06/RT/12

Páipéar Taighde Teicniúil
Research Technical Paper

Determinants of SME Loan Default:

The Importance of Borrower-Level Heterogeneity

Fergal McCann and Tara McIndoe-Calder

Banc Ceannais na hÉireann
Central Bank of Ireland
Eurosystem

Determinants of SME Loan Default: The Importance of Borrower-Level Heterogeneity

Fergal McCann and Tara McIndoe-Calder*

September 13, 2012

Abstract

Using unique borrower-level balance sheet information for a cross-section of 6,000 Irish SME loans, this paper tests the determinants of default at the micro level. Typical financial ratios, such as the ratio of the loan to total assets, the current ratio, leverage ratio, liquidity ratio and profitability ratio, are found to be significant predictors of default. Further, the length of time the borrowing firm's owner has been with the firm mitigates the likelihood of default. Conditional on the above, significant sector-level effects remain. The paper moves beyond average effects of the above-mentioned variables by repeating the analysis across seven sectors of economic activity, and across the quintiles of firm size, exposure and credit quality. The share of defaults is shown to fall as firms get larger, and to rise as loans get larger relative to assets. The results suggest that different warning signals can be identified, particularly for borrowers of different sizes and with small versus large loans. These results contribute to the literature on "fundamentals-based" modelling of corporate default risk, and represent one of very few sets of results on the determinants of default in SME lending in particular.

*E-mail: fergal.mccann@centralbank.ie and tara.mcindoecalder@centralbank.ie. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Central Bank of Ireland or the ESCB. We would like to thank Frank Browne, Trevor Fitzpatrick, Lars Frisell, Rob Kelly, Kieran McQuinn and Gerard O'Reilly for comments on an earlier draft. Any remaining errors are our own.

Non-Technical Summary

This paper exploits detailed borrower-level balance sheet information for approximately 6,000 small and medium sized enterprise (SME) loans, collected by the Central Bank of Ireland in late 2010, to present granular regression-based analyses of the determinants of default in SME lending. The study is one of very few papers to have access to micro-level data on SME loans. The paper moves beyond previous studies in the "fundamentals-based" literature on corporate default, which have generally estimated an average effect of particular borrower-level characteristics. The scope of the available data allows probit regressions to be carried out at different points of the firm size distribution, the exposure size distribution, the credit quality distribution and in different sectors of economic activity. It thus provides a contribution to both the academic and practical literature on the credit risk modelling of SME loan books.

As part of the Financial Measures Programme 2011, Irish financial institutions were required to submit their full loan books to the Central Bank of Ireland. These loan files were then used to stress-test the Irish banking system. In the SME case, the full loan files, totalling some 440,000 loans, were generally restricted to loan-level information such as the size of exposure, type of loan, date of origination, maturity, and interest rate, while lacking in information on the borrowing SME. However, a subsample of loans for which rich borrower-level information was available was also extracted from the SME books. It is the Republic of Ireland-based loans of this detailed file that form the basis of this research paper. The data contain information on 6,745 loans from unique borrowers. Information on the size of exposure as of September 2010, a dummy for Basel II default (90 days past due), sector of activity, and a wide range of balance sheet information are available for the majority of loans in this file. Apart from the exposure size variable, all borrower-level variables were collected for a "base month" by the lender. The base month for 50 percent of borrowers was in 2009, with the rest spread between 2010, 2008 and a small amount in 2007 and earlier.

This paper first emulates previous work by looking at the drivers of SME default. A number of Altman's Z-score variables are used, such as the ratio of turnover to total assets and profitability to total assets. Other explanatory variables include the size of the loan in question relative to the borrower's total assets, a measure of leverage, the current ratio and a dummy indicating whether the owner of the firm has been with the firm for more than ten years. The major contribution of the paper is in moving beyond the average effect of these variables in the data set to acknowledge borrower heterogeneity. In looking at the drivers of default separately across sectors of economic activity and quintiles of firm size, exposure and credit quality distributions, the paper points to numerous heterogeneous effects that would be missed in estimating an average effect. One example is the irrelevance of financial ratios in predicting default in small loans: in the first quintile of the exposure distribution, no borrower-level characteristic is found to significantly predict default. Similarly, among the largest firms, where the default rate is below four per cent, no borrower-level information signficantly predicts default. Further, the majority of borrower-level determinants are found to have more predictive power among smaller firms and among larger loans, suggesting that particular attention must be paid to the financial health of small firms and borrowers with

large exposures. Along with examining the data by firm size and exposure (a measure of risk) we also examine determinants of default by sector of economic activity and across the distribution of borrower quality (proxied using the current ratio). The results highlight the importance of different financial ratios in explaining default rates across the sector and credit quality distributions.

The importance of accounting for borrower heterogeneity is further evidenced in several model validation exercises including receiver operating characteristic (ROC) curves, tests for sensitivity and specificity, out of sample testing and comparative estimates of proportional expected losses (PELs). These findings can improve the predictive power of credit risk models relying on average effects across a loan book, and therefore lead to capital requirements that better reflect credit risk.

1 Introduction

This paper exploits detailed borrower-level balance sheet information for approximately 6,000 small and medium sized enterprise (SME) loans, collected by the Central Bank of Ireland in late 2010, to present granular regression-based analyses of the determinants of default in SME lending. The study is one of very few papers[1] to have access to micro-level data on SME loans. The paper moves beyond previous studies in the "fundamentals-based" literature on corporate default, which have generally estimated an average effect of particular borrower-level characteristics. The scope of the available data allows probit regressions to be carried out at different points of the firm size, exposure size and credit quality distributions and in different sectors of economic activity. It thus provides a contribution to both the academic and practical literature on the credit risk modelling of SME loan books.

Models of default in corporate lending can generally be divided into two categories: "fundamentals-based" models (as defined and surveyed in Chan-Lau (2006)), which use borrower-level balance sheet information as behavioural drivers of default, and "market outcome-based" models, such as the famous KMV model, in the spirit of Merton (1974). The latter type of model generally use market capitalization data on publicly-traded corporations to predict corporate default or bankruptcy. These models are particularly attractive to practitioners in industry, as such data are readily available from numerous sources and can be updated as regularly as is deemed relevant by the researcher.

Research on the predictors of default of SMEs must generally take a fundamentals-based approach, due to their informational opacity, i.e. a lack of market outcome-based data available for these borrowers. This presents certain difficulties as balance-sheet data is not available at anything like the same regularity or timeliness as data required for Merton-style models. In countries where data sources such as credit registers do not exist such as Ireland, the estimation of models of default for SMEs becomes even more challenging, as borrower-level data collection must be carried out by the lender.

Despite the drawbacks with respect to reporting timely probabilities of default, there are a number of advantages to fundamentals-based research on default. For example, the nature of the data used to estimate these models means that numerous borrower-level early-warning signals can be ascertained from the model. These can include levels of leverage, profitability, liquidity, turnover and relative loan size to asset levels as well as sector-specific effects and "soft" information, for example business age. This means that once forecasts can be made regarding the direction in which these variables are likely to move, predictions can then be made regarding the risk-level associated with a portfolio.

While the use of financial variables in these models has a long pedigree (Altman, 1968), more recently, "soft" information has been shown to be important (Lehmann, 2003; Berger and Frame, 2007). This may be especially useful where financial ratio variables have poor out of sample predictive power. Indeed, Berger and Frame (2007) show that adopting these information-rich models

[1] Fidrmuc and Hainz (2010) and Behr, Güttler, and Plattner (2004) also use SME loan-level data to establish determinants of default in Slovakia and Germany, respectively.

can lead to an increase in the availability of bank credit and reduce the risk and opacity of these loans for banks.

