WORKING PAPER

THE EXPERT'S EDGE? BANK LENDING SPECIALIZATION AND INFORMATIONAL ADVANTAGES FOR CREDIT RISK ASSESSMENT

Mathieu Simoens
Fabio Tamburrini

March 2025
2025/1111

GHENT UNIVERSITY

Department of Economics

Faculty of Economics and Business Administration
Campus Tweekerken, St.-Pietersplein 5, 9000 Ghent - BELGIUM

The expert's edge? Bank lending specialization and informational advantages for credit risk assessment*

Mathieu Simoens† Fabio Tamburrini‡

Abstract

We examine whether loan portfolio sectoral specialization provides informational advantages to banks, enabling better credit risk assessment. Using euro area credit register data, we compare probabilities of default assigned by specialized and non-specialized banks to the same borrowing firm several quarters before the borrower defaults. We find that banks specialized in the borrower's sector are better in predicting future defaults. This is mostly driven by specialized banks actively raising probabilities of default earlier, not by higher probabilities of default when loans are issued. As a result, specialized banks also increase provisions to these borrowers. We do not observe differences in credit risk assessment towards healthy borrowers, suggesting that the effect is not attributable to general conservatism but to more accurate evaluation of credit risk in the sectors of banks' specialization. Our results are more pronounced for smaller firms and when banks do not have long-term relationships with their defaulting borrowers.

Keywords: euro area banks, specialization, informational asymmetries, default
JEL Codes: G21, G32, D82

*Mathieu Simoens was the main contributor to the conceptualization, analysis, and writing of the paper. Fabio Tamburrini provided data access and data supervision. We thank Kris Boudt, Hans Degryse, Olivier De Jonghe, Cédric Huylebroek, Koen Inghelbrecht, Martien Lamers, Mike Mariathasan, Klaas Mulier, Alessio Reghezza, Koen Schoors and Rudi Vander Vennet, as well as participants at the 2024 Ghent University FEB Research Day, at the 2024 UCLouvain Winter Doctoral Workshop, and at seminars at the ECB, NBB, Ghent University, NEOMA Business School, and IÉSEG School of Management for helpful comments. The views expressed are those of the authors and do not necessarily represent the views of the ECB or the Eurosystem. Mathieu Simoens acknowledges financial support from the Research Foundation Flanders (FWO-Vlaanderen, grant number 1114622N).

†Ghent University. Mathieu.Simoens@UGent.be - corresponding author
‡European Central Bank. Fabio.Tamburrini@ecb.europa.eu

Non-technical summary

Informational asymmetries between lenders and borrowers are key to understand the function of banks in our economy. Borrowers have better insights into their own financial health and intentions, which implies that lenders face significant costs when assessing borrowers' creditworthiness and repayment behavior. By pooling loans and the necessary screening and monitoring activities from multiple individual lenders, banks reduce monitoring costs and gain an advantage in credit provision. In this context, the literature has argued that banks with specialized lending activities should be able to obtain better information about their borrowers. However, the inherently private nature of banks' internal risk assessments makes it difficult to formally investigate the informational advantages of specialization.

This paper overcomes this gap using granular information on banks' credit risk assessments to empirically test whether sectoral specialization leads to informational advantages for financial institutions. Using the euro area corporate credit register, we investigate whether banks are better in predicting a borrower's future default if they are more specialized in the sector of that particular borrower. To do so, we use granular data on ex-ante probabilities of default and provisions assigned by banks to individual borrowers which default later, allowing a unique look under the hood of banks' internal credit risk models. If specialization actually provides informational advantages, one would expect that a specialized bank will assign a higher probability of default and higher provisions to a later-defaulting firm, compared to a non-specialized bank.

As hypothesized, our baseline analysis shows that four quarters before the default, specialized banks assign probabilities of default which are 3.8 percentage points higher than non-specialized banks, and that this feeds through into significantly higher provisioning as well. By comparing the credit risk assessment of two banks, differing in terms of specialization, of the same defaulting firm, we ensure that our results are not driven by differences in actual riskiness between firms. We also rule out several alternative channels, including higher risk aversion by specialized banks towards all borrowers in their specialized sectors, using placebo testing.

In additional analyses, we document that our findings are stronger for lending towards smaller borrowers and for banks without long-term relationship with their borrowers, consistent with informational advantages of specialization being more important in relatively more opaque bank-firm relationships. Moreover, we shed light on the timing of the informational advantages of specialization, by showing that specialized banks do not yet deem these firms riskier when issuing new loans, but only increase probabilities of default and provisions more when the time of default approaches.

The analysis shows that informational advantages are key to banks' ability to monitor firms' behavior and manage the related risks, and that sectoral specialization can be beneficial for the safety and soundness of individual banks by improving their credit risk assessment ahead of firms' default. By revealing key details of the channel through which these benefits materialize, our analysis sheds new light on possible avenues to optimally combine the conventional financial stability benefits of diversification with the informational advantages from sectoral expertise.

1 Introduction

Informational asymmetries between lenders and borrowers are crucial in explaining the existence of banks as financial intermediaries (Leland & Pyle, 1977). Gathering information and monitoring borrowers before and during the lifetime of the loan is costly, and the traditional intermediation literature suggests that the cost advantages achieved through diversification of the lending portfolio are key to the benefits of banks' intermediation (Diamond, 1984; Ramakrishnan & Thakor, 1984; Boyd & Prescott, 1986). In contrast, building on arguments from the corporate finance literature (Jensen, 1986; Berger & Ofek, 1996), Winton (1999) argues that loan portfolio specialization may also provide benefits for banks, thanks to improved loan monitoring and better expertise when assessing borrowers' credit risk. The vast empirical literature on this topic has mainly investigated the effect of various measures of loan portfolio specialization or diversification on overall bank performance and risk (e.g., Acharya et al. (2006), Tabak et al. (2011)). More recently, the increased availability of granular loan level data has allowed researchers to also examine the implications for lending and loan performance (e.g., De Jonghe et al. (2020), Blickle et al. (2024)). This literature generally conceptualizes the potential benefits of specialization in terms of a comparative advantage in overcoming information asymmetries due to the close connection between banks and their borrowers. However, testing and quantifying these informational advantages is inherently difficult, because they are based on private information which is challenging to measure.

In our paper, we fill this gap by directly testing whether specialization in certain sectors allows banks to achieve informational advantages on the credit risk of borrowers in those sectors. We exploit supervisory data from the euro area credit register, capturing virtually all corporate loans by euro area banks. As a unique feature of this dataset, banks are required to disclose their assessment of the borrowers' credit risk by reporting and updating the probability of default (PD) assigned to each of the individual borrowers.[1] To test the hypothesis that banks have informational advantages about a borrower if they are more specialized in that borrower's sector, we focus on a sample of defaulting firms. We investigate to what extent specialized banks assign higher ex-ante PDs to these borrowers, compared to banks which are not specialized in the

[1] Only banks using the internal ratings-based approach report PDs, as discussed in more detail in Section 2.1.

defaulting borrower's sector.[2] The granularity of the dataset allows to investigate differences in PD by specialized versus non-specialized banks to the same defaulting borrower, thereby perfectly controlling for potential borrower-specific drivers of credit risk.

