Péter Bauer, Marianna Endrész

Modelling bankruptcy using Hungarian firm-level data

MNB Occasional Papers 122

2016

Péter Bauer, Marianna Endrész

Modelling bankruptcy using Hungarian firm-level data

MNB Occasional Papers 122

2016

MNB Occasional Papers 122

Modelling bankruptcy using Hungarian firm-level data*

(Csődök modellezése magyar vállalati-szintű adatok segítségével)

Written by Péter Bauer, Marianna Endrész

Budapest, April 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)

*We are grateful for comments and suggestions received from Tamás Tóth.

Contents

Abstract 5

Összefoglaló 5

1 Introduction 7

2 Data 9

2.1 Data used in the estimation 9

2.2 Identifying bankrupt firm-observations 9

3 Estimation strategy 12

4 Estimation results 15

4.1 Baseline model and the impact of macro variables 15

4.2 Robustness 17

4.3 Testing non-linearity and heterogeneity 18

4.4 Out-of-sample forecast 22

5 Some applications of the results 24

6 Conclusions 27

7 Appendix 28

8 References 35

Abstract

The ultimate aim of this paper is to generate micro-level risk measures, which can provide a useful input for further research. To this end, this paper estimates bankruptcy probabilities for Hungarian firms using probit estimation. The estimated models show reasonable performance in distinguishing surviving and failing firms. We combine macro and micro information, as the addition of macro variables is needed to capture the aggregate dynamics and level of risk, especially during the crisis period. Controlling for the non-linear impact of firm characteristics and allowing heterogeneity by firm size improves the model's performance significantly. The distributional characteristics of the micro-level risk indicators provide some interesting insights regarding the development of risk dispersion and the risk-taking of the banking sector.

JEL codes: C23, G33

Keywords: bankruptcy risk modelling, probit, micro data

Összefoglaló

A tanulmány végső célja mikro-szintű kockázati mutatók előállítása, amelyeket későbbi kutatások felhasználhatnak. Ennek céljából a tanulmány csőd valószínűségeket becsül magyar vállalatokra probit modell segítségével. A becsült modellek jelentős mértékben képesek a túlélő és a később csődöt jelentő cégek megkülönböztetésére. A mikro adatokat makro információkkal egészítjük ki, mivel ez utóbbiak szükségesek az aggregált dinamika és a kockázat szintjének megragadására, különösen a válság időszak során. A modell teljesítményét szignifikánsan növeli, ha figyelembe vesszük a vállalati jellemzők nemlineáris hatását, valamint a vállalat méret szerinti heterogenitást. A mikroszintű kockázati mutatók eloszlásának momentumai segítségével érdekes felismerésekre juthatunk a kockázat szóródásának alakulásával és a bank szektor kockázatvállalásával kapcsolatban.

1 Introduction

This paper estimates a bankruptcy model using detailed Hungarian firm level data, covering a relatively long, recent period (1996-2014) and using an administrative dataset with an almost full coverage of double book-keeping firms. The ultimate aim is to generate micro-level risk measures, which can provide a useful input for further research and facilitate the calculation of indicators capturing the heterogeneity and distributional characteristics of risk. Firm-level risk estimates can be used for example to monitor the risk-taking of the banking sector, the impact of certain policy measures or to test the risk-taking channel of monetary policy. At the same time, the distributional characteristics of risk can enhance macro analysis by providing information on its higher moments. Recently, plenty of evidence has been gathered on the importance of heterogeneity by firm types and higher moments in driving macro outcomes.

Bankruptcy modelling goes back to the seminal work of Altman (1968), who developed a Z-score model using discriminant analysis. Since then, the literature has been evolved in terms of methodology, type of information used (accounting or market, micro and/or macro) and the event modelled (default or bankruptcy). Regarding the econometric method, the logit or probit model is used for example in Ohlson (1980), Lennox (1999) and Ooghe (2002); hazard models are employed in Shumway (1999), Chava and Jarrow (2004) and Bharath and Shumway (2008) to recognise the age dependence of default and to improve on the unbiasedness inherent in static estimation. Hamerle et al. (2004), Carling et al. (2007), Bonfim (2007) and Jacobson et al. (2011) highlight the importance of combining micro and macro data. They find that the model which uses only firm-specific characteristics is outperformed by the model, which conditions also on the macro environment. Most of the papers model default events, only a few model bankruptcy (Bernhardsen (2001), Bernhardsen and Larsen (2007) Giorani et al. (2011)). This could be due to the fact, that interested stakeholder – banks, investors or central banks guarding financial stability – are primarily worry about defaults. Default is conceptually different from bankruptcy: default can be strategic, it signals payment problems from which many firm can recover while bankruptcy is final and irrevocable, default is affected by the type and terms of the loan contract as well etc. Nevertheless, modelling bankruptcy could have many advantages, because of the wider coverage of firms and the longer availability of data, which is often the case; in Hungary, for example, default data is available only from 2007, while bankruptcy data stretches back to the 1990s. The longer dataset better facilitates the investigation of the role of macro effects and its potential change between quiet and crisis periods. Another advantage of our dataset is that, because of its full coverage, we are also able to study smaller firms with potentially tighter liquidity constraints and no access to loan or capital markets.

Early papers using a small sample of Hungarian firms to model bankruptcy were Hajdu and Virág (1996) and Hajdu and Virág (2001). To our knowledge, this is the first attempt to model bankruptcy on Hungarian data which covers the entire population of double book-keeping firms.

In this paper, we estimate the one-year probability of bankruptcy by probit models. The usual indicators of firm-level performance are used (liquidity, leverage, profitability, collateral, sales), which are backward-looking, accounting-type data. The analysis is augmented by adding qualitative information on the firm's size, age, ownership, export status, industry, etc. Although the literature has documented the importance of market data, they are ignored, as they are available only for a very limited number of firms. The estimation is enhanced by adding macro variables such as GDP growth, cost of borrowing and credit growth. The goodness of fit of the models is evaluated by the area under the ROC curve and the out-of-sample predictive power.

In addition to estimating a baseline model, several alternative specifications are examined by changing the definition of bankruptcy, the explanatory variables and the sample period. The potential non-linear impact of certain variables and heterogeneity by firm size are also looked into.

Our baseline model including real and financial performance measures together with qualitative information on the firm yields a model with a good ability to differentiate between failing and surviving firms. Macro variables do not improve the goodness of fit of the model, but they are needed to better capture the aggregate level and dynamics, especially when the crisis period is included. Non-linearity is found for many variables, in the case of collateral and sales growth the impact even changes sign at large values. Signs of heterogeneity and the need to estimate by firm size are detected.

Applying our results for credit register data, we find that the pool of companies which take out new bank loan has average risk similar to that of the entire population, but banks try to lower their tail risk. That was observed also during the crisis. The development of risk dispersion measures point to an interesting feature of crisis periods and recessions, namely that during shocks not just average risk increases, but the risk distribution becomes wider as well.

In the next section, the data – in particular the identification of bankruptcy and related problems – are outlined. Then, the estimation strategy is summarised. Section 3 presents the estimation results, including the alternative models, robustness checks and discussions on non-linearity and heterogeneity. Finally some applications of the estimation results are highlighted and then we present the conclusions.

2 Data

2.1 DATA USED IN THE ESTIMATION

In this paper, we use annual financial reports of all double book-keeping firms between 1996 and 2012. To define bankrupt firm-observations, the database on court proceedings collected by the vendor OPTEN is utilised (1996-2014). The macro variables entered in the estimation are GDP growth, cost of borrowing and credit growth. The average cost of borrowing (on new loans) is calculated by the MNB and available from 2001.

The estimation sample includes only private companies – government and municipality owned firms are excluded. We drop some sectors as well – financial, tobacco, oil and energy. As very small firms could have quite different behaviour and incentives to go/not to go bankrupt, we excluded very low-activity firms. Firms, and all their observations, are dropped if their average yearly sales and total assets calculated at 2013 prices are under HUF 5 million. Many companies do not have any employees or do not report that, which significantly reduces the sample size. At the end, the baseline model is estimated on a sample of 1,585,663 firm-year observations, covering the period 1996-2012. Information in the financial reports is used to capture the financial and real performance of companies (profitability, leverage, liquidity, sales growth, productivity). Other characteristics of firms controlled for are the size, age, whether the company export or foreign owned, and the availability of collateral. The list and calculation of those variables are summarised in Table 4 in the Appendix. Outliers are treated by winsorising the variables. Bounds were chosen to eliminate apparent errors in the data or filtering out unusually low or large values. So lower bounds are usually at 0, while upper bounds are in most cases between the 95th and 99th percentile (see Table 4).

2.2 IDENTIFYING BANKRUPT FIRM-OBSERVATIONS

The Opten dataset includes information on various forms of bankruptcy/closure proceedings instituted against firms.[1] The database contains the timing of court announcements and type of proceeding. Three procedures are used to define bankruptcy:

– restructuring, initiated by the debtor, with the aim to ensure the going concern of the firm by reaching an agreement between the creditors and the firm;

– liquidation due to insolvency, initiated by the creditors;

– the company files for closure due to reasons such as shutting down or re-establishing a firm.[2]

Restructuring is very rare in Hungary, the bulk of the announcements are either liquidation or closure. Our basic definition of bankruptcy includes restructuring and liquidation only, while an alternative definition of bankruptcy considers also closures. The date of announcement is used to identify the time of bankruptcy. As liquidation can be initiated by more than one creditor repeated liquidation filings can occur, which are dropped from the sample: if a call for liquidation is followed by another one within 1 year, only the first announcement is used.

[1] During the estimation period, bankruptcy regulation went through one major change in 2006, which aimed to reshape the incentives and procedures. Other minor alterations of the law were mainly aimed to ensure transparency of procedures, to change seniority of claims (in particular that of lenders with collateral), to ensure funding of direct bankruptcy costs, etc.

[2] In Hungarian: csőd, felszámolás, végelszámolás.

To identify the variable of interest (explained variable) a dummy D is introduced:

– D takes value zero (no bankruptcy) at time t, if the firm has a recorded financial report both at t and t+1 (ongoing firm).

– D takes value 1 at time t (indicating bankrupt firm observation), if the firm submits its last financial report at time t, and there is a bankruptcy proceeding initiated against it within 2 years (at t+1 or t+2).