Further, estimating the default probabilities of SME loans as distinct to those of larger corporates (Dietsch and Petey, 2004) and retail exposures (Jacobson, Lindé, and Roszbach, 2005) is important. For example, in the absence of asset-class specific models to determine probabilities of default, Basel II lenders are required to hold higher levels of risk-weighted capital, than when these models are used. While loans to SMEs are riskier than lending to larger corporates, asset correlations in the former are typically low due to the relatively high idiosyncratic, as opposed to systematic, risk associated with these loans (Altman and Sabato, 2007). Therefore, modelling probabilities of default (PDs) separately for each of the banks' loan books reduces capital requirements and may increase bank profitability, while enhancing the ability to select better (SME) lending prospects (Altman and Sabato, 2007). These issues have policy implications for bank lending (Dietsch and Petey, 2004), perhaps especially in times of financial instability and the often associated tightening of credit standards.

The use of financial ratios to predict corporate default rates stretches back as far as the Z-score of Altman (1968), who advocated utilising a number of the financial ratios used in this paper as components of a model of corporate bankruptcy. Several paper were to the fore in incorporating cash flow, as motivated from theoretical models, as a key determinant of predicted corporate bankruptcy.[2] More recent papers in this vein include Westgaard and van der Wijst (2001) and Lennox (1999) using Norwegian corporate and UK listed firm data, respectively.

The literature on SME loans appears sparsely populated when compared to that on corporate default, which can use more readily available data sets on listed companies. Examples of SME default analyses include the following. Fidrmuc and Hainz (2010) use data on 700 SME loans in Slovakia to model loan default using probit and panel probit analyses and find that indebtedness, liquidity, profitability and sector are all significant determinants of the probability of SME loan default. Behr, Güttler, and Plattner (2004) find an effect of similar financial variables on the probability of default using data on German SMEs. Dyrberg-Rommer (2005) tests the determinants of SME default across Spain, France and Italy, using ratings agency data and finds that earnings and solvency ratios are consistent determinants of default across countries, but that other financial variables have varying effects by country.

This paper first follows previous work by examining the drivers of SME default. A number of Altman's Z-score variables are used, such as the ratio of turnover to total assets and profitability to total assets. Other explanatory variables include the size of the loan in question relative to the borrower's total assets, a measure of leverage, the current ratio (current assets: current liabilities) and a dummy indicating whether the owner of the firm has been with the firm for more than ten years. The major contribution of the paper is in moving beyond the average effect of these variables in the data set to acknowledge borrower heterogeneity. In looking at the drivers of default separately across sectors of economic activity and quintiles of the firm size, exposure and credit quality

[2] Aziz, Emanuel, and H.Lawson (1988), Gombola, Haskins, Ketz, and D.Williams (1987) and Casey and Bartczak (1985).

distributions, the paper points to numerous heterogeneous effects that would be missed in estimating an average effect.[3] One example is the irrelevance of financial ratios in predicting default for the largest firms and the smallest loans. Similarly, the majority of borrower-level determinants are found to have more predictive power where firms are smaller and where loans are larger, suggesting that particular attention must be paid to the financial health of the smallest firms and borrowers with large exposures.

The importance of accounting for borrower heterogeneity is further evidenced in several model validation exercises including receiver operating characteristic (ROC) curves, tests for sensitivity and specificity, out of sample testing and comparative estimates of proportional expected losses (PELs). These findings can improve the predictive and supervisory power of credit risk models relying on average effects across a loan book.

The paper proceeds as follows: Section 2 describes the data used, Section 3 reports results from empirical analysis and Section 5 concludes.

2 Data

As part of the Financial Measures Programme 2011, Irish financial institutions were required to submit their full loan books to the Central Bank of Ireland. These loan files were then used to stress-test the Irish banking system. In the SME case, the full loan files, totalling some 440,000 loans, were generally restricted to loan-level information such as the size of exposure, type of loan, date of origination, maturity, and interest rate, while lacking in information on the borrowing SME. However, a subsample of loans for which rich borrower-level information was available was also extracted from the SME books. It is the Republic of Ireland-based loans of this detailed file that form the basis of this research paper.

The data contain information on 6,745 loans from unique borrowers. Information on the size of exposure as of September 2010, a dummy for Basel II default (90 days past due[4]), sector of activity, and a wide range of balance sheet information are available for the majority of loans in this file. Apart from the exposure size variable, all borrower-level variables were collected for a "base month" by the lender. The base month for 50 percent of borrowers was in 2009, with the rest spread between 2010, 2008 and a small amount in 2007 and earlier.

In the Appendix, Tables A1 to A3 provide a comparison between the data in this paper, which is referred to as the "detailed" data file, and data for 444,000 loans, comprising the majority of the Irish SME lending market, i.e., the "full" data file. In terms of allocation across sectors, Table A1 shows that most sectors are similarly weighted in both datasets. The anomaly is the Agriculture

[3] Accounting for non-linearities of the explanatory variables in the probit model directly has been suggested by Fidrmuc and Hainz (2010) and Lennox (1999) however, our approach allows examination of the effects of the explanatory variables across the distribution of exposure and credit quality explicitly.

[4] Studies have shown that loans at 90 days past due need not always lead to crystallized losses for the lender. For the Irish mortgage market, Kelly (2011) shows that 7.8 per cent of impaired loans in a given time period move back to repayment, while 5 per cent move to full default. Acknowledging this fact, but given the lack of alternative data available, we proceed referring to loans 90 days past due as "defaulted".

& Food sector, which accounts for 25 percent of loans in the full file, but only 5 percent of loans in the detailed file used in this paper. There is also some over-representation in this data: Hotels & Restaurants, Manufacturing, Wholesale & Retail. On exposure size, Table A2 shows that the data used here are three times smaller in terms of the mean, or two times as small in terms of the median, in almost all sectors of activity when compared to the full file. Finally, the risk profiles of the two data files are compared in Table A3. The one year default rate among loans in the detailed file is 7.1 percent, compared to a default rate in the full file of 11.5 percent of loans. This is higher than the 1.4, 6 and 6 per cent for SMEs in the UK, USA and Slovakia found by Lennox (1999), Altman and Sabato (2007) and Fidrmuc and Hainz (2010), respectively.[5]

This skewing of the distribution of loans by sector, size and risk level must be considered when thinking about the generalisability of results in this paper. For example, the data used in this paper covers lower-risk loans of smaller volume than the average, it is possible that estimations of the drivers of default may under-estimate the effect of loan size on default. In this instance, the coefficients presented here act as a lower bound on the true effect.

Figure 1 shows the share of loans observed to be in default in the data, by sector. It is clear from this figure that there is large borrower risk heterogeneity across sectors. Sectors that were closely associated with the mid-2000s boom in Ireland are those which have the highest default rates: Hotels & Restaurants, Construction and Real Estate. This pattern is in line with what is found in the larger dataset on loan performance. Figure 2 reports identical statistics, where the share of *loan volume*, as opposed to the share of *the number of loans*, is plotted. A similar pattern is observed, with the manufacturing sector also reporting high default rates, suggesting that this sector is home to a number of high-volume defaulted loans.

Differences in default probabilities across sectors may be due to several factors including capital intensity, where capital intensity is expected to be positively correlated with dependence on external finance; as well as differential collateral and business models between sectors. However, pronounced differences in default rates across sectors may also indicate underlying structural issues (Fidrmuc and Hainz, 2010), especially in countries undergoing significant economic adjustment, for example, Ireland.

3 Empirical Models of Default

We begin by showing the default variation across sectors observed in the data. In general, sectoral heterogeneity is likely to lead to differing needs for external financing and differing collateral requirements for example, depending on relative capital intensity. In the Irish case, the nature of the credit bubble between the late 1990s and 2008 was such that disproportionate amounts of credit were allocated to speculative sectors such as Construction, Real Estate and Hotels & Restaurants.

[5]Although the time horizons (one year) for the UK, USA and Slovakia default rates are the same as that for the Irish data, the international data were gathered in different time periods, 1987-1994, 1994-2002, and 2001-2005, respectively and for SMEs drawn using differing sample selection criteria, including incorporating only non-financial firms. Further, in some papers "default" refers to corporate liquidation from balance sheet data, while in other papers it refers to defaults within a bank loan book.

