I. (1968): Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance,* 23 (4), 589–611.

Altman, E. I., G. Sabato (2007): Modeling Credit Risk for SMEs: Evidence from the U.S. Market. *Abacus,* 43 (3), 332–357.

Antunes, A., H. Gonçalves, P. Prego (2016): Firm default probabilities revisited. *Banco de Portugal Economic Studies*, 2 (2), 21–45.

Bauer, J., V. Agarwal (2014): Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking and Finance,* 40, 432–442.

Bauer, P., M. Endrész (2016): Modelling Bankruptcy using Hungarian firm-level data. *MNB Occasional Papers,* 122.

Carling, K., T. Jacobson, J. Lindé, K. Roszbach (2007): Corporate credit risk modeling and the macroeconomy. *Journal of Banking and Finance,* 31, 845–868.

Lykke, M., K. J. Pedersen, H. M. Vinther (2004): A Failure-Rate Model for the Danish Corporate Sector. *Danmarks Nationalbank Working Papers,* 2004/16.

Oravecz, B. (2007): Credit scoring modellek és teljesítményük értékelése (Credit scoring models and measuring their performance). Credit Institutes' Review, 6 (6), pp. 607–627.

Shumway, T. (2001): Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *The Journal of Business,* 74 (1), 101–124.

Sjovoll, E. (1999): Assessment of Credit Risk in the Norwegian Business Sector. *Norges Bank Working Paper,* 1999/9.

9 Appendix

9.1 METHOD FOR ANALYSING NON-LINEARITIES

The non-linearity observable in the impact of a variable was examined by dividing the relevant range of the variable's image[27] to twenty equal parts, then introducing categorical variables for the individual parts. If there were such observations, the parts outside the relevant range were included in a separate category (with the values lower than the range's minimum separated from those higher than its maximum), and in some instances, this was the case with the variable values with special meaning. With the categorised form derived in this manner from the continuous variable, we estimated the model, always choosing the category containing 0 as a reference, then we decided about the functional form to be tried in the case of the given variable depending on the log-odds[28] patterns outlined based on the coefficients we arrived at.

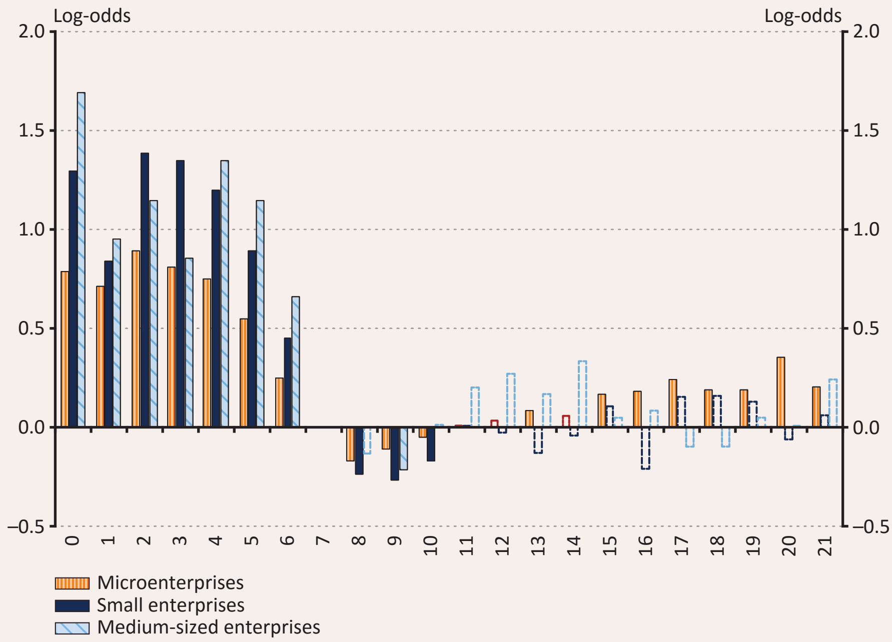
In case of a large portion of the variables, the joint inclusion of the first and the second power were deemed necessary, which was by and large confirmed in the ultimate estimation in the significance of the quadratic term. The categorised form was also a strong indication as to which observations should be regarded as outliers in the case of the individual variables, and whether these should be managed by winsorising or by the introduction of a separate dummy variable. We only introduced dummy variables if the coefficient of the outliers' category differed substantially (maybe even in its sign) from the results of the categories closest to them, and therefore merging them would have been a mistake. In other cases we strove to avoid this solution, as the dummy variables introduced for the outliers cause discontinuities in the results along the image of the variable concerned.

In order to illustrate the method (Appendix, Chart 1), we present the categorised form of the sales revenue's growth rate differentiated by size. We see different log-odds shapes for the various company sizes, therefore we will continue to use this differentiation. Based on the forms, one should include not only linear but also quadratic terms in the estimates, but this, due to the symmetry of the quadratic polynomial, cannot give a good approximation of the form we see in the case of higher growth rates. In the case of microenterprises, we see that all other things being equal, an extremely high growth in sales revenue entails greater credit risk than a positive but more moderate expansion in the sales revenue. On account of this, in the case of microenterprises, growth of over 50 per cent is indicated with a dummy variable and treated separately, while the range between -100 and 50 per cent, represented with a continuous variable, is approximated with a quadratic functional form. In the case of small and medium-sized enterprises, we find insignificant coefficients with varying signs in the categories of over 50 per cent, partly due to the relatively low number of data points in the categories. Therefore, in the case of these two company sizes, the portion above 50 per cent was winsorised, i.e. we assumed that the credit risk in the observations of sales revenue growth of over 50 per cent was not materially different from companies experiencing an expansion in sales revenue of exactly 50 per cent. Thus in the case of these two corporate categories, in contrast to microenterprises, there is no discontinuity in the log-odds shape produced during the estimation.

[27] Where based on the definition of the variable the image of the variable was not bounded on both sides, the relevant range was determined on the basis of the variable's distribution so that it covered the overwhelming majority of the variable's distribution, while making sure that in the case of an even distribution of the range, adequate number of data points were included in the individual categories.

[28] The odds is the probability of default divided by the complementary probability. Due to the shape of the logistic function, the linear combination of the variables with the coefficients gives the logarithm of the odds, therefore we can draw conclusions about the functional form to be tested in the case of the individual variables through the examination of the log-odds.

Chart 1
The categorised form of the sales revenue's growth rate differentiated by company size



Note: The 0 category represents the observations where the current sales revenue is 0. Categories 1 through 20 represent the division of the range between -100 and 200 per cent of sales revenue change into 20 equal parts, where the lower bound of the category is not part of the category, while its upper bound is. Thus Category 11, for example, contains the observations of growth rates above 50 per cent and up to 65 per cent. Category 21 includes the observations where sales revenue more than doubled year-on-year. The coloured bars represent coefficients significant at 5 per cent, while bars with only an outline show those that are insignificant at 5 per cent.

9.2 ROBUSTNESS ANALYSIS OF THE ESTIMATED COEFFICIENTS BY TAKING INTO ACCOUNT NON-FUNCTIONING COMPANIES

As the primary goal of this study is to specify models suitable for scoring, we excluded non-functioning companies from the baseline model described in Chapter 4.1. One of the reasons for this is that, as we have noted in Chapter 2.2, we believe that making a decision about lending to a company without any sales revenue should not be done based on a rule, but discretionally, i.e. based on a case-by-case assessment. On the other hand, if companies classified as non-functioning have systematically different features from functioning ones, it could bias our coefficients estimated in the scoring model. Nonetheless, as a robustness analysis, we perform an estimation by taking into consideration these as well.

It is important to emphasise that for the robustness analysis we do not specify a model for the so-called whole database that emerges with the inclusion of the non-functioning companies, but we will compare the results that the model identified for functioning companies produced about the whole database with the results about functioning firms. Since the two models only differ in the dummy variable introduced for non-functioning companies, the differences in the coefficients are caused by the observations added to the database, belonging to the companies classified as non-functioning. The estimated coefficients produced for the whole database can be found right next to the results of functioning companies (Appendix, Tables 6 and 7). Based on the estimates, it can be said that the two models are practically the same, i.e. apart from minor changes, we did not identify fundamental differences between them. We can note small changes in the significance and sign of certain variables, however, these can only be found in the case of the individual elements of complex variables. The impact of complex variables on the dependent variable only changes substantially in the case of three indicators. The change in impact of these indicators is detailed below. In addition to these, the generally greater credit risk of non-functioning companies is clearly indicated by the fact that their dummy variable is highly significant when coupled with a positive coefficient.

