



Network for Studies on Pensions, Aging and Retirement

Netspar THESES

Jan Hessel Veurink

Credit Risk Model for SME Loans in the Netherlands

MSc Thesis 2014-062

Credit risk model for SME loans in the Netherlands

Jan Hessel Veurink
Supervisor: M. Mastrogiacono

August 10, 2014

Abstract: This paper utilizes data obtained for the health check on the SME portfolios of Dutch banks by the DNB to estimate a discrete hazard model to predict the default of a SME loan in the Netherlands. Furthermore, this paper uses the macroeconomic scenario of the EBA stress test of 2014 to obtain the macroeconomic impact on SME defaults. We found that, ceteris paribus, an increase by 1.3% point in the interest rate with a GDP growth of -1.5% will translate in an increase the probability of defaults with 13%. Furthermore, this will lead to increase from 1.13 % to 1.28 % in expected losses expressed in total credit. Moreover, the effect of this macroeconomic scenario on SME is smaller compared with the impact on corporates. Therefore, we conclude that Dutch SME are less dependent on the macroeconomic environment than corporates.

1 Introduction

In the Netherlands small and medium enterprises (SME) determine for a large part the economic landscape. In 2013 SME created 66.3% of all jobs, while 99.8% of all businesses are SME. Furthermore, the Dutch SME employees generate 63.8% of the value added by the private sector, which is relatively high compared to the EU zone average. However, since the financial crisis hit Europe the Dutch SME sector had hard time. From the 2010 till 2013 the Dutch SME sector lost around the 120000 firms¹. As the outlook for the near future for the Dutch SME is still uncertain, the OECD pointed out that the financial health of the Dutch SME sector is one of the key ingredients for economic growth and innovation². Despite the great economic relevance and the enormous concerns about the defaults of many employers, little empirical work exists on the prediction of entrepreneur's defaults in the Netherlands. It is for policy makers of paramount importance to understand how firms' defaults would react to worsening economic conditions. However in the Netherlands this information is hardly available. This is mainly due to a chronic lack of micro data on the financial position and borrowing conditions of Dutch SME. The main aim of this paper is to make an attempt to fill this gap in the literature using of bank provided loan level data. This has a very practical aim. In the practice of stress tests, default multipliers are being used in order to assess the sensitivity of the SME's default rate to worsening economic scenario's. Such multipliers have never been computed empirically for SME's and are proxies based on international evidence. The research question we want to answer is indeed how Dutch SME's would react to worsening economic conditions, and whether the currently used multipliers are empirically defensible.

Under the Internal Rating Based (IRB) approach of Basel II, the Dutch banks are allowed to estimate their own risk parameters to calculate the regulatory capital. In order to gain a comprehensive overview of the performance of the IRB models used for the determining capital requirements for credit risk on loans to SME counterparties and to

¹ See SBA fact sheet 2013 by the European Commission, <http://ec.europa.eu/enterprise/policies/sme/facts->

² See Country report of the Netherlands 2014 by the OECD, <http://www.oecd.org/economy/surveys/economic-survey-netherlands.htm>

benchmark these models between banks under supervision, the De Nederlandsche Bank uses health check surveys. In 2013 De Nederlandsche Bank held the SME IRB Health Check which covered the SME portfolio from a selection of banks, covering more than 80% of SME lending. The data collected consist besides the official risk parameters (probability of default, loss given default, and exposure at default) of an extensive set of loan level data over the period 2005-2012.

Consequently this data has very interesting features. The data consists of many micro loan level determinants of SME default, such as collateral value, interest rates and turnover. In this paper this data set will be utilized to model the probability of default for SME loans. For financial institutions the probability of default is the basic input to quantify, aggregate and manage risk across industries and product lines. Therefore, these models performances are crucial for the risk management of banks. Improved performances could lead to better allocation of capital and more profits for banks. Besides the internal risk management of banks, accurate estimation of the probability of default potentially benefits supervisory oversight of banking organizations to evaluate the credit quality of banks assets. Moreover, during the recent years the outcomes of credit risk models became a key part of evaluating system wide financial risk.

The main reason for this development is the introduction of stress testing. In general, a stress test evaluates the impacts of extreme but plausible economic conditions on the health of financial institutions or the whole financial system. The outcomes of the stress test are used to decrease the likelihood of a financial system failure via altering the capital requirements for banks. This paper simulates the increase in PD's using the current a stress test on the SME portfolio. Our unfavorable macroeconomic scenario will be based on the stress test designed by the European Banking Authority (EBA) in 2014. The main purpose of this stress test is to see how vulnerable the SME loan portfolio is against these macroeconomic effects, focusing only on the stress for the PD.

Besides the importance of the credit models for financial institution and supervisors, credit risk models could mitigate one main concern for entrepreneurs in the Netherlands. Since the crisis of 2007 the lending standards in the Netherlands are sharpened

by banks (Ter Veer and Huiberichts (2013)) and especially for Dutch SME³. Moreover, Dutch entrepreneurs claim that more than 80% of their received credit is from banks⁴. The combination of these observations led to a drop in funding for SME. The OECD shows that the Dutch SME sector had hard time obtaining funds, especially for loans below the 250.000 euros¹. This result is similar to recent findings of the DNB over the last quarter of 2013. This study shows, among others, that while on average the funding of banks to SME dropped with 0.9% while the loans below 250.000 euros dropped with 1.6%. One of the main supply-side reasons for this tightening of lending standards of banks is the raised credit risk perception towards SME in the Netherlands⁵. In general, the risk perception of SME is already higher than for larger firms because of larger information asymmetries between lender and borrower. However, well performing credit risk models could eliminate part of the asymmetric information as the lender obtains more info. Most literature agree that less information asymmetries lead to better allocation of capital and therefore could possibly lead to an higher availability of funds for entrepreneurs. Overall, this implies that there is a large need in the Netherlands for accurately credit risk models for banks, supervisors and entrepreneurs.

The data from the health check is however far from ideal. Missing information on the interest rate for many loans is a main problem. Therefore, we have to implement the missing interest rates with an econometric model, before estimating the probability of default and performing a stress test. As a preview of the results we show that, ceteris paribus, an increase by 1.3% point in the interest rate with -1.5% GDP growth, will translate in an increase the probability of defaults with 13%. Furthermore, this will lead to increase from 1.13% to 1.28% in expected losses in total credit. The remainder of this paper is as follows. First, the review of literature will discuss relevant literature for default determinant, credit risk models and stress testing. Next, the data will be described and the

³ See <http://www.dnb.nl/nieuws/nieuwsoverzicht-en-archieff/dnbulletin-2013/dnb290999.jsp>

⁴ The SME employers organization (MKB Nederland) published this on the website NU.nl, the article can be found on the following URL: <http://www.nu.nl/ondernemen/3730389/banken-dominant-bij-mkb-leningen.html>

⁵ See the statistics of DNB on SME in the Netherlands: <http://www.dnb.nl/nieuws/nieuwsoverzicht-en-archieff/dnbulletin-2014/dnb309784.jsp>

model will be explained and discussed. Finally, the empirical results will interpret the findings and the concluding remarks will finish the paper.

2. Review of Literature

In order to provide an outline of the literature we will start with the two main lines of models: “structural” and “reduced form” models. All structural models descend from the paper of Merton (1974) and are based on the pricing theory of Black and Scholes (1973). Here the credit risk of a company was defining the company’s equity as a call option for the firm’s assets. Conceptually, the firms’ assets follow a certain stochastic process and the debt is honored when the assets value remain above the value of debt over remaining maturity. When the assets drop below the debt value the debt holders receive an amount that’s equals the size of the assets that are left over. Therefore the credit risk is measured as the distance between the assets and the value of debt. Structural risk models use market capitalization data from listed companies to estimate the unobserved assets prices. This has been the main barrier for SME default prediction because most SME are not listed on the stock exchange. Therefore, most of the literature with regards to SME applies reduced forms models, which do not have any theoretical assumptions on why and a default occurs. The survey of Chan-Lau (2006) showed that most utilized determinants for a firm default in reduced form models where credit balance-sheet information, and macroeconomic variables.

Outside of the default predictors surveyed by Chan-Lau (2006), economic theory suggests industrial differences are an important component for estimating credit risk. One of the major attributors to this economic theory is the fire-sales model of Schleifer and Vishny (1992). They suggest that the selling asset prices of firms in financial distress depend on the financial state of their industry peers. While those industry peers are likely to face financial problems themselves, these asset prices drop under the price of best use.

Based on this model Acharya et al. (2007) show that recoveries drop not only due to the financial distress but also because the tighter credit constraints for the peers of the defaulted firms. Lang and Stulz (1992) show with bankruptcy announcements that the contagion effect is more severe for firms with similar structure to the bankrupted firm and in higher concentrated industries. This is because investors learn about the future industry cash flows from the bankruptcy announcement. Besides the information contagion of business failures, Jorion and Zhang (2009) claim that creditors with more exposures towards distressed counterparties are more likely to experience financial problems later on. Helwege (2010) combines counterparty contagion and information contagion and shows that both have a significant impact on the industry peers. This is confirmed by the findings of McCann and McIndoe-Calder (2012) on Irish loans in 2010. Most risk held by Irish loans was concentrated in the speculative sectors such as “Construction”, “Real Estate” and “Hotels & Restaurants” in 2010. For Sloviakian SME loans between 2000 and 2005, Fidrmuc and Hainz (2010) do find that industry effect influences the probability of default. They claim that the industrial heterogeneity depends on the sector-specific capital intensity and therefore may lead to different collateral requirements and external capital demands. Furthermore, Chava and Jarrow (2005) find that even the impact of the balance sheet variables on the probability of bankruptcy varies between industries. Fortunately industry indicators are present in our dataset and will be used to explain the default probability.