Figure 1: Share of Loans in Default, September 2010

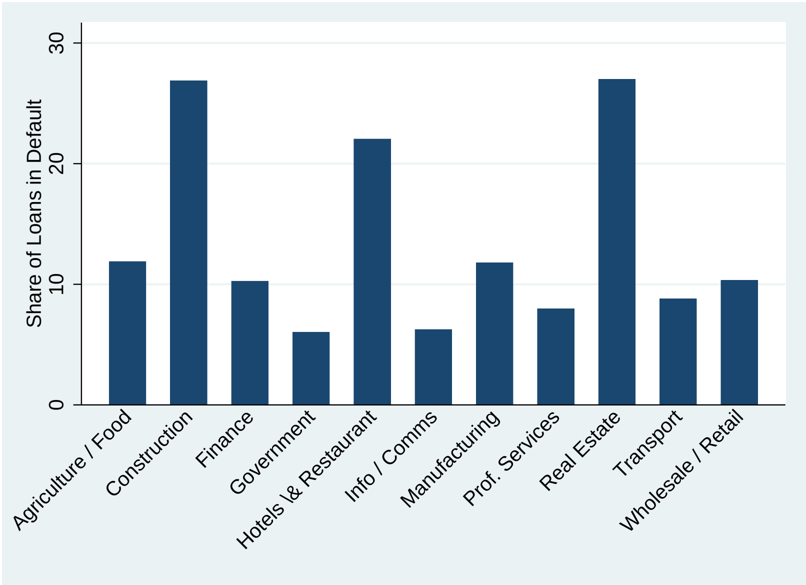
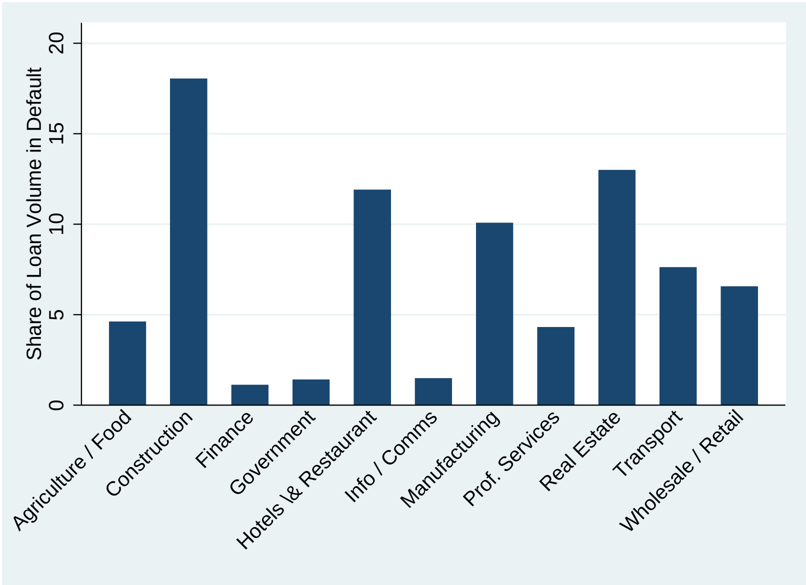


Figure 2: Share of Loan Volume in Default, September 2010.



Indeed, Table 1 shows these sectors as holding the most predictive power in a simple regression default on sector dummies. Sectors that look particularly healthy relative to Agriculture are Public / Local / Health[6] and Professional Services. The sector dummies reported in Table 1 will be included in all regressions from here on, in the interest of space coefficients may not be reported.

Table 1: Sector Dummies and Default. Y is a dummy for Basel II default. Marginal Effects Relative to Agriculture.

	(1)
Construction	0.144***
	(4.52)
Finance	-0.0163
	(-0.56)
Hotels & Restaurants	0.0972***
	(3.50)
Info / Comms	-0.0565
	(-1.27)
Manufacturing	-0.0009
	(-0.05)
Prof. Services	-0.0414**
	(-2.46)
Public / Local / Health	-0.0617***
	(-4.02)
Real Estate	0.148***
	(3.79)
Transport	-0.0311
	(-1.29)
Wholesale / Retail	-0.0160
	(-0.88)
Pseudo R^2	0.0473
N	7209

Marginal effects; t statistics in parentheses
* $p < .1$, ** $p < .05$, *** $p < .01$

The additional financial and qualitative variables of interest in this regression model are:

- The ratio of loan size (exposure) to total assets - a measures of the firm's risk to the bank.
- The ratio of turnover to total assets - a measure of the firm's activity levels.
- A dummy for whether the manager or owner has been with the borrowing firm for ten years or more - a proxy for manager experience and business age, the latter in the absence of a direct measure of this variable.
- The ratio of Earnings Before Interest, Depreciation and Tax (EBITDA) to total assets - the "Profitability Ratio".

[6] This sector includes nursing homes, schools, charities, local voluntary clubs and associations.

- The ratio of cash holdings to total assets - the "Liquidity Ratio".[7]
- The ratio of Current Assets (cash, inventory, receivables) to Current Liabilities (debt and payables) - the "Current Ratio".
- The ratio of Trade Debtors to Total Assets - a measure of soon-to-be-realised assets, i.e. anticipated liquidity.
- The ratio of Total Liabilities to Total Assets - "Leverage Ratio".[8]

Additional variables of interest include firm age, length of time with the borrower and number of borrowing relationships, however, these are unavailable in the dataset.

Table 2 reports summary statistics for each of the above mentioned variables, where the sample is restricted to observations which have data for most variables.

Table 2: Summary Statistics, Full Regression Sample.

Variable	Obs	Mean	Std. Dev.
Exposure:Assets	5,348	0.02	0.228
Turnover:Assets	5,348	2.225	21.626
Owner > 10 years	5,348	0.717	0.451
Leverage Ratio	5,348	0.719	0.528
Liquidity ratio	5,348	0.125	0.318
Current Ratio	5,348	1.471	3.215
Trade Debtors:Assets	5,348	0.187	0.329
Profitability Ratio	5,348	0.089	0.756

Table 3 gives the mean value for each variable of interest for defaulted and non-defaulted borrowers. An examination of the mean values provides motivation for the regression analysis which follows, as the two groups have clear differences in terms of all variables.

Table 3: Mean Values for Defaulted and Non-Defaulted Loans

	Non-Default	Default	P Value H_0: $\bar{Y}_0 = \bar{Y}_1$
Exposure:Assets	0.018	0.047	.0213
Turnover:Assets	2.265	1.65	.6065
Owner > 10 years	0.725	0.601	.0000
Leverage Ratio	0.702	0.953	.0289
Liquidity ratio	0.131	0.042	.0000
Current Ratio	1.54	0.496	.0000
Trade Debtors:Assets	0.19	0.148	.0207
Profitability Ratio	0.097	-0.016	.0072

[7] This is an observed rather than proxy cash flow measure and is thus superior to the measures often seen in the default literature (Gombola, Haskins, Ketz, and D.Williams, 1987).

[8] This measures total liabilities of the borrower, and should not be confused with the ratio of loan size to Total Assets, which measures only the size of the loan for which we have default information.

Table 4 provides the first evidence on the borrower-level characteristics most closely associated with default. Here we report marginal effects (D_y/D_x) calculated after probit regressions where the dependent variable takes a 1 if the loan has been classified as in default.

Table 4 reports marginal effects from regressions where each variable is included separately, with sector dummies included in each model. An examination of the Pseudo R^2 in each column provides an initial indication of the relative importance of the explanatory variables. The largest Pseudo R^2 are found to lie in Columns (4), (5) and (6), suggesting that the leverage ratio, current ratio and liquidity ratio are the three most important explanatory variables in terms of model fit.

The multivariate default model is estimated using all variables included in Table 4 and dummies for each sector included in Figures 1 and 2. Backward stepwise probit regressions, a general-to-specific procedure to establish the model of interest (Darlington, 1990), are used. The procedure begins with a full model and removes variables with a p-value above .05. Marginal effects at the mean are calculated after the stepwise procedure, which explains the appearance of variables that are statistically insignificant. The results in Table 5 report that, for the full sample, a large number of variables are important determinants of default. Firms that took out a large loan relative to total assets, and are highly leveraged overall, are more likely to default. Firms that are liquid, profitable, have managers or owners of lengthy experience and have high volumes of trade debtors are less likely to default. The sectors which exhibit strong propensity to default, unexplained by borrower characteristics in the data, are the Manufacturing, Real Estate and Construction sectors.