As hypothesized, we find that sectoral loan portfolio specialization leads to informational advantages for credit risk assessment: highly specialized banks are able to predict borrower defaults better, as the PDs they assign to later-defaulting borrowers are almost 4 percentage points higher four quarters before the default, compared to the PDs assigned to the same borrower by non-specialized banks. This difference decreases when approaching the default date, as non-specialized banks' risk assessment catches up with the borrowers' creditworthiness and the informational advantages fade. In practice, the advantages of sectoral specialization are likely to be reflected in more accurate internal credit risk models or enhanced information gathered by specialized banks and their loan experts. We also find that these informational advantages for specialized banks result in higher provisioning towards later-defaulting borrowers.

The main results are robust to controlling for various alternative channels, including banks' market power and relationship lending. In extensive placebo tests with propensity score matching, we also show that specialized banks do not assign higher PDs to non-defaulting firms, ruling out that our main findings are driven by higher risk aversion of specialized banks towards all borrowers in sectors in which they are specialized. We find stronger results for lending towards smaller borrowers, and for bank-firm pairs without long-term relationship, which is consistent with the informational advantages of specialization being larger for relatively more opaque relationships. We do not find statistically higher PDs or provisions by specialized banks towards later-defaulting borrowers when new loans are issued, and we show that the bulk of our effect is driven by specialized banks actively increasing PDs and provisions more quickly in the run-up to the default. This suggests that specialization especially allows banks to assess the true quality of their existing corporate credit portfolio earlier.

Our paper contributes to different strands of the literature. First, several papers have investigated the impact of sectoral (industry) diversification or specialization on bank level outcomes.[3]

[2] Beyhaghi et al. (2024) also use data on PDs to investigate banks' informational advantages, but they focus on the predictability of stock and bond returns, and they do not analyze the effect of specialization.

[3] While we limit our discussion to the literature on the effects of sectoral diversification or specialization, another strand of the literature has investigated the impact of functional diversification (e.g., Stiroh (2004),

Using Italian bank exposure data, Acharya et al. (2006) find that diversification is not guaranteed to produce superior performance or improve the safety of banks, and the authors argue that diversification can lead to a deterioration in monitoring effectiveness, particularly when banks enter competitive industries without prior experience. Focusing on German banks, Hayden et al. (2007) document that diversification is associated with reductions in bank profitability for the majority of banks, while Böve et al. (2010) find that sectoral specialization generally entails better monitoring quality, overcompensating the impact of higher credit concentrations for co-operative banks. Similarly, Tabak et al. (2011) find that loan portfolio specialization increases returns and reduces default risk. Using novel stock return-based measures of sectoral specialization for a worldwide sample of banks, Beck et al. (2022) show that specialization reduces individual and systemic bank risk. On the other hand, Rossi et al. (2009) and Shim (2019) demonstrate positive effects of diversification for Austrian and US banks, respectively. More recently, the focus has shifted towards investigating the impact of specialization on credit supply and loan outcomes. De Jonghe et al. (2020) show that Belgian banks facing a negative funding shock reallocate lending to sectors in which they are specialized. Using US syndicated loan data, Jiang & Li (2022) argue that banks use industry-specific knowledge developed through specialization to increase credit supply to these industries, while Bao (2022) finds that the availability of peer information from previous lending to competitor firms incentivizes banks to charge lower rates. Giometti & Pietrosanti (2022) also show that specialized banks offer less restrictive contract terms in the syndicated loan market. Regarding loan performance, Blickle et al. (2024) show that loans by specialized banks are less likely to default, and that banks use their informational advantages in the sectors in which they are specialized to lend more to smaller, more opaque firms. Finally, Goedde-Menke & Ingermann (2024) exploit a wave of early loan officer retirement in a single German bank as a shock to loan officer specialization. The authors document an increase in default rates due to less informative default risk information and excessive loan growth. We contribute to this debate by examining whether specialized banks can more accurately predict borrower defaults, thus providing a direct test of the informational advantages of sectoral specialization for credit risk assessment, based on the quasi-universe of

Stiroh (2006), Baele et al. (2007), De Jonghe (2010)) or geographical diversification (e.g., Goetz et al. (2013), Goetz et al. (2016), Chu et al. (2019), Nyola et al. (2021)) on bank performance and/or risk.

euro area bank lending to non-financial corporations (NFCs).

While our paper mainly focuses on the informational advantages of loan portfolio specialization, the literature has argued that relationship lending also plays an important role in overcoming informational asymmetries between lenders and borrowers (Berger & Udell, 1995). Several papers have shown that relationship lending can positively impact credit supply and even lead to reduced defaults (Petersen & Rajan, 1994; Degryse & Van Cayseele, 2000; López-Espinosa et al., 2017; Yildirim, 2020). At the same time, it may also enable lenders to extract rents (Sharpe, 1990; Rajan, 1992).[4] In this paper, we also contribute to this strand of the literature by testing to what extent relationship lending and sectoral specialization act as complements or substitutes in terms of information collection and loan monitoring. While we first show that both specialization and relationship lending indeed lead to better default predictions, we also document that the positive impact of specialization is smaller (but still present) in the sample of relationship loans, suggesting that they act as substitutes, but only partially.

Our findings are relevant for supervisors and policymakers, as they offer a more nuanced understanding of the financial stability benefits of specialization and diversification: while the traditional portfolio theory suggests that more specialized banks should be more risky as they are more exposed to idiosyncratic shocks, our results highlight that sectoral specialization can also enhance the safety and soundness of individual banks, through their superior credit risk assessment capabilities. By unpacking the channel through which these benefits materialize, our analysis can shed light on possible avenues to combine the benefits of diversification with the informational advantages from sectoral expertise.

The remainder of the paper is organized as follows. In Section 2, we discuss the datasets and sample, followed by the methodology in Section 3. The main results are presented in Section 4, with robustness checks covered in Section 5. We provide extensions in Section 6, before concluding in Section 7.

[4] For a detailed overview of the potential costs and benefits of relationship lending, see Boot (2000) and Kysucky & Norden (2016).