The pioneers for utilizing balance-sheet information for firm failure prediction were Beaver (1967) and Altman (1968). While Beaver (1967) used a mean comparison of the predictors of default, Altman (1968) used multivariate discriminant analyses (MDA) to search for a linear relationship of financial ratios that can distinguish between the dichotomous classification of defaulted firms and still living firms. The financial ratios were classified into five different standard categories consisting of profitability, liquidity, solvency, activity and leverage ratios. Subsequently, the combined five financial ratios with the highest prediction power entered the model. In the original paper, these five financial ratios are 1) working capital to total assets ratio, 2) retained earnings to total assets ratio, 3) earnings before interest, taxes and amortization (EBITDA) to total assets ratio, 4)

market value of equity to book value of total debt ratio, and 5) sales to total assets ratio. This method is widely applied by practitioners for measuring credit default risk and referred as Altman's Z-score⁶. Parsimony and its comfortable implementation are the main contributions to the popularity of the Z-score. Although the financial ratios are still used in the recent literature, most scholars apply the logit or probit models introduced by respectively Olson (1980) and Zmijewski (1984) instead of MDA. In our study financial ratios will not be used as the denominators such as total assets and liabilities are not available in the data set therefore we will not be able to use these as explanatory variables.

Two main reasons have led to the use of other estimation techniques than MDA. First of all, MDA has some restrictive assumption on the distribution of the explanatory variables. The variance-covariance matrices should be identical for the defaulted and the non-defaulted group. Furthermore, the predictors should be normally distributed, which limits the use of dummies (Olson (1980)) and financial ratios (Kane et al. (1998)). Besides the statistical requirements, the relative importance of the predictors is not obtained due to the ordinal ranking method MDA applies. Moreover, Lennox (1999) claims that well specified logit and probit models outperforms MDA to identify business failures. In addition, Lennox (1999) claims that probit and logit models perform relatively similar even though there are different probabilities assigned in the tails of the distributions in the logit compared to the normal distribution of the probit. Altman and Sabato (2007) apply a logit on SME bankruptcies in the United States between 1994 and 2002. They claim that practitioners should differentiate between corporates and SME with modeling credit default as their SME specific model outperformed the corporate model on SME bankruptcies. Dietsch and Petey (2004) confirm this by showing that German and French SME are on average riskier than the larger companies in the respective countries.

On the same dataset of German and French SME Dietsch and Petey (2002) find that SME are generally less depend on the "state of the economy" (unobserved systematic variable) than larger firms. On the other hand, Jacobson et al. (2005) conclude that SME are more dependent on the macroeconomic conditions. Although the literature stays

⁶ Edminister (1972), Yazdanfar (2008) and Lugovskaja (2010) apply MDA on respectively United States, Swedisch and Russian SME bankruptcies.

ambiguous on differences of macroeconomic impact between corporates and SME, the literature does support the existence of a business cycle in credit risk. Figleski et al. (2012) affirm that a favorable macroeconomic environment coincides with higher ability to repay loans, lower probabilities of default and a smaller share of non-performing loans. While during an economic downturn ratings deteriorate and loan losses increases. Some scholars claim that the procyclical trend in losses can mainly be explained by the idea that credit risk accumulates during economic upswings due to loosing credit standards of financial institutions. While the materialization of most of the risks only occurs when the economy makes a downturn. Following Udell and Berger (2004) the deterioration in the ability to distinguish good and bad borrowers of loan officers in combination with the lack of supervision of bank management explains the effect of the business cycle on lending standards. Moreover, this might explain why Ter Veer and Huiberichts (2013) observe the tightening the credit standards for banks in the Netherlands. Additionally, Jiminez and Suarina (2005) show that there is negative lagged effect of credit growth on loan losses in Spain after controlling for loan specific variables. Besides credit growth, GDP growth is generally negatively related to credit risk as is shown by Kalirai and Schneider (2002) for Austrian countries, Salas and Saurina (2002) for Spain and Pesola (2001) for the Scandinavian countries. In particular for the Netherlands, Simons and Rolwes (2009) show for that there is a strong negative effect of GDP growth on the probability of default for the whole economy. Moreover, the study performs a stress test on the GDP growth based on the methodology of Wilson (1997a,b)⁷.

Wilson (1997a,b) was the first who explicitly utilized macroeconomic factors to estimate expected and unexpected loan portfolio losses. Basically, this model consists of three steps. Firstly, the probability of default will be explained by a set of macro economic variables. Secondly, the macro economic variables independently will follow an autoregressive model in order to build macroeconomic time series out side of the sample period. Finally, the parameters are estimated and the future path of the macroeconomic variables is simulated. A Monte Carlo simulation can be used to estimate the expected and unexpected losses by assuming that conditional on the macroeconomic variables the

⁷ Originally, the model was known as *Credit Portfolio View* and developed for Mc Kinsey & Co.

defaults are independent of another. Besides Simons and Rolwes (2009), this method is among others used by Ross (2002), Virolainen (2004) for respectively Austrian and Finnish defaults. However, one critique on this method is that the macroeconomic variables are utilized independently to simulate time-series. This means, for instance, that the model treats the unemployment rate and the GDP as unrelated variables. Consequently, scholars applied vector autoregression models (VAR) to take into account the interdependencies among the multiple macroeconomic time series. The advantage of the VAR methodology is that a macroeconomic shock affects the probability of default by two channels: the direct impact and via the impact of other macroeconomic variables. To summarize, the univariate and the VAR methodology are the most commonly used models to simulate time series in stress testing. However, due to the rising importance of stress testing for banks there have been a lot of developments on this area of research. For more details on stress test methodologies, we refer to the survey of Foglia (2009).

In contrast with our paper, most of the literature regarding stress-testing banks is focused on the systemic risk component of default and therefore does not incorporate microeconomic determinants of default. Therefore, this part will focus on the related study of Eklund et al. (2001) on Norwegian loans to enterprises. Similar to the literature discussed in this chapter, Eklund et al. (2001) use bank sheet variables, financial ratios and industry variables to explain the probability of default. However, the study departs from previous studies by simulating next year's balance sheets and income statements using macroeconomic projections instead of using last year's data. Based on these estimations they predict the probability of default, the expected losses and the unexpected losses for the next year for Norwegian loans. In order to obtain these parameters notice that the time effects of changing balance sheets on probability of default should be obtained. Unfortunately, the widely used logit and probit models do not incorporate this effect. Therefore, Eklund et al. (2001) utilize similar to this study a discrete hazard model to obtain the probability of default. Discrete hazard models descend from duration analyse (survival analyse) in continuous time. Our discrete hazard model and the additional benefits of the discrete hazard model over logit and probit models will be discussed in section 4.2. However, we can't apply the simulations of balance sheets based on

macroeconomic forecast due to data opacity. Section 3 and section 4.1 will show that our data set misses too many variables and values to estimate firm specific balance sheets. In contrast to the univariate autoregression, VAR or Ecklund et al. (2001), the EBA adverse macroeconomic scenario will be used to stress the probability of defaults via the GDP growth and the interest rate.

3. Data and descriptive

3.1 Data

In 2013 DNB held the SME IRB Health Check that covered the SME portfolio from a selection of banks. The data collected by the survey consist besides the official risk parameters (PD, LGD and EAD) an extensive set of loan level and borrower level data over the period 2005-2012. The data are collected for about 4.5 million loans originated in the Netherlands, covering the vast majority of the SME lending market. In the loan level data the following loan information is present: the interest rate, month of origination, nominal amount; while from the borrower level e.g. the turnover, industry, EBITDA and the collateral value was provided. These data are annually reported every 1st of December and are linked to the defaults and the risk parameter predictions for the next year. Additionally, the defaults over the year 2013 are not reported in this dataset while the loan and borrower specific variables are reported. Therefore, the SME portfolio reported at the end of 2012 is ideal to predict the impact of unfavorable macroeconomic circumstances for the year 2013. In this section we will concentrate on the sample selection, default definition, industrial components, default indicators and finally an indication of the magnitude of the missing variables problem.

3.2 Descriptive statistics

Here we describe the steps of the sample selection and the main indicators described above, such as defaults, industry indicators, default predictors and some technical aspects related to the data collection (sample size, missing values etc.)

Originally, the data set consisted of approximately 8.5 million observations over the period 2005 till 2012 where around 4,5 million unique loans could be identified. In order to handle the data, a sample was made based on the steps shown in table 1.

Table 1: Sample selection

<i>Steps</i>	<i>Observations</i>
Dutch SME	4.344.980
Basel II SME definition: turnover \leq 50 million	2.196.526
Errors such as interest $<$ 0 or turnover $<$ 0	2.125.649
Random sample of unique loans	100.000
Random sample of unique loans + all the years of the loans	270.221

Explanatory note: we select a random sample of 100000 loans in the panel and then we augment it using all past and future information about these loans.

Before analyzing the default rates in this sample, it is valuable to know the process of defining a loan default. Besides the official Basel II definition of 90 days overdue on the material credit obligation, banks apply their internal “default triggers” to initiate the move of a loan to the pool of non performing loans. This implies that the default definition varies between banks and within our sample. Therefore, all the results represented will be controlled for financial institutions effect.

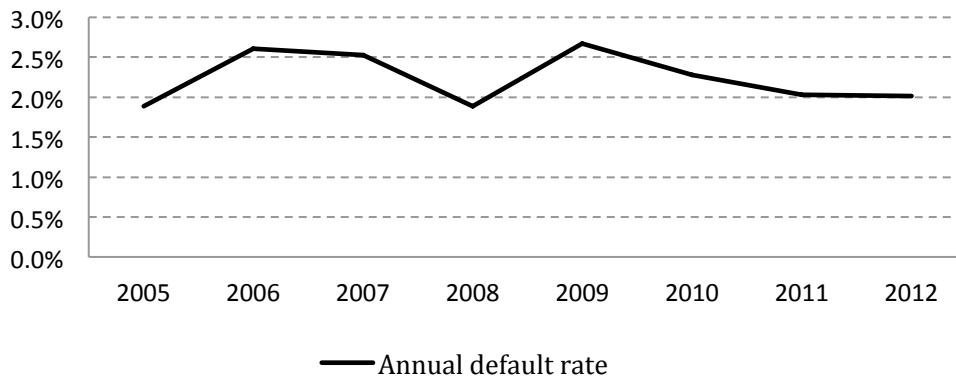
From table 2 one could conclude that in our sample there where 6048 defaults measured between 2005 and 2012. As banks report retrospective information, more recent years are more populated.