As well as presenting results on the borrower-level determinants of default, the validity of the model can also be tested in a number of ways. In categorising whether an observation is defined as defaulted or not, we observe the mean value for the default dummy separately for the sample observations for each formulation of the basic model. If the predicted value for an observation is above the mean value, we define that observation as "predicted to default". Throughout this paper, in analysing the performance of each default model, four statistics are presented.

1. Sensitivity = $Pr(\hat{D} = 1|D = 1)$, i.e. the percentage of the total number of defaults that were correctly predicted to default.

2. Specificity = $Pr(\hat{D} = 0|D = 0)$, i.e. the percentage of the total number of non-defaults that were correctly predicted not to default.

3. False Classified Positives = $Pr(D = 0|\hat{D} = 1)$, i.e. the percentage of the total number of predicted defaults that did not in fact default.

4. False Classified Negatives = $Pr(D = 1|\hat{D} = 0)$, i.e. the percentage of the total number of predicted non-defaults that did in fact default.

A model will be considered to be of higher quality if statistics (1) and (2) are higher and (3) and (4) are lower. There is a tradeoff, however, between minimising (3) and minimising (4), i.e. attempting to minimise one results in an increase in the other (unless the sample size can be increased). In fact, higher (3) is often advantageously considered to be a more conservative modelling

Table 4: Determinants of Default, Univariate Probit Regressions, Marginal Effects Reported

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Exposure to Assets	0.0232**							
	(2.32)							
Turnover:Assets		-0.00207						
		(-1.34)						
Owner > 10 years (d)			-0.0328***					
			(-4.04)					
Leverage Ratio				0.0423***				
				(8.54)				
Current Ratio					-0.0298***			
					(-14.11)			
Liquidity Ratio						-0.249***		
						(-10.49)		
Trade Debtors:Assets							-0.0620***	
							(-3.50)	
Profitability Ratio								-0.0739***
								(-6.27)
Pseudo R^2	0.0250	0.0242	0.0305	0.0488	0.0777	0.0671	0.0283	0.0378
N	5,348	5,348	5,348	5,348	5,348	5,348	5,348	5,348

Marginal effects; t statistics in parentheses

(d) for discrete change of dummy variable from 0 to 1

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 5: Determinants of Default, Backward Stepwise Probit Regression, Marginal Effects Reported

Exposure to Assets	0.0176**
	(2.41)
Owner > 10 years (d)	-0.0197***
	(-3.29)
Leverage Ratio	0.0171***
	(3.70)
Liquidity Ratio	-0.146***
	(-5.63)
Current Ratio	-0.0124***
	(-3.99)
Trade Debtors:Assets	-0.0278**
	(-2.11)
Profitability Ratio	-0.0305***
	(-3.30)
Manufacturing (d)	0.0340***
	(3.21)
Real Estate (d)	0.106***
	(3.61)
Construction (d)	0.0964***
	(5.56)
Finance (d)	0.0646
	(1.64)
Pseudo R^2	0.120
N	5,373

Marginal effects; t statistics in parentheses

(d) for discrete change of dummy variable from 0 to 1

* $p < .1$, ** $p < .05$, *** $p < .01$

of PDs from the point of view of bank capital. Table 6 presents the four statistics for the model of Table 5. The sensitivity (measure (1)) of the model, is high, with 76.64 per cent of all defaulted loans correctly predicted to default. On non-defaulters, the performance is weaker, with 65.57 per cent of all non-defaulters correctly predicted not to default. In terms of erroneous predictions, the false negative rate is only 2 per cent, i.e. a small number of the predicted non-defaulters actually did default. In the parlance of statistical testing, this is equivalent to a low rate of Type 1 errors, where the null hypothesis is that firms do not default. The weak point of the model however is its tendency to falsely predict defaults (Type 2 errors) - of all observations that were predicted to default, 86.5 per cent of those firms did not in fact default. This suggests the model is conservative in that it over-predicts default. Overall, 66.29 per cent of observations were correctly predicted.

Table 6: Measures of Model Performance. Full Sample Specification as in Table 5.

Measure	%
(1) Sensitivity	76.64
(2) Specificity	65.57
(3) Falsely Classified Positives	86.54
(4) Falsely Classified Negatives	2.43
% Correctly Classified	66.29

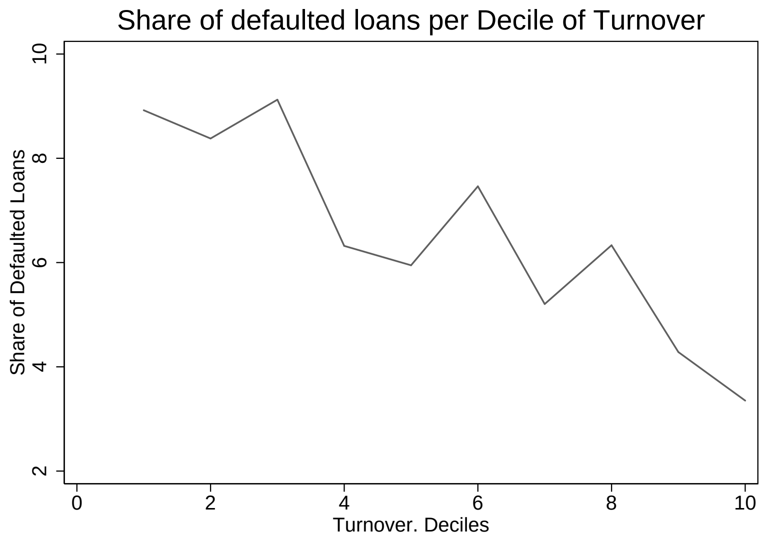
3.1 Beyond Average Effects

The results presented so far measure marginal effects at the mean of each variable of interest. It is of course possible that the effect of certain variables will differ across the distribution. That is, we may expect some explanatory variables to enter the regression non-linearly. In a regression interacting each dependent variable with dummy variables for each quintile of the exposure distribution we find that the Pseduo R^2 rises from .12 in Table 5 to .15, suggesting that accounting for quintile-specific coefficients increases the fit of the model. Motivated by this increase in information, we proceed in this section to run separate models across each quintile of turnover, to model the behaviour of firms according to their size, and each quintile of exposure, to observe the differing effects of independent variables at differing loan sizes.

Figure 3 plots the share of defaulted loans in each decile of turnover, proxying firm size. The picture is one of decreasing risk as firms become larger, with the default rate for the smallest ten per cent of firms being more than double that for the largest ten per cent of firms. This result is well established in the literature on US firms, for example Altman and Sabato (2007) and can also be found for firms in Eastern Europe (Fidrmuc and Hainz, 2010) although the result is weaker for the Slovakian as opposed to the US or Irish cases.

Table 7 reports results from backwise stepwise probit regressions for each quintile of the firm size distribution. The set of potential explanatory variables is identical to that in Table 5. The heterogeneous impact of firm characteristics is evident. Among the smallest sixty per cent of firms,

Figure 3: Default across the distribution of Turnover.



firm risk measures such as liquidity, the current ratio and profitability are shown to matter in two out of three quintiles each. Further, the size of the loan relative to firm assets, the ratio of turnover to assets, and the leverage ratio are significant default predictors in the second quintile, while a dummy for the construction sector is significant across all three quintiles. In the fourth quintile of the turnover distribution, the liquidity and current ratios are found to predict default. Noticeably, the marginal effect is significant for no explanatory variable at the top of the exposure distribution, where only three per cent of loans have defaulted. This suggests that small and large firms display differing behavioural patterns where default is concerned. On the basis of these results, when modelling default risk among the largest firms, it is incumbent on banks to rely on a different information set to that presented here. The explantory power of the model, as measured by the Pseudo R^2, rises as firms get larger, which appears counter-intuitive given that less firm-level control variables are significant for larger firms. This finding suggests that, despite the success of a number of right-hand-side variables in explaining default, further, potentially "soft" information may be crucial for banks in providing more accurate default predictions.