2 Data

2.1 Overview of the data sources

As our main dataset, we make use of the euro area credit register of the European System of Central Banks (AnaCredit). This credit register contains data on all individual loans by euro area banks to firms with a total commitment amount of at least EUR 25,000, since 2018Q3.[5] Because of the relatively low threshold, the dataset captures almost the entire universe of euro area corporate lending by banks, including loans to small and medium-sized enterprises (SMEs). The credit register contains information on the identity of the bank and the borrower, including its sector, location and size[6], as well the loan's outstanding amount, interest rate, maturity, collateral, loan type, etc. The dataset also includes the PDs assigned by banks to each of the borrowers and the associated provisioning, as well as the number of days a borrower is past the due date on his loans and whether or not the borrower has defaulted on the loan. Crucially, PDs are only reported by banks adopting the internal rating-based (IRB) approach for calculating risk-weighted assets for credit risk, not for banks following the standardized approach (SA).[7]

To obtain additional bank data for the analysis, AnaCredit is matched to ECB supervisory bank data reported under the common reporting (COREP) and financial reporting (FINREP) frameworks.[8] This matching is done at the level of the ultimate parent in the Single Supervisory Mechanism (SSM), and allows to include information on banks' profitability, size, CET1 buffer, total loan portfolio size, and share of non-performing loans. While not necessary for the main analysis, we also use Orbis (Bureau van Dijk) to obtain additional firm data for the borrowers in our sample. More specifically, we include data on firms' leverage, profitability and size for the propensity score matching exercise, as described in more detail in Section 4.2.[9]

[5]We only include EUR-denominated term loans, revolving credit and other credit lines to NFCs. We omit multi-creditor and multi-borrower loans, as well as borrowers which are assigned to multiple sectors. More info on the euro area credit register: https://www.ecb.europa.eu/stats/ecb_statistics/anacredit/html/index.en.html

[6]Borrowers are classified as micro, small, medium-sized, or large enterprises, in accordance with the Annex to Recommendation 2003/361/EC.

[7]The Basel III standards allow banks to adopt the SA or the IRB approach for their credit risk assessment. Banks that opt for the IRB approach apply their own (regulatory-approved) models to estimate the credit risk parameters, including PD, used to calculate risk weights and subsequent capital requirements. The SA applies fixed risk weights per borrower category. IRB banks account for approximately 80% of total loans to NFCs. For a more detailed discussion on IRB banks, we refer to Bruno et al. (2023).

[8]More info: https://www.ecb.europa.eu/stats/supervisory_prudential_statistics/html/index.en.html

[9]Since not all AnaCredit firms can be matched to Orbis, and because this match is only necessary for a subset of the analysis, we do not remove AnaCredit observations from the main analysis if no Orbis match can be found.

2.2 Loan portfolio specialization

In this subsection, we discuss our measure of loan portfolio specialization, which is the main explanatory variable in the analysis. We use AnaCredit loan level exposure data on banks' total corporate loan portfolio to construct this measure. We define our bank-sector-quarter level specialization variable in Equation (1), following, among others, Paravisini et al. (2023) and Blickle et al. (2024). For the baseline analysis, we define sectors at the NACE2 level, resulting in 85 sectors.

$$Specialization_{b,s,t} = \frac{Exposure_{b,s,t}}{\sum_s Exposure_{b,s,t}} - \frac{\sum_{b \in c} Exposure_{b,s,t}}{\sum_{b \in c} \sum_s Exposure_{b,s,t}} \tag{1}$$

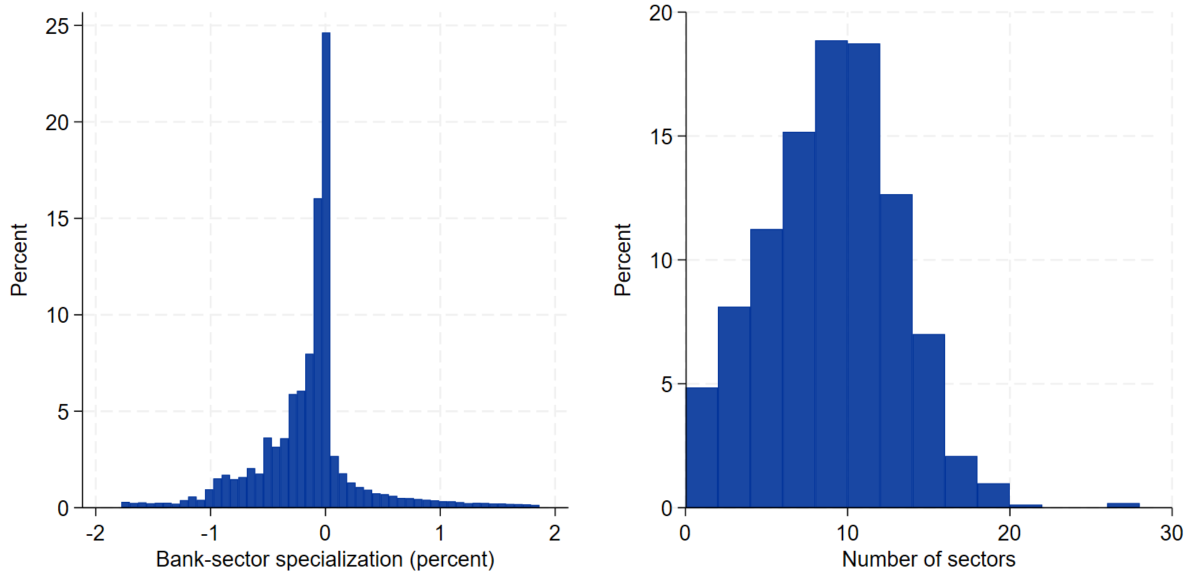
The first term in Equation (1) captures to what extent the total corporate loan portfolio of bank b is exposed to sector s in quarter t, as a percentage of the total loan exposure of bank b to all sectors in that quarter. In line with Paravisini et al. (2023) and Blickle et al. (2024), we adjust this 'pure' exposure measure by correcting for the size of the sector. Indeed, ideally, we should capture the relative degree to which a lender is over- or under-exposed to a sector, without this measure being affected by sector size. Therefore, we include a second term in Equation (1), to correct for the total lending directed to that particular sector by all banks in the same country.[10] Intuitively, this implies that a high (low) specialization measure will reflect that a bank is relatively more (less) exposed to a particular sector than its peers. In the baseline, we correct by subtracting the percentage exposure of all banks in the same country as proposed by Blickle et al. (2024), although we also implement the original measure of Paravisini et al. (2023) – dividing the bank's exposure by the percentage exposure of all banks in the same country, instead of subtracting – in a robustness check.[11] Importantly, Equation (1) implies that specialization is determined at the bank-sector-quarter level. Although, for simplicity, we sometimes discuss our findings using the concept "specialized banks", this is strictly-speaking

[10] Alternatively, including sector-time or firm-time fixed effects will result in a similar correction, as in Jiang & Li (2022) and De Jonghe et al. (2020, 2024). However, in the baseline, we focus on the proposed solution by Paravisini et al. (2023) and Blickle et al. (2024), because we construct high specialization dummy variables based on the full distribution of the specialization measure before saturating the specification with fixed effects, which implies that the dummies could still be influenced by sector size if the correction is not applied. We discuss alternative approaches in Section 5.