Table 2: Data and default frequency

<i>Years</i>	<i>2005</i>	<i>2006</i>	<i>2007</i>	<i>2008</i>	<i>2009</i>	<i>2010</i>	<i>2011</i>	<i>2012</i>
<i>Total loans</i>	52	10867	11924	29003	39960	43589	43789	41330
<i>Defaulted loans</i>	1	328	351	703	1348	1213	1042	1062

Furthermore, graph 1 shows that the default is relatively stable over time. In addition, comparing with other studies on SME default such as Altman and Sabato (2007), Fidrmuc and Hainz (2010) and McCann and McIndoe-Calder (2012) the default rates are relatively low. They observe respectively 6%, 6% and 7.1 % annual default rates for US, Slovakia and Irish SME loans.

Graph 1: Annual SME default rate in the Netherlands



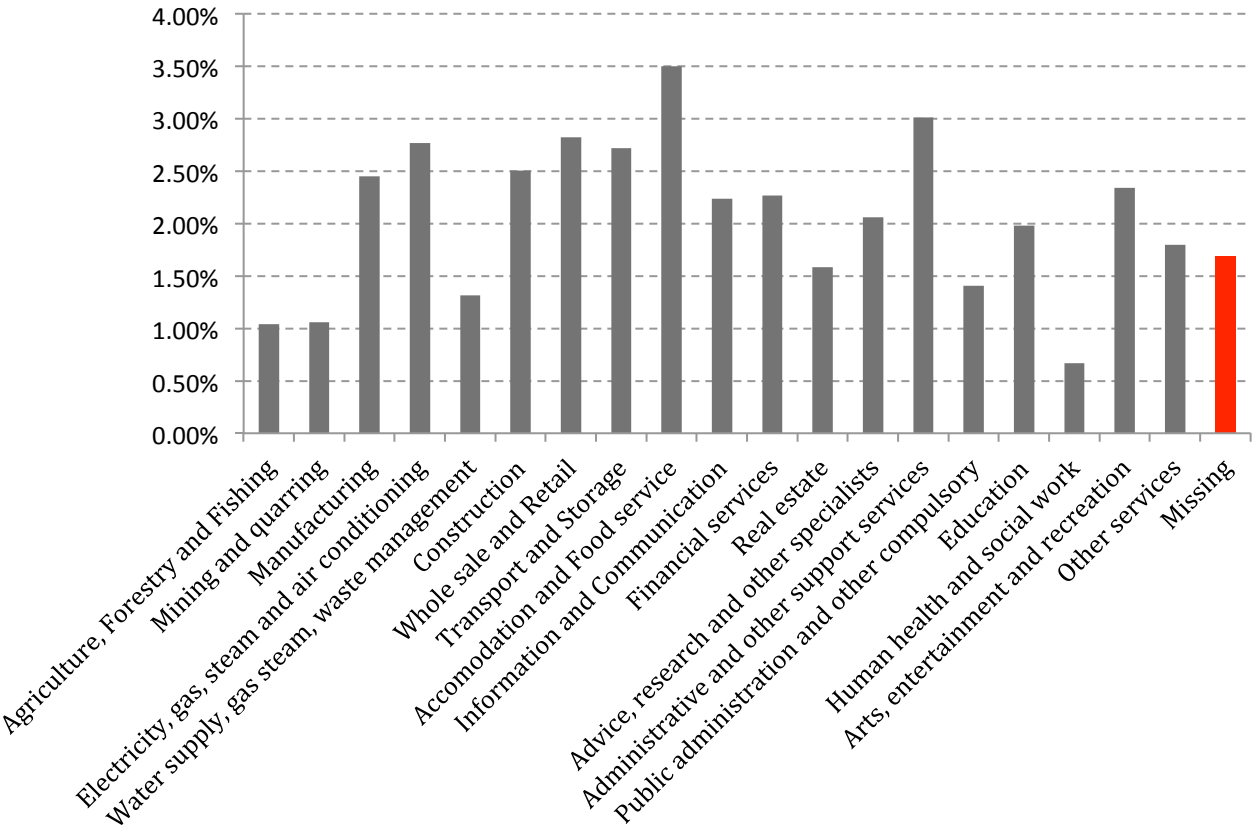
The observation of the stable default rate is contradicting the raised risk perception of banks for SME noted in the introduction. However, one could argue that the credit tightening of banks has prevented the raise of default rates. Additionally, the risk perception of financial institutions is likely to be very dependent on the outstanding exposures.

In order to explore the loan exposures, the loans will be classified in 21 main groups based on the Standaard Bedrijfs Indeling (SBI) of 2008. The SBI is the Dutch industrial classification for economic activities, which is based on the European classification, NACE, and the American classification, ISIC. Moreover, the 21 main groups of the SBI correspond with NACE classification. In the appendix, graph 8.1 and 8.2 provide pie charts

with respectively the average decomposition of the exposures towards the industries in the sample and specifically the industrial portfolio of the year of 2013. These graphs show that that the largest 4 sectors, “Whole sale and retail”, “Real estate”, “ Manufacturing” and “Construction” obtain more than 50% of all the loans supplied. This implicates that, besides the economic theory to differentiate between industries, a uniform probability of default across sectors could lead to large aggregation mistakes with calculating the expected losses of the SME portfolio under a stress model.

The need to differentiate between sectors for credit risk models is strengthens by graph 2, as it shows that actual default rates show large heterogeneity between sectors⁸.

Graph 2: Default rates by industry

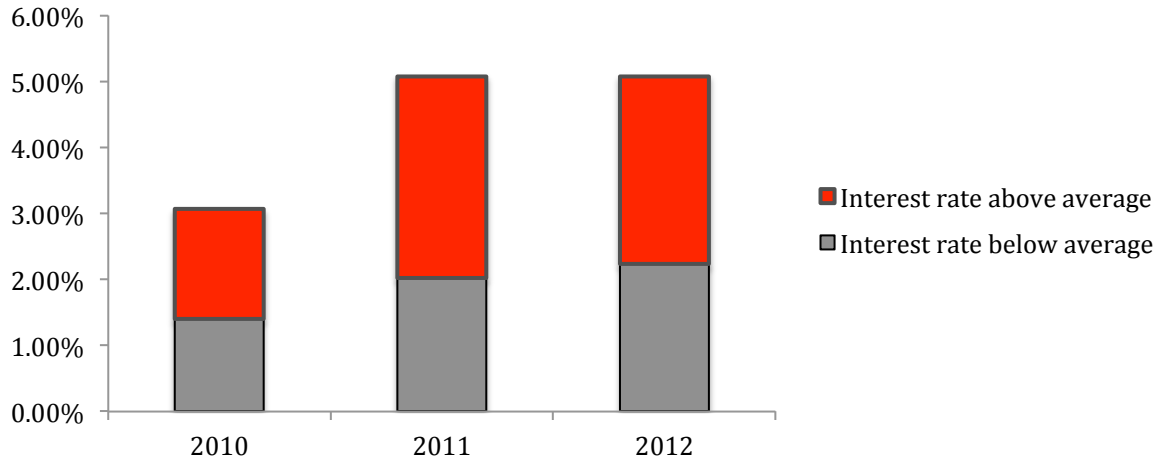


⁸ The sectors “Households as employers” and “Extra-territorial institutions” are excluded from the graph as no defaults were observed on a total of 12 observations.

From graph 2 one could notice that the sectors “Accommodation and food service activities” and “Administrative and other support services” are the only two services that have a default rate above 3%. Furthermore, we observe that “Agriculture, Forestry and Fishing”, “ Mining and quarrying” and “Human health and social work” default less compared to the rest of the SME portfolio. The difference between the lowest, “Human health and social work”, and the highest default rate, “Accommodation and food service activities” is no less than 2.8%. Combined with the economic literature, these observed heterogeneity has led us to control in all credit default models by dummies. The marginal effects of these industries are represented in table 8.1 of the appendix where the reference group consists out of the sectors “Households as employers”, “Extra-territorial institutions” and the group where the industry codes were missing. From this table one could that the results are more the less comparable with the graph where the reference group is indicated by the red color. In the interest of space these coefficients are not reported for the rest of the estimations.

