



BUSINESS CREDIT INFORMATION SHARING AND DEFAULT RISK OF PRIVATE FIRMS.

MAIK DIERKES

Finance Center Münster, University of Münster

CARSTEN ERNER

Finance Center Münster, University of Münster

THOMAS LANGER

Finance Center Münster, University of Münster

LARS NORDEN

Rotterdam School of Management, Erasmus University

MoFiR working paper n° 64

March 2012

Business credit information sharing and default risk of private firms

Maik Dierkes^a, Carsten Erner^a, Thomas Langer^a, Lars Norden^{b,*}

^a Finance Center Münster, University of Münster, 48143 Münster, Germany

^b Rotterdam School of Management, Erasmus University, 3000 DR Rotterdam, the Netherlands

Abstract

We investigate whether and how business credit information sharing helps to better assess the default risk of private firms. Private firms represent an ideal testing ground because they are smaller, more informationally opaque, riskier, and more dependent on trade credit and bank loans than public firms. Based on a representative panel dataset that comprises private firms from all major industries, we find that business credit information sharing substantially improves the quality of default predictions. The improvement is stronger for older firms and those with limited liability, and depends on the sharing of firms' payment history and the number of firms covered by the local credit bureau office. The value of soft business credit information is higher for smaller and less distant firms. Furthermore, in spatial and industry analyses we show that the higher the value of business credit information the lower the realized default rates. Our study highlights the channel through which business credit information sharing adds value and the factors that influence its strength.

JEL classification: D82; G21; G32; G33

Keywords: Credit risk; Asymmetric information; Credit bureau; Hard and soft information; Private firms

* Corresponding author (L. Norden). Address: Department of Finance, Rotterdam School of Management, Erasmus University, Burgemeester Oudlaan 50, 3000 DR Rotterdam, the Netherlands. Tel.: +31-10-4082807; fax: +31-10-4089017; *E-mail address:* lnorden@rsm.nl.

1. Introduction

Theoretical and empirical research has examined various effects of credit information sharing on lenders, borrowers, and economic activity. Most of these studies document significantly positive effects of information sharing, such as an increase in the supply of credit by banks, a decrease in the costs of credit and realized default rates, and an increase in GDP growth. However, almost entirely missing from these studies is a direct empirical examination of the channel through which these positive effects of information sharing occur. Does information sharing raise the overall accuracy of default predictions in an economy? Which factors influence the potential improvement in default prediction accuracy associated with information sharing? What is the link between credit information sharing, the quality of ex ante default predictions, and ex post default rates? In this paper we make an attempt to fill this gap by investigating whether business credit information sharing helps to better assess the default risk of private (unlisted) firms and which factors influence the magnitude of its marginal value. Our evidence is based on the improvement of default prediction accuracy, which we study in aggregate, firm-specific, industry-specific, and spatial analyses.

Theory has highlighted several important aspects of credit information sharing (e.g., Milion and Thakor, 1985; Pagano and Jappelli, 1993; Padilla and Pagano, 1997, 2000; Karapetyan and Stacescu, 2010). Later empirical work has provided evidence in support of these theories based on international and country-specific analyses. A first set of empirical studies documents benefits of credit information sharing in a cross-country context (e.g., Jappelli and Pagano, 2002; Miller, 2003; Djankov et al., 2007; Brown et al., 2009; Tsai et al., 2011). A second set of studies investigates the effects of credit information sharing within countries (e.g., Kallberg and Udell, 2003; Doblaz-Madrid and Minetti, 2010; Hertzberg et al., 2011). For example, Kallberg and Udell (2003) analyze data from the US credit bureau Dun & Bradstreet on 2,723 retailing firms from the late 1980s to study whether business credit informa-

tion is valuable in assessing borrower quality. They show that information on firms' payment history, the PAYDEX score of Dun & Bradstreet, is significantly related to the firm's probability of survival. We complement and extend the study of Kallberg and Udell (2003) in several dimensions. We analyze the marginal benefit of business credit information sharing by comparing the default prediction accuracy of models that include and do not include business credit information. Moreover, we study which factors influence the potential improvement in default prediction accuracy and differentiate between the value-added of hard and soft business credit information. Our analysis is based on a large representative dataset that includes firms from all major industries. Moreover, Doblus-Madrid and Minetti (2010) examine the case of information sharing among equipment finance companies. They analyze equipment loans to almost 4,000 US businesses and find that the staggered entry of firms in the credit bureau is associated with a drop in firms' delinquency rates by 13% on average (37% for the maximum). This finding can be largely attributed to the disciplining effect of credit information sharing. It becomes more costly for firms to default or to be past due with payments when credit information is shared not only among current but also potential future lenders. For comparison, we take a different perspective by studying the link between the improvement in ex ante default prediction accuracy due to business credit information sharing and ex post default rates in a spatial and industry analysis, respectively. Another important question is whether credit information sharing has unambiguously positive effects. The study of Hertzberg et al. (2011) provides evidence on this issue. Based on a natural experiment in Argentina where the public credit registry was expanded to all firms, the study documents an amplification of coordination effects among lenders when negative credit information was shared in response to the credit registry reform. This study is one of the rare that documents significant negative real effects of credit information.

Our study is based on a representative panel dataset from the largest commercial credit bureau in Germany (Creditreform). The biggest European economy is the ideal testing ground to study effects of credit information sharing since the vast majority of all German firms are private and small firms (Federal Statistical Office, 2009). For example, 96% of all German firms are considered as small- and medium-sized enterprises according to the definition of the European Commission (European Commission, 2006). The randomly drawn sample comprises more than 25,000 firms over the period from 2002 to 2005 (resulting in approximately 100,000 firm-year observations). Creditreform regularly screens the official German commercial register for new entries to ensure a full coverage of all firms. We focus on private firms because they are smaller, more informationally opaque, riskier, and more dependent on trade credit and bank loans than public firms (e.g., Petersen and Rajan, 1994; Norden and Weber, 2010; Saunders and Steffen, 2011). We expect that potential benefits of business credit information sharing are particularly relevant for private firms. Moreover, private firms, especially SMEs, are of key importance for economic activity, employment and innovation in many countries.

We conduct three sets of empirical analyses to identify the benefits of business credit information sharing and factors that influence its magnitude. First, we compare the accuracy ratio of default prediction models that include and do not include business credit information. The accuracy ratio represents a widely used measure in the credit risk literature and credit rating industry and indicates the aggregate default prediction accuracy in a sample (e.g., Cantor and Mann, 2003; Engelmann et al., 2003). Second, we define a measure that indicates the likelihood of improvement in firm-specific default predictions due to the addition of business credit information and investigate in probit regression models which factors determine this measure. We also examine the differential impact of firm-specific factors on the value-added of hard and soft information, such as payment history or business outlook (for parallels to the

banking literature, see Berger and Udell, 2002; Petersen, 2004; Berger et al., 2005; Grunert et al., 2005). Third, we test for potential real effects of business credit information sharing. In a spatial and industry analysis, we investigate whether there is a relation between the improvements in the accuracy of ex ante default risk assessments due to business credit information sharing and realized default rates.