The comparison of the models suggests that the more substantial differences are in the original maturity, the liquidity indicator (Appendix, Chart 11), and the proportion of export activities relative to the sales revenue (Appendix, Chart 12).[29] In the model expanded with non-functioning companies, we arrived at a coefficient for the original maturity's variables that was also negative, but greater in absolute value than what we saw in the model for functioning companies. This is because ceteris paribus the default risk of non-functioning companies without sales revenue may be reduced more by a lower repayment burden.

Compared to the model for functioning companies (Appendix, Chart 6), it can be observed in the case of the liquidity indicator that around the -1 variable value, the effects of the indicator's variables by size categories on the dependent variable are in the moderately negative range, close to the zero log-odds (Appendix, Chart 11). The modification of the indicator's effect at this point can be attributed to the fact that the -1 value of the liquidity indicator typically comprises companies with negative equity.[30] And among non-functioning companies, there are relatively more firms with negative equity, thus by expanding the database with non-functioning companies, the effect is modified at this point. The direction and extent of the impact, however, should only be evaluated together with the coefficient of the dummy variable controlling for companies with negative equity and possibly capturing a portion of the effect.

In the case of the export activities relative to sales revenue, compared to the effects deduced from the model for functioning companies (Appendix, Chart 8), export activities reduce the default risk less in all size categories of the whole database (Appendix, Chart 12). This makes the impact of the export ratio on the dependent variable insignificant in the case of small and medium-sized enterprises.

Our success indicators such as the goodness of fit (measured with the pseudo-R^2) and classification accuracy deteriorate when non-functioning companies are excluded from the estimation (Appendix, Table 1). This is understandable if we assume that companies that became non-functioning overwhelmingly terminated their activities due to operational problems or payment issues. This presumed effect is captured in the model specification with a dummy variable. As we find companies just defaulting in much greater proportion among non-functioning companies, the model can classify them much more accurately.

Table 1
The goodness of fit and the classification accuracy indicators for functioning companies and for the whole database

	Functioning companies	Whole database
Number of observations	2,165,166	2,629,468
Pseudo-R^2 with trend and macro variables	11.2%	13.0%
Pseudo-R^2 with quarter dummies	11.3%	13.1%
AUROC	76.2%	77.4%
Cutoff	5.5%	6.5%
Sensitivity	71.4%	70.9%
Classification accuracy	67.5%	69.9%

[29] In Table 6 of the Appendix we can also see that the signs and significance levels of the subvariables of the indicator representing the stock of fixed assets relative to the sum of the outstanding loans' contractual loan amounts are markedly different. This, however, is not reflected substantially in the combined effect of the variable's components. Therefore, the major differences in coefficients can only be caused by the fact that in certain cases, the curves that are only slightly different can be represented by polynomials with vastly different coefficients.

[30] The -1 value of the liquidity indicator corresponds to a company with zero stock of liquid assets, financing its operation exclusively from short-term liabilities. But actually, as the variable was winsorised at -1, companies with negative equity and short-term liabilities in excess of their balance sheet total are also clustered in the -1 value.

9.3 ADDITIONAL TABLES

Table 2
Definitions of the continuous explanatory variables

Name of variable	Definition	Handling of missing values	First winsorising	
			Minimum	Maximum
Leverage for companies with positive equity	1 - (Equity / Total assets)	0, if Equity < 0	0	
Sales revenue growth rate	(Sales revenue (t) - Sales revenue (t-4)) / Sales revenue (t-4)	0, if Sales revenue (t-4) = 0		5
After-tax ROA	After-tax return / Total assets		–2	1
Change in after-tax ROA	After-tax ROA (t) - After-tax ROA (t-4)		–3	3
Original maturity	Average original maturity of the company's outstanding contracts, weighted by contractual loan amount (year)			30
Liquidity indicator	(Liquid assets - Short liabilities) / Total assets		–1	1
Potential coverage of loans	Fixed assets / Contractual loan amount			50
Change in fixed assets	(Fixed assets (t) - Fixed assets (t-4)) / Total assets (t-4)		–1	5
Export ratio (relative to sales revenue)	Export sales revenue / Sales revenue	0, if Sales revenue = 0		
Proportion of FX loans	Share of the FX-denominated portion in the company's outstanding loans, weighted by contractual loan amount			

Table 3
Winsorising and truncation during modelling

Name of variable	Further winsorising or truncation		Truncation (dummy variable)
	Minimum	Maximum	
Leverage for companies with positive equity			
Sales revenue growth rate		0,5	large values for microenterprises
After-tax ROA	-0,5		
Change in after-tax ROA		0,4	
Original maturity			
Liquidity indicator			
Potential coverage of loans		6	
Change in fixed assets	-0,5	0,3	large values for small enterprises
Export ratio (relative to sales revenue)			separation of small and large positive values at 15 per cent for microenterprises
Proportion of FX loans			

Table 4
Descriptive statistics of the continuous explanatory variables

Name of variable	Examined population	Number of observations	Average	Standard deviation	Minimum	5th percentile	10th percentile	25th percentile	Median	75th percentile	90th percentile	95th percentile	Maximum
Leverage for companies with positive equity	Functioning companies	1,874,355	0.57	0.25	0.00	0.13	0.21	0.37	0.59	0.78	0.90	0.95	1.00
	Microenterprises	1,424,281	0.57	0.26	0.00	0.11	0.20	0.37	0.59	0.79	0.91	0.95	1.00
	Small enterprises	366,708	0.57	0.23	0.00	0.17	0.24	0.39	0.57	0.75	0.87	0.93	1.00
	Medium-sized enterprises	83,366	0.57	0.23	0.00	0.19	0.26	0.40	0.58	0.75	0.88	0.93	1.00
Sales revenue growth rate	Functioning companies	2,165,166	0.23	0.92	–1.00	–0.58	–0.38	–0.14	0.02	0.25	0.81	1.80	5.00
	Microenterprises	1,693,544	0.25	0.99	–1.00	–0.64	–0.43	–0.16	0.01	0.27	0.95	2.10	5.00
	Small enterprises	385,162	0.15	0.65	–1.00	–0.38	–0.26	–0.10	0.04	0.20	0.49	0.88	5.00
	Medium-sized enterprises	86,460	0.14	0.61	–1.00	–0.31	–0.20	–0.07	0.05	0.18	0.40	0.71	5.00
After-tax ROA	Functioning companies	2,165,166	0.00	0.30	–2.00	–0.41	–0.18	–0.01	0.03	0.09	0.20	0.29	1.00
	Microenterprises	1,693,544	–0.02	0.32	–2.00	–0.48	–0.22	–0.03	0.02	0.09	0.21	0.30	1.00
	Small enterprises	385,162	0.04	0.18	–2.00	–0.16	–0.06	0.01	0.03	0.09	0.17	0.24	1.00
	Medium-sized enterprises	86,460	0.03	0.14	–2.00	–0.14	–0.06	0.00	0.03	0.08	0.15	0.20	1.00
Change in after-tax ROA	Functioning companies	2,165,166	–0.02	0.27	–3.00	–0.34	–0.19	–0.06	0.00	0.03	0.13	0.25	3.00
	Microenterprises	1,693,544	–0.02	0.30	–3.00	–0.38	–0.22	–0.07	0.00	0.04	0.15	0.28	3.00
	Small enterprises	385,162	–0.01	0.16	–3.00	–0.19	–0.11	–0.04	0.00	0.02	0.08	0.15	3.00
	Medium-sized enterprises	86,460	–0.01	0.13	–2.34	–0.14	–0.09	–0.03	0.00	0.02	0.07	0.11	3.00
Liquidity indicator	Functioning companies	2,165,166	–0.05	0.42	–1.00	–0.93	–0.61	–0.28	–0.03	0.20	0.46	0.63	1.00
	Microenterprises	1,693,544	–0.05	0.44	–1.00	–1.00	–0.67	–0.31	–0.03	0.23	0.50	0.67	1.00
	Small enterprises	385,162	–0.05	0.31	–1.00	–0.58	–0.44	–0.23	–0.04	0.14	0.33	0.44	1.00
	Medium-sized enterprises	86,460	–0.06	0.28	–1.00	–0.52	–0.40	–0.23	–0.06	0.11	0.27	0.39	1.00
Original maturity for companies with positive equity	Functioning companies	1,874,355	5.32	4.89	0.00	0.99	1.00	2.52	4.34	6.03	9.95	13.89	30.00
	Microenterprises	1,424,281	5.51	5.14	0.00	0.98	1.00	2.52	4.53	6.08	10.00	15.00	30.00
	Small enterprises	366,708	4.70	3.87	0.00	0.99	1.25	2.51	4.00	5.58	8.22	10.02	30.00
	Medium-sized enterprises	83,366	4.88	4.16	0.00	1.00	1.44	2.63	4.07	5.68	8.46	10.72	30.00
Original maturity for companies with negative equity	Functioning companies	290,811	7.81	6.86	0.00	1.00	1.98	3.76	5.52	9.52	24.99	25.02	30.00
	Microenterprises	269,263	7.98	6.95	0.00	1.00	1.99	3.93	5.77	9.89	25.02	25.02	30.00
	Small enterprises	18,454	5.80	5.34	0.00	0.99	1.42	2.96	4.68	6.33	10.03	18.07	30.00
	Medium-sized enterprises	3,094	5.08	4.63	0.08	0.98	1.07	2.43	4.17	5.99	8.96	13.24	30.00