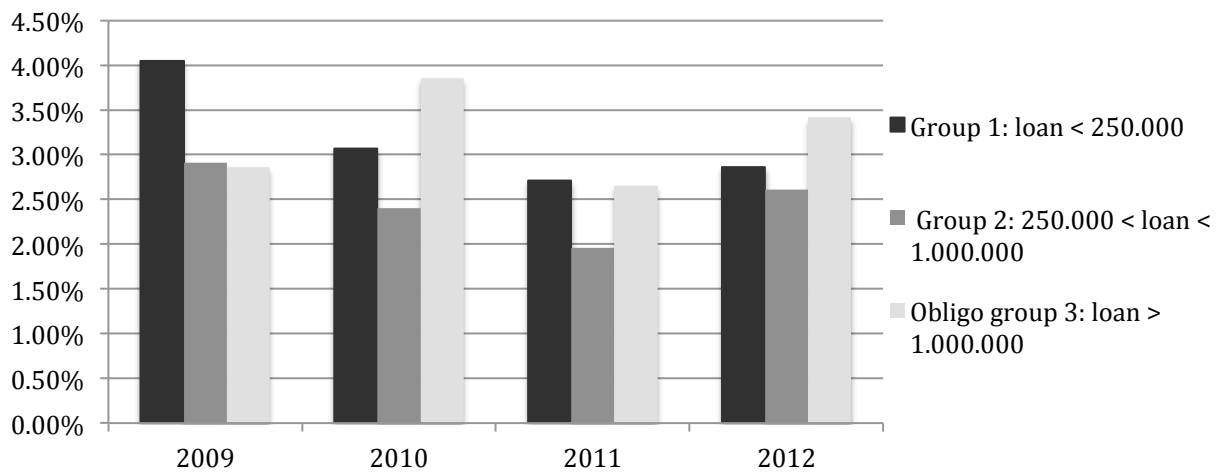
Further, a small look into the data is enough to observe that differences in default rates can be related to loan specific and borrower specific variables. In graph 3 the loans are divided into two groups; loans with an above average interest rate and loans with a lower than average interest rate. From graph 3 it is easy to observe that loans with a higher interest rate have a larger default rate. Even more, in 2011 and 2012 the default rate was almost three times higher for the group with an above average interest rate

Graph 3: Annual default rate by interest rate groups



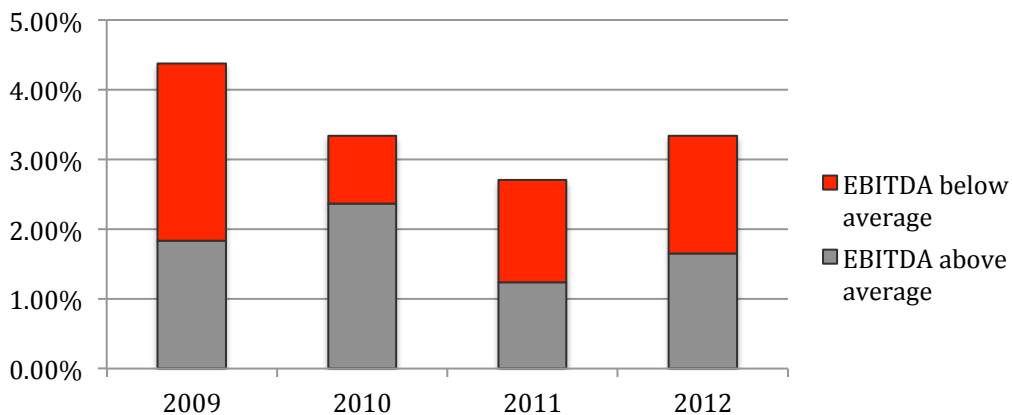
Besides the interest rate, it is interesting to observe if the amount of the loan as influence on the default rate. Therefore, the loans are divided in to three groups: loans with an amount lower than 250.000 euros, loans with a loan size between 250.000 euros and 1.000.000 and finally loans higher than 1.000.000 euros. The annual default rates of those groups are presented in graph 4. From graph 4 shows that the default rates of the group with the smallest loans are decreasing over time. This may be explained by the fact that banks funded fewer loans below the 250.000 since the crisis.

Graph 4: Annual default rate by loan size groups



In addition to the loan specific variables, the annual default rate and the borrower specific variables are clearly related. Graph 5 divides the loans in to two similar groups by the EBITDA similar as the interest. Group 1 consists out the borrowers with an EBITDA below the average while group 2 represents borrowers with lower than average EBITDA. From graph 4 it is clear that the borrowers with a lower EBITDA have a higher annual default rate in each year shown.

Graph 5: Annual default rate by EBITDA groups



To summarize, the graphs show that the loans and borrower specific variables have impact on the default rates in our sample. Therefore, it is clear that there are default predictors that can be included in a credit risk model. However, before presenting the model the missing variables and values in our data set has to be discussed.

The review of literature showed that most of the scholars pick financial ratios from the following accounting areas; profitability, liquidity, solvency, activity and leverage. However, the only variables available in our dataset concerning the financial health of the lender are the SME EBITDA and turnover. Furthermore, the most used denominators of the ratios such as assets and liabilities are absent. Therefore, this study won't utilize financial ratios as explanatory variables for the probability of default. As financial health based on accounting ratios is the main indicator of defaulting probabilities in studies regarding SME, this incompleteness could lead to lower accuracy of the models. In addition, the lack of

balance sheet data unfortunately does not allow us to use Ecklund et al. (2001) method of predicting balance sheet variables based on macroeconomic forecasts. On top of the absence of the data, one can see from table 3 that the missing variables are substantial for key variables that are received in the survey. Additionally, the variables of table 3 are combined known for only 2302 observations, which is less than 1% of the dataset. The next section will continue on this issue and shows that we the utilized an interest rate model to implement for missing values and that we control with dummies for the EBITDA and turnover.

Table 3: Missing variables

<i>Variable</i>	<i>Observation</i>	<i>Missing %</i>
<i>Interest rate</i>	37125	86.3
<i>EBITDA</i>	43647	84.2
<i>Turnover</i>	101856	62.3
<i>Collateral value</i>	193106	28.5

4. Model

In this section, the two steps model of estimating the probability of default for SME loans and the stress test methodology will be presented and explained. First, the rational behind the choosing an implementation technique for the interest rate (instead of dropping all observations with missing values) will be discussed. Secondly, the interest rate model will be explained in detail. After the implementation method, both the practical and theoretical considerations of utilizing a discrete hazard model over the widely used logit models will be pointed out. This resolves in the choice for our macroeconomic variable and in the selection of variables. This paragraph will end with the outline of our stress test methodology.

4.1 Interest rate imputation

The first step of the analysis is to overcome the data incompleteness. As noted in the section 3, the interest rate, EBITDA and turnover combined are complete for only known for less than 1% of the observations. One way to overcome this problem is to drop all the observations with missing variables as if those are missing completely at random (MCAR). However, the amount of missing data varies by bank. In other words, if we would drop all these incomplete observations we would threat the data as if the combined portfolio of a selection of banks is representative for the whole SME portfolio. In this paper the problem is handled differently, because the allocation of capital towards Dutch SME varies too much between banks to consider that the selection of banks with complete data would be a representative sample for the whole SME sector. As a result, a prediction of the interest rate will be imputed for the missing, as this variable is out of borrowers direct control but largely depends on market conditions. Even though the imputation overcomes the selection bias it does assume that each bank sets the interest rates for SME with a similar model. However, banks might set their interest rates differently based on the same information. Despite this assumption we believe that doing so is preferred over dropping all the observations with missing variables and having then to deal with the selection issue. As for the other missing variables, we use them in the analysis up to a dummy that takes value 1 if the variable is not missing. So the interpretation of these variables will depend on the analysis of two coefficients.

In order to predict the missing interest rate we assume that the interest rate setting of the banks can be explained by the following function for firm i at time t

$$r_{i,t} = g(\theta_t, z_{i,t}, \mathbf{x}_{i,t}) \quad (1)$$

Where $r_{i,t}$ is the nominal interest rate, θ_t represents the funding costs of the banks, $z_{i,t}$ is the risk premium for and individual lender and $\mathbf{x}_{i,t}$ is an vector with loan specific variables. This model is based on a simplistic view of the business model of banks where the spread between the funding costs and the interest rate on lending are the earnings for the bank.

First, the main components of the overall funding of banks are generally deposits and debt securities. A significant part of the interest paid on these components is variable. Therefore, the existing empirical literature uses money market rates as a proxy for the funding costs of banks. In line with the literature the annual average of Euro Interbank Offered Rate (Euribor) on a maturity of 12 months will serve as a proxy of the funding cost, θ_t ⁹. The Euribor rate is based on the average interest rates at which a large panel of European banks borrows unsecured funds from one another. The Euribor is the reference rate for euro-denominated products such as mortgages and deposits. Considering that banks are not completely price takers in the SME credit market, a rise in the Euribor would increase the interest rates of the loan.

Secondly, as the validation of the IRB models of the banks for the PD, LGD was the initial goal of the survey of the banks, the risk premium, $z_{i,t}$, will be measured by the PDs that the banks provided in the survey. Despite that this might be a near perfect proxy of credit risk, one might notice that the probability of default is a proxy for the risk profile conditioned on the fact that lender obtained credit. Consequently, the probability of default is an *ex post* observation while interest rates are determined *ex ante*. Unfortunately, this is a limitation of this proxy even though it only holds for fixed rate loans.

Additionally, the vector of loan specifics, $\mathbf{x}_{i,t}$, includes the nominal amount of the loan and the collateral value. Banks face fixed costs for a loan independent of the size of the loan. Therefore, with a higher loan amount lenders are more able to spread these fixed costs over the loan. Under perfect competition in the lenders market this results in lower interest rates for higher loan amounts. Similar to the nominal amount the collateral has a negative relationship with the interest rate. Collateral is an asset pledged to the lender to secure the loan which is used to overcome the asymmetric information problem between borrower and lender (Chan and Kantas (1985). In exchange for the securitization of the lender the borrower demands lower interest rates which explains the negative relationship between the two variables. In accordance with the theory, Breit and Arano (2009) show that lower nominal amounts and collateral values of loans leads to higher interest rates for

⁹ Obtained from European Money Markets Institute (EMMI) database.

United States SME. Unfortunately, the maturity of the loans can't be included in $x_{i,t}$ because it isn't obtained in the survey. This is definitely a limitation of the prediction as we assume that the interest is independent of the maturity of the loan. In order to give a clear overview, table 4 shows the variables used and the sign of the relationship is presented.

Table 4: Variables and the relationship with the interest rate

Euribor	+
Probabillity of default	+
Nominal Amount	-
Colleteral value	-

This section will end with the notion that regression prediction will make a perfect fit without taking account for regression errors (residuals). Consequently, we will predict the residuals and randomly assign a residual to the interest prediction. Therefore, a Monte Carlo simulation will be applied to receive the unbiased estimators of the interest rate coefficient in the discrete hazard model and the expected losses in the stress test¹⁰.