All the other observations are dropped from the sample.

The "within 2 years" condition calls for some justification. Usually risk measures use a 1-year horizon (default or bankruptcy within 1 year). However the filing of bankruptcy often follows the last financial reporting with a large gap. The Table below shows the distribution of bankruptcy filings according to the time elapsed since the last financial report submitted by the firm. To capture the majority of filings, but at the same time to ensure consistency over time, data is analysed up to 2007 with a 7-year cap on the time gap. The table shows average figures for the period 1996-2007.

Table 1
Distribution of bankruptcies by the time elapsed between the last financial report and the announcement

	1 year	2 years	3 years	4 years	5-7 years	up to 7 years
1996-2007	33.8%	21.1%	14.7%	11.5%	19.0%	100%

Source: Opten and NAV, own calculations. Only bankruptcies taking place up to the 7th year following the last submitted report are considered. Averages over 1996 - 2007. Calculations are based on the merged sample, where public companies and certain sectors are already dropped. However the sample is not identical with the final estimation sample.

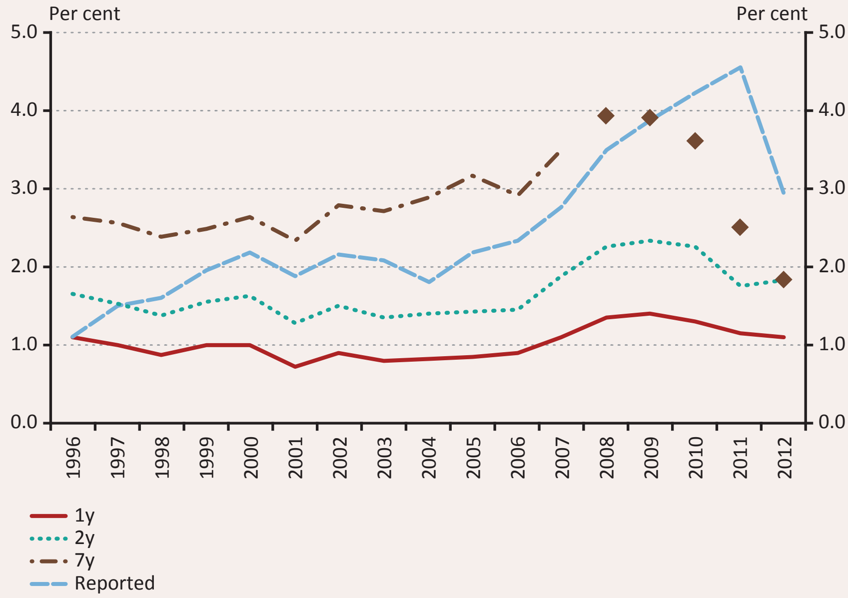
As the data demonstrates, only 34% of bankruptcies are filed within 1 year. The rest of the proceedings stretch out over time. The bankruptcy filings on average take place 2.8 years following the last financial report. To choose the maximum time lag when bankruptcy observations are identified, one needs to compromise: to cover as many bankruptcies as possible, but also to allow estimation for the recent years. Having as long time series as possible is important if one intends to capture the effect of the macro environment and also include the crisis period, which might have an impact on the results. This is why we decided to use the 2-year cap, which allows us to extend the estimation sample up to 2012.

Considering further, more distant filings would have made the interpretation of risks over time and the specification of the model hard as well. For example, taking financial reports from 2005, if we would consider even firms going bankrupt in 2008 or 2009, the aggregate risk measure calculated for 2005 would be very misleading. Observations identified as bankrupt in 2005 are very likely to go bankrupt because of the already poor performance of the firm in 2005. Nevertheless, the macro environment of the following (including the crisis) years also might play a role in the firm ending up bankrupt. That would complicate the model specification and the calculation of ex ante risk measures too; as we will see later, leads of macro variables are needed in order to fit the level of aggregate risk.

As shown by the next chart, the actual bankruptcy rates, using different time gaps show similar dynamics. This is rather reassuring. On the other hand, as our favoured bankruptcy rate uses just about 56% of all bankruptcies, there is a large difference in the level of risk measures (compare the 2y and 7y definition).

The dispersed nature of bankruptcy filings has another implication illustrated in Chart 1: our aggregate bankruptcy rates are different from the "reported" rate based on the time of bankruptcy announcement. The latter is calculated by dividing the number of bankruptcies registered in a year by the total number of firms in the previous year. Then it is lagged, to make it comparable with the other rates, where a firm is qualified bankrupt at the time of its last financial report. The comparison of the two types of bankruptcy rates reveals, that the dispersed nature of bankruptcy filing makes the "reported" bankruptcy rate give a misleading picture on risk, both in terms of level and dynamics (signalling turning points).

Chart 1
Bankruptcy rate using different time gaps



Note: The reported rate is lagged to make it consistent with the calculation of the other bankruptcy rates

To consider further the consequences of our definition of bankruptcy, we run regression analysis and compare mean characteristics for the group of bankrupt firms omitted vs included. This exercise uncovers some differences. Firms which go bankrupt with a larger time lag (omitted from our estimation sample) typically show better financial and real performance, while they tend to be smaller, younger, have less collateral and leverage, and be more likely to be foreign owned. Nevertheless, they still underperform the going-concern firms by all measures. Detailed results are reported in Table 5 and 6 in the Appendix.

3 Estimation strategy

A frequent approach modelling default risk is using discrete choice models (see e.g. Bonfim, 2007 or Jacobson et al., 2011). We follow this approach as we model the probability of bankruptcy by using binary dependent variable panel models. As Shumway (2001) showed, this model class is equivalent to discrete time hazard models. Shumway emphasises that it is important that these models have a time dimension and time-varying explanatory variables as models using only cross sectional data (so-called static models) may give biased results. We use probit model, but we note that the logit model would provide virtually the same results.

As usual, the probit model is constructed as follows. Let

$$y_{it}^{*} = x_{it}\beta + z_t\gamma + \varepsilon_{it}, \tag{1}$$

where

$$y_{it} = \begin{cases} 1: y_{it}^{*} \geq 0 \\ 0: y_{it}^{*} < 0 \end{cases}$$

We observe y but not y*, which is the latent dependent variable. y is 1 if firm *i* is defined as bankrupt at time *t*, and 0 otherwise. The error term, ε_{it} is normally distributed and assumed that it is not correlated with the explanatory variables, i.e. our explanatory variables are exogenous. Moreover, we assume that they are independent between firms and take the strong assumption that they are independent between time periods. As a consequence, we do not use firm-level fixed or random effects in the model.[3] See Jacobson at al. (2011) for example, for a discussion of these assumptions. As a robustness check in one of our specifications, we estimate the model with standard errors robust to clustered error structure (i.e. that the error term may be correlated between time periods for the same firm). We note that the above description of the model is equivalent to writing the probability of bankruptcy as the following:

$$P\left(y_{it} = 1 \middle| x, z\right) = \Phi\left(x_{it}\beta + z_t\gamma\right),$$

where $\Phi(.)$ is the normal c.d.f.

Our baseline specification is the following:

$$y_{it}^{*} = x_{it}\beta_1 + \Delta x_{it}\beta_2 + z_t\gamma_1 + z_{t+1}\gamma_2 + \delta_{st} + \varepsilon_{it}.$$

Where y* is the latent variable. As regards the right hand side variables, x_{it} and Δx_{it} denote firm level variables and their difference $(x_{it}-x_{it-1})$, z_t are macro variables and their leads, and δ_{st} are the sector level dummy variables. As fixed or random effects are not included in the model, the panel structure of the data is used only for generating the difference of the firm-level variables.

We control for several firm characteristics.[4] As a measure of size, the *log of total assets* and *log of number of employees* are used. *Sales growth* measures the ability of the firm to generate revenue. Note, that when

[3] There is another reason why we do not use firm-level fixed effect: using FE automatically omits firms that do not go bankrupt in the sample, as in this case there is no variation in the dependent variable. However, these firms constitute the majority of our sample.

[4] See details on the calculation of each variable in the Appendix.

this variable and its lag are used firms with age of 1 and 2 years are omitted. To measure productivity, *labour productivity* is defined as real value added/employment. We include a number of financial variables: *collateral* is defined as fixed assets/total assets. This variable captures liquidity constraints, by showing the availability of collateral. *Profitability* is defined as (net income + depreciation)/total assets. *Leverage* basically shows the indebtedness of the firm, and is defined as 1-(equity/total assets). As there are several operating (sometimes even profitable) firms with *negative equity*[5] (which in itself implies insolvency), we include a dummy for them. *Liquidity* is defined as liquid assets/short term liabilities. We use a dummy for *foreign* owned firms (with a foreign share above 50 percent). A dummy for *exporting* firms is also included where an exporting firm is defined as a firm with an export revenue exceeding 10 percent of its total revenues. The methodology we use is able to model the duration dependence detected by the empirical literature – i.e. probability of bankruptcy is dependent on the length of time that the firm has not gone bankrupt - only if we explicitly include a variable for firm age. For that reason we always control for *age*.

The deteriorating performance of the firm and the build-up of vulnerabilities, which finally leads to bankruptcy, is likely to be a longer-term process. This is why for all the firm-level variables contemporaneous values should be used together with their *one-year lag.*[6] On the negative side, using lags excludes the first year of our sample as well as firms that are too young to have lags of some firm-level variables. For easier interpretation of the results, we use differences of variables instead of their lags. The two models are analytically equivalent, because

$$x_{it}\beta + x_{it-1}\beta_2 = x_{it}\left(\beta_1 + \beta_2\right) + \Delta x_{it}\left(-\beta_2\right),$$

From the formula it can be seen that when differences are used the coefficient of x_{it} captures directly the overall level effect of the variable. It is also more convenient as standard errors and significance of the level effect are calculated directly in the regression.

As for macro variables, we chose *GDP growth* (in percent, compared to the previous year) and *credit growth*. We use also the *one-year lead* of GDP growth.[7] Macro variables explain the part of the probability of bankruptcy that firm-level variables do not fully capture. In principal, firm-level variables should already include all relevant information from the macro environment. However, there could be a couple of channels through which macro variables affect bankruptcy even if we control for firm-level variables. First, macro variables can capture spill-over effects from other firms (e.g. through supply chains). Second, macro variables can reflect the expectation channel. This way firms can react faster than justified by their own firm-level attributes. Third, as we include the lead of macro variables we capture the effects of shocks that are not yet present in firm-level variables. As our definition of bankruptcy is forward looking, it is a valid approach. Finally, macro variables turn out to be important for fitting the aggregate yearly bankruptcy rate, as we will see in the Results section.