Table 4 (continued)
Descriptive statistics of the continuous explanatory variables

Name of variable	Examined population	Number of observations	Average	Standard deviation	Minimum	5th percentile	10th percentile	25th percentile	Median	75th percentile	90th percentile	95th percentile	Maximum
Potential coverage of loans	Functioning companies	2,165,166	2.67	6.89	0.00	0.00	0.05	0.36	0.94	1.98	4.67	9.37	50.00
	Microenterprises	1,693,544	2.46	6.67	0.00	0.00	0.03	0.30	0.87	1.82	4.19	8.23	50.00
	Small enterprises	385,162	3.17	6.96	0.00	0.10	0.22	0.58	1.23	2.61	6.16	12.04	50.00
	Medium-sized enterprises	86,460	4.40	9.87	0.00	0.10	0.21	0.59	1.26	2.84	8.83	24.62	50.00
Change in fixed assets	Functioning companies	2,165,166	0.08	0.48	–0.50	–0.18	–0.12	–0.04	0.00	0.06	0.27	0.54	5.00
	Microenterprises	1,693,544	0.09	0.52	–0.50	–0.20	–0.13	–0.05	0.00	0.06	0.30	0.62	5.00
	Small enterprises	385,162	0.05	0.29	–0.50	–0.10	–0.07	–0.03	0.00	0.06	0.19	0.33	5.00
	Medium-sized enterprises	86,460	0.05	0.28	–0.50	–0.08	–0.05	–0.02	0.00	0.05	0.14	0.25	5.00
Export ratio (relative to sales revenue)	Functioning companies	2,165,166	0.04	0.16	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.33	1.00
	Microenterprises	1,693,544	0.03	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.15	1.00
	Small enterprises	385,162	0.08	0.20	0.00	0.00	0.00	0.00	0.00	0.03	0.27	0.59	1.00
	Medium-sized enterprises	86,460	0.22	0.32	0.00	0.00	0.00	0.00	0.03	0.34	0.85	0.96	1.00
Proportion of FX loans	Functioning companies	2,165,166	0.24	0.38	0.00	0.00	0.00	0.00	0.00	0.44	1.00	1.00	1.00
	Microenterprises	1,693,544	0.24	0.39	0.00	0.00	0.00	0.00	0.00	0.41	1.00	1.00	1.00
	Small enterprises	385,162	0.25	0.36	0.00	0.00	0.00	0.00	0.00	0.45	0.98	1.00	1.00
	Medium-sized enterprises	86,460	0.32	0.39	0.00	0.00	0.00	0.00	0.09	0.68	1.00	1.00	1.00

Table 5
Distribution of the categorical variables in the sample of functioning companies

	Typical corporate size			Total
	Microenterprises	Small enterprises	Medium-sized enterprises	
Ownership				
Domestic ownership of minimum 50 per cent	1,637,994	357,480	66,436	2,061,910
Foreign ownership over 50 per cent	55,550	27,682	20,024	103,256
Equity				
Positive or zero	1,424,281	366,708	83,366	1,874,355
Negative	269,263	18,454	3,094	290,811
Actual corporate size				
Microenterprises	1,625,551	36,568	617	1,662,736
Small enterprises	66,459	337,896	7,900	412,255
Medium-sized enterprises	1,427	10,643	76,007	88,077
Large enterprises	107	55	1,936	2,098
NACE category				
A - Agriculture, forestry and fishing	69,876	27,730	8,228	105,834
C - Manufacturing	167,540	90,462	32,596	290,598
F - Construction	194,933	49,090	5,728	249,751
G - Wholesale and retail trade; repair of motor vehicles and motorcycles	475,048	108,746	20,880	604,674
H - Transportation and storage	102,917	23,861	3,995	130,773
I - Accommodation and food service activities	55,517	17,323	2,386	75,226
J - Information and communication	62,982	10,112	1,849	74,943
L - Real estate activities	142,498	12,348	1,881	156,727
M - Professional, scientific and technical activities	172,481	17,435	2,223	192,139
N - Administrative and support service activities	77,528	16,979	4,927	99,434
P - Education	18,323	1,539	122	19,984
Q - Human health and social work activities	111,125	4,597	854	116,576
R - Arts, entertainment and recreation	19,625	1,562	288	21,475
S - Other service activities	23,151	3,378	503	27,032
Region where the company is registered				
Budapest	467,850	105,229	26,517	599,596
Western Transdanubia	163,339	37,563	8,605	209,507
Central Transdanubia	168,240	38,351	8,014	214,605
Southern Transdanubia	146,800	32,443	6,448	185,691
Central Hungary	238,424	47,008	9,209	294,641
Northern Hungary	140,416	30,306	6,301	177,023
Northern Great Plain	183,998	44,655	11,028	239,681
Southern Great Plain	184,477	49,607	10,338	244,422
Age of company				
1 year	73,220	7,496	1,276	81,992
2 years	119,097	12,969	2,035	134,101
3 years	128,508	15,131	2,313	145,952
4 years	128,015	16,406	2,512	146,933
5 years	120,852	17,370	2,565	140,787
6 years	111,153	17,719	2,848	131,720
7 years	102,351	18,141	3,079	123,571
8 years	94,915	18,534	3,335	116,784
9 years	88,978	18,505	3,444	110,927
10 years	87,671	20,116	3,840	111,627
11 years	85,424	21,219	4,182	110,825
12 years	80,836	21,442	4,328	106,606
13 years	75,384	21,060	4,450	100,894
14 years	71,207	21,152	4,824	97,183
15 years	66,396	21,157	5,125	92,678
16 years	63,457	22,315	5,811	91,583
17 years	55,185	21,560	5,925	82,670
18 years	43,592	18,633	5,575	67,800
19 years	33,443	15,502	4,908	53,853
20 years	24,970	12,621	4,260	41,851
21 years	17,276	9,600	3,406	30,282
22 years	11,378	6,828	2,465	20,671
23 years	5,542	3,873	1,434	10,849
24 years	2,164	1,620	637	4,421
25 years and above	2,530	4,193	1,883	8,606
Total	**1,693,544**	**385,162**	**86,460**	**2,165,166**

Table 6
Estimates for functioning companies and for the whole database – Coefficients of selected company-specific variables