The model for each exposure quintile is validated in Table 8. In line with the finding in Table 7 that the explanatory power of the model increases with turnover, the sensitivity, specificity and falsely classified negatives of the model are found also to improve monotonically as firms get larger.

Figure 4 plots the share of defaulted loans in each decile of exposure to assets (Figure 4a). The figure suggests that default risk does increase decile by decile, i.e. larger loans are more risky. As identified in the figures, the relationship does not appear to be linear, with large jumps in the share of defaulted loans in the top 30 per cent of exposure to assets. As one would expect, exposures that account for the largest shares of the value of the borrowing firm are the most risky. Figure 4b documents the marginal effect of exposure at different points in the exposure to assets (Figure

Table 7: Determinants of Default Across Quintiles of Turnover, Backward Stepwise Probit Regression. Marginal Effects reported

	1st	2nd	3rd	4th	5th
Exposure to Assets		0.729***			
		(2.65)			
Turnover:Assets		-0.0138***			
		(-2.97)			
Leverage Ratio		0.0348***			
		(3.06)			
Liquidity Ratio	-0.139**		-0.0998**	-0.111**	
	(-2.56)		(-1.96)	(-2.37)	
Current Ratio		-0.0283***	-0.0246***	-0.0166***	-0.00734
		(-4.61)	(-4.29)	(-4.02)	(-1.60)
Profitability Ratio	-0.0581**		-0.0432***		
	(-2.39)		(-2.71)		
Construction (d)	0.0901**	0.0834**	0.0990***	0.0504**	0.0252
	(2.20)	(2.05)	(2.82)	(2.02)	(1.58)
Real Estate (d)		0.170**	0.0902	0.115	0.0198
		(2.09)	(1.45)	(1.43)	(0.88)
Manufacturing (d)			0.0586**	0.0186	
			(2.33)	(1.59)	
Pseudo R^2	0.0362	0.129	0.171	0.162	0.233
N	1075	1033	1074	1032	1074

Marginal effects; t statistics in parentheses

(d) for discrete change of dummy variable from 0 to 1

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 8: Model Performance Across Quintiles of Turnover, statistics for each specification of Table 9.

Quintile	1	2	3	4	5
Sensitivity	66.67	71.08	76.39	87.1	85.37
Specificity	58.66	64.42	67.17	61.96	71.93
Falsely Classified Positives	86.75	84.14	85.68	87.23	89.23
Falsely Classified Negatives	5.11	3.77	2.46	1.31	0.8
% Correctly Classified	59.35	64.96	67.78	63.47	72.44

4b) distributions. The non-monotonic effect is evident immediately - at larger values of exposure, a larger marginal effect of exposure on default is estimated (with less precision as exposure to assets gets larger). Given that the share of defaulted loans varies across the distribution, one might also posit that the drivers of default differ. For this reason, the model of Table 5 is extended, by running an identical regression separately for the quintiles of the exposure distribution.[9]

Figure 4: Default across the distribution of Exposure:Assets.

(a) Defaulted Loans by Exposure:Assets

(b) Marginal Effects by Exposure:Assets

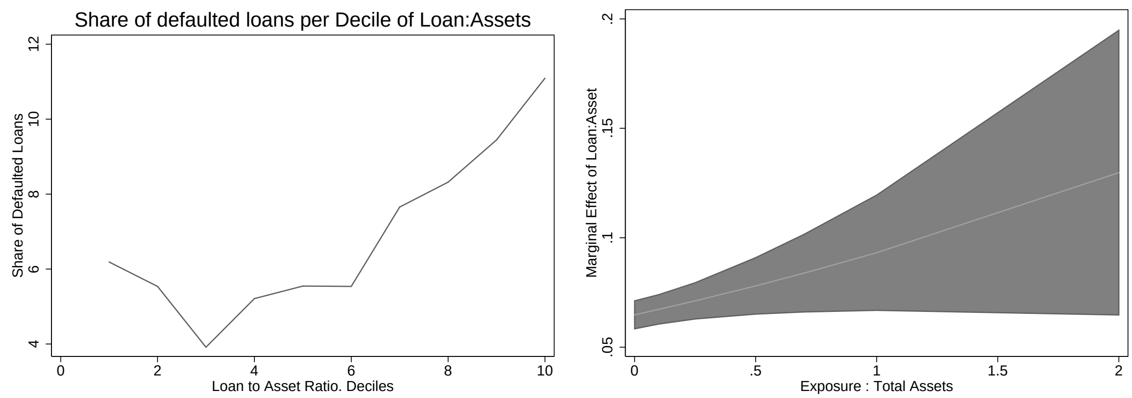


Table 9 confirms that moving beyond the average effect of each variable on default uncovers a large amount of heterogeneity. The effect on the probability of default of a large loan to asset ratio is shown to be driven mainly, as one would expect, by large loan to asset ratios in the 5th quintile of exposure.The effect of this variable on default in all other quintiles is either insignificant, multiple times smaller, or even negative, i.e. in the first quintile. This result points immediately to the importance of moving beyond average effects when analysing the drivers of default.

The firm-level characteristics that are found to most consistently mitigate default are the leverage, liquidity and current ratios, each of which are significant in three quintiles. The Pseudo R^2 of the model is largest in the fifth quintile. This indicates that a large number of indicators of financial health (high turnover to assets, low exposure to assets, low leverage ratio, high liquidity ratio, high current ratio, owner with longevity) have simultaneously most predictive power when exposure size is largest. This suggests a behavioural pattern, whereby firms that engage in large volume borrowing are those for which financial health is most important for predicting PDs. At lower quintiles of the distribution of exposure, less of these characteristics are simultaneously significant. Indeed, for very small loans, i.e. the first quintile of exposure, we find that no indicator of financial health is a significant determinant of default, suggesting that monitoring of borrowers in this quintile is of relatively low importance.

The model for each exposure quintile is validated in Table 10. In line with the finding in Table 9 that the explanatory power of the model increases with loan size, the specificity of the model is

[9]Quintiles rather than deciles are used for the regression analysis in order to preserve sample sizes in each part of the exposure distribution.

Table 9: Determinants of Default Across Quintiles of Exposure, Backward Stepwise Probit Regression. Marginal Effects reported

	1st	2nd	3rd	4th	5th
Exposure to Assets	-0.795				0.375**
	(-1.58)				(2.41)
Turnover:Assets		0.00212**		-0.00823***	-0.00882**
		(2.00)		(-2.65)	(-2.08)
Owner $>$ 10 years (d)	-0.00322		-0.0281*		-0.0279*
	(-0.91)		(-1.81)		(-1.83)
Leverage Ratio	0.00370		0.0277**	0.0276**	0.0305***
	(1.09)		(2.32)	(2.34)	(2.80)
Liquidity Ratio		-0.121***	-0.252***		-0.267***
		(-2.81)	(-4.50)		(-3.56)
Current Ratio	-0.00179	-0.0150***		-0.0221***	-0.0165***
	(-0.92)	(-3.45)		(-4.83)	(-2.58)
Profitability Ratio			-0.0780***		
			(-2.59)		
Real Estate (d)	0.0198	0.0603	0.130*	0.112	0.0971
	(0.89)	(1.33)	(1.93)	(1.45)	(1.45)
Manufacturing (d)	0.00585			0.0573**	0.0393*
	(0.86)			(2.14)	(1.69)
Construction (d)	0.0112	0.0789**	0.0746*	0.175***	0.110**
	(0.97)	(2.54)	(1.92)	(3.19)	(2.16)
Finance (d)		0.141			
		(1.24)			
Hotels & Restaurants (d)				0.0575**	
				(1.98)	
Pseudo R^2	0.104	0.147	0.115	0.154	0.164
N	1,087	1,066	1,047	1,064	1,073

Marginal effects; t statistics in parentheses

(d) for discrete change of dummy variable from 0 to 1

* $p < .1$, ** $p < .05$, *** $p < .01$

found to do likewise, although not in a monotonic way. The model correctly predicts more of the true defaulted loans to default in quintiles 2 and 3 than in the 1st quintile, and in improves again in quintiles 4 and 5 relative to the previous three. The share of falsely classified defaults also falls with loan size up to the 4th quintile, but still remains high at 84.69 percent in the 5th quintile.