Instead of macro variables, one might use time fixed effects. We do not include time fixed effects in our baseline model, because we aim to calculate ex ante risk measures for firms. Time fixed effects are necessarily ex post values, while for the lead of macro variables we can exploit macroeconomic forecasts, i.e. we can substitute real-time macro forecasts. Moreover out-of-sample exercises cannot be done with time fixed effects.

Finally, we use 2-digit level NACE *sector dummies* to capture sector specific effects, including the differences in average firm characteristics.

For some of the above explanatory variables, we have ex ante expectations about their effect on the probability of bankruptcy. We expect that sales growth, labour productivity, profitability, collateral, and liquidity decrease the probability of bankruptcy, while leverage and negative equity increase it. The effect of firm size is not clear

[5] The large number of observations with negative equity prevents us from using equity as a denominator, although in many cases that would be an alternative.

[6] Except for age and the foreign dummy, which is justified by the fact that age is perfectly correlated with its lag, while ownership changes are rare events.

[7] It is important to emphasise that the parameter estimates for the macro variables capture only their partial effects. The impact of the macroeconomy on firm-level variables is not modelled, and thus our approach cannot be used to conduct macro stress tests.

ex ante. Some of the variables may capture the strength of the liquidity constraints the firm faces. For example, if a firm has enough collateral or a foreign parent, these can help adjustment to liquidity or profitability shocks by ensuring access to bank or group funding. Regarding the macro variables, we expect that higher GDP growth lowers the probability of bankruptcy, while large credit growth could signal a credit boom, the bursting of which can aggravate a crisis or recession.

After the estimation of the baseline model, we make some robustness checks concerning the sample period (namely including or excluding the crisis period) and the definition of some explanatory variables. We also experiment with two alternative dependent variables: (1) considering only bankruptcies announced within 1 year instead of 2 years; (2) including self-liquidation in addition to bankruptcy.

Then, we extend our analysis to discuss possible nonlinearity and heterogeneity. The issues we consider are the role of age and the size of firms, and the non-linearity of all the other firm characteristics.

To assess the performance of each model we use the pseudo R^2 and the area under the ROC curve.[8] The so-called Receiver Operating Characteristics gives a measure of accuracy for models, where there are two states of potential outcomes. In those cases, we face a trade-off: both the share of true hits (bankruptcy is predicted and the firm actually goes bankrupt) and false alarms (bankruptcy is predicted but the firm survives) are increasing with the threshold used to decide on the signal. The ROC curve plots the true hit rate against the false alarm rate by changing the thresholds. The closer the area under the ROC curve is to 1,[9] the better the model is at distinguishing bankrupt and surviving firms. The advantage of this measure is that the model diagnostic becomes independent from threshold selection.

In addition to in-sample prediction ability, we also analyse the out-of-sample performance of the estimated models, in order to avoid over-fitting.

[8] For a discussion on the choice between ROC and CAP analysis in the case of default modelling, see Irwin and Irwin (2012).

[9] More precisely, the area between the ROC curve and the diagonal divided by the area above the diagonal.

4 Estimation results

4.1 BASELINE MODEL AND THE IMPACT OF MACRO VARIABLES

The estimation yields a baseline model with good ability to differentiate between surviving and failing firms (see Table 1). The pseudo-R^2 of the baseline model is 0.233, while the area under the ROC is 0.859. Both suggest a good fit of the model. The goodness-of-fit is markedly high compared to the results of similar papers. For example, in Bonfim (2007), who models default, the best R^2 is 0.062. In Jacobson et al. (2011), the R^2 is higher, above 0.3, but that paper uses a wide definition of default (including various kind of defaults, bankruptcies and other events) and utilises firm-level information that are not available in other studies. Bernhardsen (2001) also uses bankruptcy data. The paper does not report the R^2, but the area under the ROC curve reported is slightly higher than ours, which suggest a similarly high R^2. Apparently, models utilising information on bankruptcy as opposed to default show better goodness of fit. One likely reason for that could be that bankruptcy is more influenced by the past real and financial performance of the firm. At the same time, the decision on default could be strategic, more dependent on the expectation of the firm or on other non-observed characteristics (such as contract details, future sales prospects, etc.).

Basically, all the firm characteristics are significant at the 1 percent level. Regarding the level effects, our ex ante expectations are justified by the results: larger sales growth, labour productivity, profitability, collateral and liquidity decrease the probability of bankruptcy, while firms with large leverage and negative equity are more likely to fail. Moreover, foreign owned firms are less likely to go bankrupt. For size we get mixed results, because firms with larger total assets have a higher probability of bankruptcy, while firms with higher employment have a lower probability of bankruptcy.[10]

We emphasise that our results should not be interpreted as causal effects due to possible omitted variable bias. They are indicative of correlations between the firm-level variables and the probability of bankruptcy.

In the Appendix (Table 11), we show the results for an alternative specification, where lags are used instead of differences. Lagged firm characteristics are almost always significant, suggesting that vulnerabilities build-up over time. Omitting the lags or differences would make the model's performance deteriorate (ROC measure dropping to 0.83).

As regards the effect of macro variables, while GDP growth lowers the probability of bankruptcy, credit growth increases it. When macro variables are added, the fit of the model slightly increases, mainly when the lead value of GDP growth is included (see Table 1). We experimented with adding the cost of borrowing as well, but that does not improve the fit of the model, the ROC remains 0.858.[11] Not surprisingly, we get the best fit when the macro variables are replaced by year dummies (see Table 1). Altogether the improvement attributed to the macro variables is small, which is in contrast to the findings of Bonfim (2007) and Jacobson et al. (2011).

---

[10] Strangely, *ceteris paribus* exporters face a higher probability of bankruptcy. One reason why exporters may have a higher bankruptcy probability is likely to be the different dynamics of bankruptcy proceedings. Among "all" the bankruptcies (occurring within 7 years) 67% are initiated in the first 2 years in the case of exporters. The same figure for the entire population is just 55%. It is not clear why we observe this pattern. It could be due to differences in the behavior of the owners or the creditors. Nevertheless, this difference between exporters and the entire population can explain why we estimate a positive marginal effect for the export indicator variable. Moreover including/omitting the export dummy as an explanatory variable changes the results only slightly. Goodness-of-fit and predicted probabilities remain practically the same.

[11] Nor does this help to improve the fit of the aggregate rate, as discussed later. Since the borrowing cost is available only from 2001, the fit is compared to the baseline estimated over 2001-2012.

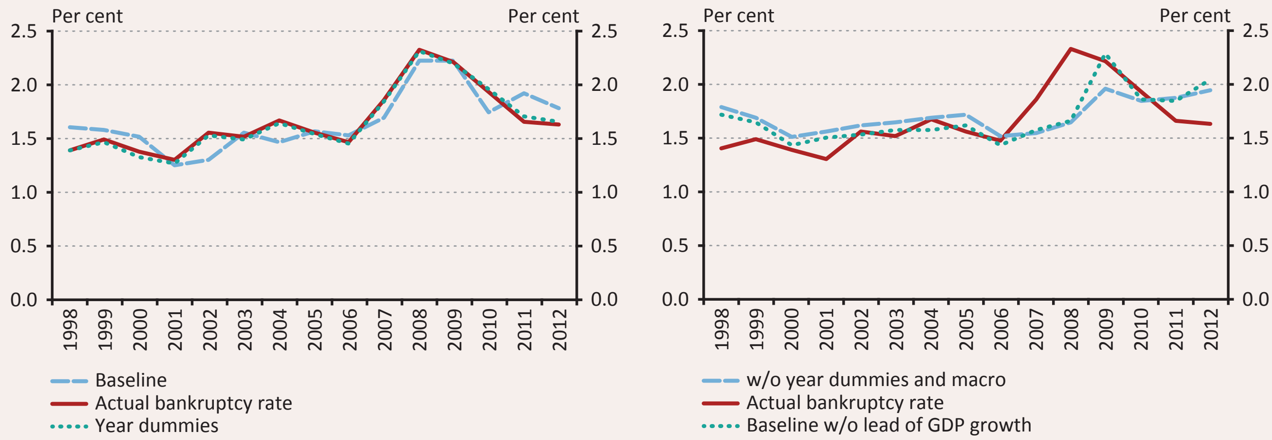
Table 2
Baseline model and the role of macro variables

	w/o year dummies and macro	baseline	w/o lead of GDP	year dummies	baseline from 2001	interest rate from 2001
log_TA	0.1605***	0.1558***	0.1576***	0.1544***	0.1578***	0.1586***
D.log_TA	–0.1832***	–0.1779***	–0.1804***	–0.1763***	–0.1761***	–0.1767***
log_emp	–0.0614***	–0.0536***	–0.0566***	–0.0520***	–0.0560***	–0.0572***
D.log_emp	–0.1471***	–0.1521***	–0.1495***	–0.1543***	–0.1529***	–0.1518***
sales_growth	–0.0781***	–0.0773***	–0.0753***	–0.0755***	–0.0698***	–0.0703***
D.sales_growth	–0.0192***	–0.0189***	–0.0189***	–0.0197***	–0.0228***	–0.0225***
collateral	–0.5215***	–0.5154***	–0.5168***	–0.5189***	–0.5077***	–0.5092***
D.collateral	–0.0626***	–0.0670***	–0.0655***	–0.0651***	–0.0652***	–0.0653***
age	–0.0265***	–0.0283***	–0.0278***	–0.0284***	–0.0299***	–0.0297***
profit1	–0.7477***	–0.7539***	–0.7457***	–0.7552***	–0.7509***	–0.7514***
D.profit1	0.1877***	0.1963***	0.1887***	0.1959***	0.1821***	0.1823***
leverage	0.0861***	0.0843***	0.0851***	0.0847***	0.0834***	0.0838***
D.leverage	–0.0021	0.0000	–0.0011	–0.0002	0.0001	–0.0002
neg_equity	0.3150***	0.3141***	0.3135***	0.3159***	0.2937***	0.2943***
D.neg_equity	0.1636***	0.1612***	0.1631***	0.1583***	0.1728***	0.1723***
foreign_dummy	–0.4189***	–0.4128***	–0.4161***	–0.4110***	–0.4116***	–0.4116***
liquidity	–0.0726***	–0.0726***	–0.0728***	–0.0728***	–0.0723***	–0.0723***
D.liquidity	0.0266***	0.0264***	0.0266***	0.0266***	0.0274***	0.0274***
labprod	–0.0410***	–0.0421***	–0.0413***	–0.0424***	–0.0428***	–0.0429***
D.labprod	0.0042***	0.0044***	0.0044***	0.0044***	0.0042***	0.0043***
export_dummy	0.0468***	0.0518***	0.0472***	0.0546***	0.0629***	0.0629***
D.export_dummy	–0.0445***	–0.0508***	–0.0468***	–0.0535***	–0.0527***	–0.0521***
gdp_growth		–0.0110***	–0.0107***		–0.0107***	–0.0110***
fgdp_growth		–0.0160***			–0.0143***	–0.0143***
creditgr		0.0050***			0.0057***	0.0052***
intrate						0.0077***
fintrate						0.0039
sector fixed effect	yes	yes	yes	yes	yes	yes
year fixed effect	no	no	no	yes	no	no
Observations	1,585,914	1,585,914	1,585,914	1,585,914	1,391,274	1,391,023
Pseudo-R2	0.230	0.233	0.231	0.234	0.229	0.229
ROC	0.857	0.859	0.858	0.860	0.858	0.858