	Functioning companies	Whole database
Leverage for companies with positive equity (t-1)	0.56 ***	–0.14 **
Square of leverage for companies with positive equity (t-1)	0.91 ***	1.31 ***
Sales revenue growth rate (t-1)		
micro (truncated)	–1.04 ***	–1.30 ***
micro, square (truncated)	–0.09 ***	–0.42 ***
micro, dummy above truncation point	0.06 ***	0.09 ***
small	–1.75 ***	–2.18 ***
small, square	–0.27 ***	–1.19 ***
medium-sized	–1.92 ***	–2.71 ***
medium-sized, square	–1.02 ***	–1.80 ***
After-tax ROA (t-1)	–0.82 ***	–1.02 ***
Square of after-tax ROA (t-1)	–0.58 ***	0.24 ***
Change in after-tax ROA (t-1)		
micro	–0.31 ***	–0.22 ***
micro, square	–0.19 ***	–0.13 ***
small	–1.21 ***	–1.18 ***
small, square	–0.65 ***	–0.57 ***
medium-sized	–3.47 ***	–3.31 ***
medium-sized, square	–2.27 ***	–1.88 ***
Majority foreign ownership (t-1)	–0.26 ***	–0.25 ***
Original maturity for companies with positive equity (t-1)	–0.03 ***	–0.07 ***
Original maturity for companies with negative equity (t-1)	–0.07 ***	–0.10 ***
Liquidity indicator (t-1)		
micro	0.37 ***	0.35 ***
micro, square	0.28 ***	0.11 ***
small	–0.47 ***	–0.33 ***
small, square	–0.27 ***	–0.49 ***
medium-sized	–1.38 ***	–1.06 ***
medium-sized, square	–0.68 ***	–1.16 ***
Potential coverage of loans (t-1)		
micro	–0.10 ***	0.01
micro, square	0.00	–0.03 ***
small	–0.18 ***	–0.05 ***
small, square	0.01	–0.02 ***
medium-sized	–0.07	0.03
medium-sized, square	–0.03 **	–0.06 ***
Negative equity (t-1)	1.69 ***	1.13 ***
Export ratio (relative to sales revenue) (t-1)		
micro, dummy for values under 15 per cent	–0.12 ***	–0.10 ***
micro, dummy for values of 15 per cent and above	0.09 ***	0.14 ***
small	–0.78 ***	–0.14
small, square	0.83 ***	0.21
medium-sized	–1.56 ***	–0.39
medium-sized, square	1.24 ***	0.20
Proportion of FX loans (t-1)	3.38 ***	3.86 ***
Square of proportion of FX loans (t-1)	–3.09 ***	–3.53 ***

*Note: The table contains the estimated coefficients of the model. The ***, **, and * next to the coefficients mean that the coefficients differ from zero at a confidence level of 99, 95 and 90 per cent, respectively.*

Table 7
Estimates for functioning companies and for the whole database – Coefficients of further variables

	Functioning companies	Whole database
Actual corporate size (Reference: microenterprises)		
small enterprises	–0.29 ***	–0.33 ***
medium-sized enterprises	–0.83 ***	–0.84 ***
large enterprises	–1.15 ***	–1.17 ***
no submitted tax return or zero headcount and sales revenue	–	0.62 ***
Region (Reference: Budapest)		
Western Transdanubia	–0.04 ***	–0.06 ***
Central Transdanubia	–0.02	–0.04 ***
Southern Transdanubia	0.13 ***	0.05 ***
Central Hungary	0.02 **	0.03 ***
Northern Hungary	0.15 ***	0.08 ***
Northern Great Plain	0.06 ***	–0.02 *
Southern Great Plain	0.09 ***	0.06 ***
Age of company (year)		
1	0.21 ***	0.40 ***
2	0.31 ***	0.47 ***
3	0.30 ***	0.44 ***
4	0.28 ***	0.40 ***
5	0.22 ***	0.31 ***
6	0.17 ***	0.24 ***
7	0.14 ***	0.16 ***
8	0.07 ***	0.06 ***
9	0.01	–0.02
10	Reference	Reference
11	–0.03	–0.04 **
12	–0.05 **	–0.05 ***
13	–0.08 ***	–0.06 ***
14	–0.12 ***	–0.07 ***
15	–0.16 ***	–0.13 ***
16	–0.17 ***	–0.15 ***
17	–0.20 ***	–0.17 ***
18	–0.26 ***	–0.21 ***
19	–0.16 ***	–0.14 ***
20	–0.23 ***	–0.20 ***
21	–0.16 ***	–0.18 ***
22	–0.20 ***	–0.23 ***
23	–0.14 **	–0.13 ***
24	–0.31 ***	–0.30 ***
25+	–0.35 ***	–0.33 ***
Trend 2007 Q2 - 2010 Q2	–0.16 ***	–0.13 ***
Growth rate of GDP deflator (t-1)	–1.74 ***	–1.91 ***
Growth rate of 3-month BUBOR (t-1)	0.49 ***	0.47 ***
Constant	–3.23 ***	–2.93 ***

Table 8
Estimates of the scoring models for the different company sizes – Coefficients of further variables

	Micro	Small	Medium-sized
Region (Reference: Budapest)			
Western Transdanubia	–0.04 ***	–0.01	–0.10
Central Transdanubia	–0.01	–0.06 *	0.14
Southern Transdanubia	0.12 ***	0.15 ***	–0.06
Central Hungary	0.02 **	0.00	0.18 **
Northern Hungary	0.12 ***	0.27 ***	0.25 ***
Northern Great Plain	0.06 ***	0.00	0.18 **
Southern Great Plain	0.08 ***	0.15 ***	–0.14
Age of company (year)			
1	0.28 ***	0.14 **	
2	0.36 ***	0.17 ***	0.13
3	0.34 ***	0.05	
4	0.32 ***	0.06	
5	0.25 ***	0.10 *	–0.13
6	0.19 ***	0.10 **	
7	0.17 ***	–0.02	
8	0.09 ***	–0.06	Reference
9	0.03	–0.09	
10	Reference	Reference	
11	–0.01	–0.12 **	
12	–0.02	–0.13 **	
13	–0.08 ***	–0.14 **	–0.22 ***
14	–0.13 ***	–0.09 *	
15	–0.17 ***	–0.18 ***	
16	–0.16 ***	–0.24 ***	
17	–0.20 ***	–0.23 ***	
18	–0.28 ***	–0.21 ***	–0.20 ***
19	–0.16 ***	–0.21 ***	
20	–0.24 ***	–0.23 ***	
21	–0.15 ***	–0.21 ***	
22	–0.11 **	–0.51 ***	
23	–0.07	–0.48 ***	0.01
24	–0.22 *	–0.51 ***	
25+	–0.44 ***	–0.23 *	
Trend 2007 Q2 - 2010 Q2	–0.17 ***	–0.15 ***	–0.09 ***
Growth rate of GDP deflator (t-1)	–1.66 ***	–1.52 *	–6.33 ***
Growth rate of 3-month BUBOR (t-1)	0.51 ***	0.45 ***	0.29 ***
Constant	–2.79 ***	–4.50 ***	–5.94 ***

Table 9
Assessment of the temporal stability of scoring models for the different company sizes – AUROC values

			2014 Q3	2014 Q1	2013 Q3	2013 Q1
Micro-enterprises	Fit of model estimated over the whole time period	up to the quarter in heading	74.7%	74.9%	75.3%	75.6%
		after the quarter in heading		69.5%	69.5%	70.1%
	Fit of model estimated over the time period up to the quarter in heading	in-sample		74.9%	75.3%	75.6%
		out-of-sample		69.3%	69.2%	69.7%
		over the whole time period (both in and out-of-sample)		74.7%	74.7%	74.7%
Small enterprises	Fit of model estimated over the whole time period	up to the quarter in heading	79.4%	79.7%	80.0%	80.4%
		after the quarter in heading		74.8%	74.4%	74.9%
	Fit of model estimated over the time period up to the quarter in heading	in-sample		79.7%	80.0%	80.4%
		out-of-sample		74.6%	74.1%	74.4%
		over the whole time period (both in and out-of-sample)		79.4%	79.4%	79.3%
Medium-sized enterprises	Fit of model estimated over the whole time period	up to the quarter in heading	84.4%	84.7%	85.0%	85.0%
		after the quarter in heading		77.1%	78.7%	81.4%
	Fit of model estimated over the time period up to the quarter in heading	in-sample		84.7%	85.0%	85.1%
		out-of-sample		76.5%	77.9%	80.4%
		over the whole time period (both in and out-of-sample)		84.4%	84.4%	84.3%

Table 10
Equations of the models estimated by economic sectors – Coefficients and goodness-of-fit indicators