4.2 Discrete hazard model

After the interest rate imputation, a discrete hazard model will estimate the probability of default for SME loans. Before explaining the discrete hazard model, the "ordinary logit" will be introduced to show the improvements of using a duration model for estimating the probability of default. In addition, the choice for the macroeconomic variable will be discussed. This section will end with the selection of the predictors of default.

The single period logit model tries to predict the probability of default with a binary dependent variable y_i , which takes the value 1 for firm i if it defaulted and 0 if firm i is still

¹⁰ All the results in the study are based on 500 simulations

healthy, conditioned on a set of covariates \mathbf{z}_i such as the financial ratios of the Z-score. Modeling this with a logit makes the regression equation:

$$\Pr(y_i = 1|\mathbf{z}_i) = \frac{1}{1+e^{-\beta\mathbf{z}_i}} \quad (2)$$

One should note that in equation 2 the probability of default is independent of time. However, bankruptcy's and credit defaults occur infrequently over the time; therefore most scholars use multiple years to estimate this single period prediction model. One drawback of applying this "static" model is that a selection bias arises. As the firm covariates \mathbf{z}_i are based on the year prior to default for defaulted loans while one of the available firm covariates \mathbf{z}_i is used for the solvent loans, this ignores the years where the solvent loan was at default risk but remained healthy. As result of this selection bias, Shumway (2001) shows that the estimators of the static logit are positively biased and inconsistent. Secondly, the static model cannot incorporate time-variant variables, such as macroeconomic variables. This implies that the probability of default does not depend directly on the business cycle, because the static models measure the average effect over the time sample used. Especially for the purpose of stress testing this is a main objection to applying a static model. Consequently, Shumway (2001) suggest using a discrete hazard model to overcome these methodological concerns of the static models.

The discrete hazard model is a specific form of survival analysis. In the traditional continuous survival model, the timespan of observations is denoted by t where T represents the time of the occurrence of some event, in our case credit default. Time T is a nonnegative continuous random variable with a density function $f(x)$ and a distribution function of $F(x)$. The *survival function* $S(t)$ is defined by

$$S(t) = \Pr(T \geq t) = \int_t^{\infty} f(x)dx = 1 - F(t) \quad (3)$$

In words, the survival function represents the probability that a loan survives over the time span t . This makes the *hazard rate* the probability of default conditional on that it survived t . The hazard function the ratio of the probability density function and the survival function.

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} \quad (4)$$

The remaining difficulty of the survival analysis is the specification of the hazard rate. In the earlier studies towards business failure with survival analysis, such as Luoma and Laitinen (1991), used the semi parametric proportional hazard specification of Cox (1972).

$$\lambda(t | \mathbf{z}_i) = \lambda(t|0) e^{\beta \mathbf{z}_i} \quad (5)$$

Where $\lambda(t | \mathbf{z}_i)$ represents the probability of default at time t for loan i conditioned on the loan specific set of variables x_i . Here $\lambda_0(t)$ is called the baseline hazard function, which is time variant. However, the firm specific covariates \mathbf{z}_i are still independent of time. Therefore, this model has the same inefficiencies as the static models. Therefore, the more flexible specification of the continuous hazard model which takes into account time-varying covariates which can be written as

$$\lambda(t | \mathbf{z}_{i,t}) = \lambda(t|0) e^{\beta \mathbf{z}_{i,t}} \quad (6)$$

Shumway (2001) explicitly show that in discrete time the hazard model of (6) is equivalent to a multi-period logit model that is presented below

$$\Pr(y_{i,t} = 1 | \mathbf{z}_{i,t}) = \lambda(t | \mathbf{z}_{i,t}) = \frac{1}{1 + e^{-(\alpha_t + \beta \mathbf{z}_{i,t})}} \quad (7)$$

Where α_t is the time-varying baseline and $\mathbf{z}_{i,t}$ are the loan specific variables at time t .

Becker et al. (1998) use dummies κ_t as baseline hazard α_t for which takes the value 0 for the sequence of observations before failure¹¹. They argue that one can not add systemic variables that are the same across units but vary over time, because they will be highly collinear with the dummies κ_t . In contrast to Becker et al. (1998) who apply this

¹¹ This artificial example serves as an extra explanation of the dummies κ_t . In the table below loan 1 was observed in the period 2008 and 2010 and defaulted in 2010, while loan 2 survived the period 2009-2011 observed in our data set. Therefore, in the empirical results these dummies will be referred as periods.

Years	2008	2009	2010	2011	Dummies κ_t	1	2	3
Loan 1	0	0	1			0	0	1
Loan 2		0	0	0		0	0	0

model on a peace study, most default models make use of a macroeconomic variable as baseline hazard instead of the dummies κ_t . This is because temporal dependence alone doesn't explicitly explain business failure. Shumway (2001) suggest the use of the natural logarithm of the firm's age. However, this baseline hazard is independent of the economic conditions at any particular time. Therefore, Hillegeist et al. (2004) use the economic wide bankruptcy rate of previous year as a proxy for the macroeconomic environment in their bankruptcy prediction. Furthermore, they confirm Becker et al. (1998) theory that adding the dummies κ_t has no significant impact. Although this method incorporates a time dependent baseline hazard, Nam et al. (2008) argue that due to the autoregressive specification the out of sample forecast has no prediction power because it is not dependent on underlying fundamental macroeconomic structures. Consequently, Nam et al. (2008) let the volatility of the foreign exchange rate serve as the baseline hazard without considering the dummies κ_t . This choice is mainly based on the Asian crisis in 1997 that affected most listed firms on the Korea Stock Exchange. For modeling credit default for Dutch SMEs the foreign exchange rate volatility would be a far fetched baseline hazard as only 38% of the SME in Netherlands are involved in international activities, such as investment and trade. Furthermore, the main trading partner of Dutch SME is Germany which has the same currency, euro, and attributes to no cross country differences in the volatility of the foreign country exchange rates which could directly explain the SME defaults of loans. To overcome these drawbacks for the purpose of our SME study, we examined the GDP growth and the dummies κ_t . In contrast with Hillegeist et al. (2004) we find that including dummies κ_t does have a positive impact on the model.

Apart from the baseline hazard rate, one should note that loans that existed before entering the timeframe of the data were already at risk for default but unobserved. This issue is referred as 'left censoring' or 'left truncated' in duration analyze. Consequently, the age of the loan before entering the sample will be added as a loan specific variable \mathbf{z}_i to mitigate this problem. Moreover, we expect that this variable has a negative relationship with the probability of default because the borrowers history proves that the borrower is able to honour the loan payments. Most of other loan specific variables to influence credit risk, \mathbf{z}_i , are already employed in the interest rate imputation. Besides the predicted interest

rate, none of these variables will be included as covariates to avoid multicollinearity. Furthermore, the interest rate is cost of the loan for the lender therefore it's positively related with credit risk. In addition to the age of the loan and the interest rate, the available financial indicators utilized are the EBITDA and the turnover of the firm. The EBITDA is profitability measure and will be corresponding negatively with credit risk. However, we expect that turnover alone has a positive impact for the firm but after controlling for the EBITDA it might become negative. The reason for this is that turnover is most of all a size variable instead of a profitability measure. For this reason *ceteris paribus* firms become less efficient if the turnover rises but the EBITDA remains constant. Finally, in order to measure the effect of missing variables the dummies of turnover and EBITDA are included. Table 5 summarizes the expected signs of the coefficients in our final model.

Table 5: Variables and the relationship with the probability of default

GDP growth	-
Interest rate	+
Age of the loan	-
Ebitda	-
Turnover	+

4.3 Stress test methodology

As mentioned in section 2, the main methodology for stress testing is to estimate the effects of an artificial macroeconomic scenario on the probability of default compared with the baseline scenario. In contrast with the literature we do not apply a univariate auto regression or VAR estimations to simulate the future macroeconomic time series. However, this study will base the unfortunate macroeconomic scenario on the EBA macroeconomic adverse scenario of 2014. Although this paper does not perform a VAR or a univariate autoregressive model, the outcomes from the adverse scenario of the EBA are based on

macroeconomic simulations. In addition, adopting this scenario adds simplicity and leaves room for cross-country comparison.

In order to create the adverse scenario of the EBA the European Systemic Risk Board (ESRB) pointed out the pertinent threats to the financial stability of banks: increasing global bond yields, further deterioration of credit quality in vulnerable countries, stalling policy reforms, poor bank balance sheet repairs. Based on these risks, the adverse scenario of the EBA starts with an increase of the long term interest rates of the US. This shock leads to an overall adverse scenario projection for multiple macroeconomic variables for all EU countries for the years 2014-2016. The forecasts of the variables of our interest under the stress scenario in the Netherlands are shown in table 6.

Table 6: Adverse scenario Netherlands 2014-2016

	<i>2013 (baseline)</i>	<i>2014</i>	<i>2015</i>	<i>2016</i>
<i>GDP growth</i>	0.9	-0.5	-1.5	0.5
<i>Interest on 10 year bond</i>	2.0	3.3	3.3	3.4

Although it would be simple to incorporate this scenario completely, the baseline year of 2013 are the forecasts of the European Commission. These projections were accurate for all macroeconomic factors except for the GDP growth. According to the data of the Central Bureau voor de Statistiek (CBS) and the OECD the GDP growth of 0.9 should be -0.8 in retrospective. Therefore, we can't apply exactly the same scenario of the year 2014. Consequently, we will use the GDP growth in 2015 of -1.5 and we assume that interest rate of SME loans increase with the same value as the interest on the Dutch 10-year bond. Table 7 summarizes the adverse macroeconomic shocks applied on the variables of our model.