Note: D. denotes difference of the variable; fgdp_growth is the lead of gdp_growth, fintrate is the lead of intrate.

Although the impact of macro variables on R^2 or ROC is not so large, the macro variables are needed to fit the level of risk, which we capture by the aggregate yearly bankruptcy rate – see Chart 2. Not surprisingly, the model with time fixed effects allow an almost perfect fit.[12] Otherwise, adding the lead GDP growth causes the largest improvement, which is likely to be attributed to the forward-looking definition of bankruptcy and the occurrence of the crisis. The macro variables are mostly needed during the recent crisis, where the discrepancy between the prediction of the model with firm-specific controls only and the actual bankruptcy rate is the largest.

Chart 2
Actual and estimated aggregate bankruptcy rate



Note: All the calculations are based on the estimation sample. Aggregate bankruptcy rates are calculated as the average of predicted values.

4.2 ROBUSTNESS

Some papers question the assumption that error terms are uncorrelated (see Das et al., 2007). Therefore, the baseline model is re-estimated by using clustered standard errors. The estimation results (not reported) remain practically unchanged.

Next, we discuss the robustness of the estimation results to the estimation period. The model estimated on the pre-crisis period (up to 2006) does a better job distinguishing between bankrupt and surviving firms, and the area under the ROC curve rises to 0.864 (compared to that of the baseline: 0.859). The role of firm-specific variables is virtually unchanged, but that does not apply for the macro variables. GDP growth has a positive contemporaneous impact and a smaller negative with a lead. The impact of credit growth is still positive but much smaller (0.0014 instead of 0.005, see Table 7 in the Appendix). Thus, the model is robust to changes in the estimation period as long as firm-specific effects are concerned, but the macro impact is sensitive to the inclusion of the crisis period.

Two alternative definitions of profitability are also used: after tax profit and operational profit[13] instead of the baseline cash-flow type profit (after tax profit + depreciation). The results are similar, although the fit of the model deteriorates slightly (see Table 7).

[12] Despite the better fit of the model with time FE, we use the model with macro variables as a baseline, as in future applications we intend to calculate ex ante risk measures. When we examine for example the risk-taking behavior of the banking sector, we need the ex ante risk measures, which can be calculated by using real-time GDP forecasts.

[13] During the crisis, firms suffered losses on the net balance of financial expenditures and expenses. That could be one reason why the use of total profit is a better measure of profitability.

Because of the issues related to the definition of our left hand side variable, in the next robustness check we replaced the indicator variable with the one where only bankruptcies with a strict (technical) one-year horizon are considered. While the estimated parameters – both of firm specific and macro variables – are very similar, the performance of the model improves considerably. The area under the ROC curve rises from 0.859 to 0.89, and so does the R^2 (from 0.233 to 0.282).

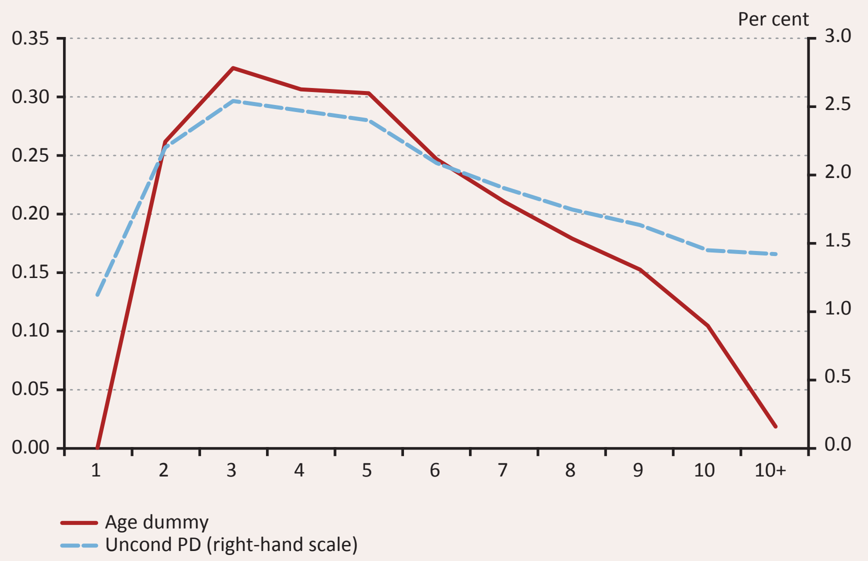
When we extend the definition of bankruptcy by adding self-liquidation/closure, the ability of the model to distinguish bankrupt and surviving firms deteriorates, as the ROC measure drops to 0.818. The role of firm's characteristics also changes, although the direction of impact is the same, the estimated sensitivities vary. This demonstrates that it is indeed better to differentiate self-liquidation from other type of bankruptcy procedures.

4.3 TESTING NON-LINEARITY AND HETEROGENEITY

4.3.1 The role of age

The first non-linearity we consider is that of the impact of age. To include firms of all ages, we modify the baseline model by dropping the term of change in firm characteristics and sales growth and adding age dummies. Indicator variables capturing age[14] produce an inverted U-shape. The unconditional bankruptcy ratio (calculated over the estimation sample) shows very similar pattern.

Chart 3
Bankruptcy by age – unconditional frequency and dummy parameter estimates



Newly funded firms go bankrupt with a rather low probability. The PD keeps increasing up to the age of 3, and then gradually decreases. Our baseline model, omitting 1-year and 2-year old firms, only captures the monotone decreasing part of this age dependence. Therefore, using the discrete variable of age seems sufficient.

[14] Age is winsorised at 11 years. Firms older than 10 years provide the control group.

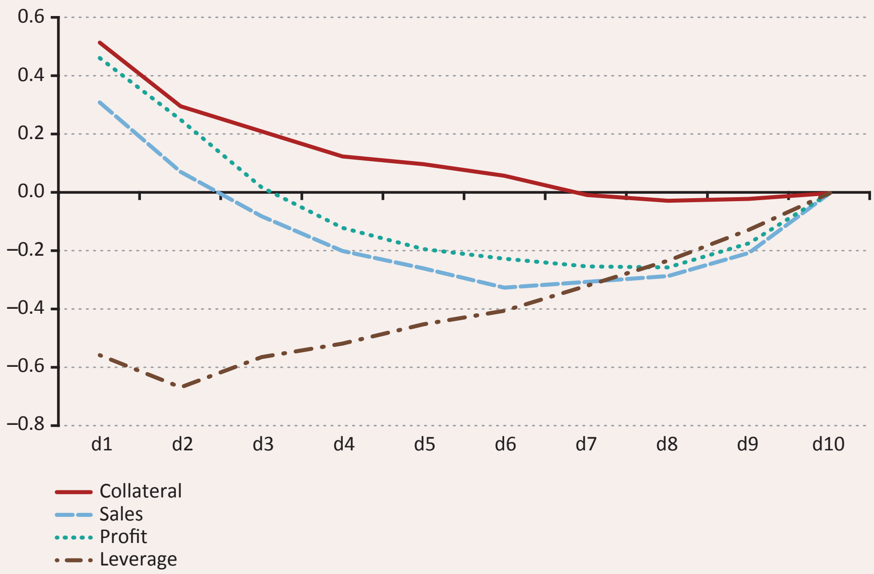
4.3.2 Monotonicity of other firm characteristics

The second issue we consider is the monotonicity of the impact of other firm characteristic, such as leverage, profitability or liquidity. For example, Giordani et al. (2013) fits a non-linear spline model and find threshold effects and sign inversion for leverage – leverage increases risk only above a threshold, but also increases PD below a level.

To test monotonicity, we first employ a simple method, introducing dummies for each decile of the distribution of the variable being considered. We conduct a partial analysis, since we add one set of dummies only at a time.

We found signs of a U or inverted U-shape relationship for some indicators as demonstrated in Chart 4. The increase of profitability and sales growth lowers risk only up to a point, and then becomes indifferent, while for larger values – which might indicate risky strategies – they increase the probability of bankruptcy. PD increases with leverage only from the second deciles, where the mean ratio is about 20%. At the same time, collateral does not seem to matter above a threshold. For the other variables considered (size, liquidity, productivity), the relationship stays largely monotone.[15]

Chart 4
Non-linearity of firm characteristics



Note: Parameter estimates for each decile-indicator. Separate estimation for each variable.

Results slightly change when we move from the simple partial analysis to fully capture the possible non-linearity of variables by adding quadratic (squared) terms to the baseline model (see Table 3). According to the results, all quadratic terms are highly significant while linear terms remain significant as well. The goodness-of-fit of the model improves markedly due to the quadratic terms (ROC goes up to 0.874).

[15] Estimated parameters and mean values of the variable by deciles are reported in the Appendix in Table 8 and Table 9.