Table 10: Model Performance Across Quintiles of Exposure, statistics for each specification of Table 9.

Quintile	1	2	3	4	5
Sensitivity	65	74.19	73.97	77.33	76.54
Specificity	65.34	62.15	66.02	72.4	65.42
Falsely Classified Positives	90.13	89.2	85.97	82.48	84.69
Falsely Classified Negatives	3.03	2.5	2.87	2.32	2.84
% Correctly Classified	65.32	62.85	66.57	72.74	66.26

As with differences across the exposure distribution in Table 9, one should also expect differing effects across sectors of economic activity. Table 11 runs the same regression as that in Table 9 separately for each sector. Due to their small sample sizes, the Finance, Real Estate and Transport sectors have all been subsumed into categories with larger samples. As expected, there is a large degree of variation in the drivers of default. In high-risk sectors such as Construction, Financial/Real Estate/Professional Services and Hotels & Restaurants, large exposures are found to be drivers of default. In the Financial, Professional and Real Estate category, the current ratio, profitability ratio and turnover to assets are all important mitigating factors. In Construction, liquidity and turnover are again important, as is the longevity of the owner or manager. There is no particular variable that consistently determines default across all sectors of activity, with most explanatory variables having statistical power in between two and four sectors.

Table 12 reports the model validation statistics as previously for each sector-specific regression. As in all previous models, the sensitivity and low rate of false negatives are the most impressive features of each model, with specificity being poorer than sensitivity in each case, and a high rate of falsely classified defaults prevalent throughout. There is a large degree of variation on overall model performance, with the share of correctly classified observations ranging from 53 per cent in the combined Finance, Real Estate and Professional Service category to 75 per cent in Agriculture.

Along with examining the data by turnover (a measure of firm size), exposure (a measure of risk) and sector of economic activity, we finally divide the data into quintiles of the borrower quality, to observe whether the explanatory variables driving default differ for low versus high quality borrowers. Observing the Pseudo R^2 of Table 4, the most appropriate proxy for borrower quality is deemed to be the current ratio. In the spirit of Table 9, the backward stepwise regression model is run separately for each quintile of the current ratio distribution. The results reported in Table 13 suggest that among the highest and lowest quality borrowers, little to none of the drivers of default have explanatory power. There are no variables that stand out as having predictive power across a broad range of quality of borrower. The only variable found to drive default among the highest

Table 11: Determinants of Default by Sector, Backward Stepwise Probit Regression. Marginal Effects reported

	(1) Agri/ Food	(2) Construction	(3) Finance	(4) Hotels & Rests.	(5) Manuf.	(6) Prof. Services	(7) Community/ Health	(8) Real Estate	(9) Transport	(10) Wholesale/ Retail
Exposure to Assets		4.467***		0.396*				13.99**		
		(2.85)		(1.76)				(2.10)		
Turnover:Assets		-0.0287**		-0.0144**						
		(-2.13)		(-2.13)						
Owner > 10 years (d)		-0.141***			-0.0192		-0.0262			
		(-3.43)			(-1.56)		(-1.63)			
Leverage Ratio		0.0894**		0.0273*			0.0217**			0.0160**
		(2.50)		(1.74)			(1.99)			(2.28)
Liquidity Ratio		-0.413***				-0.0854***	-0.136***			-0.289***
		(-3.01)				(-2.66)	(-2.75)			(-6.84)
Current Ratio	-0.000208			-0.0741***	-0.0276***				-0.0221	
	(-0.29)			(-5.35)	(-4.75)				(-1.04)	
Trade Debtors:Assets				0.202**			-0.115*			
				(2.42)			(-1.79)			
Profitability Ratio			-0.322**		-0.0376*	-0.0346**				-0.0500**
			(-2.31)		(-1.95)	(-2.50)				(-2.10)
Pseudo R^2	0.262	0.160	0.199	0.106	0.160	0.0461	0.201	0.0771	0.164	0.100
N	261	435	93	503	716	1022	472	158	132	1,511

Marginal effects; t statistics in parentheses

(d) for discrete change of dummy variable from 0 to 1

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 12: Model Performance, statistics for each specification of Table 11.

	(1) Agri/ Food	(2) Construction	(3) Finance	(4) Hotels & Rests.	(5) Manuf.	(6) Prof. Services	(7) Community/ Health	(8) Real Estate	(9) Transport	(10) Wholesale/ Retail
Sensitivity	91.67	73.13	50	75.56	92.16	84.62	78.26	41.67	66.67	83.33
Specificity	73.9	70.65	82.76	56.77	62.71	49.03	73.94	78.36	67.46	52.97
Falsely Classified Positives	85.53	68.79	83.33	85.34	84.07	93.82	86.67	74.36	91.11	91.2
Falsely Classified Negatives	0.54	6.47	4	4.06	0.95	1.23	1.48	11.76	2.3	1.68
% Correctly Classified	74.71	71.03	80.65	58.45	64.8	50.39	74.15	72.78	67.42	54.53

quality of borrower is the leverage ratio.

Table 13: Determinants of Default Across Quintiles of Current Ratio, Backward Stepwise Probit Regression. Marginal Effects reported

	(1) 1st	(2) 2nd	(3) 3rd	(4) 4th	(5) 5th
Exposure to Assets		0.0455**		0.0277**	
		(2.27)		(2.35)	
Owner > 10 years (d)				-0.0250**	
				(-2.02)	
Leverage Ratio			0.0318**		0.0205**
			(2.49)		(2.54)
Liquidity Ratio			-0.266***	-0.0699***	
			(-4.31)	(-2.78)	
Current Ratio		-0.167**			
		(-2.00)			
Profitability Ratio		-0.150***		-0.0334**	
		(-4.51)		(-2.21)	
Manufacturing (d)	0.119*	0.158***			
	(1.79)	(3.73)			
Construction (d)	0.198***	0.105**	0.162***	0.0350	0.0358*
	(3.59)	(2.46)	(3.62)	(1.48)	(1.65)
Real Estate (d)	0.153**		0.336***	0.0550	
	(2.27)		(2.76)	(1.22)	
Finance (d)			0.235*		
			(1.78)		
Pseudo R^2	0.0307	0.0849	0.142	0.129	0.0493
N	1,075	1,067	1,080	923	856

Marginal effects; t statistics in parentheses
(d) for discrete change of dummy variable from 0 to 1
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 14 reports the model diagnostics as previously. The share of correctly classified observation ranges between 70 and 80 per cent in each quintile, with a wide degree of variation in the ability to correctly identify defaulters (sensitivity). Compared with the delineation of the data across quintiles of loan exposure of Table 10, the results here show that a breakdown in terms of borrower quality leads to a higher share of correctly classified observations across the distribution.

4 Further Model Validation

4.1 Model Validation Using ROC Curves

An additional method to validate a binary outcome model is to plot the model's Receiver Operating Characteristic (ROC) curve, widely used to validate ratings systems (Lehmann, 2003). This curve plots the model's sensitivity (ability to correctly predict defaults) against (one minus) its specificity

Table 14: Model Performance Across Quintiles of Current Ratio, statistics for each specification of Table 13.

Quintile	1	2	3	4	5
Sensitivity	30.82	56.00	63.93	75.00	60.00
Specificity	86.76	76.53	79.10	74.42	78.23
Falsely Classified Positives	73.21	80.21	84.52	92.74	93.81
Falsely Classified Negatives	11.14	5.61	2.66	0.89	1.21
% Correctly Classified	79.16	74.60	78.24	74.43	77.80

(ability to correctly predict non-defaults). As the area under the curve gets larger, the model becomes more accurate in predicting defaults. That is, the distribution of ratings is expected to be higher for non-defaulters than for defaulters in well-designed ratings systems (Behr, Güttler, and Plattner, 2004), increasing the area under the ROC curve. In this way models are not evaluated on a single threshold (above or below the average) but on the whole range of thresholds for default/non-default (Lehmann, 2003).