Our first result is that business credit information sharing substantially improves the accuracy ratio of default predictions for private firms. The additional consideration of this information increases the accuracy ratio of default predictions by approximately 20 percentage points. We confirm this result in out-of-the sample tests and by means of a type I and II error analysis. It is noteworthy that this finding is present in most industries (although its magnitude varies substantially), which suggests that the effect is more general than suggested by earlier studies. In a second set of analyses we provide evidence on the factors that influence the magnitude of the improvement. Default predictions become more accurate for older firms and those with limited liability. This finding is mainly driven by the sharing of information on firms' payment history and the extent of firm concentration on local credit bureau offices. Interestingly, the value of soft business credit information decreases in firm size and distance from the corresponding local credit bureau office. Third, in a spatial and industry analysis, we find that stronger improvement in the ex ante accuracy of default predictions of private firms leads to lower realized default rates in the same geographic area. Various additional empirical checks confirm that our results are robust and not the product of particular choices of samples, methods, or model specifications.

Our study contributes to the literature on credit information sharing and the growing field of research on private firms in three ways. We highlight a not yet explicitly analyzed channel through which the positive effects of business credit information sharing occur: the improvement in the accuracy of aggregate and firm-specific default predictions. None of the related

papers has directly examined the existence and functioning of this channel. We also provide comprehensive evidence on the factors that influence the strength of the channel. This channel is indeed effective: more accurate default predictions due to business credit information are associated with lower future default rates.

The remainder of the paper is organized as follows. In Section 2 we describe the institutional background and data. In Section 3 we report our findings on the value of business credit information sharing and on the factors that influence its strength. In Section 4 we investigate the link between ex ante prediction accuracy and ex post default risk in a spatial and industry analysis. In Section 5 we summarize several additional empirical checks that confirm the robustness of our findings. Section 6 concludes.

2. Institutional Background and Data

We obtain the data from Creditreform, the largest commercial credit bureau in Germany. Creditreform provides a wide range of services, such as business data collection and exchange, credit scoring based on various default prediction models, factoring, receivables management, dunning, and debt collection (see Creditreform, 2007). Its main product is the Credit Status Index, which is an aggregated credit score that reflects a firm's probability of default. To generate the Credit Status Index, Creditreform exploits all public sources of information (e.g., financial and business media, trade registers, registers of residents, courts, etc.) as well as various non-public sources. The latter comprise information obtained from interviews with the company's management and data on the firms' payment behavior obtained from the firm's business partners. Whenever a firm requests a commercial report about a business partner from the credit bureau, that firm itself is requested to provide its experiences with and information about its business partners to the credit bureau in return. This information exchange based on the principle of reciprocity also applies to firms' payment history. Creditreform's

business model is similar to the one of Dun & Bradstreet in the United States (see Kallberg and Udell, 2003), and its database comprises more than 3.5 million entries. Creditreform regularly screens the official German commercial register for new entries, thereby ensuring a full coverage of all firms. Firms that are included in Creditreform's database do not self-select but are added by the credit bureau itself or the firms' suppliers. 96% of all firms in Germany are private and small firms (see Federal Statistical Office 2009), making the country an ideal testing ground for investigating factors that influence the value of credit information sharing on private firms.¹

The dataset comprises a representative, randomly drawn panel of 25,344 private firms spanning the period from 2002 to 2005. The final sample consists of 97,174 firm-year observations. For each of the years from 2002 to 2005 we observe the attributes that Creditreform considers to generate a firm's score for the credit status index. These attributes include, among others, the firm's annual sales, age, legal form, ZIP code, and industry. Moreover, we have for each firm four attributes that reflect business credit information. These attributes are based on a firm's payment history with business partners (PAYMENT), its creditworthiness (CREDIT), information about the business outlook (OUTLOOK), and its order book (ORDER), which is gathered and evaluated by business analysts of Creditreform. We further use information about the physical distance between the firm's headquarter and the credit bureau office that covers the firm. We gather the latter information manually from Creditreform's website (www.creditreform.de) and then calculate the geodesic distance between a firm's ZIP code and the corresponding Creditreform office.

¹ Similar to other countries there is also a public credit registry in Germany. Banks have to report all new bank loans that exceed 1.5 million Euros to the Deutsche Bundesbank and only banks have access to this registry. However, two characteristics limit the scope of the public credit registry. First, because of the reporting threshold, the vast majority of loans to micro and small businesses are not included. Second, the public credit registry covers only bank debt. Thus, sharing credit information via the largest private commercial credit bureau Creditreform is particularly important for small and private firms and trade credit.

We also consider the number of employees of the credit bureau office that covers a firm. We gather the total number of employees for each Creditreform office from the Amadeus database and then calculate the ratio of firms covered to the number of employees (FIRMS_PER_EMPLOYEE). Due to limited availability of data we were able to calculate this ratio for 50 of the 119 Creditreform offices. Table 1 presents descriptive statistics.

(Insert Table 1 here)

Panel A of Table 1 shows that one third of all firms have LIMITED_LIABILITY. The distribution of AGE is skewed to the right with a mean of 16.5 and a median of 12.5 years. The logarithm of SALES amounts to 12.8, which corresponds to roughly 400,000 Euros. A firm's DISTANCE to its corresponding Creditreform office averages to 22 kilometers and varies greatly with a standard deviation of about 19.5 kilometers. However, about half of all firms exhibit a distance of less than 16 kilometers to their Creditreform office. The mean of FIRMS_PER_EMPLOYEE is 10. We further report statistics of the four business credit information factors used by Creditreform (PAYMENT, CREDIT, ORDER and OUTLOOK), which are measured on a scale from 1 to 70 (higher numbers indicate higher default risk). We observe 3,069 default events (bankruptcy filings), which corresponds to an average default rate of 3.16% (DEF).² This number is consistent with related studies (e.g., Dietsch and Petey, 2004). Except for these firm defaults and/or recoveries, the panel is balanced. Panel B of Table 1 shows the means of the default rates and the four business credit information factors by terciles of firm and credit bureau characteristics (except for LIMITED_LIABILITY, which is a dummy variable). It can be seen that both default rates and business credit information fac-

² Our default indicator variable (DEF) equals one if a firm defaults in the subsequent year, and zero otherwise. DEF equals one for 624 firms in 2002, 756 firms in 2003, 712 firms in 2004, and 977 firms in 2005.

tors vary by tercile categories. We differentiate by terciles here and in the subsequent analyses to be able to detect potentially non-monotonic patterns.

3. Empirical Analysis

3.1. Business credit information and aggregate default prediction accuracy

In a first set of analyses, we investigate whether business credit information helps to improve default predictions based on an aggregate prediction accuracy measure. We estimate one default prediction model that includes public information and business credit information and another model that excludes the business credit information. Specifically, we estimate the following two probit regression models with DEF as dependent variable to obtain the probability of default over a one year horizon for all firms in our sample.³

$$P(DEF = 1) = f(\text{base factors}) \quad (1)$$

$$P(DEF = 1) = f(\text{base factors}, \text{hard bci factors}, \text{soft bci factors}) \quad (2)$$

The explanatory variables in these models are the same as in Creditreform's model for the Credit Status Index. The first model includes only the base factors that the credit bureau collects from public sources of information (equation 1, "baseline model"). Base factors capture effects due to firms' legal form, age, size, industry, and other dimensions. The second model includes all factors considered by the credit bureau (equation 2, "full model"), including the four attributes that reflect business credit information (hard bci factors: PAYMENT and

³ The one year prediction horizon is the industry standard and also required by banking regulations (Basel II; see Basel Committee on Banking Supervision, 2006). We report results from cross sectional-time series pooled probit models. We obtain similar results when we estimate a probit random effects model. We cannot estimate logit and probit fixed effects models because of computational problems, especially a lack of convergence.