Table 3
Quadratic model – estimation results

	baseline	quadratic model	
			sq term
log_TA	0.1558***	0.4241***	–0.0134***
D.log_TA	–0.1779***	–0.3655***	0.0100***
log_emp	–0.0536***	0.0335***	–0.0033**
D.log_emp	–0.1521***	–0.0892***	–0.0032
sales_growth	–0.0773***	–0.5198***	0.2251***
D.sales_growth	–0.0189***	0.1236***	–0.0576***
collateral	–0.5154***	–1.2910***	0.9209***
D.collateral	–0.0670***	–0.8763***	0.9639***
age	–0.0283***	–0.0562***	0.0013***
profit1	–0.7539***	–0.4918***	0.2339***
D.profit1	0.1963***	0.1583***	–0.0244
leverage	0.0843***	0.1825***	–0.0094***
D.leverage	0.0000	0.0336***	–0.0037***
neg_equity	0.3141***	0.2080***	
D.neg_equity	0.1612***	0.1808***	
foreign_dummy	–0.4128***	–0.4139***	
liquidity	–0.0726***	–0.1285***	0.0061***
D.liquidity	0.0264***	0.0566***	–0.0031***
labprod	–0.0421***	–0.0248***	0.0001
D.labprod	0.0044***	–0.0222***	0.0026***
export_dummy	0.0518***	0.0601***	
D.export_dummy	–0.0508***	–0.0294**	
gdp_growth	–0.0110***	–0.0097***	
fgdp_growth	–0.0160***	–0.0169***	
creditgr	0.0050***	0.0058***	
Observations	1,585,914	1,585,914	
Pseudo-R2	0.233	0.256	
ROC	0.859	0.875	

Note: D. denotes difference of the variable; fgdp_growth is the lead of gdp_growth.

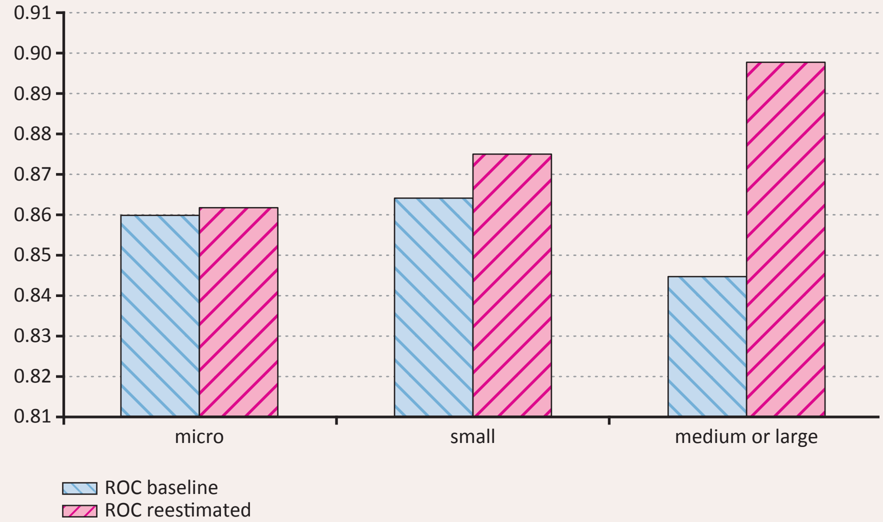
For all the variables the quadratic term has a sign opposite to the linear term, suggesting that the impact of the variable weakens. For example, profitability in general lowers the probability of bankruptcy, but the impact becomes smaller at larger values (the quadratic term has a positive sign). Moreover, for most of the explanatory variables, the effects remain monotonic on the domain. There are two exceptions, where the sign of the impact changes. The effect of collateral has a minimum at 0.65 while the domain is [0, 1], and for more than 10 percent of the firm-observations in the sample the collateral share is higher than that. The other variable is sales growth where the minimum effect is at 1.15, and about 10 percent of the observations are larger than that. Accordingly, in the case of collateral and sales growth we found a change in the sign of the effect on bankruptcy for a relevant part of the domain, or in a less stark interpretation: there is a threshold above which there is no further effect of the increase in these variables. A possible interpretation of the result is that sales growth above a level rather signals increasing risk-taking, while having collateral lowers liquidity constraint only up to a point.[16]

[16] The findings are very similar when the quadratic model is estimated by size groups. Except that (1) for collateral the actual minimum value is much smaller for medium-large firms (0.51 instead of 0.71 and 0.74); while in case of sales growth more micro firms are affected.

4.3.3 Heterogeneity: the role of firm size

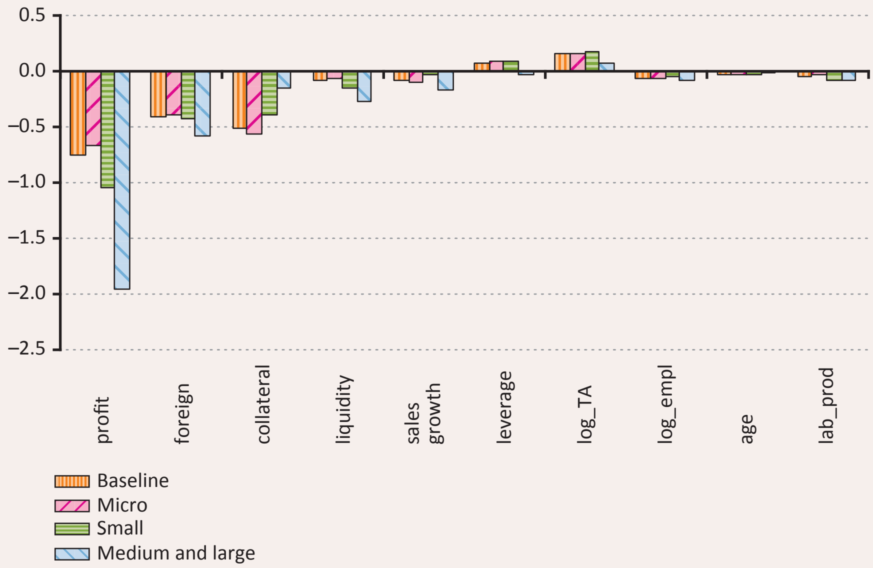
Apparently, the performance of the bankruptcy model varies by the size of the firm. Medium and large firms were put in one group because of the small number of bankruptcy of large firms. The model shows the best performance in the micro and small firm categories and the worst in the group of medium and large firms.

Chart 5
Performance of the baseline model by firm size



To deal with the suggested heterogeneity, we re-estimate the baseline specification for each size group. The second columns in Chart 5 shows that the re-estimated models do a much better job of separating bankrupt and surviving firms, especially in the medium and large category. Not just the performance of the re-estimated models differs, but the estimated parameters vary as well. Detailed estimation results by size can be found in the Appendix in Table 10. To highlight the heterogeneity, some variables were picked and their overall "size" effect is compared in the next chart.

Chart 6
Impact of firm characteristics – size effect



Notes: Value of estimated parameters, when the same specification is estimated for 3 groups with different firm size.

Large variation can be observed, and the differences are not always correlated with size. Size of balance sheet and age seem to have less importance for larger firms, while their financial and real performance (profitability, liquidity or sales growth) seem to be more influential in driving risk. At the same time, having enough collateral matters more for smaller firms. These results suggest the presence of size-related liquidity constraints.

The chart also makes it clear that the parameter estimates of the baseline model are dominated by the large presence of micro firms – micro firms account for 77% of the estimation sample.

According to the results, applying the baseline model for medium and large firms is not recommended.

4.4 OUT-OF-SAMPLE FORECAST

We calculated the pseudo R^2 and ROC to assess the goodness-of-fit of differently specified models. In this section, we examine how the models perform out-of-sample, and whether the improvement achieved by changing the model specification is still observed out-of-sample. For the latter, we focus on the baseline and quadratic model specification.[17] Moreover, we discuss the robustness of the results with respect to the crisis period through the lens of out-of-sample performance.

First, we assess the performance of the baseline model and the quadratic model by calculating the one-year-ahead out-of-sample ROC values for 2007-2012. We estimate the model until 2006, 2007, ..., 2011 and then we predict the probability of bankruptcy for 2007, 2008, ..., 2012, respectively, using the estimated model parameters and actual values for the explanatory variables (as actual values are not available in real-time, this exercise is not a real forecast). Based on the results (see Chart 7), the out-of-sample ROC values do not decrease much compared to the in-sample ones, which suggests that our models do not over-fit the data. Comparing the two models, the quadratic model outperforms the baseline not only in-sample but also out-of-sample.

Chart 7
One-year-ahead out-of-sample ROC values

Next, we extend the out-of-sample window, in order to check the robustness of our models with respect to the crisis period (see Charts 12 and 13 in the Appendix). First, we estimate the baseline and the quadratic model from 1998 until 2004 and calculate the out-of-sample ROC values for the following years (2005-2012). Second, we estimate the same models for 2006 – 2012, then do backward prediction and calculate ROC values

[17] We do not discuss the out-of-sample predictions with firm-size specific models, because the number of actual bankruptcies in the large group is very small (less than 10 in some years).

for 1998,1999,...,2005. The out-of-sample performance is still reasonably good, but the crisis period yields better out-of-sample record. While models estimated during the crisis period seem adequate even for the years before the crisis, the models estimated well before the crisis show worse performance during the crisis years (out-of-sample ROC compared to the in-sample ROC of the baseline model).[18] The average fall in ROC – with respect to the baseline – is 0.5% in the former case and 0.8% in the latter case. Again, these differences (in-sample versus out-of-sample) are smaller than the difference between the baseline and quadratic model, or the variance observed in the performance of the baseline model by years. Apparently, capturing non-linearity is more important than covering a long period in the estimation. Moreover, estimation on a crisis period has better out-of-sample performance for tranquil periods than vice-versa.

[18] This applies not just for the performance by ROC, but also for the ability to capture the level of risk. The model estimated on the crisis period does a better job to fit the aggregate bankruptcy rate in the period of 1998-2004 than the other way around. Moreover, adding macro variables to the model estimated over 1998-2004 results in even worse performance out-of-sample than the model estimated without macro variables. See Chart 13 in the Appendix.