Figure 5 plots the ROC curve and the 45 degree line, which indicates what a purely randomly selected model would look like, for the full sample model of Table 5. The area under this ROC curve is 0.7749. ROC results here, especially at the higher ends of the exposure and credit quality distributions, are in line with those in the literature that use this technique for corporate or SME default modelling. For example, Behr, Güttler, and Plattner (2004) find an ROC of 85.61 per cent for German SMEs and the ROC estimates in Lehmann (2003) range from 0.718 to 0.812.[10]

In the interests of space, we do not plot any further ROC curves, but rather document the area underneath the curve for each model presented in Section 3.1 (Table 15).

Table 15: Area under the ROC curve for the models of Tables 9 and 13

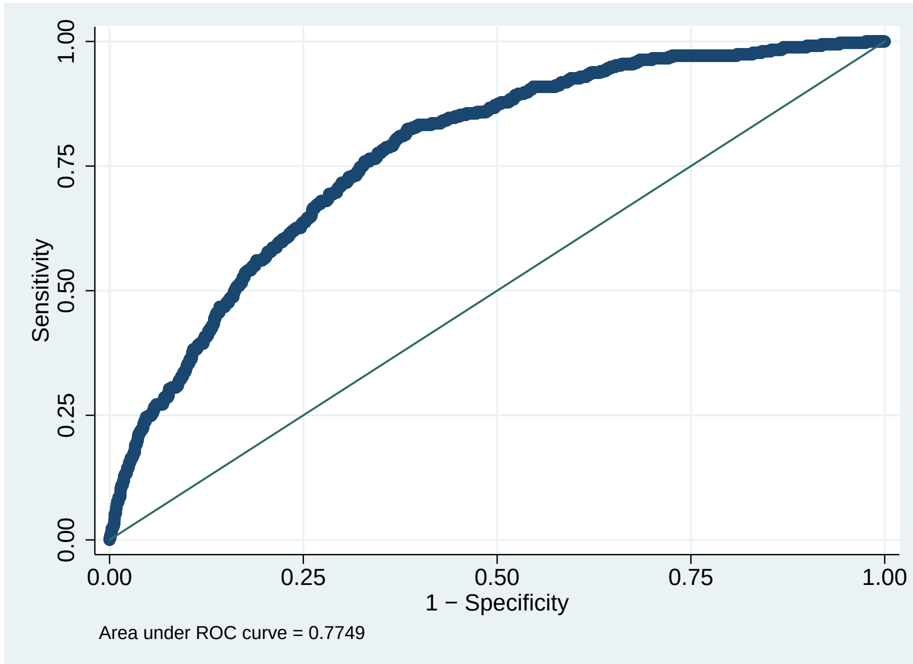
Quintile	1	2	3	4	5
Exposure (Table 9)	0.762	0.773	0.769	0.797	0.805
Current Ratio (Table 13)	0.590	0.719	0.801	0.810	0.749

4.2 Out of Sample Validation

To complement the model validation within sample, an out-of-sample (OOS) model validation is also carried out. We begin with a process that returns uniformly distributed random variates on the interval $[0, 1)$. We then remove the top quintile of this randomised distribution, and perform an estimation identical to the model of Table 5 on the first four quintiles of the random distribution. Results of this estimation are reported in Table 16, which when compared with those in Table 5, show that randomly removing one fifth of the data does not change the results significantly. One difference

[10] Although the usual caveats apply when comparing statistics on samples used in other studies.

Figure 5: ROC curve for the full model of Table 5



of note is that the ratio of trade debtors to total assets loses significance, while some sector dummies change significance. The coefficients on all other borrower-level explanatory variables retain very similar coefficients. We then retrieve the coefficients on each variable in Table 16 and apply these coefficients to the out of sample data in the fifth quintile of the randomly generated distribution. An analysis similar to that in Table 6 can then be carried out, in comparing the predicted and actual defaults, where a predicted default is an observation with a predicted value above the mean of the default variable in the fifth randomised quintile, .07.

Table 17 reports the model validation results. Compared to Table 6, the number of overall correctly specified observations is slightly improved at 69.7 per cent. Similar problems arise in falsely predicting defaults - of all observations predicted to default, 83.9 per cent did not default, although this is lower than the 86.5 per cent in-sample. The specificity of the model, along with the share of falsely classified non-defaults, is also marginally improved out of sample compared to the in-sample tests of Table 6.

4.3 Extension: Measuring Proportional Expected Losses

While the scope of this paper is not to provide a direct stress-testing application for the Irish SME lending market, the tools used thus far allow illustrations of how proportional expected losses (PEL) vary for active (non-defaulting) SMEs across sectors and points in the exposure distribution. Firms observed to have defaulted in the data have been excluded from the analysis of PEL to facilitate the validation of our model when comparing active and defaulting firms as well as comparing active

Table 16: Determinants of Default, Backward Stepwise Probit Regression. Four Quintiles of Randomly Generated Distribution. Marginal Effects reported

	(1)
Exposure to Assets	0.0168**
	(2.07)
Owner > 10 years (d)	-0.0179***
	(-2.68)
Leverage Ratio	0.0155***
	(3.01)
Liquidity Ratio	-0.134***
	(-4.77)
Current Ratio	-0.0135***
	(-4.12)
Profitability Ratio	-0.0271***
	(-2.59)
Transport (d)	-0.0281***
	(-3.40)
Real Estate (d)	0.0682**
	(2.55)
Wholesale / Retail (d)	-0.0180***
	(-3.12)
Construction (d)	0.0659***
	(3.95)
Prof. Services (d)	-0.0185***
	(-2.92)
Pseudo R^2	0.120
N	4279

Marginal effects; t statistics in parentheses
(d) for discrete change of dummy variable from 0 to 1
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 17: Measures of out-of-sample model performance. Sample Coefficients from Table 16 Imposed on remaining one fifth of the data.

Measure	%
(1) Sensitivity	84.5
(2) Specificity	68.3
(3) Falsely Classified Positives	83.9
(4) Falsely Classified Negatives	1.6
% Correctly Classified	69.7

firms under alternative PDs.

Our benchmark is a “Model PD” provided in the data set which was generated by the banks using a Basel II-compliant credit-risk model. It is expected that these “internal” PDs are based on “through the cycle” Basel II models. This means that the models are designed to minimise the requirement to alter their capital levels through the business cycle. Through the cycle models result in higher capital requirements in economic upturns and vice versa, than would be seen if pro-cyclical models were used. Capital smoothing is an attractive characteristic of these standardised models. PDs generated using our model, however, are estimated during a deep and protracted economic contraction and are thus expected to reflect current economic circumstances much more closely.

Our calculation for PEL is extremely simple, and assumes a recovery rate of zero:

$$EL = \frac{\sum PD^{*}Exposure}{\sum Exposure} \tag{1}$$

i.e. it is the sum of the exposure of each loan multiplied by its estimated probability of default from the models presented earlier in the paper, divided by the total exposure. This is calculated twice, once using the PDs estimated in this paper, and once using the PDs (Model PD) provided in the dataset.

Table 18 presents the PEL of the model of Table 5, broken down by sector and quintile of exposure. Column (2) provides PEL associated with the PDs calculated in Table 11 in separate models for each sector of activity. Column (3) provides the total PEL in each sector using the internal PDs (given in the data). It should be noted that these are the PEL for all non-defaulted loans. In line with the pattern presented in Figures 1 and 2, the PEL estimates calculated from the models of Table 11 predict large expected losses in the Construction, Hotels and Restaurants and Real Estate sectors. Similarly, column (3) suggests the internal PDs are derived from a “through the cycle” credit rating procedure, as expected.