[11] Blickle et al. (2024) argue that the original measure of Paravisini et al. (2023) is more likely to introduce large right tails.

not correct, as the same bank can (and is likely to) be highly specialized in one sector but less specialized in another sector. Therefore, throughout the paper, "specialized banks" should be interpreted as "banks specialized in the sector of the particular borrower".

Figure 1: Specialization measure



A: Full distribution

B: Number of highly specialized sectors per bank

In Panel A of Figure 1, we show the distribution of the specialization measure at the bank-sector (NACE2) level.[12] The vast majority of the observations is situated between -1% and 1%, with a mean of 0% and a median of -0.07%. In the remainder of the analysis, we will use a high specialization dummy in the baseline setup which captures whether the bank-sector specialization measure is in the highest decile of the distribution. As can be seen in Panel B, all banks have between 0 and 28 sectors in which they are highly specialized. Figure 1 shows descriptives for the full sample of euro area banks for which loan level data is available for the year 2022 in AnaCredit. The baseline analysis, however, requires the availability of PD data and will therefore only include loans from IRB banks. Hence, the main focus of this paper will be on a smaller sample of 77 IRB banks. We therefore also present the same descriptives regarding the specialization measure for this subset of banks in Figure A1 and A2 in the Appendix.

[12]We present the average bank-sector value over the four quarters in 2022. For readability, we only show a histogram of the specialization measure between the 5^{th} and 95^{th} percentile.

2.3 Sample and descriptive statistics

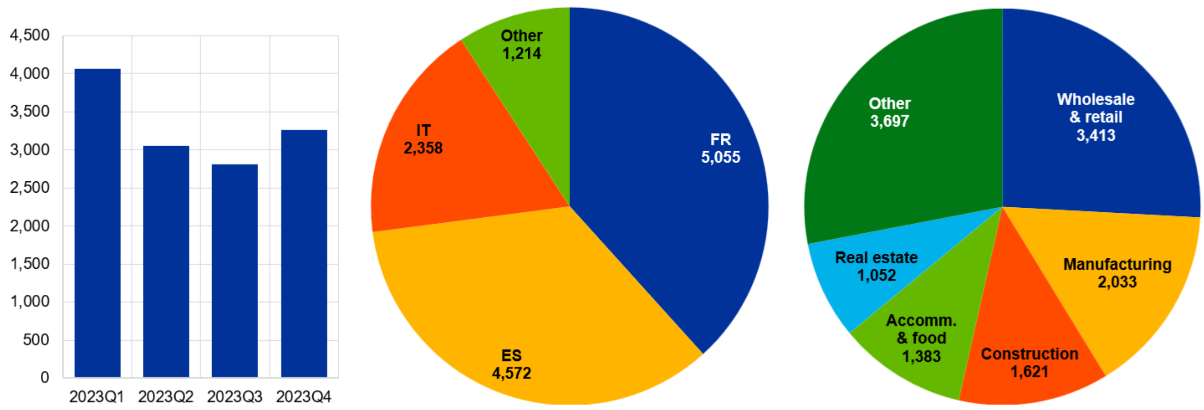
In our analysis, we focus on a sample of defaulting borrowers. To determine this defaulting sample, we use the standardized default indicator reported by the banks in AnaCredit, which is defined following the EU Capital Requirements Regulation (Article 178 of Regulation (EU) No 575/2013). For every borrower, we determine the first quarter since 2018Q3 (the start of AnaCredit) in which at least one of its banks reports a loan as being in default. As a robustness check, we also consider an alternative default indicator which we construct ourselves, based on whether the bank reports a borrower as being at least 90 days past the due date on one of its outstanding loans. To examine whether specialized banks have informational advantages and can therefore predict defaults better, we use the PD as main variable of interest. This PD, reported by banks following the IRB approach, represents the borrower's probability of default over a one-year horizon, established based on the reporting bank's internal model. In AnaCredit, PDs are reported at the bank-firm-quarter level, i.e., they are not loan-specific. Therefore, we also collapse all loan level data to the bank-firm-quarter level (e.g., total outstanding loan amount, weighted average interest rate).

For the baseline analysis, we focus on all borrowers for which the first default occurs in one of the four quarters of 2023, and which did not default before in any quarter since 2018Q3. We investigate defaults in 2023 to avoid abnormal influences from the Covid-19 pandemic and aftermath. During the Covid-19 period, national authorities and supervisors have intervened considerably, implementing loan moratoria and public government guarantee schemes, which may have impacted both PDs as well as actual defaults during those periods. To investigate whether sectoral specialization leads to informational advantages in advance of the default, we focus our main analysis on PDs in the four quarters before the first quarter of default. Thus, including defaults during all quarters in 2023 implies that the first observations are from 2022Q1.

A key element of the identification strategy consists of comparing the PDs given to the same defaulting firm by (at least) two banks which differ in terms of their degree of specialization in the defaulting firm's sector, i.e., a within-firm between-bank analysis.[13] Technically, as will be discussed in more detail in Section 3, this will be achieved by saturating the specification with

[13]The requirement that a firm borrows from at least two banks (irrespective of whether these banks follow the SA or IRB approach) reduces the sample from approximately 86,000 to 22,000 defaulting firms.

Figure 2: Number of defaulting firms



A: Per quarter | B: Per country | C: Per sector (NACE1)

firm-quarter fixed effects. Combined with the fact that the baseline analysis focuses on PDs (which are reported by IRB banks, cf. Section 2.1), this implies that we only include borrowers with loans from at least two different IRB banks. The resulting baseline sample consists of 129,501 bank-firm-quarter observations between 2022Q1 and 2023Q4, encompassing 77 euro area banks and 13,199 firms which default during 2023.[14] As can be seen in Panel A of Figure 2, the number of defaults is quite well spread across the four quarters. Panel B shows that most defaulting firms are headquartered in France, Spain and Italy. While approximately 15% of the defaulting firms are active in the wholesale and retail sector, Panel C nevertheless shows that the distribution of firms over sectors (presented at NACE1 level) is quite heterogeneous. Descriptive statistics for the most important variables in the analysis can be found in Table 1, with detailed definitions of the variables in Table A1 in the Appendix.

3 Methodology

The key contribution of this paper is that we examine to what extent banks have informational advantages regarding borrowers in sectors in which the bank is specialized. To test this empirically, we focus on defaulting firms and investigate whether banks can predict this default better

[14]For the majority of the analysis, we use data between 2022Q1 and 2023Q4, with a few exceptions such as the propensity score matching (cf. Section 4.2) and some extensions (cf. Section 6.2).