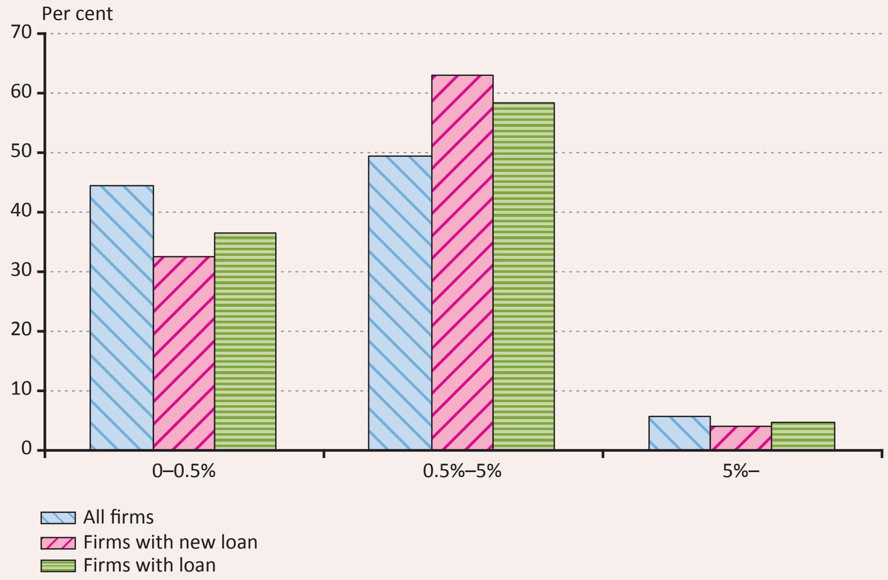
5 Some applications of the results

In this section, we use the probability of bankruptcy estimates to obtain some insight into the development and usefulness of higher moments of risk. We also compare the risk composition of the group of firms with a new bank loan to those of the entire population (the estimation sample). It is an interesting question to what extent banks finance a representative pool of firms in terms of riskiness and how their risk-taking changes through time.

Both the actual figures and our estimated probability of bankruptcy suggest that on average firms with a new bank loan are almost as risky (1.76%[19]) as the entire population of firms (1.84%), while those which have a loan are slightly riskier (1.87%) – however the latter could be just an artifact of bad loans/firms staying in banks' book for many years.

Comparing the risk distribution of the three groups (see Chart 8) two features stand out. Banks extend a loan to medium-risk firms, while the share of low (below 0.5%) and large (above 5%) risk firms is smaller in their portfolio than in the entire population. The former is likely to be driven by demand factors (low risk firms with enough internal sources do not apply for loan or do not want to grow any more), the latter suggests that banks do not undertake large risks. The Chart shows data for 2005, although the same findings apply for all the other years, except that risks are higher in the crisis years.

Chart 8
Risk distribution of firms in the estimation sample and in banks' portfolio (2005)



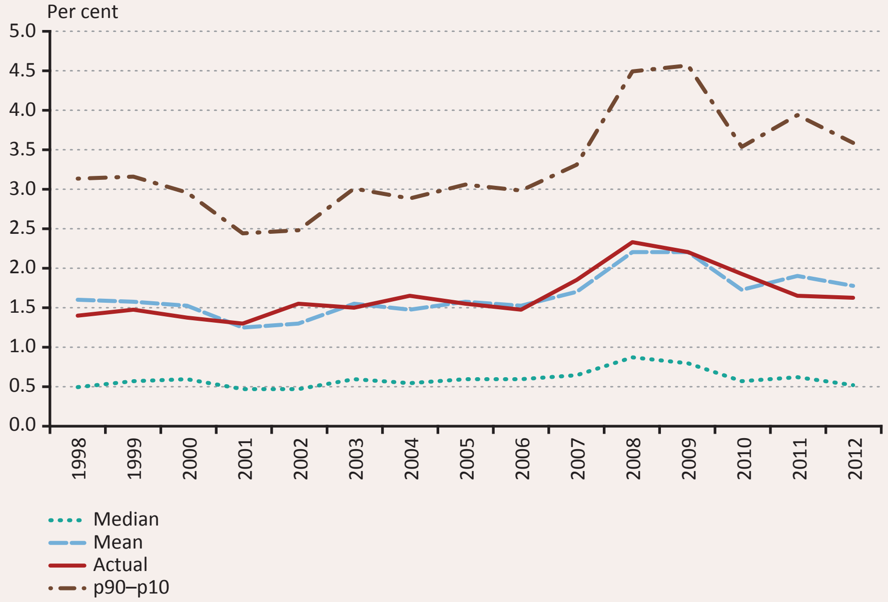
Note: The distributions are defined from the estimated values.

The second issue we examine is the evolution of various distributional features of risk through time. In particular, we analyse the spread of risk in addition to the mean or median values. We use the p90-p10 ratio (the difference between the 90th and 10th percentile) as a measure of risk dispersion, capturing the relative riskiness of low and high quality firms.

[19] Calculated for those in the estimation sample, average over 2005-2011.

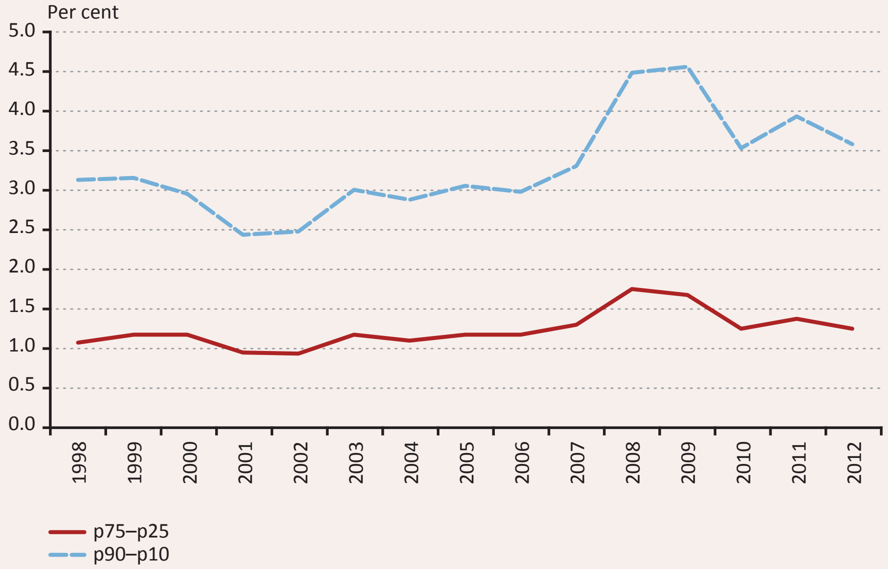
According to the Chart below, the development of average risk (mean and median calculated from the estimated probabilities of bankruptcy) is fairly correlated with the actual bankruptcy rate. They remain at a rather low level in the first half of the sample, increase following the onset of the crisis, and start decreasing again after 2-3 years. The dispersion follows a similar path, indicating that not just the average risk increases, but the dispersion, the differences between high and low risk firms widens as well.

Chart 9
Development of risk measures – actual and estimated



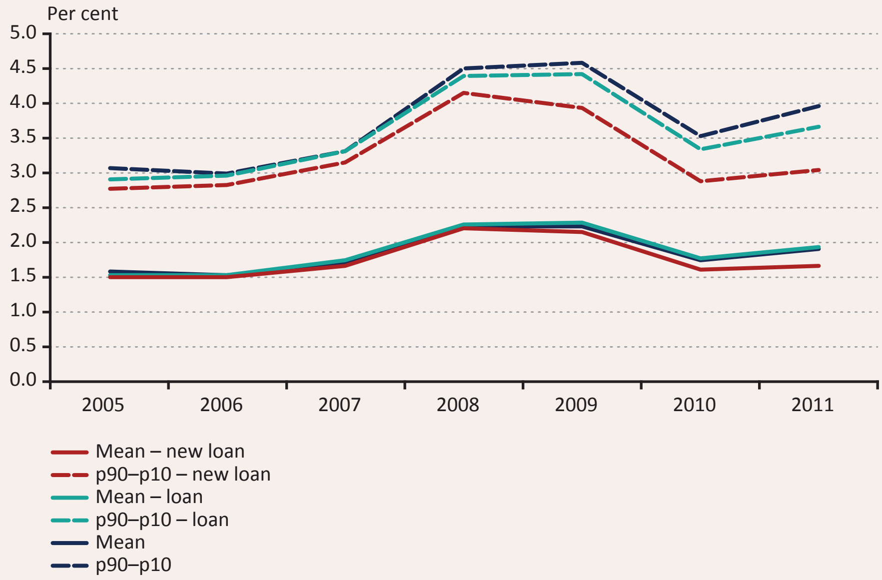
We see a very similar picture when we examine the p75-p25 ratio, instead of the p90-p10 ratio. To summarize, following a shock not just the mean but the dispersion of risk also increases. The measures of dispersion might contain useful information even for macro analysis.

Chart 10
Two dispersion measures



To study the development of the distributional characteristics of firms with a bank loan, we have to restrict the analysis to the years 2005-2011, because credit register data is available only for that period. Average risk measures follow a highly correlated path, although the gap between the mean of new loan takers and the other two groups widens at the end of the sample. As banks try and manage to limit their exposure to high-risk borrowers, the dispersion of all their clients does not widen as much as that of the full population. As for new clients, the dispersion is even more contained during the crisis.

Chart 11
Average risk and risk dispersion of firms with a bank loan



6 Conclusions

In this paper, we modelled the probability of bankruptcy on Hungarian data which covers the entire population of double book-keeping firms during 1996-2012. We estimated the probability of bankruptcy by panel probit model, exploiting both the cross sectional and time dimension of our data. In our baseline model, the usual indicators of firm-level performance are used (liquidity, leverage, profitability, collateral, sales and labour productivity) as explanatory variables. The analysis is augmented by adding information on firm size, age, ownership, export status and industry. We also use macro variables as explanatory variables.

Most of the estimated parameters of our baseline model are highly significant and their sign is intuitive. Sales growth, labour productivity, profitability, collateral and liquidity decrease the probability of bankruptcy, while leverage and negative equity increase the probability. Moreover, foreign-owned firms are less likely to go bankrupt. While GDP growth lowers the probability of bankruptcy, credit growth increases it.

The fit of our baseline model is good compared to similar studies based on both the R^2 (0.233) and the AUROC (0.859). The model also has a good out-of-sample forecasting performance – suggesting that the good in-sample performance is not the result of over-fitting the model on our large dataset. Macro variables do not significantly improve the goodness of fit of the model, but are needed to better capture the aggregate level and dynamics, especially when the crisis period is included.