CREDIT; soft bci factors: ORDER and OUTLOOK). We include year fixed effects as controls in both models.

Based on these estimations we calculate the accuracy ratio (AR). The AR is a widely used measure of the aggregate quality of default prediction models and can be derived from cumulative accuracy profile curves. To obtain the latter, firms are ordered according to their estimated probability of default on the x-axis (from high to low). The percentage of correctly classified defaulters is displayed on the y-axis. The cumulative accuracy profile curve of a random default prediction model corresponds to the 45 degree line. The AR, defined in the interval $[0, 1]$, equals the ratio of the area between the cumulative accuracy profile of the rating model and the random model over the area between the cumulative accuracy profile of a perfect model and the random cumulative accuracy profile. The higher the AR the higher is the prediction accuracy of the model (for details, see, e.g. Cantor and Mann, 2003; Engelman, Hayden, and Tasche, 2003). We then examine the difference in the ARs to assess the improvement in default predictions. If the difference is positive, then the addition of business credit information helps to improve default forecasts. Note that for confidentiality reasons we cannot report the marginal effects or coefficients of the base factors in both models. Table 2 reports the results.

(Insert Table 2 here)

We obtain two key results. First, three of the four business credit information factors (PAYMENT, CREDIT and ORDER) in model (2) are positive and highly significant (i.e., higher values for these factors suggest a higher default risk). This finding indicates that credit business information helps predicting future firm defaults and confirms the evidence on US retailing firms provided by Kallberg and Udell (2003). Second and more important in our

context, the aggregate accuracy ratio AR increases by 19.80 percentage points (from 42.89% to 62.69%) when we compare model (1) and (2). Moreover, the adjusted McFadden- R^2 more than triples when we add business credit information factors to the base factors (from 0.052 to 0.161). This finding supports the view that business credit information helps to improve the accuracy of default predictions.⁴ The difference in the accuracy between the full model and the baseline model corresponds to the area between the two cumulative accuracy profiles plotted in Figure 1.

(Insert Figure 1 here)

We also investigate whether the increase in prediction accuracy is robust across industry sectors. For this purpose, we take advantage of the fact that our sample covers all 53 industries in Germany (based on the first two digits of the industry classification code). Calculating the difference in the accuracy ratio between the full and the baseline model for each industry separately shows that in 38 of 45 industries (84.4%) the additional consideration of business credit information increases the accuracy ratio.⁵ The improvement ranges from 2.9 to 73.5 percentage points. On the one hand, public sector firms and rental companies display only small improvements in the accuracy ratio (2.9 and 3.9 percentage points, respectively). Public sector companies may largely depend on the politicians' willingness to cover losses which makes their default behavior more idiosyncratic. Rental companies pursue a very specific business model (e.g., long-term leasing by car rental companies). On the other hand, engineering and construction firms and car manufacturers exhibit differences of 33.2 and 35.8 percentage points, respectively. Both heavily depend on trade credit provided by various suppliers.

⁴ Furthermore, we conduct an in-the-sample vs. out-of-the-sample analysis and examine the type I and II errors. Both analyses lead to similar conclusions. We report the results of these additional empirical checks in Section 5.

⁵ We exclude eight of the two-digit industry categories because there were no default events during our sample period in these categories.

We note that the influence of business credit information on the default prediction accuracy varies substantially across industries. This finding is novel since previous studies have either focused on a cross-country perspective or firms from single industries.

3.2. Factors that influence the aggregate default prediction accuracy

We now investigate whether and how firm and firm-credit bureau characteristics influence the improvement in aggregate default prediction accuracy. For this purpose, we split the firms in our sample into terciles based on the following characteristics: LIMITED_LIABILITY (binary split), AGE, SALES, DISTANCE, and FIRMS_PER_EMPLOYEE. We apply a tercile split instead of a median split to detect potential non-monotonic effects. We then enter the firm's characteristics in the baseline and full model and calculate for each tercile sample the difference in the accuracy ratios. Note that we do not recalibrate the baseline and full model on the tercile samples since Creditreform calibrates its model also on the full database and not on subsamples. Table 3 reports the results.

(Insert Table 3 here)

Most important, we confirm the aggregate result for all tercile subgroups, i.e., all reported differences in default prediction accuracy are positive and economically significant. We find that business credit information is more valuable for firms with limited liability since the difference in accuracy ratio with and without business credit information is almost six percentage points higher for these firms (26.26 vs. 20.65). If there is unlimited liability, a firm's default risk largely depends on its owner's willingness and ability to let the firm survive, creating more ambiguity about the likelihood of default. For the variables AGE and DISTANCE we find a monotonically increasing value of business credit information. One possible expla-

nation for this result is that there are more cross-sectional and time-series data available (i.e., a higher number of business partners and a higher number of repeated interactions with the same business partner) for older firms than for younger firms, allowing the older ones to establish a more reliable credit history. Our interpretation for the result on DISTANCE is that business credit information sharing reduces the information asymmetry arising from distance via the exchange of hard information (e.g., information on the firm's payment history with its suppliers, and not soft information). We revisit the differential role of hard and soft business credit information in the next section. For SALES, our proxy for firm size, we find improvement of default prediction accuracy in all terciles which is consistent with the related literature that shows that credit information sharing is more beneficial for relatively small and opaque firms. The value of business credit information is highest for medium-sized firms. Recall that all firms in our sample are private firms and predominately small and medium-sized enterprises (SMEs). Thus, the non-monotonic pattern indicates a relative size effect among all private firms. This pattern can be explained with the trade-off between firms' transparency and firms' reliance on trade credit. On the one hand, relatively large firms might be less sensitive to credit information sharing since there is more public information on these firms available and they have a higher number of bank relationships than relatively small firms (e.g., voluntary disclosure, size-based accounting and disclosure requirements). On the other hand, relatively small firms may only have a single supplier and/or do not use trade credit but pay their supplies by using cash or bank debt, which renders credit information sharing either unnecessary or less efficient. This reasoning predicts that medium-sized firms benefit the most from credit information sharing. For FIRMS_PER_EMPLOYEE (the ratio of the number of firms covered by a Creditreform office to the number of employees of that office), we find a strong peak in the upper tercile while the first two terciles do not differ much. This result in-

dicates that the prediction accuracy increases when credit bureau analysts cover a substantial number of firms, which can be explained with learning effects and an increase in experience.