	Construction	Trade	Manufacturing	Agriculture
Leverage for companies with positive equity (t-1)	1.03 ***	–0.45 ***	1.57 ***	1.84 ***
Square of leverage for companies with positive equity (t-1)	0.49 ***	1.92 ***	0.37 **	–0.66
Sales revenue growth rate (t-1)				
micro (truncated)	–0.89 ***	–1.27 ***	–1.13 ***	–0.66 ***
micro, square (truncated)		–0.18 ***	–0.21 **	0.60 ***
micro, dummy above truncation point	–0.02	0.10 ***	0.12 ***	0.20 ***
small	–1.14 ***	–2.36 ***	–1.74 ***	–1.13 ***
small, square	0.61 ***	–0.84 ***		
medium-sized	–1.25 ***	–1.49 ***	–2.46 ***	–2.73 ***
medium-sized, square	–1.72 **		–1.69 ***	
After-tax ROA (t-1)	–0.31 ***	–0.82 ***	–1.36 ***	–1.28 ***
Square of after-tax ROA (t-1)		–0.60 ***	–1.23 ***	–1.23 ***
Change in after-tax ROA (t-1)				
micro	–0.43 ***	–0.58 ***		–0.39 **
micro, square	–0.29 ***	–0.26 ***		–0.30 **
small	–1.78 ***	–2.04 ***	–1.19 ***	
small, square	–1.10 ***	–0.95 ***	–0.82 ***	
medium-sized	–5.59 ***	–5.30 ***	–4.08 ***	–6.33 ***
medium-sized, square	–4.00 ***	–3.42 ***	–2.48 ***	
Majority foreign ownership (t-1)	–0.16 **	–0.53 ***	–0.37 ***	
Original maturity for companies with positive equity (t-1)	–0.04 ***	–0.02 ***	–0.02 ***	0.01 ***
Original maturity for companies with negative equity (t-1)	–0.09 ***	–0.08 ***	–0.08 ***	–0.06 ***
Liquidity indicator (t-1)				
micro	0.33 ***	0.48 ***	0.44 ***	0.42 ***
micro, square	0.22 ***	0.33 ***	0.37 ***	0.21 **
small	–0.68 ***	–0.42 ***	–0.49 ***	–1.20 ***
small, square	0.40 ***	–0.52 ***		–1.24 ***
medium-sized	–1.75 ***	–2.29 ***	–2.09 ***	–2.05 ***
medium-sized, square	–1.17 **	–1.95 ***	–1.33 ***	
Potential coverage of loans (t-1)				
micro	–0.16 ***	–0.08 ***	–0.07 ***	–0.16 ***
micro, square				0.01 **
small	–0.19 ***	–0.20 ***	–0.11 ***	
small, square				
medium-sized		–0.34 ***	–0.15 ***	
medium-sized, square				
Negative equity (t-1)	1.72 ***	1.73 ***	2.16 ***	2.00 ***
Export ratio (relative to sales revenue) (t-1)				
micro, dummy for values under 15 per cent		–0.20 ***	–0.10 ***	
micro, dummy for values of 15 per cent and above		0.15 ***	0.15 ***	
small			–1.41 ***	–1.42 **
small, square			1.61 ***	1.87 **
medium-sized			–1.32 ***	–7.65 ***
medium-sized, square			0.77 *	8.19 **
Proportion of FX loans (t-1)	3.63 ***	3.56 ***	3.24 ***	2.86 ***
Square of proportion of FX loans (t-1)	–3.20 ***	–3.32 ***	–3.08 ***	–2.64 ***

Table 10 (continued)
Equations of the models estimated by economic sectors – Coefficients and goodness-of-fit indicators

	Construction	Trade	Manufacturing	Agriculture
Actual corporate size (Reference: microenterprises)				
small enterprises	–0.35 ***	–0.36 ***	–0.28 ***	–0.34 ***
medium-sized enterprises	–0.86 ***	–1.09 ***	–0.82 ***	–0.94 ***
large enterprises	–0.46	–1.67 ***	–1.55 ***	
Region (Reference: Budapest)				
Western Transdanubia	–0.14 ***	–0.01	–0.12 ***	–0.46 ***
Central Transdanubia	–0.09 ***	–0.02	–0.06 *	–0.36 ***
Southern Transdanubia	0.00	0.19 ***	0.03	–0.39 ***
Central Hungary	–0.01	0.02	–0.14 ***	–0.21 **
Northern Hungary	0.06 *	0.19 ***	0.18 ***	–0.25 ***
Northern Great Plain	0.08 ***	0.01	–0.05	–0.47 ***
Southern Great Plain	0.11 ***	0.10 ***	–0.04	–0.51 ***
Age of company (year)				
1	0.57 ***	0.08 *	0.13 *	
2	0.44 ***	0.26 ***	0.29 ***	0.19 ***
3	0.43 ***	0.18 ***	0.26 ***	
4	0.35 ***	0.18 ***	0.27 ***	
5	0.27 ***	0.16 ***	0.27 ***	0.13 ***
6	0.22 ***	0.09 **	0.17 ***	
7	0.23 ***	0.13 ***	0.04	
8	0.12 **	0.06 *	–0.01	0.05
9	0.02	–0.03	0.00	
10	Reference	Reference	Reference	
11	0.05	–0.05	–0.11 *	
12	0.00	–0.07 *	–0.06	Reference
13	0.01	–0.12 ***	–0.06	
14	–0.04	–0.16 ***	–0.11 *	
15	–0.14 **	–0.20 ***	–0.25 ***	
16	–0.25 ***	–0.25 ***	–0.17 ***	
17	–0.21 ***	–0.23 ***	–0.18 ***	–0.13 ***
18	–0.33 ***	–0.33 ***	–0.31 ***	
19	–0.25 ***	–0.19 ***	–0.11 *	
20	–0.38 ***			
21	–0.36 ***			
22	–0.59 ***			
23	–0.54 ***	–0.17 ***	–0.07	–0.21 ***
24	–0.50 *			
25+	–0.59 **			
Trend 2007 Q2 - 2010 Q2	–0.18 ***	–0.15 ***	–0.16 ***	–0.16 ***
Growth rate of construction sector's production volume (t-2)	–0.47 **			
Growth rate of manufacturing sector's implicit price index (t-2)			–0.63 ***	
Growth rate of GDP deflator (t-1)		–1.64 ***		
Growth rate of 3-month BUBOR (t-1)	0.47 ***	0.52 ***	0.49 ***	0.46 ***
Constant	–3.28 ***	–3.31 ***	–3.65 ***	–3.53 ***

Table 10 (continued) Equations of the models estimated by economic sectors – Coefficients and goodness-of-fit indicators				
	Construction	**Trade**	**Manufacturing**	**Agriculture**
Number of observations	249,751	604,674	290,598	105,786
Pseudo-R^2 with trend and macro variables	10.6%	11.9%	12.6%	10.7%
AUROC	75.0%	76.8%	77.7%	76.5%
Cutoff	6.6%	5.3%	5.0%	4.0%
Sensitivity	71.1%	71.0%	71.7%	72.2%
Classification accuracy	65.8%	68.8%	69.3%	67.2%

9.4 ADDITIONAL CHARTS

Charts 2–9 and 11–20 in the Appendix show the impact of the given variable on the logarithm of the odds (log-odds) ceteris paribus.