The robustness of our estimation was analysed along a number of dimensions. The parameter estimates of firm-level variables are very similar if the model is estimated on different time periods, but the calculated effects of macro variables are sensitive to the inclusion/exclusion of the crisis period. The importance of including the crisis period is also highlighted when the out-of-sample performances are compared.

We found some evidence on the non-linear impact of firm characteristics. Adding quadratic terms improves the performance significantly. However, the impact remained monotone for most of the variables with the exception of collateral and sales growth. Indicating that the increase in collateral or sales growth lowers risk only up to a point.

Apparently, the specification of the best model varies by firm size. Since the performance of the baseline model was much worse for medium and large firms, we re-estimated the baseline model for each size-group of firms. After re-estimation, the models' ability to distinguish failing and surviving firms improved significantly.

Finally, we showed some applications of our model, utilising estimated bankruptcy probabilities. First, we compared the risk composition of the group of firms with a bank loan to that of the entire population. Banks' clients come primarily from middle-risk firms, while the proportion of low and high risk firms is lower than in the entire population. Second, firm-level risk measures give the opportunity to analyse not just the mean of risk, but its other moments as well. The dispersion, defined as the difference in riskiness of the best and worst firms, indicates that during the crisis not just average risk increases, but the dispersion of firms widens as well. At the same time, the changes in risk dispersion of firms with a new loan remained more muted, suggesting that banks tried to limit their exposure to high-risk borrowers even during the crisis.

7 Appendix

Table 4
Summary statistics on firm characteristics

	Definition	Total estimation sample		Bankrupt observations		Winsorisation	
		mean	median	mean	median	lower	upper
sales_growth	sales/sales in previous year	14.3%	2.6%	–17.8%	–33.0%	–	3
profit	(net income + depreciation)/total assets	8.4%	8.8%	–29.7%	–14.0%	–1	1
liquidity	liquid assets/short term liabilites	2.16	1.06	0.88	0.43	0	10
leverage	1-(equity/total assets)	0.73	0.60	2.32	1.10	0	10
collateral	fixed assets/total assets	33.2%	26.8%	25.2%	11.9%	0	1
lab_prod	real value added/ employment	3.63	2.64	1.55	0.69	p5	p95
neg_equity	1 if equity<0, 0 otherwise	13.1%	0.0%	55.7%	100.0%	–	–
log_TA	log(total assets)	10.3	10.1	9.9	10.0	–	–
log_empl	log(employment)	1.44	1.10	1.42	1.10	–	–
exporter	1 if export sales/total sales>0.1, 0 otherwise	8.9%	0.0%	8.6%	0.0%	–	–
foreign	1 if foreign ownership exceeds 50 percent of shareholder's equity, 0 otherwise	7.5%	0.0%	5.2%	0.0%	–	–
age	Number of years since birth of the firm	9.8	9.0	8.5	7.0	–	–

Table 5
Parameter estimate for the dummy, indicating bankruptcies over 2 years. Only bankrupt firm-observations. Sector and year fixed effects

negative equity	–0.229***	exporter	–0.0162***	collateral	–0.0132***
sales growth	0.338***	foreign dummy	0.0868***	leverage	–1.249***
profitability	0.157***	total assets	–0.488***		
liquidity	1.550***	age	–1.558***		

Table 6
Bankrupt firms' characteristics
(mean in 2006)

	negative equity	sales growth	profita-bility	liquidity	collateral	leverage	exporter	foreign dummy	log_total assets	age
bankrupt 1y	61.8%	–11.6%	-32.3%	1.50	19.0%	3.15	4.0%	3.8%	8.49	7.47
bankrupt 2y	37.9%	25.1%	-11.5%	2.49	18.1%	1.76	3.9%	4.4%	8.74	5.76
bankrupt over 2y	32.9%	32.5%	-8.1%	3.22	15.7%	1.56	1.9%	10.3%	8.18	5.54
total sample	23.2%	29.0%	4.3%	3.12	27.1%	1.19	3.9%	7.0%	8.80	7.86

Table 7
Robustness checks: estimation results

	baseline	profit2	profit3	until 2006	Dcsod2	Dcsod1_1y
log_TA	0.1558***	0.1644***	0.1633***	0.1375***	0.0866***	0.1677***
D.log_TA	–0.1779***	–0.1777***	–0.1768***	–0.1800***	–0.2117***	–0.2243***
log_emp	–0.0536***	–0.0582***	–0.0565***	–0.0353***	–0.0431***	–0.0435***
D.log_emp	–0.1521***	–0.1512***	–0.1550***	–0.1459***	–0.1725***	–0.1731***
sales_growth	–0.0773***	–0.0778***	–0.0839***	–0.0628***	–0.1010***	–0.1591***
D.sales_growth	–0.0189***	–0.0191***	–0.0159***	–0.0141***	–0.0042	0.005
collateral	–0.5154***	–0.5885***	–0.5753***	–0.6165***	–0.4999***	–0.4339***
D.collateral	–0.0670***	–0.0022	–0.007	–0.0669**	–0.1499***	–0.0559**
age	–0.0283***	–0.0279***	–0.0279***	–0.0219***	–0.0175***	–0.0222***
profit1	–0.7539***			–0.7988***	–0.7395***	–0.7847***
D.profit1	0.1963***			0.2215***	0.2425***	0.2147***
leverage	0.0843***	0.0928***	0.1007***	0.0938***	0.0567***	0.0834***
D.leverage	0	0.0012	0.0064	–0.0047	–0.0104***	–0.0031
neg_equity	0.3141***	0.3145***	0.3442***	0.3576***	0.2828***	0.3681***
D.neg_equity	0.1612***	0.1677***	0.1916***	0.1217***	0.1316***	0.1735***
foreign_dummy	–0.4128***	–0.4088***	–0.4126***	–0.3900***	–0.1911***	–0.4611***
liquidity	–0.0726***	–0.0727***	–0.0727***	–0.0817***	–0.0180***	–0.0848***
D.liquidity	0.0264***	0.0261***	0.0250***	0.0301***	0.0138***	0.0281***
labprod	–0.0421***	–0.0461***	–0.0467***	–0.0435***	–0.0364***	–0.0447***
D.labprod	0.0044***	0.0063***	0.0074***	0.0090***	–0.0028**	0.0012
export_dummy	0.0518***	0.0515***	0.0558***	0.0654***	0.0819***	0.0642***
D.export_dummy	–0.0508***	–0.0516***	–0.0544***	–0.0636***	–0.0796***	–0.0492***
gdp_growth	–0.0110***	–0.0110***	–0.0117***	0.0472***	–0.0043***	–0.0121***
fgdp_growth	–0.0160***	–0.0159***	–0.0162***	–0.0072**	–0.0123***	–0.0153***
creditgr	0.0050***	0.0048***	0.0044***	0.0014**	0.0038***	0.0062***
profit2		–0.6762***				
D.profit2		0.1559***				
profit3			–0.5053***			
D.profit3			0.0766***			
sector fixed effect	yes	yes	yes	yes	yes	yes
year fixed effect	no	no	no	no	no	no
Observations	1,585,914	1,585,914	1,585,914	775,083	1,593,436	1,575,978
Pseudo-R2	0.233	0.23	0.227	0.242	0.185	0.282
ROC	0.859	0.858	0.855	0.864	0.818	0.890

Note: D. denotes difference of the variable; fgdp_growth is the lead of gdp_growth.

Table 8
Average mean characteristics by deciles
(estimation sample of the baseline model)

	log_TA	log_empl	sales_ growth	collateral	profit	liquidity	leverage	lab_prod
d1	7.9		–72%	0%	–48%	7%	7%	–0.6
d2	8.7		–36%	3%	–6%	24%	21%	0.6
d3	9.2	0.0	–19%	8%	2%	45%	33%	1.3
d4	9.5	0.7	–9%	14%	5%	68%	44%	1.8
d5	9.9		–1%	22%	7%	93%	54%	2.4
d6	10.3	1.2	6%	32%	11%	121%	65%	3.0
d7	10.7	1.7	15%	42%	15%	164%	75%	3.8
d8	11.3	2.1	28%	54%	20%	243%	85%	5.0
d9	12.0	2.7	54%	69%	28%	449%	98%	7.4
d10	13.6	4.0	176%	88%	51%	952%	252%	11.5

Table 9
Monotonocity: Estimation results for deciles of firm characteristics – separate estimations

	log_TA	log_empl	sales_ growth	collateral	profit	liquidity	leverage	lab_prod
d1	–0.8525***		0.3144***	0.5152***	0.4610***	0.6142***	–0.5561***	0.5188***
d2	–0.6801***		0.0706***	0.2972***	0.2513***	0.5985***	–0.6682***	0.4195***
d3	–0.5679***	0.1847***	–0.0775***	0.2141***	0.0184	0.5848***	–0.5645***	0.3504***
d4	–0.4626***	0.1120***	–0.2029***	0.1277***	–0.1220***	0.5698***	–0.5209***	0.2637***
d5	–0.3973***		–0.2581***	0.0961***	–0.1937***	0.5621***	–0.4513***	0.2150***
d6	–0.3405***	0.0728***	–0.3271***	0.0586***	–0.2294***	0.5104***	–0.4045***	0.1907***
d7	–0.2751***	0.0916***	–0.3082***	–0.0108	–0.2558***	0.4648***	–0.3161***	0.1666***
d8	–0.1857***	0.0812***	–0.2894***	–0.0277*	–0.2567***	0.3779***	–0.2337***	0.1244***
d9	–0.0988***	0.0714***	–0.2098***	–0.0224	–0.1764***	0.1892***	–0.1270***	0.0852***