3.3. Factors that influence the firm-specific default prediction accuracy

The accuracy ratio, which we considered in our first set of analyses, is an aggregate measure of default prediction accuracy. It is not possible to calculate this measure for individual firms, which means that a multivariate analysis based on firm-year observations cannot be conducted. Therefore, we define a second measure that indicates improvements in default accuracy predictions at the individual firm level, using the probit models outlined in the previous section. Equation (3) provides the definition of this measure:

$$\Delta PD = \begin{cases} 1 & \text{if } (PD^{wBCI} > PD^{woBCI}) \text{ and } DEF = 1 \\ 1 & \text{if } (PD^{wBCI} < PD^{woBCI}) \text{ and } DEF = 0 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

The dummy variable ΔPD equals one if the firm-specific estimated probability of default from the full model with business credit information (PD^{wBCI}) is higher than the estimated probability of default from the baseline model without business credit information (PD^{woBCI}) for defaulters, or if the firm-specific estimated probability of default from the model with business credit information (PD^{wBCI}) is lower than the estimated probability of default from the model without business credit information (PD^{woBCI}) for non-defaulters, and zero otherwise. In other words, ΔPD indicates whether the additional incorporation of business credit information changes the estimated probability of default for a firm in the right direction. The mean of this variable in the full sample is 0.78, suggesting that the additional consideration of business credit information helps adjusting the estimated probability of default in the correct direction for three out of four firms. We also calculate the mean improvement in the firm-

specific default prediction by terciles of firm characteristics. We find that the mean improvement in firm-specific default predictions is positively related to `LIMITED_LIABILITY` and `AGE`, and u-shaped in `FIRMS_PER_EMPLOYEE`.

We now continue with analyzing how firm and firm-credit bureau characteristics influence firm-specific default prediction improvements in a multivariate setting. To do so, we estimate a probit model with the dummy variable ΔPD as dependent variable and the following explanatory variables: `LIMITED_LIABILITY` (dummy), `AGE`, `SALES`, and `DISTANCE`. In an alternative specification we additionally include `FIRMS_PER_EMPLOYEE` since we do not observe this variable for all firms. Similar to the analysis in the previous section, we use tercile dummy variables (with the lower tercile, T1, as reference category) to allow for non-monotonic effects. We control for year, industry and regional fixed effects in all models. Table 4 reports the regression results.

(Insert Table 4 here)

The analysis yields several interesting insights. In model (1) we find that business credit information significantly helps to improve the default prediction if a firm has `LIMITED_LIABILITY` and `AGE` is high. Both variables show significantly positive marginal effects. The effect is monotonic in `AGE`, i.e., business credit information sharing improves defaults predictions the more the older the firm. For the variables `SALES` and `DISTANCE`, we do not obtain significant results. Thus, the corresponding effects shown in Table 3 do not survive in a multivariate setting.

We further find in model (2) that the more firms are covered per Creditreform office employee the more likely that the default prediction accuracy is improved as the effect of the upper tercile dummy for `FIRMS_PER_EMPLOYEE` is significantly positive. Recall that we

have the number of Creditreform employees only for 50 out of 119 offices, which results in a lower number of observations when we include this variable. The peak in the upper tercile is consistent with learning effects for the credit bureau analysts. This result resembles findings on the behavior of loan officers who exhibit learning in the number of interactions with borrowers in the cross-section and over time (e.g., Scott 2004; Alessandrini et al., 2009; Liberti and Mian, 2009). However, while loan officers rotate after a number of years to mitigate incentives for under-, over-reporting or collusion (e.g., Hertzberg et al., 2010), these problems are largely absent for Creditreform business analysts because the credit bureau does not lend to the firm that is analyzed and the compensation of the analysts is fixed. All these effects are statistically significant at the 0.01-level. Moreover, we confirm the previous results by estimating random effects panel regression models. The effects for LIMITED_LIABILITY and AGE remain unchanged in model (3). In model (4), we further confirm the impact of FIRMS_PER_EMPLOYEE. However, the inclusion of business credit information now leads to a more accurate estimate of the probability of default for the largest firms in our sample (SALES_T3).

We conclude that the limited liability, age and learning of the credit bureau's analysts have a significant and robust influence on the value of business credit information for default predictions.

3.4. Determinants of the value-added of hard and soft business credit information

In the above analyses we considered all four attributes of business credit information, without differentiating between the type of information included in these attributes. However, there is evidence from the empirical banking literature that suggests that the type of information, especially hard vs. soft information, matters for its value-added (e.g., Berger and Udell, 2002; Petersen, 2004; Berger et al., 2005; Grunert et al., 2005; Liberti and Mian, 2009; Hertz-

berg et al., 2010; Norden and Weber, 2010). We exploit in our study that some of the attributes of business credit information are based on hard information while others are based on soft information. The remainder of this section deals with the differential impact of hard and soft business credit information.

As described beforehand, the credit bureau considers the firm's payment history and overall creditworthiness as hard information, while it considers the order book situation and general business outlook as soft information. This distinction between hard and soft is largely based on the way the information is generated: the former two factors are mostly based on quantitative information about a firm's payment history provided by the firm's suppliers, while the latter two factors are qualitative assessments made by the credit bureaus' business analysts on the basis of interviews with the firm management, onsite visits, and additional written information provided by the firm.

To investigate the marginal value of different types of business credit information, we now gradually extend our analysis, using (i) a baseline model without business credit information, (ii) the baseline model plus two factors that reflect hard business credit information (payment history and overall creditworthiness), and (iii) a full model, which includes the baseline factors, hard business credit information, and soft business credit information (order book and business outlook). A comparison of the cases (ii) and (i) informs about the marginal value of hard information and a comparison of the cases (iii) and (ii) informs about the extra marginal value of soft information. The motivation for this specific order of gradually adding hard and then soft business credit information is that hard information is usually easy to collect, almost costless, easy to store and to assess. Therefore, consistent with wide-spread practice in the banking and credit rating industry, default prediction models typically first include public hard information and are then extended by including private hard and private soft in-

formation (e.g., Treacy and Carey, 2000; Grunert et al., 2005). Table 5 reports the regression results.

(Insert Table 5 here)

The value of hard business credit information fully mirrors the results for the previous regression models, in which we included all four attributes of business credit information (columns (1) and (2)). Put differently, the hard business credit information is a key factor that drives our previous results. This finding has an interesting parallel to the recent research on sources of private information in banking (e.g., Mester et al., 2007; Norden and Weber, 2010). Related studies document that proprietary information on credit line usage and checking account activity help banks to better predict borrower defaults and manage their credit relationships with firms. Norden and Weber (2010) show that these benefits are most pronounced in the case of SMEs and individuals but rather weak for large firms. That result is consistent with our finding on the importance of hard business credit information, which initially was private information of the firm, its supplier, and their banks before it was shared with others through the credit bureau. We again find a positive impact of `FIRMS_PER_EMPLOYEE` on the marginal value of hard business credit information in model (2), indicating effects from learning and experience.

Moreover, we see that the marginal value of soft information is also related to firm characteristics. It is higher for firms with limited liability and older firms. Furthermore, soft business credit information becomes more useful for smaller firms, as indicated by the negative coefficient of `SALES`. Similarly, the value of soft business credit information decreases in `DISTANCE`. Both results are consistent with findings from the banking literature (e.g., Petersen and Rajan, 2002; Degryse and Ongena, 2005; DeYoung et al., 2008; Alessandrini et al.,

2009; Liberti and Mian, 2009). As we control for regional fixed effects in all regressions, the fact that credit bureau offices are more likely to be located in urban areas where firms and/or industries tend to be more profitable is an unlikely explanation. Unlike for hard information, we obtain a significantly negative coefficient for the mid-tercile dummy (T2) of `FIRMS_PER_EMPLOYEE`. This result indicates a non-monotonic u-shaped relation, suggesting that there are benefits from soft information production arising from focus and specialization (T1) as well as from learning and experience (T3). We note that such non-monotonic pattern is not uncommon and has been documented in various other contexts (e.g., Bandiera and Rasul, 2006).