Table 1: Descriptive statistics

Variable	**N. obs**	**Mean**	**Stdev**	**Min**	**P25**	**P50**	**P75**	**Max**
Dependent variables:								
PD	129,501	21.95	33.23	0.00	1.91	6.00	21.04	100.00
Provisions (% outstanding)	128,261	5.40	10.83	0.00	0.19	0.96	5.04	62.71
Specialization variables:								
Specialization	129,501	0.30	3.48	-40.29	-0.30	-0.01	0.30	96.18
$D_{90}^{Specialization}$	129,501	0.18	0.38	0.00	0.00	0.00	0.00	1.00
$D_{75}^{Specialization}$	129,501	0.49	0.50	0.00	0.00	0.00	1.00	1.00
$\text{Specialization}_{alt}$	129,501	5.75	10.41	0.00	0.97	2.01	5.53	97.68
$\text{Specialization}_{NACE1}$	129,501	0.53	4.99	-40.29	-1.52	-0.04	1.13	87.50
$\text{Specialization}_{rel}$	129,501	105.52	45.86	16.97	84.75	99.41	118.11	857.02
Control variables:								
Bank ROA	129,359	0.46	0.26	-0.33	0.29	0.42	0.62	2.00
Bank size	129,359	13.37	1.01	6.63	12.69	13.57	14.27	14.79
Bank CET1 buffer	129,273	2.70	1.85	-0.87	1.24	2.63	3.46	15.21
Bank loans/assets	129,359	61.76	7.03	40.80	57.65	61.96	66.32	86.84
Bank NPL/loans	129,359	2.83	0.98	0.68	2.14	2.92	3.43	8.78
Maturity	128,683	6.82	1.00	3.33	6.51	7.02	7.39	8.98
Exposure	129,492	0.00	0.02	0.00	0.00	0.00	0.00	0.82
Interest rate	122,770	2.95	2.01	0.00	1.48	2.53	4.12	10.45
$D^{Protection}$	129,501	0.91	0.28	0.00	1.00	1.00	1.00	1.00
$D^{Same\ country}$	129,501	0.93	0.26	0.00	1.00	1.00	1.00	1.00
$D^{Past\ due\ date}$	129,431	0.16	0.36	0.00	0.00	0.00	0.00	1.00
$D^{First\ default}$	129,501	0.45	0.50	0.00	0.00	0.00	1.00	1.00
$D^{Market\ power}$	129,501	1.00	0.07	0.00	1.00	1.00	1.00	1.00
$D_{2018}^{Relationship}$	129,501	0.43	0.49	0.00	0.00	0.00	1.00	1.00
$D_{2019}^{Relationship}$	129,501	0.54	0.50	0.00	0.00	1.00	1.00	1.00

This table shows the number of observations, mean, standard deviation, minimum, 25^{th} percentile, median, 75^{th} percentile and maximum for the main variables in the analysis. Except for bank size (natural logarithm), maturity (natural logarithm) and the dummy variables, all variables are expressed in percentage.

if they are specialized in that firm's sector. More specifically, we examine whether specialized banks already assign higher PDs to these defaulting borrowers several quarters before the default, compared to non-specialized banks. The baseline specification is presented in Equation (2).

$$Y_{b,f,s,t} = \sum_{q=-4}^{0} \beta_q \cdot Specialization_{b,s,t} \cdot RelTime^q_{f,t} + \gamma \cdot Z_{b,f,t} + \eta_{f,t} + \alpha_{b,t} + \epsilon_{b,f,s,t} \tag{2}$$

$$\text{with } t \in [quarter_f^{dflt} - 4, quarter_f^{dflt}]$$

In Equation (2), the dependent variable ($Y_{b,f,s,t}$) is the one-year-ahead PD assigned by bank b to firm f, which is active in sector s, in quarter t. In other specifications, we use provisions as a percentage of the outstanding loan amount as an alternative dependent variable. Each bank-firm lending relationship is included in at most five time periods: the quarter in which the firm defaults for the first time on one of its loans, as well as the four preceding quarters. The main explanatory variable of interest is the interaction between a specialization measure ($Specialization_{b,s,t}$) and a relative timing dummy ($RelTime^q_{f,t}$), which indicates the quarter relative to the default to which the bank-firm observation belongs. By estimating five separate β_q coefficients on this interaction (one for each quarter relative to the default), we are able to observe to what extent the effect of specialization changes dynamically over time. We hypothesize a positive β_q, although we expect the effect to gradually decrease over time, as also non-specialized banks should be able to predict the default better as they approach the quarter of default and information on the creditworthiness of the borrowing firm becomes more readily available. The specialization variable is defined at the bank-sector-quarter level. In the baseline setup, we use a high specialization dummy variable ($\mathrm{D}_{90}^{\mathrm{Specialization}}$) which is one if the specialization measure, as constructed in Section 2.2, is in the highest decile in that quarter.[15]

In the most saturated version of Equation (2), we include firm-quarter ($\eta_{f,t}$) and bank-quarter ($\alpha_{b,t}$) fixed effects. The firm-quarter fixed effects are key for the identification, as they allow to compare the PD (or provisioning) assigned by a specialized versus a non-specialized bank to the same defaulting firm. Thus, this within-firm (between-bank) setup allows to perfectly abstract from firm-specific differences in risk which are likely to drive differences in PDs. Using bank-quarter fixed effects, we also control for all time-varying bank characteristics. In alternative

[15]While this dummy approach facilitates the interpretation of the results, we provide ample robustness in Section 5 to various alternative approaches, including the use of another threshold and several continuous specialization measures. Because the specialization measure is included in interaction with all five relative timing indicators (one for each of quarter), the individual specialization variable is omitted due to perfect collinearity.

versions, we replace the bank-quarter fixed effects by a combination of bank fixed effects and time-varying bank controls, capturing profitability (measured using return on assets, ROA), size (natural logarithm of total assets), the capital buffer (actual CET1 ratio minus overall capital requirements), the size of the loan portfolio (as percentage of total assets), and the share of non-performing loans in the total loan portfolio. In the baseline specification, we also include several loan characteristics, collapsed at the bank-firm-quarter level ($Z_{b,f,t}$). We control for (the natural logarithm of) the average maturity and the weighted average interest rate of the loans, and include dummy variables to capture whether the loans are secured by collateral, whether bank and firm are headquartered in the same country, and whether the borrower is already past the due date on some of its loans at that bank. Moreover, to avoid that our measure of sectoral specialization captures the effect of the bank's exposure to the individual firm rather than exposure to the broader sector, we add a separate variable measuring the percentage exposure of the bank to the firm. Finally, since borrowers do not necessarily default on their loans with all their banks in the same quarter (cf. Section 2.3), we also include a dummy to indicate whether the first default observation occurred at that particular bank-firm pair or not. In the baseline setup, we cluster standard errors at the bank and firm level, in line with Ferreira & Matos (2012), Fraisse et al. (2020) and Jiang & Li (2022).