Table 10
Heterogeneity: Results of estimation by firm size

	baseline	baseline micro	baseline small	baseline medium and large
log_TA	0.1558***	0.1638***	0.1727***	0.0697***
D.log_TA	–0.1779***	–0.1866***	–0.1925***	–0.0775**
log_emp	–0.0536***	–0.0625***	–0.0365**	–0.0772***
D.log_emp	–0.1521***	–0.1830***	–0.0141	0.1276***
sales_growth	–0.0773***	–0.0987***	–0.0302*	–0.1677***
D.sales_growth	–0.0189***	–0.0084**	–0.0389***	–0.0252
collateral	–0.5154***	–0.5669***	–0.3810***	–0.1425**
D.collateral	–0.0670***	–0.0788***	–0.0451	0.1452
age	–0.0283***	–0.0309***	–0.0215***	–0.0086***
profit1	–0.7539***	–0.6595***	–1.0426***	–1.9544***
D.profit1	0.1963***	0.1900***	0.1800***	0.3802***
leverage	0.0843***	0.0922***	0.0981***	–0.0307
D.leverage	0.0000	–0.0005	–0.0127	0.0421
neg_equity	0.3141***	0.2780***	0.4483***	0.3810***
D.neg_equity	0.1612***	0.1309***	0.1863***	0.1414**
foreign_dummy	–0.4128***	–0.3859***	–0.4154***	–0.5830***
liquidity	–0.0726***	–0.0661***	–0.1448***	–0.2768***
D.liquidity	0.0264***	0.0222***	0.0566***	0.0887***
labprod	–0.0421***	–0.0311***	–0.0771***	–0.0851***
D.labprod	0.0044***	0.0027*	0.0049	–0.0083
export_dummy	0.0518***	0.006	0.1326***	0.1563***
D.export_dummy	–0.0508***	–0.0394**	–0.0443	–0.1754***
gdp_growth	–0.0110***	–0.0122***	–0.0084***	0.003
fgdp_growth	–0.0160***	–0.0153***	–0.0183***	–0.0314***
creditgr	0.0050***	0.0046***	0.0059***	0.0040*
sector fixed effect	yes	yes	yes	yes
Observations	1,585,914	1,224,515	291,918	67,367
Pseudo-R2	0.233	0.232	0.273	0.322
ROC	0.859	0.862	0.875	0.898

Chart 12
ROC values based on models estimated over 1998-2004

Chart 13
ROC values based on models estimated over 2006-2012

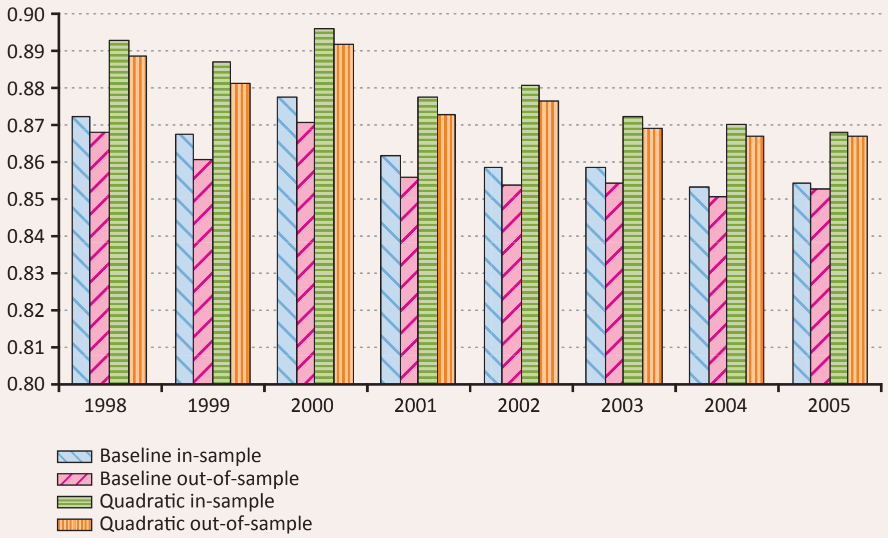


Chart 14
Out of sample aggregate bankruptcy rates

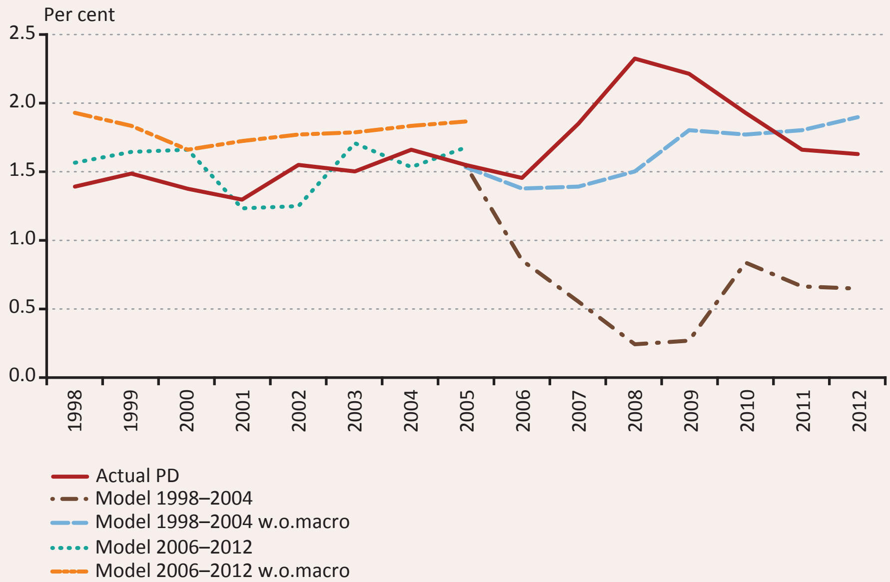


Table 11
Baseline model with differences and lags

	baseline with differences	baseline with lags
log_TA	0.1558***	–0.0220***
D or L.log_TA	–0.1779***	0.1779***
log_emp	–0.0536***	–0.2058***
D or L.log_emp	–0.1521***	0.1521***
sales_growth	–0.0773***	–0.0962***
D or L.sales_growth	–0.0189***	0.0189***
collateral	–0.5154***	–0.5824***
D or L.collateral	–0.0670***	0.0670***
age	–0.0283***	–0.0283***
profit1	–0.7539***	–0.5575***
D or L.profit1	0.1963***	–0.1963***
leverage	0.0843***	0.0843***
D or L.leverage	0.0000	0.0000
neg_equity	0.3141***	0.4753***
D or L.neg_equity	0.1612***	–0.1612***
foreign_dummy	–0.4128***	–0.4128***
liquidity	–0.0726***	–0.0462***
D or L.liquidity	0.0264***	–0.0264***
labprod	–0.0421***	–0.0377***
D or L.labprod	0.0044***	–0.0044***
export_dummy	0.0518***	0.001
D or L.export_dummy	–0.0508***	0.0508***
gdp_growth	–0.0110***	–0.0110***
fgdp_growth	–0.0160***	–0.0160***
creditgr	0.0050***	0.0050***
sector fixed effect	yes	yes
Observations	1,585,914	1,585,914
Pseudo-R2	0.233	0.233
ROC	0.859	0.859

Note: D. denotes difference of the variable; fgdp_growth is the lead of gdp_growth.

8 References

Altman, E. (1968): Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. Journal of Finance.

Bernhardsen (2001): A Model of Bankruptcy Prediction. Norges Bank Working Paper 2001/1.

Bernhardsen, E. and K. Larsen (2007): Modeling credit risk in the enterprise sector – further development of the SEBRA model. Economic Bulletin 3/2007 Vol. 78.

Bharath, S.T. and T. Shumway (2008): Forecasting default with the Merton distance to default model. The Review of Financial Studies, Volume 21, Issue 3.

Bunn, P. and V. Redwood (2003): Company accounts based modeling of business failures and implications for financial stability. Bank of England, Working Paper 210.

Bonfim, D. (2007): Credit risk drivers: Evaluating the contribution of firm level information and macroeconomic dynamics. Banco de Portugal Working Papers 7/2007.

Carling, K., T. Jacobson, J. Linde and K. Roszbach (2007): Corporate credit risk modeling and the macroeconomy. Journal of Banking & Finance 31.

Chava, S. and R.A. Jarow (2004): Bankruptcy prediction with industry effects. Review of Finance Volume 8, Issue 4.

Das, S., D. Duffie, N. Kapadia, and L. Saita (2007): Common failings: How corporate defaults are correlated. Journal of Finance, 62, 93-117.

Giordani, P., T. Jacobson, E. von Schedvin and M. Villani (2013): Taking the twist into account: Predicting firm bankruptcy risk with splines of financial ratios. Sveriges Riskbank Working Paper Series. No. 256.

Hajdu, O. and M. Virág (1996): Pénzügyi mutatószámokon alapuló csődmodell számítások. Bankszemle, XV. évf. 5. sz. pp. 42–53.

Hajdu, O. and M. Virág (2001): A Hungarian Model for Predicting Financial Bankruptcy. Society and Economy in Central and Eastern Europe, Vol. 23. No. 1–2. pp. 28–46.

Hamerle, A. , T. Liebig and Scheule, H. (2004): Forecasting Credit Portfolio Risk. Discussion Paper Series 2: Banking and Financial Supervision No 01/2004.

Hamerle, A., T. Liebig, and D. Rösch (2003): Credit Risk Factor Modeling and the Basel II IRB Approach. Discussion Paper Series 2: Banking and Financial Supervision, No 02/2003.

Irwin, R. J. and T. C. Irwin (2012): Appraising credit ratings: Does the CAP fit better than the ROC? IMF WP/12/122

Jacobson, T., J. Lindé, K. Roszbach (2003): Internal Rating Systems, Implied Credit Risk and the Consistency of Banks' Risk Classification Policies. Sveriges Risksbank Working Paper Series no.168.

Larsen, K. and E. Bernhardsen (2001): Model for analyzing credit risk in the enterprise sector. Norges Bank, Economic Bulletin 2001, Q3.

Ooghe, H. and S. Balcaen (2002): Are failure predicting models transferable from one country to another? An empirical study using Belgian financial statements. Vlerick Working Papers 2002/5.

Syversten (2004): How accurate are credit risk models in their predictions concerning Norwegian enterprises? Norges Bank, Economic Bulletin 2004, Q4.

Shumway, T. (2001): Forecasting bankruptcy more accurately: A simple hazard model. The Journal of Business, Vol. 74, No.1. January 2001.

Tóth, E. (2006): Csődelőrejelzés pénzügyi mutatószámok alapján. BCE Szakdolgozat.

MNB OCCASIONAL PAPERS 122

MODELLING BANKRUPTCY USING HUNGARIAN FIRM-LEVEL DATA

April 2016

Print: Prospektus–SPL consortium

6 Tartu u., Veszprém H-8200