4. Improvement in default prediction accuracy and realized credit risk

The previous literature suggests that credit information sharing has positive real effects. For instance, Jappelli and Pagano (2002) use survey data in cross-country analysis to show that credit information sharing leads to a reduction in credit risk. In this section we investigate whether there is a link between the accuracy of ex ante credit quality assessments (including business credit information) and realized default rates. An improvement in default prediction accuracy represents a macro-economic efficiency gain because the overall allocation of credit in the economy becomes more efficient. If default prediction accuracy is improved, then more healthy firms and less unhealthy firms should obtain credit because of a reduction of type I and II errors. In the extreme case, i.e., under a perfect default prediction model (everything else equal), the credit allocation would be optimal; only healthy firms (ex post non-defaulters) would obtain credit while unhealthy firms (ex post defaulters) do not obtain credit. Surprisingly, none of the related empirical studies on information sharing has directly tested the hypothesis that a higher quality of ex ante default predictions leads to lower ex post default rates in an economy. Unlike cross-country studies, we can use firm default data because we inves-

tigate the hypothesis based on a representative dataset from one large country, which circumvents the comparability problems of default risk measures between countries.

To investigate the potential link between prediction accuracy and realized default rates empirically, we create subsamples based on spatial and industry segmentation. We ensure to have (1) a sufficient number of segmentation units for the regression analysis, (2) a sufficient number of firms per segmentation unit to calculate meaningful accuracy ratios, and (3) a sufficient variation in realized default rates and default prediction accuracy across segmentation units. It is virtually impossible to conduct such analysis at the individual firm-level because default events are rare events and our sample period spans only four years.⁶

We start with a spatial segmentation of our dataset in two ways. First, we create pools of firms from a 2-digit ZIP Code area, and, second, pools of firms within a 50 kilometer radius around the center of every 5-digit ZIP Code area. The first criterion leads to a non-overlapping spatial segmentation (small sample size), while the second criterion leads to an overlapping spatial segmentation (large sample size). For the latter, we use a relatively small radius because the business structure in Germany is predominantly local. We exclude those pools for which the estimation of our models is not possible or not reliable (e.g., too few observations, negative ARs; in total, less than 5% of all observations).

We then regress the yearly realized default rate per region (DEF_RATE , calculated from our default indicator variable DEF) on the year-specific difference of the accuracy ratios of the full model and baseline model (ΔAR) in the same region. In this regression we control for the year-specific level of default prediction accuracy of the baseline model (without business credit information; $AR^{Baseline}$), yearly regional GDP growth, and year fixed effects. If the coefficient of our key variable ΔAR is negative and statistically significant, we will be able to

⁶ Even a very long time series per firm would not solve this problem because it would come along with other serious problems (e.g., structural economic, socio-economic, legal, or political breaks).

conclude that there is a link between the improvement in the accuracy of default predictions and realized default rates in the same geographic area. Table 6 reports the regression results.

(Insert Table 6 here)

Panel A of Table 6 shows the findings for the spatial segmentation. Model 1 refers to the non-overlapping 2-digit ZIP code segmentation and Model 2 to the overlapping 5-digit ZIP code segmentation. The result is clear and consistent across both spatial segmentations. The greater the improvement of the accuracy ratio due to the additional consideration of business credit information the lower is the realized default rate. The negative coefficient of ΔAR is significant at the 0.01-level. This result establishes the link between our previous results and evidence from related studies. We document that the improvement in ex ante credit quality assessments that arises from business credit information sharing is associated with lower realized default rates in the same geographic area. We also find that the coefficients of the control variables $AR^{Baseline}$ and GDP_GROWTH are significantly negative, which is consistent with our expectation. The negative coefficient of $AR^{Baseline}$ suggests that default rates are lower the better the default prediction of the baseline model (before adding business credit information). Moreover, the negative impact of GDP_GROWTH is consistent with evidence that default rates decline in economic upturns.

We rerun the same analysis using an industry segmentation based on the 2-digit industry code. Panel B of Table 6 reports the results. We again find ΔAR to be negative and highly significant. Thus, an improvement of the default prediction accuracy due to the additional consideration of business credit information in an industry leads to significantly lower realized default rates. The effect of $AR^{Baseline}$ is not significant with this industry-based segmentation. Our interpretation is that some industries are too opaque or too idiosyncratic to have a

reasonable level of default prediction accuracy based on public information. This fact once more highlights the value of business credit information documented in Section 3.1.

Moreover, in unreported robustness checks, we confirm our results for the geographic and industry-based segmentation by orthogonalizing ΔAR on AR^{Baseline} . A similar approach has been chosen by Agarwal and Hauswald (2010) to eliminate the informational overlap between public and private credit assessments made by banks. The main reason for this additional test is that the value of business credit information is correlated with the level of the default prediction accuracy without such information (i.e., there is a significantly negative correlation between ΔAR and AR^{Baseline}). It turns out that the results remain similar.

5. Additional empirical checks

We conduct several additional empirical checks to ensure that our previous results are robust and not the product of particular choices of samples, methods, or model specifications. Specifically, we compare the in- vs. out-of-the sample prediction accuracy, consider type I and II errors as alternative evaluation criteria, provide further evidence on the dynamics of our default predictions, and study the link between default prediction accuracy and realized default rates by industries.

First, to examine the robustness of the increase in prediction accuracy measured by increases in the accuracy ratio, we conduct a bootstrap simulation. In each simulation run we randomly split the sample in two equally large parts, the calibration and the prediction set. The calibration set is used to fit the above mentioned two models. We then calculate the accuracy ratios for both models, separately for the calibration set (in-the-sample) and prediction set (out-of-the-sample). Using these accuracy ratios, we obtain two differences in prediction accuracy between the full and baseline model, one in-the-sample and one out-of-the-sample. This procedure is repeated for 1,000 simulation runs. The mean of the in-the-sample differ-

ences is 19.56 percentage points, while the mean of the out-of-the-sample differences is even higher with 21.61 percentage points. Both values are economically large and significantly different from zero. We conclude that the value of business credit information is robust and confirmed by the out-of-the-sample analysis.

Second, an alternative approach to assess the default prediction accuracy is an examination of the type I and type II error. Type I error refers to cases where firms are predicted as non-defaulters but they default, while type II error refers to cases where firms are predicted as defaulters but they do not default. These errors are based on a binary classification for default and non-default. Given the estimated probability of default from the probit regression models, the binary classification can be derived by applying a cut-off value. If the firm's estimated probability of default lies above the cut-off value, we predict that the firm will default and vice versa. To derive the optimal cut-off value we apply a cost function weighing the number of false-negative and false-positive with different costs. The related literature has emphasized that false-negatives lead to substantially higher costs than false-positives (e.g., Grunert, Norden and Weber, 2005). Therefore, we assign higher error costs to type I errors than to type II errors by setting the cost ratio of false-negatives over false-positives to 20:1. Assuming an interest rate of 5%, a non-predicted default would lead to a loss of the face value of 100 Euro (false-negative) and a denying of a non-default credit (false-positive) would lead to a loss of the potential interest payment of 5 Euro. Based on this cost ratio we derive the optimal cut-off points for the full and the baseline model. For these optimal cut-off points the full model reduces the number of false-negatives by 25.8% and the number of false-positives by 25.0%. Hence, the analysis of type I/II errors confirms that business credit information helps to improve the default prediction accuracy.