Table 19 presents PEL by sector of activity, where the model is run separately within each sector as in Table 11. The PEL estimates in Column (1) do not differ greatly from those in Table 18. In both Tables, in all sectors, the estimated PEL in Column (1), calculated from the model is two to three times as large as that in Column (3). These results, with expected losses from this paper’s model reflect actual default rates in each sector more closely than an internal PD model. In a severe economic downturn, our approach provides support for a detailed granular approach which accounts for heterogeneity of borrowers in different sectors of activity.

5 Conclusion

This paper aims to contribute to the literature on “fundamentals-based” models of corporate default. It is one of a small number of papers that exploit micro-level loan data on defaults among SME borrowers. Using unique borrower-level balance sheet information for a cross-section of 6,000 Irish SME loans, this paper tests the determinants of default at the micro level. Typical financial ratios,

Table 18: Share of Proportional Expected Losses in Total Exposure, Full Sample of Non-Defaulter Loans. Model of Table 5

Sector	Model PEL	Internal PEL
Agriculture & Food	0.041	0.020
Construction	0.136	0.031
Finance	0.044	0.012
Hotels & Restaurant	0.063	0.045
Info & Comms	0.040	0.016
Manufacturing	0.065	0.025
Prof. Services	0.037	0.024
Community / Health	0.044	0.018
Real Estate	0.158	0.042
Transport	0.043	0.023
Wholesale & Retail	0.052	0.029
Quintiles of Exposure		
1	0.068	0.035
2	0.064	0.036
3	0.059	0.033
4	0.057	0.031
5	0.055	0.026

Table 19: Share of Proportional Expected Losses in Total Exposure for each sector, Sector-Specific Model of Table 11

Sector	Model PEL	Internal PEL
Agriculture / Food	0.032	0.020
Construction	0.124	0.031
Finance	0.029	0.012
Hotels & Restaurant	0.086	0.045
Manufacturing	0.059	0.025
Prof. Services	0.036	0.024
Public / Local / Health	0.035	0.018
Real Estate	0.145	0.042
Transport	0.044	0.023
Wholesale / Retail	0.050	0.029

such as the ratio of the loan to total assets, the current ratio, leverage ratio, liquidity ratio and profitability ratio, are found to be significant predictors of default. Further, the length of time the borrowing firm's owner or manager has been with the firm mitigates the likelihood of default. Conditional on the above, significant sector-level effects remain.

The paper moves beyond average effects of the above-mentioned variables by repeating the analysis across seven sectors of economic activity, and across the quintiles of the firms' size, loan exposure and credit quality distributions. The results suggest that different warning signals can be identified, particularly for smaller firms and borrowers with small versus large loans.

The importance of accounting for borrower heterogeneity is further evidenced in several robustness tests including receiver operating characteristic (ROC) curves, tests for sensitivity and specificity, out of sample testing and comparative estimates of proportional expected loss (PEL). These findings can improve the predictive and supervisory power of credit risk models relying on average effects across a loan book, especially during severe economic downturns, and contribute to the literature on "fundamentals-based" modelling of corporate default risk.

Repeated sampling of the Irish SMEs loans in the future may allow regressions on quintiles of PDs predicted here for updated versions of the data.

References

Altman, E. I. (1968): "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," *Journal of Finance*, 23(4), 589–609.

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

Aziz, A., D. C. Emanuel, and G. H.Lawson (1988): "Bankruptcy Prediction - An Investigation Of Cash Flow Based Models," *Journal of Management Studies*, 25(5), 419–437.

Behr, P., A. Güttler, and D. Plattner (2004): "Credit Scoring and Relationship Lending: The Case Of German SME," .

Berger, A. N., and W. S. Frame (2007): "Small Business Credit Scoring and Credit Availability," *Journal of Small Business Management*, 45(1), 5–22.

Casey, C., and N. Bartczak (1985): "Using Operating Cash Flow Data to Predict Financial Distress: Some Extensions," *Journal of Accounting Research*, 23(1), pp. 384–401.

Chan-Lau, J. A. (2006): "Fundamentals-Based Estimation of Default Probabilities: A Survey," Working Paper Series 06/149, International Monetary Fund.

Darlington, R. (1990): *Regression and Linear Models*. Wiley, New York.

DIETSCH, M., AND J. PETEY (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*, 28, 773–788.

DYRBERG-ROMMER, A. (2005): "A Comparative Analysis of the Determinants of Financial Distress in French, Italian and Spanish firms," Working Paper Series 26, Danmarks Nationalbank.

FIDRMUC, J., AND C. HAINZ (2010): "Default rates in the loan market for SMEs: Evidence from Slovakia," *Economic Systems*, 34(2), 133–147.

GOMBOLA, M. J., M. E. HASKINS, J. E. KETZ, AND D. D.WILLIAMS (1987): "Cash Flow in Bankruptcy Prediction," *Financial Management*, 16(4), pp. 55–65.

JACOBSON, T., J. LINDÉ, AND K. F. ROSZBACH (2005): "Credit Risk versus Capital Requirements under Basel II: Are SME Loans and Retail Credit Really Different?," *Journal of Financial Services Research*, 28, 43–75.

KELLY, R. (2011): "The Good, The Bad and The Impaired - A Credit Risk Model of the Irish Mortgage Market," Research Technical Papers 13/RT/11, Central Bank of Ireland.

LEHMANN, B. (2003): "Is It Worth the While? The Relevance of Qualitative Information in Credit Rating," Working paper, EFMA 2003 Meetings.

LENNOX, C. (1999): "Identifying failing companies: a re-evaluation of the logit, probit and DA approaches," *Journal of Economics and Business*, 51(4), 347–364.

MERTON, R. C. (1974): "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates," *Journal of Finance*, 29(2), 449–70.

WESTGAARD, S., AND N. VAN DER WIJST (2001): "Default probabilities in a corporate bank portfolio: A logistic model approach," *European Journal of Operational Research*, 135(2), 338–349.

A Supplementary Tables

Table A1: Share of Sectors in Total Number of Loans, Detailed versus Full file.

	Detailed Data		Full Sample	
	Count	%	Count	%
Agriculture / Food	322	4.8	110,251	24.8
Construction	529	7.8	35,519	8
Financial	147	2.2	33,062	7.4
Hotels & Restaurants	683	10.1	22,184	5
Manufacturing	817	12.1	28,571	6.4
Professional & Real Estate	1,530	22.7	83,483	18.8
Public, Community & Local	747	11.1	37,581	8.5
Transport & Comms	206	3.1	20,331	4.6
Wholesale & Retail	1,730	25.7	73,102	16.5
Total	6,742	100	444,084	100

Table A2: Mean and Median Exposure Size, Detailed versus Full file.

	Mean Exposure		Median Exposure	
	Detailed	Full	Detailed	Full
Agriculture & Food	13,127	32,464	5,325	10,200
Construction	7,448	22,659	2,980	8,253
Finance	25,232	113,321	4,570	9,034
Hotels & Restaurant	16,296	164,125	5,860	11,035
Manufacturing	10,146	61,712	4,460	12,091
Professional & Real Estate	9,432	33,121	4,165	6,250
Public, Community & Local	55,289	47,988	6,570	12,882
Transport & Comms	15,236	35,138	3,845	10,000
Wholesale & Retail	11,010	64,757	5,030	12,697
Total	16,334	53,034	4,700	10,000

Table A3: Default Rates, Detailed versus Full file.

	Detailed Data			Full Sample		
	Performing	Default	Sample	Performing	Default	Sample
Agriculture & Food	0.941	0.059	322	0.952	0.048	105,887
Construction	0.843	0.157	529	0.780	0.220	35,197
Financial	0.939	0.061	147	0.820	0.180	24,408
Hotels & Restaurants	0.896	0.104	683	0.837	0.163	20,969
Manufacturing	0.924	0.076	817	0.864	0.136	27,163
Professional & Real Estate	0.939	0.061	1,530	0.890	0.110	81,378
Public, Community & Local	0.960	0.040	747	0.920	0.080	35,469
Transport & Comms	0.956	0.044	206	0.864	0.136	19,916
Wholesale & Retail	0.943	0.057	1,730	0.864	0.136	71,642
Total	0.929	0.071	6,742	0.885	0.115	422,029