4 Results

4.1 Specialization and probabilities of default

To investigate whether loan portfolio specialization leads to informational advantages for credit risk assessment, we test the hypothesis that banks specialized in the sector of a defaulting borrower assign higher PDs to this borrower already several quarters before the default, compared to non-specialized banks. To test this empirically, we estimate Equation (2), in which we focus on the coefficients β_q on the interaction between the specialization measure and the relative timing dummy.

In Figure 3, we show the point estimates and confidence intervals of β_q in the baseline specification. We find that sectoral loan portfolio specialization indeed leads to informational advantages, as PDs assigned to defaulting borrowers by specialized banks are 3.8 percentage points higher than PDs by non-specialized banks to the same defaulting borrower, four quarters before the default. The gap between specialized and non-specialized banks decreases closer to the quarter of default: PDs by specialized banks are still 2.6 percentage points higher two quarters before the default, but we find no significant difference anymore in the quarter before the default or the quarter of the default itself. This is in line with our predictions, as specialization should provide most informational advantages in terms of early detection of the potential default, whereas also non-specialized banks are expected to realize that the borrower has an increased chance of defaulting as they approach the default date.

Figure 3: Effect of sectoral specialization on PDs

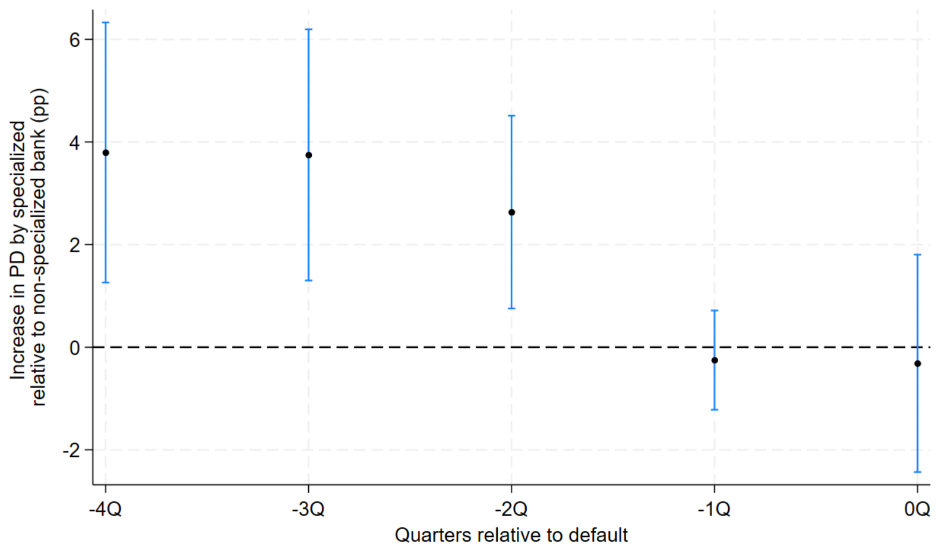


Figure showing the point estimates and 90% confidence intervals of the β_q coefficients estimated in Equation (2). The estimation corresponds to column (4) in Table 2.

In Table 2, we present variations of Equation (2), differing with respect to the included control variables and fixed effects. In column (1), we start with a minimalist regression, including only bank and firm-quarter fixed effects besides the specialization variables, while time-varying bank controls are added in column (2). Both columns confirm that banks assign relatively higher PDs to defaulting borrowers if they are specialized in that borrower's sector, especially between four and two quarters before the default. Moreover, column (2) shows that larger and more profitable banks, as well as banks with a larger distance to the capital buffer requirements typically assign higher PDs to defaulting borrowers. In column (3), the bank fixed effects and bank controls are replaced by more granular bank-quarter fixed effects, controlling for all potential bank-time-varying differences in PDs. Column (4) presents the baseline specification, for which the β_q coefficients were shown in Figure 3, including bank-quarter and firm-quarter fixed effects, as well as bank-firm-quarter controls. While adding bank-firm-quarter controls does not lead to meaningful changes in the coefficients of interest, this column also shows that banks assign lower PDs to borrowers headquartered in the same country. Unsurprisingly, if a borrower is already behind on one of its loan payments (past due date), this is reflected in higher PDs. Ex-ante PDs are also higher for the bank-firm pairs where the default will occur first.

In column (5), the specification is extended with loan type-quarter and interest rate type-quarter fixed effects, which capture general time-varying differences in PDs between different loan types or interest rate types. We distinguish between borrowers with only term loans, only revolving credit, only other credit lines, or a mix of these loan types, and between borrowers with only fixed rate loans, only floating rate loans, or a mix of both. Including these fixed effects implies no changes for the main findings. In columns (6) to (8), additional control variables are added. First, as argued by Jiang & Li (2022), it is important to distinguish between specialization and market share (market power) since both measures are based on the bank's exposure to a sector.[16] Indeed, if banks extract higher rents from borrowers in sectors in which they have higher market shares (Giannetti & Saidi, 2019; De Jonghe et al., 2024), this could make those borrowers more prone to defaulting and thus lead to higher PDs. Even though we already control for, e.g., the weighted average interest rate charged to the borrower, we construct a bank-sector-quarter measure for banks' market share to each sector and add a dummy capturing whether the market share is in the top decile of the distribution (similar to the main specialization measure) to the specification in column (6). The coefficient on the market share dummy is not significant and the effect of specialization on PDs does not change in sign or magnitude. Second, banks with a longer-term relationship with their borrowers are also likely to have informational advantages which may enable them to predict defaults better. To avoid that this is captured by our specialization measure, we add a variable in columns (7) and (8)

[16] Specialization captures the exposure as percentage of the bank's total exposure to all sectors, whereas market share is calculated as a percentage of all banks' exposures to that particular sector.