Third, to study the dynamic aspects of the firm-specific default prediction accuracy in a different way we carried out the following test. Based on the variable ΔPD that we defined in

Section 3.3 we create a new dependent variable that takes a value of one if the default prediction accuracy decreases in each of the four years, two if the default prediction accuracy is increased in some and decreased in other years, and three if the default prediction accuracy is increased in all of the four years. We then collapse the dataset by calculating the means of the explanatory variables used in Table 4 and conduct a cross-sectional ordered probit regression with industry and regional fixed effects. We obtain findings that are very similar to those reported in Section 3.3.

6. Conclusion

In this study we provide a direct examination of whether and how business credit information sharing helps to better assess the default risk of private firms. The analysis is based on a representative panel dataset from the largest commercial credit bureau in Germany and includes firms from all major industries.

We obtain three main results. First, we find that business credit information sharing substantially improves the accuracy of aggregate and firm-specific default predictions. We interpret our result as novel and direct evidence for the channel that explains why credit information sharing exerts a positive influence on credit availability, cost of credit and realized credit risk. In other words, through this channel (i.e., the improvement in default prediction accuracy associated with business credit information sharing) it is possible to achieve a better credit allocation in the economy. While the effect is found in most industries we also measure a substantial heterogeneity in the value of business credit information between industries. This finding is also new since previous studies are either conducted at the country level or based on firms from single industries. Second, we provide evidence on the factors that influence the magnitude of the value of business credit information sharing for private firms. The default prediction accuracy is improved for older firms and those with limited liability, and it depends

on the sharing of firms' payment history and the number of firms covered by a local credit bureau office. The value of soft business credit information sharing is higher for smaller and less distant firms. Third, we show that the higher the value of credit business information the lower the realized default rates. This result is confirmed in spatial and industry analyses and provides direct evidence that the improvement in default prediction accuracy due to credit information sharing serves as a channel that leads to a more efficient credit allocation.

We extend and complement the existing literature by providing new evidence on the channel through which business credit information sharing adds value and on the factors that influence its strength. Because private firms, especially SMEs, are of key importance for economic activity, employment and innovation in many countries, we believe that our study may have broader implications about the impact of business credit information sharing.

References

- Agarwal, S., Hauswald, R., 2010. Distance and Private Information in Lending. *Review of Financial Studies* 23, 2757-2788.
- Alessandrini, P., Presbitero, A., Zazzaro, A., 2009. Banks, Distances and Firms' Financing Constraints. *Review of Finance* 13, 261-307.
- Bandiera, O., Rasul, I., 2006. Social Networks and Technology Adoption in Northern Mozambique. *Economic Journal* 116, 869-902.
- Basel Committee on Banking Supervision, 2006. International Convergence of Capital Measurement and Capital Standards. Bank for International Settlements, June 2006.
- Berger, A., Miller, N., Petersen, M., Rajan, R., Stein, J., 2005. Does function follow organizational form? Evidence from the lending practices of large and small banks. *Journal of Financial Economics* 76, 237-269.
- Berger, A., Udell, G., 2002. Small Business Credit Availability and Relationship Lending: The Importance of Bank Organisational Structure. *Economic Journal* 112, F32-F53.
- Brown, M., Jappelli, T., Pagano, M., 2009. Information sharing and credit: Firm-level evidence from transition countries. *Journal of Financial Intermediation* 18, 151-172.
- Cantor, R., Mann, C., 2003. Measuring the Performance of Corporate Bond Ratings, Special Comment, Report #77916. Moody's Investor's Service, 04/2003.
- Creditreform, 2007. Our Company. Neuss, Germany.
- Degryse, H., Ongena, S., 2005. Distance, Lending Relationships, and Competition. *Journal of Finance* 60, 231-266.
- DeYoung, R., Glennon, D., Nigro, P., 2008. Borrower-lender distance, credit scoring, and loan performance: Evidence from informational-opaque small business borrowers. *Journal of Financial Intermediation* 17, 113-143.

- Dietsch, M., Petey, J., 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 and Finance* 28, 773-788.
- Djankov, S., McLiesh, C., Shleifer, A., 2007. Private credit in 129 countries. *Journal of Financial Economics* 84, 299-329.
- Doblas-Madrid, A., Minetti, R., 2010. Sharing Information in the Credit Market: Contract-Level Evidence from U.S. Firms. Working Paper, Michigan State University.
- Engelmann, B., Hayden, E., Tasche, D., 2003. Measuring the Discriminative Power of Rating Systems, Discussion paper 01/2003, Series 2: Banking and Financial Supervision, Deutsche Bundesbank.
- European Commission. 2006. European business: Facts and figures 1995-2005.
- Federal Statistical Office, 2009. Business Register. Local units with employees liable to pay social insurance contributions in 2007. Wiesbaden, Germany.
- Grunert, J., Norden, L., Weber, M., 2005. The Role of Non-Financial Factors in Internal Credit Ratings. *Journal of Banking and Finance* 29, 509-531.
- Hertzberg, A., Liberti, J., Paravasini, D., 2010. Information and incentives inside the firm: Evidence from loan officer rotation. *Journal of Finance* 65, 795-828.
- Hertzberg, A., Liberti, J., Paravasini, D., 2011. Public Information and Coordination: Evidence from a Credit Registry Expansion. *Journal of Finance*, 379-412.
- Jappelli, T. and Pagano, M., 2002. Information Sharing, Lending and Defaults: Cross-Country Evidence. *Journal of Banking and Finance* 26, 2017-2045.
- Kallberg, J., Udell, G., 2003. The Value of Private Sector Business Credit Information Sharing: The US Case. *Journal of Banking and Finance* 27, 449-469.
- Liberti, J. and Mian, A., 2009. Estimating the Effect of Hierarchies on Information Use. *Review of Financial Studies* 22, 4057-4090.

- Mester, L., Nakamura, L., Renault, M., 2007. Transaction accounts and loan monitoring. *Review of Financial Studies* 20, 529-556.
- Miller, M., 2003. Credit reporting around the globe. In: Miller, M.J. (Ed.), *Credit Reporting Systems and the International Economy*. MIT Press, Cambridge.
- Millon, M., Thakor, A., 1985. Moral Hazard and Information Sharing: A Model of Financial Information Gathering Agencies. *Journal of Finance* 40, 1403-1422.
- Norden, L., Weber, M., 2010. Credit line usage, checking account activity, and default risk of bank borrowers. *Review of Financial Studies* 23, 3665-3699.
- Padilla, A., Pagano, M., 1997. Endogenous Communication Among Lenders and Entrepreneurial Incentives. *Review of Financial Studies* 10, 205-236.
- Padilla, A., Pagano, M., 2000. Sharing Default Information as a Borrower Discipline Device. *European Economic Review* 44, 1951-1980.
- Pagano, M., Jappelli, T., 1993. Information Sharing in Credit Markets. *Journal of Finance* 43, 1693-1718.
- Petersen, M., 2004. *Information: Hard and Soft*. Working Paper, Kellogg Graduate School of Management, Northwestern University.
- Petersen, M., Rajan, R., 1994. The Benefits of Lending Relationships: Evidence from Small Business Data. *Journal of Finance* 49, 3-37.
- Petersen, M., Rajan, R., 2002. Does Distance Still Matter? The Information Revolution in Small Business Lending. *Journal of Finance* 57, 2533-2570.
- Saunders, A., Steffen, S., 2011. The Costs of Being Private: Evidence from the Loan Market. *Review of Financial Studies*, forthcoming.
- Scott, J., 2004. Small Business and the Value of Community Financial Institutions. *Journal of Financial Services Research* 25, 207-230.