Table 7: Baseline and adverse scenario

	<i>2013 (baseline)</i>	<i>2013(adverse)</i>
<i>GDP growth</i>	-0.8	-1.5
<i>Interest rates SME</i>	0	1.3

As can be observed from table 7, we apply the stress test only for a one-year horizon. Pointed out by Duffie et al. (2007) extrapolation of models for one-year-ahead default probabilities to longer time horizon fails if one does not incorporate the long-term dynamics of the underlying firm specific covariates. These dynamics, for instance the Markov Chain utilized by Duffie et al. (2007), are not incorporated in our model. Consequently, the stress test isn't extended to longer periods than the one-year horizon.

After estimating the macroeconomic effects on our probability of defaults. This impact will be put on the original expected losses obtained by the banks. These expected losses, conditional on the macroeconomic environment, are based on the following formula:

$$EL_i = PD_i * LGD_i * EAD_i \quad (8)$$

Where EL_i are the expected losses of loan i , LGD_i is the loss given default and EAD_i is the exposure at default. The results will be given as a percentage of total outstanding credit. Before ending this paragraph, one should notice that from this method only the expected losses could be calculated while the unexpected losses (mostly obtained by estimating the Value at Risk (VaR) or Expected Shortfall) cannot be obtained because there are no uncertainties about the macroeconomic future in this model. Even though unexpected loss calculations are not possible; it is interesting to see the impact of the stress test on the expected losses for the SME portfolio. The magnitude of the effects on the SME portfolio will be compared with the internal models on corporates of the DNB. In order to obtain this comparison paragraph 5 will continue to discuss the empirical results of the models explained in paragraph 4.

5. Empirical results

5.1 Prediction of the interest rate

The interest rate model (1) in section 4.1 is estimated with a linear regression where the natural logarithm of the interest rate $r_{i,t}$ will be the independent variable. Notice that the prediction of the interest rate isn't the key interest of this study. The key interest of this model is to maximize the predictive power and the model fits better with the logarithm function of the interest rate. Furthermore, the probability of default is explained better by the level effect of the interest rate than the absolute effect. As a consequence, model (1) will be modeled with the following equation

$$\ln(r_{i,t}) = \alpha + \beta_1 EU_t + \beta_2 PD_{i,t} + \beta_4 \ln(NA_{i,t}) + \beta_5 D_{CV_{i,t}} + \beta_6 D_{CV_{i,t}} CV_{i,t} + \varepsilon_{i,t} \quad (8)$$

Where $r_{i,t}$ is the nominal interest rate, EU_t is the Euribor rate on a maturity of 12 months, $PD_{i,t}$ is the probability of default the banks provided in the survey, $\ln(NA_{i,t})$ is the natural logarithm of the nominal amount. From the descriptive statistics in table 8.2 of the appendix one can observe that the distribution of the nominal amount is positively skewed, therefore the nominal amount is transformed with the natural logarithm. $D_{CV_{i,t}}$ is a dummy that obtains the value 1 when the debt is securitized and $D_{CV_{i,t}} CV_{i,t}$ is the interaction effect of the collateral value in 100.000 euro. Table 6 provides the results of the model (8).

Table 6: Estimates of the interest rate model

Variable	Coefficient	Standard error
EU_t	0.06456	0.0034399
$PD_{i,t}$	1.35448	0.0362878
$\ln(NA_{i,t})$	- 0.11636	0.0014064
$D_{CV_{i,t}}$	- 0.30933	0.0061979
$D_{CV_{i,t}} CV_{i,t}$	- 0.00378	0.0003919

α	2.98911	0.0150027
Observations	26188	
R^2	0.4969	

One notable observation from table 6 is that the level effect of the collateral is miniscule with a decrease of 0.38% if the collateral value increases with 100.000 euros. While the coefficient of - 0.30933 shows that on average the collateralized loans approximately pay 31% less interest compared with the unsecured loans. As a result, one could conclude that the interest rates for SME do not depend on the size of the collateral given that the lender pledged collateral. Furthermore, the rest of the results are in accordance with expectations of table 4. The size of the loan does lower the interest rate while both the Euribor rate and the probability of default increases the interest rate. Finally, the predicted interest rates of this model will be used as explanatory variable for the defaults of the SME loans in the next section.

5.2 Results of discrete hazard model

In this section the results of our discrete model of function (6) in section 4.2 will be evaluated. This will start by explaining the addition of the dummies κ_t on our baseline hazard variable of GDP growth. Furthermore, the results will be compared with the expected relationships shown in table 5. Finally, the prediction power of the model will be assessed with a receiver operating characteristic (ROC) curve. All the estimation results are presented with table 7.

Table 7: Estimations of discrete hazard model

	(1)	(2)	(3)	(4)	(5)
<hr/>					
Variables					
GDP growth	0.0399*** (0.00576)	0.0628*** (0.00602)	0.0546*** (0.00658)	0.0624*** (0.00695)	0.0636*** (0.00710)
Interest rate			0.449*** (0.0285)	0.363*** (0.0321)	0.347*** (0.0331)
Age loan			-0.0754*** (0.00355)	-0.0721*** (0.00358)	-0.0716*** (0.00360)
Turnover missing				-0.502*** (0.185)	-0.702*** (0.203)
Turnover				0.0152 (0.0127)	0.0329** (0.0142)
EBITDA missing					2.106*** (0.380)
EBITDA					-0.197*** (0.0330)
Periode 2		-0.0303 (0.0332)	0.0563 (0.0360)	0.0711* (0.0378)	0.0757** (0.0386)
Periode 3		-0.196*** (0.0411)	-0.0838* (0.0441)	-0.0666 (0.0452)	-0.0659 (0.0461)
Periode 4		-0.417*** (0.0557)	-0.185*** (0.0577)	-0.212*** (0.0594)	-0.213*** (0.0605)
Periode 5		-0.262*** (0.0645)	-0.0231 (0.0663)	-0.0392 (0.0669)	-0.0348 (0.0674)
Periode 6		-0.521*** (0.0822)	-0.261*** (0.0844)	-0.295*** (0.0850)	-0.299*** (0.0852)
Periode 7		-1.073*** (0.111)	-0.756*** (0.114)	-0.810*** (0.114)	-0.817*** (0.115)
Control dummies					
Industries	yes	yes	yes	yes	yes
Financial institutions	yes	yes	yes	yes	yes
Constant	-3.942*** (0.0415)	-3.940*** (0.0427)	-4.285*** (0.0758)	-4.059*** (0.111)	-4.098*** (0.114)
Observations	225,557	225,557	169,094	160,200	157,852

Standard errors in parentheses,
*** p<0.01, ** p<0.05, * p<0.1

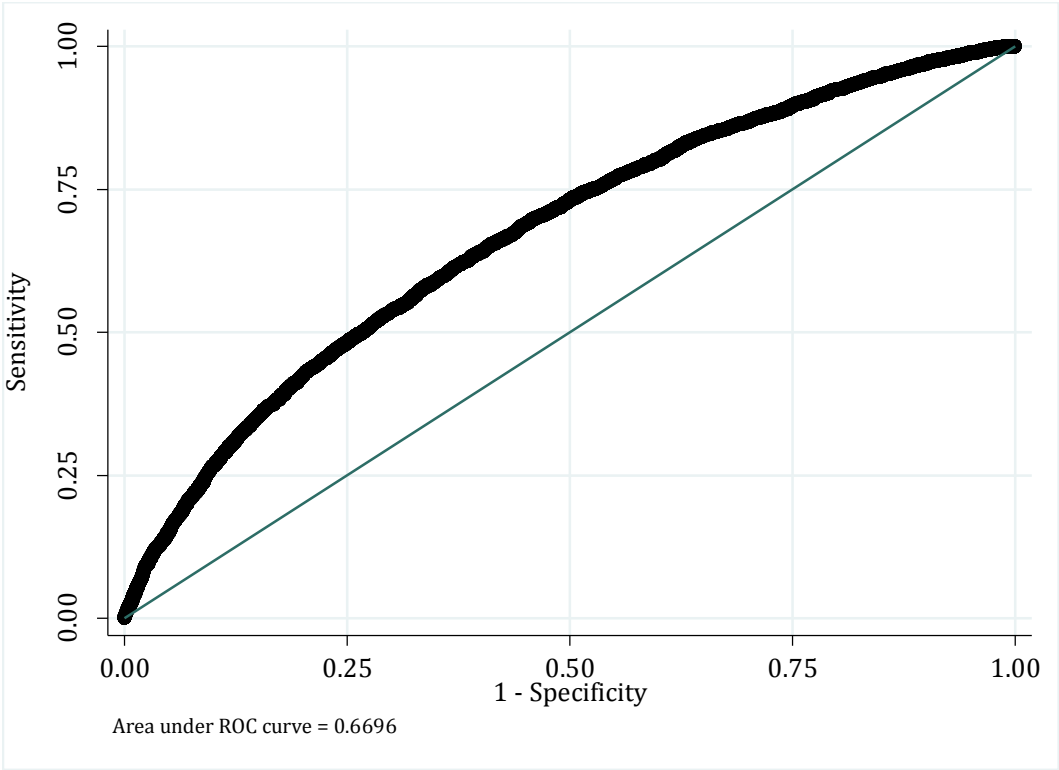
From table 7 it is clear that our baseline hazard of GDP growth has the expected positive coefficient in all the specifications. However, in contrast with Hillegeist et al. (2004) the dummies κ_t do have a significant impact on the estimations. The dummies κ_t are represented by period 2 till period 7 in table 7, where period 1 is the reference group. After the inclusion of the dummies in specification 2 we observe that all these coefficients are significant except period 2. Moreover, the effect of the GDP growth on the probability of default increases from 0.0399 to 0.0628. In addition, most periods stay significant in all the specifications. This means that the effect of the GDP is dependent on including the dummies κ_t . Therefore, the dummies κ_t has a significant impact on our discrete hazard model and will be included in all our specifications.