Chart 2
Leverage for companies with positive equity – Functioning companies

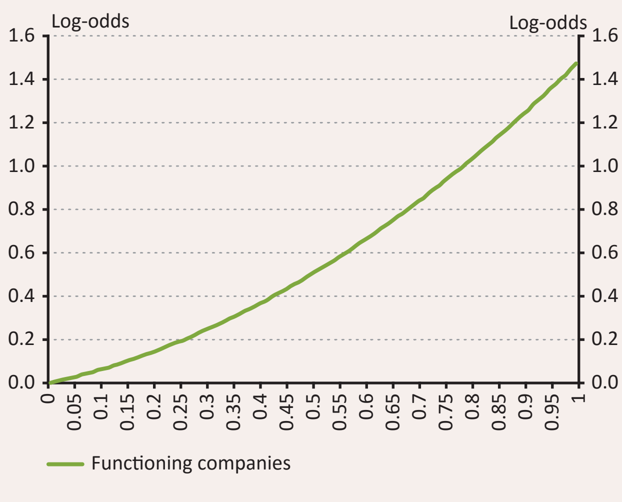


Chart 3
Growth rate of the sales revenue – Functioning companies

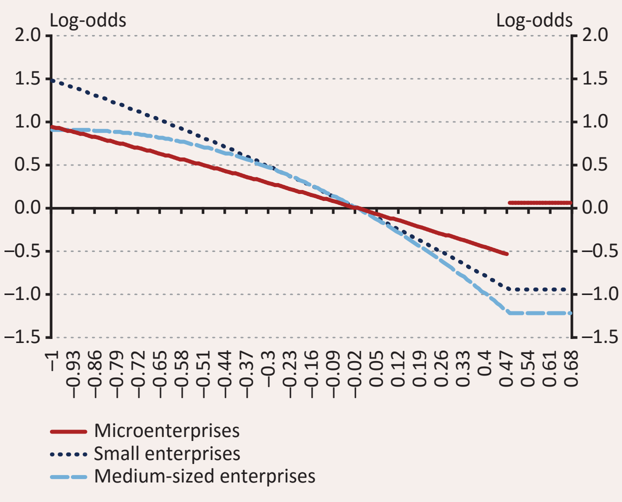


Chart 4
After-tax ROA – Functioning companies

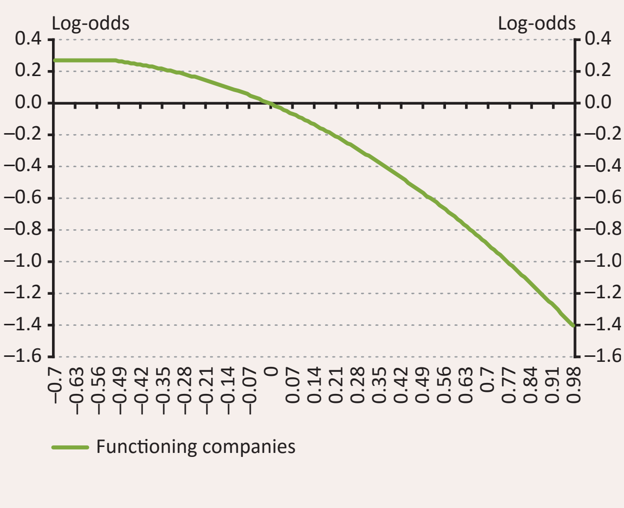


Chart 5
Change in the after-tax ROA – Functioning companies

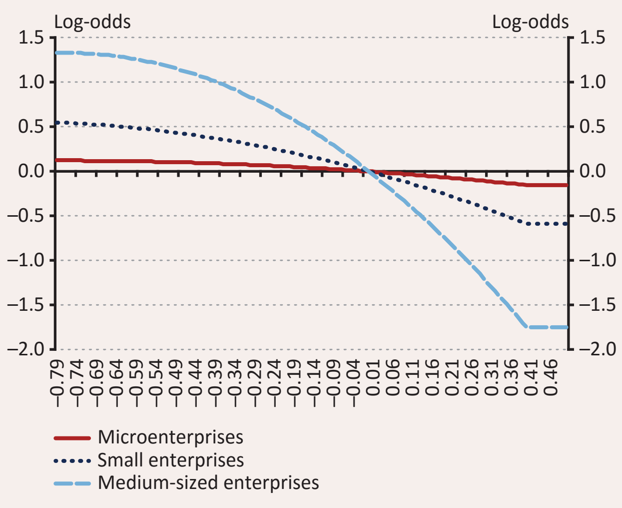


Chart 6
Liquidity indicator – Functioning companies

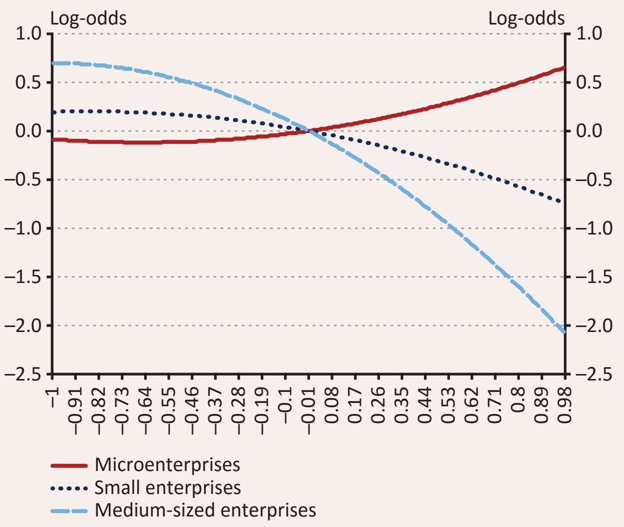


Chart 7
Potential coverage of loans – Functioning companies

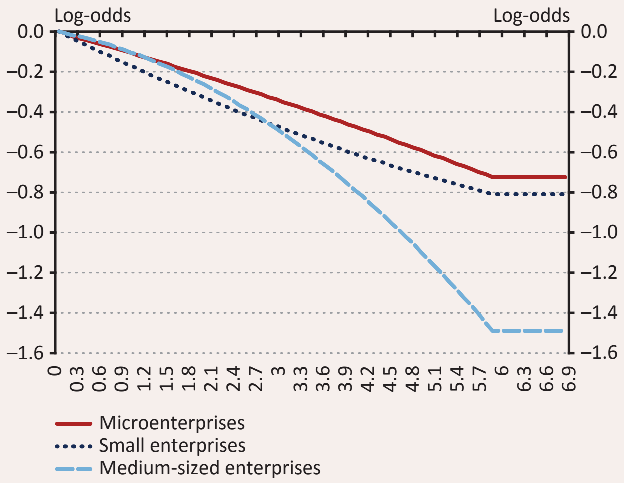


Chart 8
Export ratio (relative to sales revenue) – Functioning companies

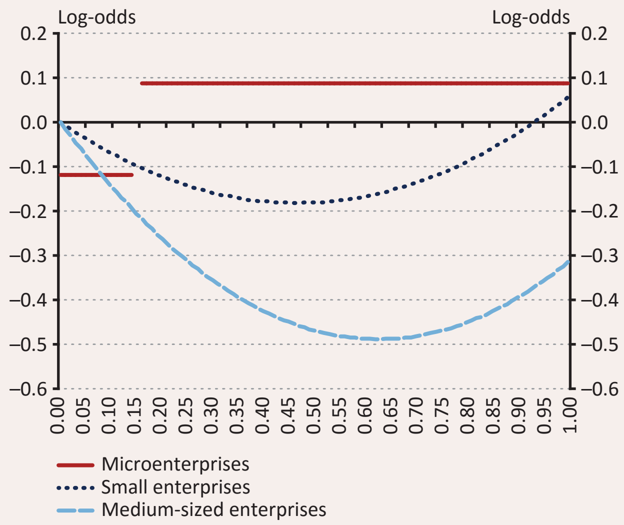


Chart 9
Proportion of FX loans – Functioning companies

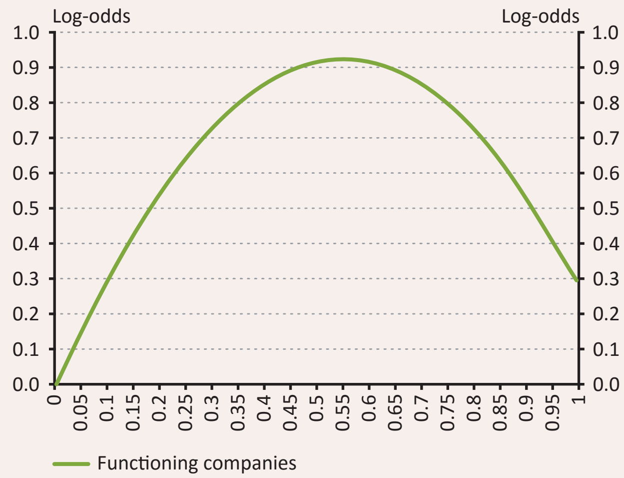


Chart 10
Rate of correct rejection, rate of correct acceptance and classification accuracy as a function of the cutoff – Functioning companies