Table 2: Effect of sectoral specialization on PDs

Dependent variable:	Probability of default							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
-4Q $\times$ $D_{90}^{\text{Specialization}}$	3.264**	3.822**	3.833**	3.794**	3.736**	3.784**	3.729**	3.739**
	(0.043)	(0.041)	(0.044)	(0.015)	(0.018)	(0.016)	(0.016)	(0.016)
-3Q $\times$ $D_{90}^{\text{Specialization}}$	3.129**	3.535**	3.614**	3.748**	3.680**	3.738**	3.709**	3.718**
	(0.044)	(0.035)	(0.038)	(0.013)	(0.016)	(0.014)	(0.014)	(0.014)
-2Q $\times$ $D_{90}^{\text{Specialization}}$	2.517*	2.538*	2.605**	2.634**	2.589**	2.621**	2.608**	2.611**
	(0.063)	(0.052)	(0.045)	(0.022)	(0.025)	(0.024)	(0.024)	(0.024)
-1Q $\times$ $D_{90}^{\text{Specialization}}$	0.535	0.285	0.053	-0.251	-0.272	-0.263	-0.265	-0.267
	(0.357)	(0.597)	(0.918)	(0.667)	(0.646)	(0.650)	(0.645)	(0.643)
0Q $\times$ $D_{90}^{\text{Specialization}}$	-0.485	-1.288	-1.464	-0.314	-0.373	-0.331	-0.345	-0.345
	(0.791)	(0.432)	(0.324)	(0.805)	(0.769)	(0.795)	(0.788)	(0.787)
Bank ROA		2.859*						
		(0.071)						
Bank size		51.715***						
		(0.001)						
Bank CET1 buffer		1.555***						
		(0.002)						
Bank loans/assets		0.262						
		(0.197)						
Bank NPL/loans		1.583						
		(0.368)						
Maturity				0.902	0.455	0.903	0.945	0.948
				(0.129)	(0.286)	(0.128)	(0.109)	(0.108)
Exposure				-7.170	-3.864	-7.442	-8.077	-7.820
				(0.310)	(0.613)	(0.302)	(0.263)	(0.290)
Interest rate				0.278	0.323	0.279	0.251	0.269
				(0.179)	(0.141)	(0.177)	(0.201)	(0.188)
$D^{\text{Protection}}$				-0.238	0.091	-0.242	-0.554	-0.609
				(0.683)	(0.872)	(0.678)	(0.378)	(0.335)
$D^{\text{Same country}}$				-3.055**	-2.801**	-3.044**	-3.359***	-3.310***
				(0.016)	(0.026)	(0.016)	(0.007)	(0.009)
$D^{\text{Past due date}}$				14.622***	14.443***	14.620***	14.558***	14.547***
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$D^{\text{First default}}$				16.725***	16.723***	16.725***	16.634***	16.638***
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$D^{\text{Market power}}$						1.812		
						(0.496)		
$D_{2018}^{\text{Relationship}}$							2.014**	
							(0.013)	
$D_{2019}^{\text{Relationship}}$								2.060***
								(0.009)
Bank FE	Yes	Yes	No	No	No	No	No	No
Bank x Quarter FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Firm x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loan type x Quarter FE	No	No	No	No	Yes	No	No	No
Rate type x Quarter FE	No	No	No	No	Yes	No	No	No
N obs	129,493	129,117	129,426	117,608	117,550	117,608	117,608	117,608
R^2	0.6337	0.6355	0.6429	0.6937	0.6947	0.6937	0.6941	0.6941

This table shows the results of the estimation of Equation (2). Standard errors are clustered at bank and firm level. The numbers in parentheses are p-values. *, **, and *** indicate significance at 10%, 5%, and 1% respectively.

which indicates whether or not the bank-firm pair already had an outstanding loan at the end of 2018, or at the end of 2019, respectively. The coefficients on the relationship variables are indeed positive and significant, suggesting that an established lending relationship improves the bank's ability to predict defaults, but the effect of specialization on PDs remains unchanged.

4.2 Placebo tests

The results in the previous subsection show that banks assign relatively higher PDs to defaulting borrowers, already several quarters before the default, if they are more specialized in the sector of that firm. While we attribute these findings to informational advantages, an important alternative hypothesis is that banks are generally more risk-averse towards sectors in which they are specialized, and therefore assign higher PDs to all firms in those sectors, irrespective of whether or not the firm is more likely to default. To test this potential alternative channel, we estimate the same Equation (2) on several samples of non-defaulting firms. If these placebo tests do not show significant results, we can rule out this alternative channel.

To construct a sample of non-defaulting firms for the placebo test, we use propensity score matching (PSM) as our main approach. For each country and quarter of default, we first estimate probit regressions predicting whether firms default or not based on firm characteristics five quarters before the default, i.e., one quarter earlier than the first quarter included in the main regressions.[17] As explanatory variables, we include the firm's equity ratio (equity as percentage of total assets), size (natural logarithm of total assets), and profitability (ROA). Subsequently, for each defaulting firm, we select the 10 nearest non-defaulting neighbors (with replacement) based on the propensity score.[18] We do not include non-defaulting firms as potential nearest neighbor if they default in any quarter in our full dataset (between 2018Q3 and 2023Q4). Descriptive statistics for the PSM placebo sample are shown in Table A2 in the Appendix.

Using PSM, we employ a regression-based approach to assign non-defaulting firms as nearest neighbors of defaulting firms, based on their similarities in terms of (pre-default) characteristics

[17] To ensure sufficient observations for a meaningful regression, we only estimate the probit regression if there are at least 25 defaulting firms in the country-quarter. Note that we restrict the sample to only include firms with at least two banks, because single-bank firms would drop later due to the inclusion of firm-quarter fixed effects in Equation (2).

[18] Because of the inclusion of the relative timing dummy in Equation (2), we remove non-defaulting firms in the placebo test if they are a nearest neighbor for two or more firms which default in different quarters.

associated with a default. However, with this PSM approach, there is no guarantee that the defaulting firm firm will be, e.g., in the same sector as the non-defaulting nearest neighbor. Therefore, we also design an alternative approach, where we first impose a hard restriction that nearest neighbors belong to the same sector (defined at NACE2 level), same region (defined at NUTS2 level) and same size bucket (micro, small, medium-sized, or large firm) as the defaulting firm. Subsequently, the nearest neighbors are selected based on the closest distance regarding one of the following variables: profitability (ROA), leverage, total outstanding loan amounts, or weighted average interest rate.[19] This alternative approach thus generates an additional four samples of non-defaulting firms to be used in the placebo tests.

Table 3: Placebo tests

Dependent variable:	Probability of default					
	(1)	(2)	(3)	(4)	(5)	(6)
-4Q × $D_{90}^{\text{Specialization}}$	4.517**	-0.030	-0.100	-0.026	0.059	-0.098
	(0.026)	(0.695)	(0.265)	(0.725)	(0.262)	(0.124)
-3Q × $D_{90}^{\text{Specialization}}$	4.623**	0.002	-0.120*	-0.039	0.069	-0.050
	(0.016)	(0.978)	(0.091)	(0.651)	(0.236)	(0.418)
-2Q × $D_{90}^{\text{Specialization}}$	3.646**	-0.001	-0.108	-0.066	0.077	-0.004
	(0.015)	(0.986)	(0.166)	(0.435)	(0.136)	(0.956)
-1Q × $D_{90}^{\text{Specialization}}$	-0.265	-0.024	-0.076	-0.045	0.068	0.020
	(0.741)	(0.741)	(0.398)	(0.606)	(0.350)	(0.712)
0Q × $D_{90}^{\text{Specialization}}$	1.284	0.047	-0.125	-0.058	-0.001	-0.073
	(0.272)	(0.634)	(0.235)	(0.583)	(0.987)	(0.243)
Controls	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ
Bank x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
N obs	81,239	425,950	250,713	270,605	507,041	465,606
R^2	0.6859	0.6218	0.5948	0.5947	0.6023	0.6011
Sample	Default	Non-default	Non-default	Non-default	Non-default	Non-default
Matching	Has match	PSM	ROA	Leverage	Amount	Rate

This table shows the results of the estimation of Equation (2) for defaulting firms (column (1)) or (matched) non-defaulting firms (columns (2) to (6)). All specifications include bank-firm-quarter (BFQ) control variables. Standard errors are clustered at bank and firm level. The numbers in parentheses are p-values. *, **, and *** indicate significance at 10%, 5%, and 1% respectively.