Furthermore, from specification 3 in table 7 onwards the loan and firm specific variables are included into the model. The interest rate, obtained from the outcomes of the prediction model of section 5.2, has a positive effect on the probability of default. This is in accordance with the theory, where increasing borrowing costs leads to more difficulties in meeting financial obligations. In addition, a higher loan age before entering the sample significantly decreases the probability of default. This implies that the history of a borrower can be used to predict defaults. In contrast to interest rate and the age of the loan, the turnover in specification 6 has the wrong positive sign according to the theory. Following the theory the turnover is more a size variable than a profitability measure as mentioned in section 4.2. However, the probability of default is normally negatively related with a size variable because larger firms default less than smaller firms. Therefore, the insignificant coefficient of -0.0721 contradicts the theory. After controlling for the EBITDA the turnover becomes significant and positive sign in specification 6, which is in accordance with the theory summarized in table 5. EBITDA itself has as expected a negative sign. Overall, this means that specification 6 has the expected signs and will therefore used in the stress test. Before discussing the prediction power of the model, the dummies for the missing variables will be discussed. While the dummy of the turnover is negatively related with a small coefficient, the dummy of the EBITDA has a large positive coefficient of 2.106. This means that if the banks do not provide the EBITDA the default rate increases even after controlling for financial institutions. This suggests that banks believe that the EBITDA

is not a relevant variable for risky loans. An explanation can be that the ability to pay the loan becomes more important than the actual earnings of the borrowing firm therefore banks might focus more on liquidity variables for evaluating risky loans. Although these missing variables do not have an exact economic impact on the model, controlling for these variables does prevent a selection bias and help to predict the defaults more accurate.

To assess the accuracy of this binary model we use the area under the ROC curve. The ROC is plotting the fraction of true positives out of the total actual on the y-axis and the fraction of false positives out of the total actual negatives on the x-axis. Here, a line of 45 degrees would under the law of large numbers represent a random binary classification model. Therefore, the area under the ROC curve is used as a measure for the performance of binary classification models. Under a perfect model this area will be 1 will under a random model this will be 0.5. The ROC curve and the area under the ROC curve for specification 6 of table 7 are presented in graph 6.

Graph 6: ROC curve



From the ROC curve in graph 7 one can see that our model outperforms the random binary classification model because it is above the 45 degrees line. Furthermore, it shows that the area under the ROC curve is 0.6696. To compare this performance with the SME loan literature, we use the results of McCann and McIndoe-Calder (2012). They find an area under the ROC curve of 0.7749. However, their model is based on a data set with full information of the balance sheet of the SME, this leads consequently to a model with higher prediction power. Moreover, we believe that the main issue of our relatively low prediction power is the data opacity. To improve the credit risk models in the future the main point to improve is the data collection. Improved data collection and therefore more accurately models would lead to more robust stress test results, which all combined could result in better supervision of banks, improved allocation of capital and reducing informational asymmetries in lending. Even though our model does not predict defaults perfectly, the stress test result represented in the next section could be a stepping-stone towards future research.

5.3 Results of the stress test

In our stress test methodology the interest rates in 2013 increased with 1.3% while the GDP growth decreased from -0.8% till -1.5%. Compared with our baseline projections this results in an increase of 13% of the probability of defaults in 2013. Most of this effect is created by the interest rate increase due to the larger coefficient in the regression specification. Only the interest rate increase of 1.3% results in an increase of 9% in the probability of defaults. Although these results cannot be compared within the literature, for the same macroeconomic scenario of the EBA the increase of the probability of defaults for SMEs is lower than for corporates modeled internally by the DNB. This means that our results agree with Dietsch and Petey (2002), who find that SME are less dependent on the “state of the economy”. On the other hand, the results are contradicting Jacobson et al. (2005) findings. Furthermore, we observed that the expected losses increases from 1.13 % of total credit to SME to 1.28 % of total credit under the EBA scenario. The impact of the

macroeconomic scenario on the expected losses for the SME portfolio is difficult to compare due to the lack of similar studies. However, these numerical results can be seen as a starting point for future research on portfolio losses of SME loans.

6 Conclusion

This paper aims to contribute to the literature of SME loan default prediction and the macroeconomic influence on the default of SME loans by using unique data from DNB's health check survey on SME portfolios of Dutch banks.

However, the data is far from ideal and therefore our model consisted out of three steps. First, an econometric model was utilized to implement the interest rates. Secondly, the probability of default was related with these interest rate and loan and borrower specific variables such as the EBITDA, turnover, age of the loan. Finally, the macroeconomic dependence was measured for the year 2013. We found that, *ceteris paribus*, an increase by 1.3% point in the interest rate with -1.5% GDP growth, will translate in an increase the probability of defaults with 13%. Furthermore, this will lead to increase from 1.13% to 1.28% in expected losses in total credit. Moreover, the effect of this macroeconomic scenario is smaller compared with corporates. Therefore, we conclude that Dutch SME are less dependent on the macroeconomic environment than corporates.

However, there are some limitations to our model. The interest rate model assumes that all banks set the interest rate based on uniform model, while in reality one might expect heterogeneity between interest rate policies of banks. Furthermore, the data opacity causes low prediction power of the model. As final remark, the chronic lack of data on SME firms has to be improved in order to establish future research. Especially considering the importance of SME in the economy and the benefits of accurate credit risk models such as improved bank supervision, allocation of capital and the reducing informational asymmetries between lenders and borrowers.

7 References

Altman, Edward I., 1968.

"Financial Ratios, Discriminant Analysis And The Prediction Of Corporate Bankruptcy," *Journal of Finance*, American Finance Association, vol. 23(4), pages 589-609, September.

Altman, Edward I. & Sabato, Gabriele, 2007.

"Modelling Credit Risk for SMEs: Evidence from the U.S. Market," *Abacus*, Accounting Foundation, University of Sydney, vol. 43(3), pages 332-357.

Acharya, Viral V. & Bharath, Sreedhar T. & Srinivasan, Anand, 2007.

"Does industry-wide distress affect defaulted firms? Evidence from creditor recoveries," *Journal of Financial Economics*, Elsevier, vol. 85(3), pages 787-821, September.

Beaver, William H, 1966.

"Financial Ratios as Predictors of Failure", *Journal of Accounting Research*, Vol. 4, Issue Empirical Research in Accounting: Selected Studies, pages 71-111.

Beck, Nathaniel & Katz, Jonathan N., Tucker, Richard, 1998.

"Taking time seriously: Time-Series-Cross-Section Analyses with Binary Dependent Variable" *American Journal of Political Science*; vol. 42(4), pages 1260 - 1288

Berger, Allen N. & Udell, Gregory F., 2004.

"The institutional memory hypothesis and the procyclicality of bank lending behavior," *Journal of Financial Intermediation*, Elsevier, vol. 13(4), pages 458-495, October.

Boss, M., 2002

"A Macroeconomic Credit Risk Model for Stress Testing the

Austrian Credit Portfolio" *Financial Stability Report 4*, Oesterreichische Nationalbank.

Breit, Emily & Arano, Kathleen, 2009.

"Reduction in Information Asymmetries in Small Business Lending: Evidence from the Survey of Small Business Finances" *The Journal of Financial and Economic Practice*.

Black, Fischer & Scholes, Myron S, 1973.

"The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, University of Chicago Press, vol. 81(3), pages 637-54, May-June.

Chan, Yuk-Shee & Kanatas, George, 1985.

"Asymmetric Valuations and the Role of Collateral in Loan Agreements," *Journal of Money, Credit and Banking*, Blackwell Publishing, vol. 17(1), pages 84-95, February.

Chan-Lau, Jorge A, 2006.

"Fundamentals-Based Estimation of Default Probabilities: A Survey,"

IMF Working Papers 06/149, International Monetary Fund.

Chava, Sudheer, and Robert Jarrow, 2004.

"Bankruptcy prediction with industry effects", *Review of Finance*, vol. 8 , 537–569.

Dietsch, Michel & Petey, Joel, 2002.

"The credit risk in SME loans portfolios: Modeling issues, pricing, and capital requirements," *Journal of Banking & Finance*, Elsevier, vol. 26(2-3), pages 303-322, March.

Dietsch, Michel & Petey, Joel, 2004.

"Should SME exposures be treated as retail or corporate exposures? A comparative analysis of default probabilities and asset correlations in French and German SMEs," *Journal of Banking & Finance*, Elsevier, vol. 28(4), pages 773-788, April.

Duffie, Darrell & Saita, Leandro & Wang, Ke, 2007.

"Multi-period corporate default prediction with stochastic covariates," *Journal of Financial Economics*, Elsevier, vol. 83(3), pages 635-665, March.

Eklund, Trond & Larsen, Kai & Berhardsen, Eivind, 2001.

"Model for analysing credit risk in the enterprise sector", *Economic Bulletin*, vol. 72 (3), Norges Bank

Fidrmuc, Jarko & Hainz, Christa, 2010.

"Default rates in the loan market for SMEs: Evidence from Slovakia," *Economic Systems*, Elsevier, vol. 34(2), pages 133-147, June.

Figlewski, S., Frydman, H., & Liang, W., 2012.

"Modeling the Effects of Microeconomic factors on Corporate Default and Credit Rating Transitions" ,*International review of Economics and Finance*, vol. 21, pages 87 – 105.

Foglia, A. 2008.

"Stress testing credit risk: a survey of authorities' approaches." Occasional Paper No. 37, Bank of Italy.

Helwege, Jean, 2010.

"Financial firm bankruptcy and systemic risk," *Journal of International Financial Markets, Institutions & Money*, vol 20, pages 1-12

Hillegeist, Stephen A. & Keating, Elizabeth K. & Cram, Donald P. & Lundstedt, Kyle G., 2004.