Treacy, W., Carey, M., 2000. Credit Risk Rating Systems at Large US Banks. *Journal of Banking and Finance* 24, 167-201.

Tsai, H., Chang, Y., Hsiao, P., 2011. What drives foreign expansion of the top 100 multinational banks? The role of the credit reporting system. *Journal of Banking and Finance* 35, 588-605.

Figure 1

Cumulative Accuracy Profiles

This figure shows the cumulative accuracy profiles of different default prediction models. The baseline model only includes publicly available information while the full model additionally includes the four attributes for business credit information. The full model yields an accuracy ratio of 62.69% and the baseline model yields an accuracy ratio of 42.89%. Using the cumulative accuracy profile, the accuracy ratio is defined as the ratio of the area between the respective model and the random model over the area between the perfect model and the random model.

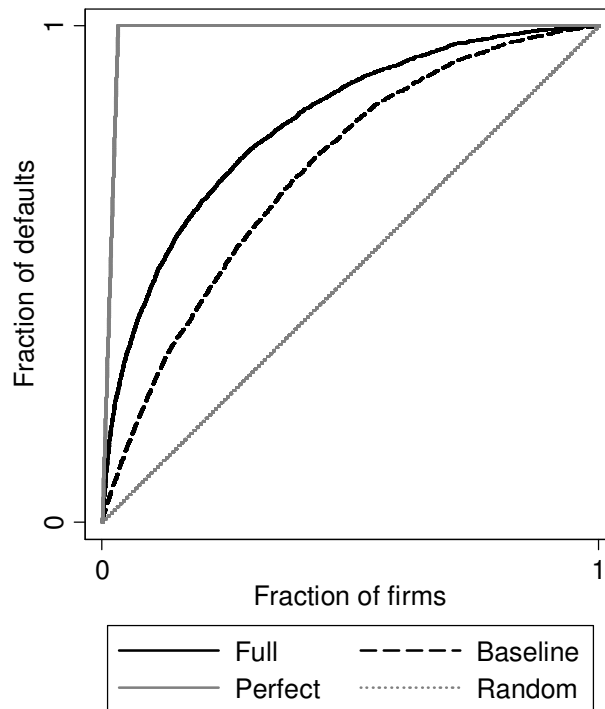


Table 1

Summary statistics

We obtain our data from Creditreform, which is the largest business credit bureau in Germany. The sample includes data on 25,344 private firms spanning the period from 2002 to 2005 (97,174 firm-year observations). Panel A reports descriptive statistics of the main variables. Panel B shows the mean default rates and mean business credit information assessments by Creditreform differentiated by terciles of firm and credit bureau characteristics.

Panel A: Descriptive statistics of the main variables

Variable name	Variable description	Full sample			Obs.
		Mean	Median	St. Dev.	
LIMITED_LIABILITY	Indicator variable which equals one if the firm has limited liability and zero otherwise	0.33	-	-	97,174
AGE	Age of the firm in years	16.46	12.50	15.23	97,174
SALES	Logarithm of a firm's annual sales in Euros	12.89	12.63	1.52	97,174
DISTANCE	Distance between a firm's head office and the corresponding Creditreform office in kilometers	21.76	16.13	19.44	97,174
FIRMS_PER_EMPLOYEE	Ratio of the number of firms covered by a Creditreform office to the total number of employees of that office	10.24	8.93	4.79	43,216
PAYMENT	Assessment payment history (scale from 1 to 70)	22.52	21.00	5.93	97,174
CREDIT	Assessment creditworthiness (scale from 1 to 70)	24.03	21.00	5.16	97,174
ORDER	Assessment order book situation (scale from 1 to 70)	30.39	31.00	3.32	97,174
OUTLOOK	Assessment business outlook (scale from 1 to 70)	30.75	31.00	3.46	97,174
DEF	Indicator variable which equals one if the firm has filed for bankruptcy and zero otherwise	0.03	-	-	97,174

Table 1 (continued)

Panel B: Default rates and business credit information by terciles

Variable	Default rates	Business credit information factors			
		Hard information		Soft information	
	DEF	PAYMENT	CREDIT.	ORDER	OUTLOOK
LIMITED_LIABILITY					
=0	0.039	23.00	24.37	30.66	30.96
=1	0.016	21.57	23.35	29.84	30.33
AGE					
Lower tercile	0.042	22.75	24.44	30.45	30.71
Mid tercile	0.030	22.71	24.36	30.39	30.79
Upper tercile	0.022	22.08	23.26	30.32	30.74
SALES					
Lower tercile	0.045	23.08	24.69	30.91	31.11
Mid tercile	0.032	23.03	24.12	30.62	30.97
Upper tercile	0.018	21.40	23.25	29.62	30.14
DISTANCE					
Lower tercile	0.036	22.48	23.59	30.25	30.56
Mid tercile	0.030	22.43	24.01	30.43	30.79
Upper tercile	0.028	22.65	24.48	30.48	30.90
FIRMS_PER_EMPLOYEE					
Lower tercile	0.026	21.71	22.50	30.27	30.08
Mid tercile	0.029	22.34	23.20	30.33	30.84
Upper tercile	0.038	22.59	23.81	30.39	30.89

Table 2

Estimation results for default prediction models

This table reports results from cross sectional-time series pooled probit regressions. Model (1) includes the base factors from Creditreform (legal form, size, industry, and other variables). Because of confidentiality reasons we cannot report the estimated effects of the base factors. Model (2) includes the base factors and the four business credit information factors from Creditreform (PAYMENT, CREDIT, ORDER and OUTLOOK). The sample includes data on 25,344 private firms spanning the period from 2002 to 2005. We include year fixed effects in both models. ***, **, * indicate that marginal effects are statistically significant at the 1%, 5%, and 10%-level, using robust standard errors clustered within firms. Marginal effects (mfx) are reported as average marginal effects.

	(1) Baseline model (Probit Pooled) mfx	(2) Full model (Probit Pooled) mfx	t-stat.	
Base factors	Yes	Yes		
PAYMENT	No	0.0023	28.44	***
CREDIT	No	0.0025	22.57	***
ORDER	No	0.0006	2.67	***
OUTLOOK	No	0.0002	0.93	
Year fixed effects	Yes	Yes		
Adj. McFadden R ²	0.052	0.161		
Accuracy Ratio (AR)	42.89%	62.69%		
Number of observations	96,320	96,320		

Table 3

Improvement of the aggregate default prediction

This table reports differences in the default prediction accuracy, measured by the accuracy ratio (in percentage points), between the full model and the baseline model. All model comparisons are performed on subsamples derived from tercile splits of the reported variables to allow for non-monotonic effects (except LIMITED_LIABILITY which is a dummy variable). The analysis is based on 25,344 private German firms spanning the period from 2002 to 2005 (97,174 firm-year observations).