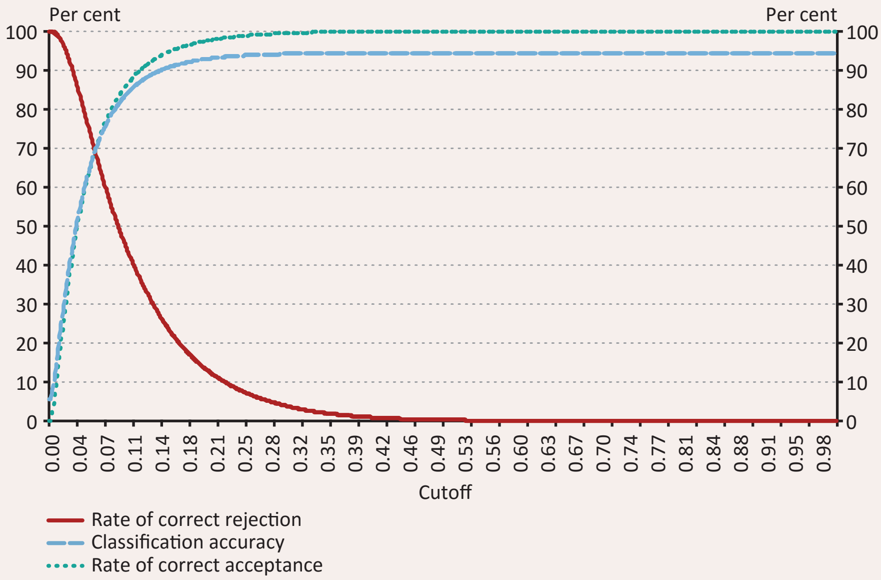


Chart 11
Liquidity indicator – Whole database

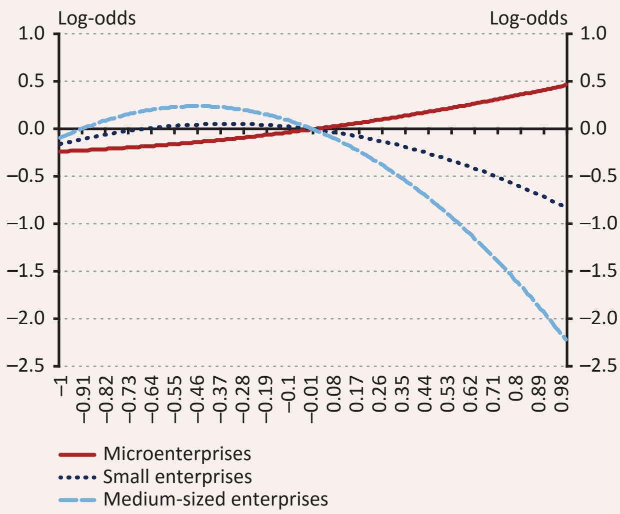


Chart 12
Export ratio (relative to sales revenue) – Whole database

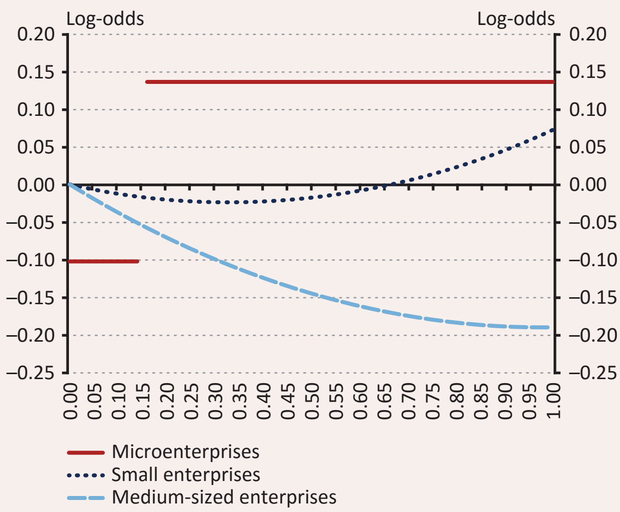


Chart 13
Leverage for companies with positive equity – Models differentiated by company size

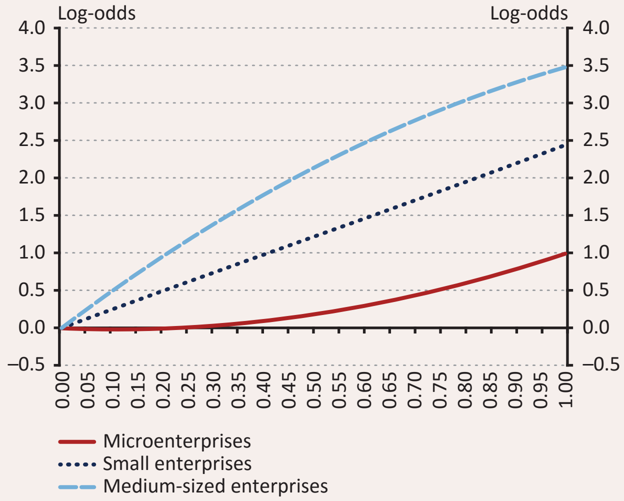


Chart 14
Growth rate of the sales revenue – Models differentiated by company size