"Assessing the Probability of Bankruptcy", *Review of Accounting Studies*, vol. 9(1).

Jacobson, Tor & Lindé, Jesper & Roszbach, Kasper, 2005.

"Credit Risk Versus Capital Requirements under Basel II: Are SME Loans and Retail Credit Really Different?", *Journal of Financial Services Research*, Springer, vol. 28(1), pages 43-75, October.

Jiménez, Gabriel & Saurina, Jesús (2006)

"Credit cycles, credit risk, and prudential regulation," *International Journal of Central Banking*, vol. 2, pages 65-98.

Jorion, Philippe & Zhang, Gaiyan, 2009.

"Credit Contagion from Counterparty Risk," *Journal of Finance*, American Finance Association, vol. 64(5), pages 2053-2087, October.

Kalirai, H & Scheicher, M. (2002)

"Macroeconomic Stress Testing: Preliminary Evidence for Austria" *Financial Stability Report 3*, Oesterreichische Nationalbank

Kane, Gregory D. & Richardson, Frederick M. & Meade, Nancy L. 1998.

" Rank Transformations and the Prediction of Corporate Failure" , *Contemporary Accounting Research*, vol. 15(2), pages 145-166

Lang, Larry H.P. & Stulz, René M., 1992.

"Contagion and Competitive Intra-Industry Effects of Bankruptcy Announcements" *Journal of Financial Economics*, vol. 32, pages 45-60.

Lennox, Clive, 1999.

"Identifying failing companies: A re-evaluation of the Logit, Probit, and DEA approaches", *Journal of Economics and Business*, vol.51 (4) pages 347-364, July-August.

Luoma, M, Laitinen, E.K.,1991.

Survival analysis as a tool for company failure prediction, *Omega*, Vol. 19(6) Pages 673-678.

McCann, Fergal & McIndoe-Calder, Tara, 2012.

"Determinants of SME Loan Default: The Importance of Borrower-Level Heterogeneity,"
Research Technical Papers 06/RT/12, Central Bank of Ireland.

Merton, Robert C, 1974.

"On the Pricing of Corporate Debt: The Risk Structure of Interest Rates," *Journal of Finance*, American Finance Association, vol. 29(2), pages 449-70, May.

Nam, Chae Woo & Kim , Tong Suk & Park, Nam Jung & Lee, Hoe Kyung, 2008.

"Bankruptcy prediction using a discrete-time duration model incorporating temporal and macroeconomic dependencies," *Journal of Forecasting*, John Wiley & Sons, Ltd., vol. 27(6), pages 493-506.

Ohlson, James. A. 1980.

"Financial ratios and the probabilistic prediction of bankruptcy", *Journal of Accounting Research*, vol. 18, pages 109-131.

Pesola, Jarmo, (2001).

"The Role of Macroeconomic Shocks in Bank Crises", Discussion Paper No. 6.2001, Bank of Finland

Salas, V., &J. Saurina. 2002.

"Credit Risk in Two Institutional Regimes: Spanish Commercial and Savings Banks." *Journal of Financial Services Research*, vol. 22 (3), pages 203–24.

Shleifer, Andrei & Vishny, Robert W, 1992.

" Liquidation Values and Debt Capacity: A Market Equilibrium Approach," Journal of Finance, American Finance Association, vol. 47(4), pages 1343-66, September.

Simons, Dietske, & Rolwes, Ferdinand, 2009.

"Macroeconomic Default Modeling and Stress Testing." International Journal of Central Banking, vol. 5 (3).

Shumway, Tyler, 2001.

"Forecasting Bankruptcy More Accurately: A Simple Hazard Model," The Journal of Business, University of Chicago Press, vol. 74(1), pages 101-24,

Van der Veer, Koen & Hoeberichts, Marco, 2013.

"The Level Effect of Bank Lending Standards on Business Lending," DNB Working Papers 396, Netherlands Central Bank, Research Department.

Virolainen, K. 2004. "Macro Stress Testing with Macroeconomic Credit Risk Model for Finland." Discussion Paper No. 18, Bank of Finland.

Wilson, Thomas C. 1997a.

"Credit Portfolio Risk (I)." RISK, October.

Wilson, Thomas C. 1997b.

"Credit Portfolio Risk (II)." Risk, November

Zmijewski, Mark E., 1984.

"Methodological issues related to the estimation of financial distress prediction models", Journal of Accounting Research, vol. 22, pages 59-82.

Graph 8.2: Pie chart of the exposures by industry for the year 2013

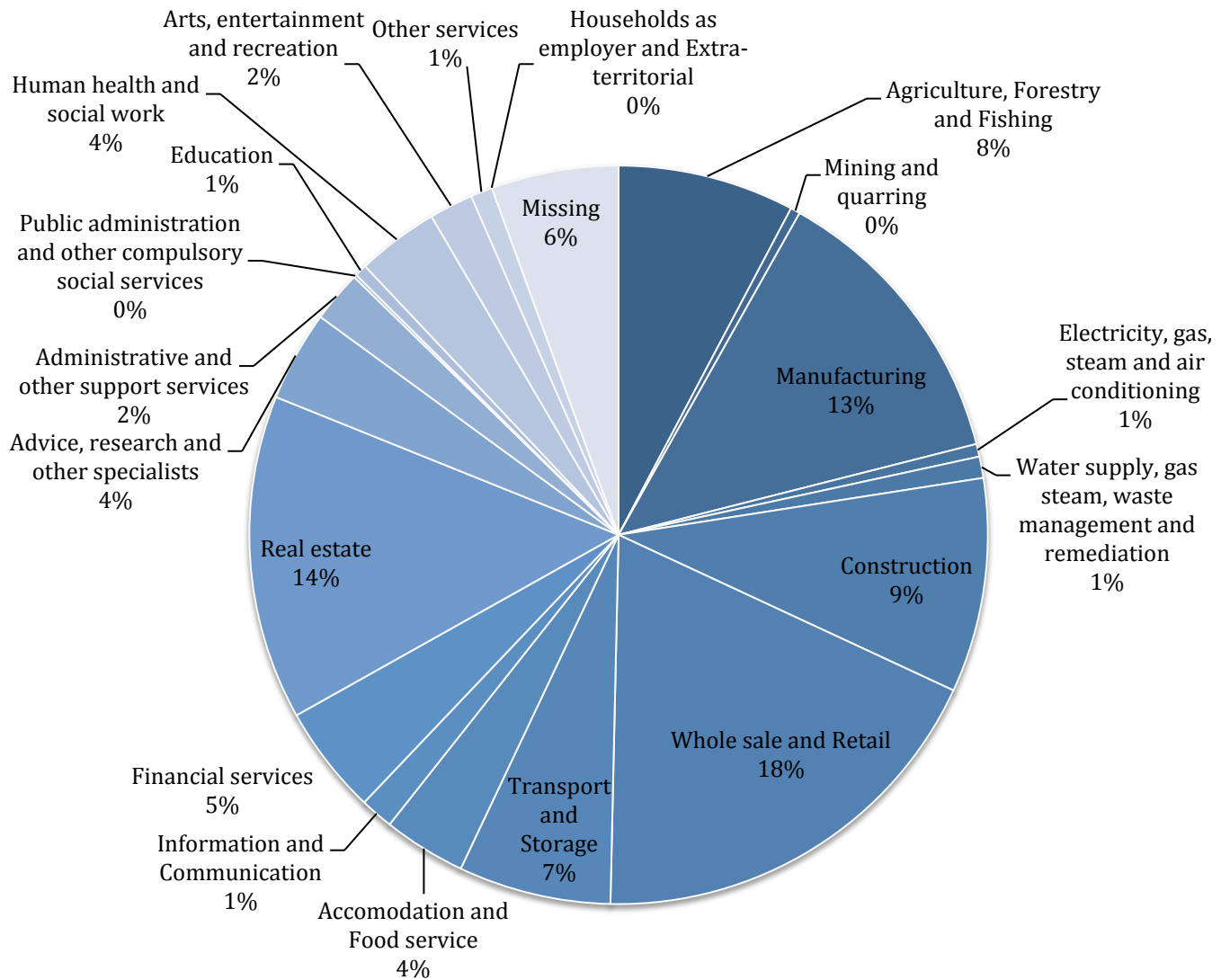


Table 8.1: The industry marginal effects corresponding with specification 1) of table 7

(1)

Industries	
Agriculture, Forestry and Fishing	-0.469*** (0.151)
Mining and quarrying	-0.339 (0.413)
Manufacturing	0.432*** (0.0634)
Electricity, gas, steam and air conditioning	0.529 (0.362)
Water supply, gas steam, waste management and remediation	-0.173 (0.714)
Construction	0.423*** (0.0576)
Whole sale and Retail	0.555*** (0.0525)
Transport and Storage	0.527*** (0.0711)
Accommodation and Food service	0.765*** (0.0677)
Information and Communication	0.338*** (0.0905)
Financial services	0.297*** (0.0756)
Real estate	0.0232 (0.0946)
Advice, research and other specialists	0.231*** (0.0621)
Administrative and other support services	0.633*** (0.0735)
Public administration and other compulsory social services	-0.0609 (1.010)
Education	0.189* (0.113)
Human health and social work	-0.885*** (0.134)
Arts, entertainment and recreation	0.362*** (0.0982)
Other services	0.120 (0.0935)

Table 8.2: Descriptive statistics of variables used in the models

Variable	Mean	Median	Min	Max	Skewness
Interest rate	6.958	7.33136	0	30.48	0.2848
PD	0.038	0.01196	0.0003	0.9397	3.673
NA	156725.6	15000	0	2.19E+08	64.297
CV	323356.6	11307	0	8.95E+08	193.725
EBITDA	168472.4	94000	-2.51E+08	5.52E+08	50.754
Turnover	2773290	348000	0	5.00E+07	4.103
Age loan	6.245	5	0	4.10E+01	1.865