Variable	Lower tercile	Mid tercile	Upper tercile
LIMITED_LIABILITY	20.65	-	26.26
AGE	17.46	22.68	23.46
SALES	20.77	23.03	20.87
DISTANCE	18.58	19.30	22.26
FIRMS_PER_EMPLOYEE	18.57	18.89	27.50

Table 4
Improvement of firm-specific default predictions

This table reports estimation results from probit regression models. The dependent variable ΔPD equals one if the estimated probability of default increases for firms that default or decreases for firms that do not default, and zero otherwise, when we add four factors on business credit information to the baseline default prediction model. Explanatory variables are split into tercile dummies (T2, T3) with tercile T1 as reference category to allow for non-monotonic effects (except `LIMITED_LIABILITY` which is a dummy variable). We include dummy variables for years, the first two digits of the industry classification code and the first two digits of the ZIP Code in all models to control for year, industry and regional fixed effects. ***, **, * indicate that marginal effects are statistically significant at the 1%, 5%, and 10%-level, using robust standard errors clustered within firms. Marginal effects (mfx) are reported as average marginal effects.

Dependent Variable: ΔPD Model	(1) Probit Pooled			(2) Probit Pooled			(3) Probit Panel Random Effects			(4) Probit Panel Random Effects		
	mfx	t-stat.		mfx	t-stat.		mfx	t-stat.		mfx	t-stat.	
<code>LIMITED_LIABILITY</code>	0.026	4.81	***	0.026	3.52	***	0.027	6.07	***	0.017	3.87	***
<code>AGE_T2</code>	0.017	3.58	***	0.017	2.54	**	0.018	4.87	***	0.011	2.72	***
<code>AGE_T3</code>	0.036	6.91	***	0.043	6.00	***	0.030	6.68	***	0.023	5.05	***
<code>SALES_T2</code>	-0.008	-1.52		0.001	0.14		-0.006	-1.35		0.007	1.53	
<code>SALES_T3</code>	0.000	0.07		0.011	1.30		0.004	0.88		0.017	3.25	***
<code>DISTANCE_T2</code>	0.001	0.10		0.006	0.81		0.007	1.46		0.007	1.40	
<code>DISTANCE_T3</code>	0.005	0.84		0.013	1.45		0.009	1.53		0.009	1.48	
<code>FIRMS_PER_EMPLOYEE_T2</code>				0.009	0.61					-0.008	-0.76	
<code>FIRMS_PER_EMPLOYEE_T3</code>				0.041	3.04	***				0.027	3.00	***
Year fixed effects	Yes			Yes			Yes			Yes		
Industry fixed effects	Yes			Yes			Yes			Yes		
Regional fixed effects	Yes			Yes			Yes			Yes		
Adj. McFadden R ²	0.036			0.039			-			-		
Number of observations	96,320			42,919			96,320			43,077		

Table 5

Improvement of firm-specific default predictions with hard and soft information

This table reports estimation results from probit regression models. In columns (1) and (2) we compare a model with hard business credit information to a model with base factors. In columns (3) and (4) we compare the full model with hard and soft business credit information to a model with base factors and hard information. The dependent variable ΔPD equals one if the estimated probability of default increases for firms that default or decreases for firms that do not default, and zero otherwise, when we add four factors on business credit information to the baseline default prediction model. Explanatory variables are split into tercile dummies (T2, T3) with tercile T1 as reference category to allow for non-monotonic effects (except LIMITED_LIABILITY which is a dummy variable). We include dummy variables for years, the first two digits of the industry classification code and the first two digits of the ZIP Code in all models to control for year, industry and regional fixed effects. ***, **, * indicate that marginal effects are statistically significant at the 1%, 5%, and 10%-level, using robust standard errors clustered within firms. Marginal effects (mfx) are reported as average marginal effects.

Dependent Variable: ΔPD Model	(1)			(2)			(3)			(4)		
	Probit Pooled			Probit Pooled			Probit Pooled			Probit Pooled		
	Marginal impact of HARD business credit information			Marginal impact of HARD business credit information			Marginal impact of SOFT business credit information			Marginal impact of SOFT business credit information		
	mfx	t-stat.		mfx	t-stat.		mfx	t-stat.		mfx	t-stat.	
LIMITED_LIABILITY	0.023	4.39	***	0.025	3.34	***	0.038	5.80	***	0.024	2.53	**
AGE_T2	0.019	4.05	***	0.018	2.67	***	0.012	2.25	**	0.014	1.71	*
AGE_T3	0.038	7.31	***	0.043	5.97	***	0.084	13.40	***	0.093	10.15	***
SALES_T2	-0.007	-1.34		0.003	0.41		-0.107	-17.48	***	-0.123	-13.45	***
SALES_T3	0.004	0.59		0.017	1.99	**	-0.199	-26.82	***	-0.213	-19.22	***
DISTANCE_T2	0.002	0.34		0.007	0.93		-0.015	-2.33	**	-0.010	-1.05	
DISTANCE_T3	0.007	1.19		0.014	1.56		-0.048	-6.39	***	-0.040	-3.55	***
FIRMS_PER_EMPLOYEE_T2				0.010	0.70					-0.068	-3.78	***
FIRMS_PER_EMPLOYEE_T3				0.042	3.15	***				0.009	0.50	
Year fixed effects	Yes			Yes			Yes			Yes		
Industry fixed effects	Yes			Yes			Yes			Yes		
Regional fixed effects	Yes			Yes			Yes			Yes		
Adj. McFadden R ²	0.039			0.042			0.112			0.124		
Number of observations	96,320			42,921			96,320			43,001		

Table 6

Analysis of the link between ex ante default prediction accuracy and realized default rates

This table reports results from a cross-sectional OLS regression. Panel A shows the results for the spatial segmentation (models (1) and (2)). We consider firms from 2-digit ZIP Codes (non-overlapping areas) and firms from areas with a radius of 50 kilometers around the five-digit ZIP Code (overlapping areas). Panel B shows the results for the industry segmentation we use the first two digits of the industry code. The dependent variable (DEF_RATE) indicates the realized default rate per area (Panel A) or industry (Panel B), calculated from the default indicator DEF. Explanatory variables are the difference in the accuracy ratios of the full and baseline model (ΔAR) and the accuracy ratio of the baseline model (AR^{Baseline}). We control for regional GDP growth in spatial segmentation models (models (1) and (2)) and year fixed effects in all models. ***, **, * indicate that coefficients are statistically significant at the 1%, 5%, and 10%-level, using standard errors clustered within areas (industries).

Panel A: Spatial segmentation

Dep. Variable: DEF_RATE	2-digit ZIP Code area (non-overlapping spatial segmentation)			50 kilometer radius around 5-digit ZIP Code (overlapping spatial seg- mentation)		
		(1)		(2)		
	Coeff.	t-stat.		Coeff.	t-stat.	
ΔAR	-0.018	-3.06	***	-0.006	-6.85	***
AR^{Baseline}	-0.035	-6.48	***	-0.025	-32.44	***
GDP_GROWTH	-0.470	-3.32	***	-0.357	-27.04	***
Year fixed effects	Yes			Yes		
R^2	0.233			0.253		
Number of areas	90			5,586		
Number of area-year observations	304			21,380		

Table 6 (continued)

Panel B: Industry segmentation

Dep. Variable: DEF_RATE	Industry segmentation (2-digit industry code)	
	Coeff.	t-stat.
Δ AR	-0.053	-2.05 **
AR ^{Baseline}	0.009	0.24
GDP_GROWTH		
Year fixed effects	Yes	
R ²	0.122	
Number of areas	44	
Number of area-year observations	140	