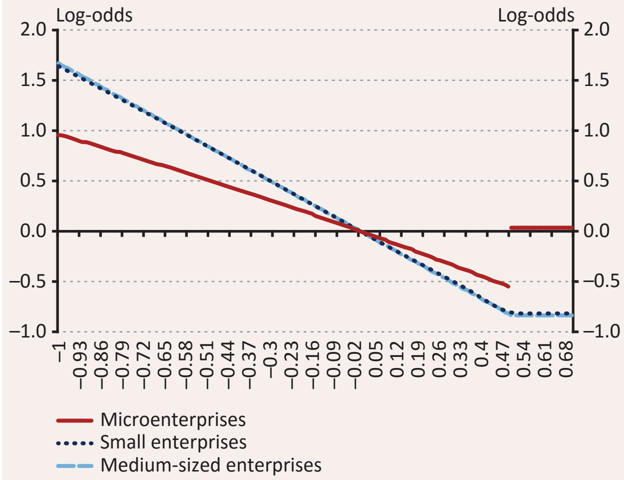


Chart 15
After-tax ROA – Models differentiated by company size

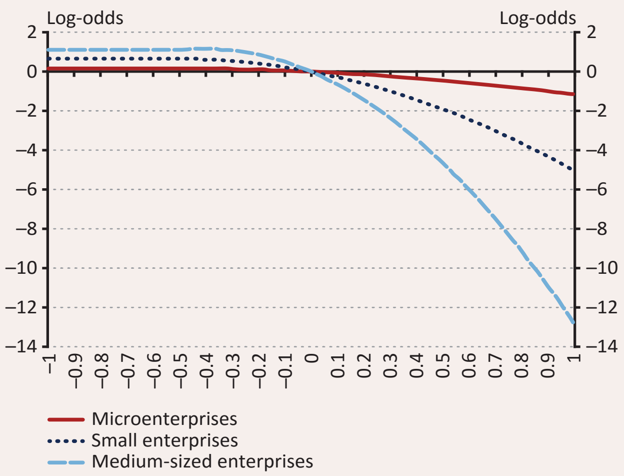


Chart 16
Change in the after-tax ROA – Models differentiated by company size

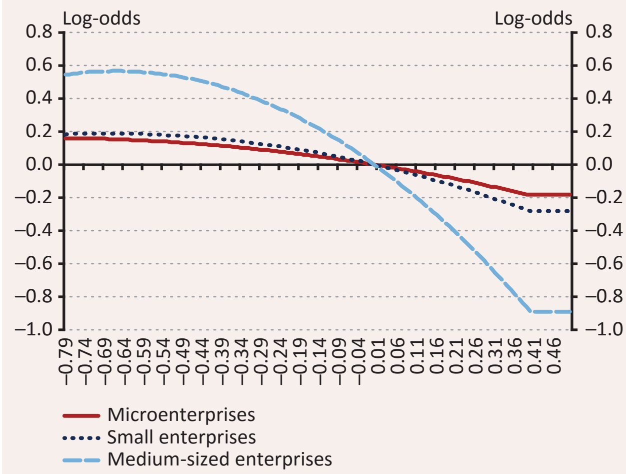


Chart 17
Potential coverage of loans – Models differentiated by company size

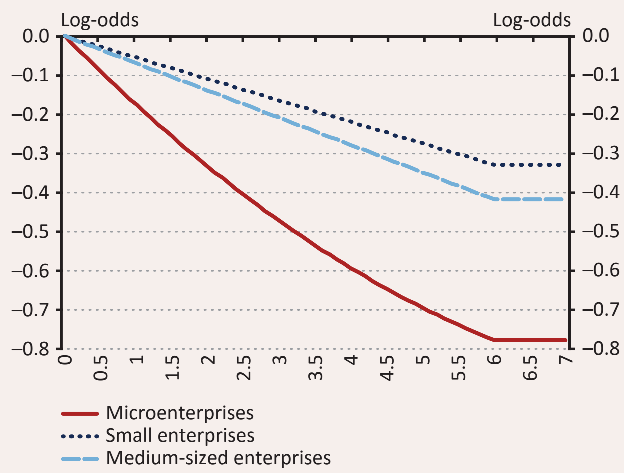


Chart 18
Change in fixed assets – Models differentiated by company size

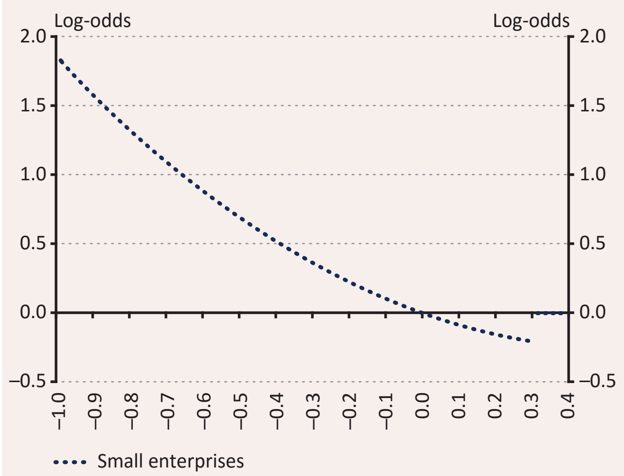


Chart 19
Export ratio (relative to sales revenue) – Models differentiated by company size

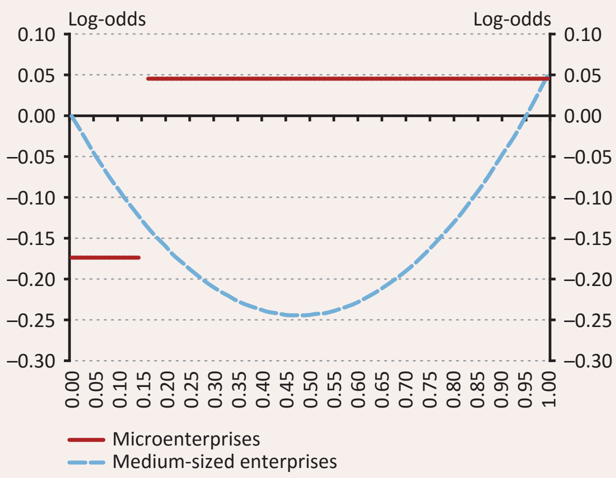


Chart 20
Proportion of FX loans – Models differentiated by company size

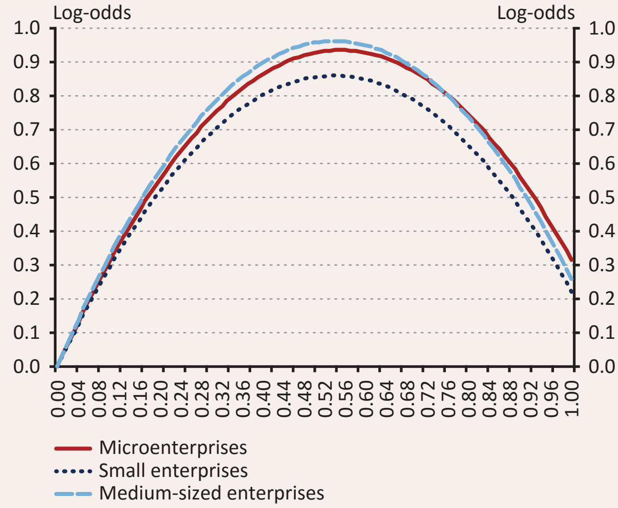


Chart 21
The distribution of estimated PDs and their fit to the default rate – Microenterprises

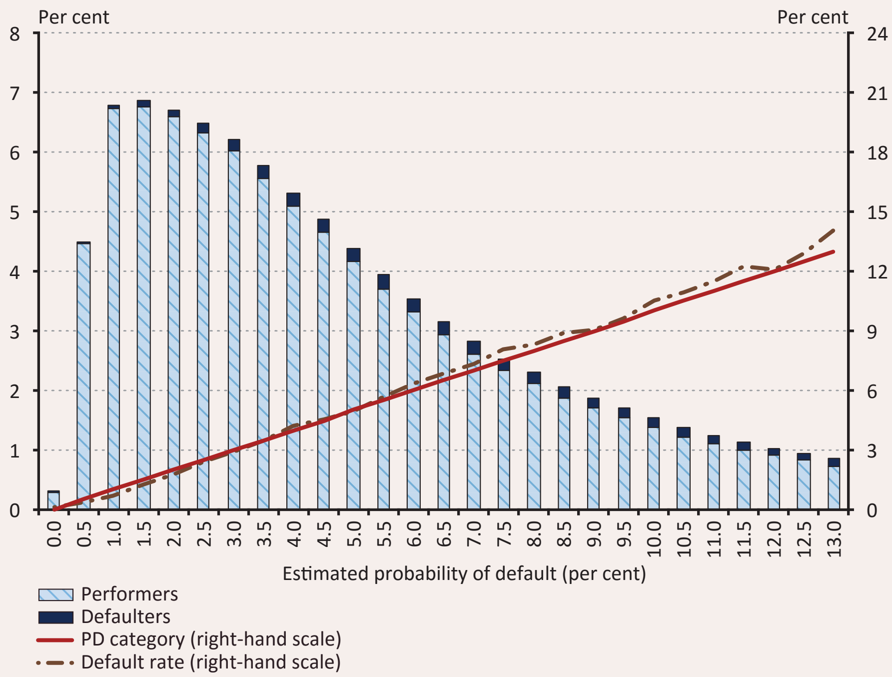


Chart 22
The distribution of estimated PDs and their fit to the default rate – Small enterprises

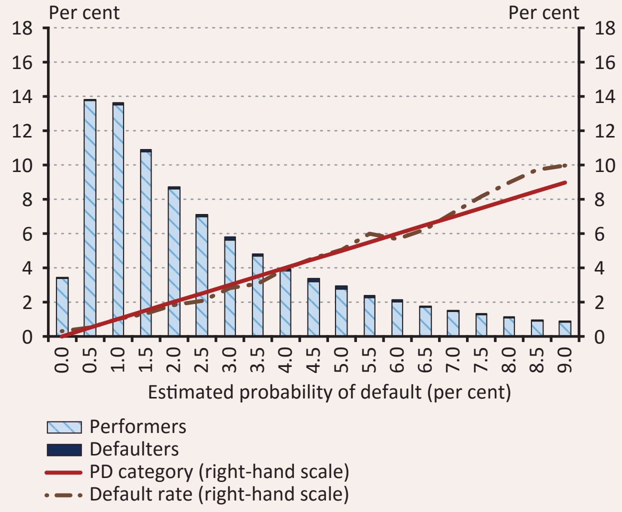
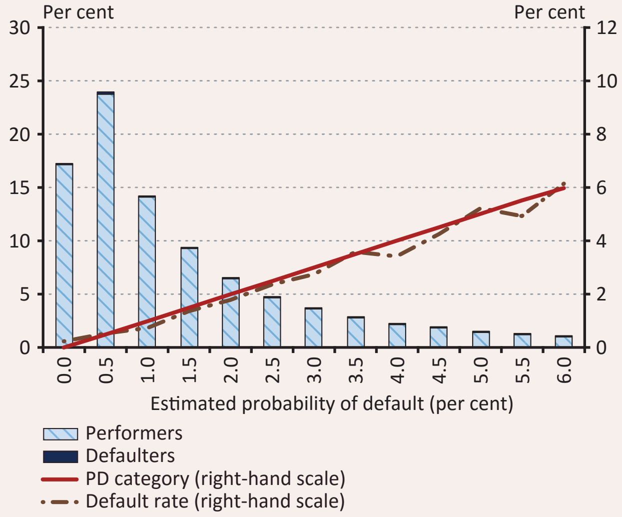


Chart 23
The distribution of estimated PDs and their fit to the default rate – Medium-sized enterprises



MNB OCCASIONAL PAPERS 123

MODELLING THE CREDIT RISK OF THE HUNGARIAN SME SECTOR

December 2016

Print: Prospektus–SPL consortium

6 Tartu u., Veszprém H-8200