In Table 3, we show the results of the placebo tests. In column (1), we first repeat the baseline specification for the sample of defaulting firms, but now on a slightly restricted sample of defaulting firms for which a non-defaulting PSM match can be found. We can therefore

[19]These firm variables are again measured five quarters before the default.

directly compare the placebo tests with the results in this column, without, e.g., selection bias concerns. Next, we estimate the same Equation (2) on the different samples of non-defaulting nearest neighbors. In column (2), we show the results for the non-defaulting firms selected using PSM. Columns (3) to (6) show the results for the alternative matching approach, matching on profitability, leverage, total outstanding loan amounts, and weighted average interest rate, respectively. While the results in column (1) confirm the baseline results, showing a significantly positive effect of specialization on PDs in the first quarters for defaulting firms (and even somewhat stronger in magnitude), none of the placebo tests show any significant positive effect of specialization on PDs assigned to non-defaulting firms. We can therefore reject the alternative hypothesis that banks assign higher PDs to all firms in the sectors in which they are specialized.

4.3 Specialization and provisioning behavior

Under IFRS 9, banks' provisioning is tightly linked to expected credit losses, which are determined by a combination of the PD, the loss given default (LGD), and the exposure at default (EAD). Therefore, we would expect that the informational advantages of specialization are also reflected in higher provisioning by specialized banks towards later-defaulting firms.[20] We investigate this hypothesis formally in this subsection. As mentioned before, the sample used in Section 4.1 is limited to all firms for which a PD is available from at least two different banks, which implies that only loans by banks using the IRB approach for their credit risk assessment are included. When focusing on provisions, however, a broader sample can be used, because provisions are also reported by SA banks. We provide descriptive statistics for this broader sample in Table A3 in the Appendix.

We examine the impact of specialization on banks' provisioning behavior by estimating Equation (2) using the bank's ex-ante provisioning towards the loans of the defaulting borrower (as a percentage of the total outstanding loan amount towards that borrower) as dependent variable. Figure 4 represents the point estimates and confidence intervals of the estimated β_q. As hypothesized, we find that the informational advantages of sectoral specialization are also visible in banks' levels of provisioning, with provisions around 0.5 percentage points higher four

[20] For more details on the link between PDs, expected losses, and provisioning, we refer to Behn & Couaillier (2023).

quarters before the default for banks specialized in the sector of the defaulting firm. As before, the relative advantage of specialization decreases when approaching the default date: while specialized banks' provisions are still almost 0.4 percentage points higher two quarters before the default, the effect becomes insignificant afterwards.

Figure 4: Effect of sectoral specialization on provisioning

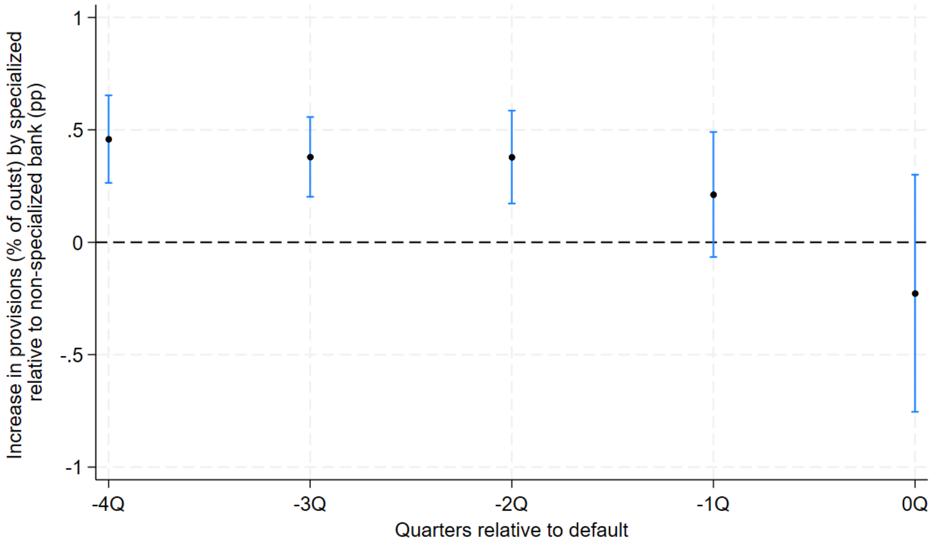


Figure showing the point estimates and 90% confidence intervals of the β_q coefficients estimated in Equation (2), using provisions (as percentage of outstanding loans) as dependent variable. The estimation corresponds to column (4) in Table 4.

In Table 4, we again show variations of Equation (2), differing regarding the control variables and fixed effects. We start in column (1) with a specification which, besides the specialization variables, only includes bank fixed effects and firm-quarter fixed effects. Subsequently, this specification is extended with bank controls (column (2)), bank-quarter fixed effects (column (3)), and bank-firm-quarter control variables (column (4), which corresponds to the results in Figure 4). Columns (5) to (8) expand the main specification by including loan type-quarter and interest rate type-quarter fixed effects, or additional controls for the bank's market share or the existence of a longer-term relationship with the borrower. None of these extensions causes meaningful changes to the main result. In terms of control variables, we find that higher exposures, higher interest rates, and the absence of collateral are associated with higher provisioning. Moreover, as was the case for PDs, we document higher provisions for borrowers which are already behind on their payments, bank-firm pairs where the first default occurs, and bank-firm pairs with a longer relationship. As before, we also use placebo tests to show that our results are not driven by higher risk aversion of banks towards all borrowers in sectors in which they are specialized. The results of these placebo tests can be found in Table A4 in the Appendix. In column (1), we repeat the baseline specification on the slightly restricted sample of defaulting firms for which a non-defaulting PSM match can be found. Next, in columns (2) to (5), we estimate the same Equation (2) on the different samples of non-defaulting nearest neighbors, again allowing us to reject the alternative hypothesis.

The observation that specialized banks assign higher provisions to later-defaulting firms raises the question whether this effect is purely driven by a mechanical relationship between PDs and provisions. To shed additional light on this issue, we explore two angles. First, we exploit the fact that the sample investigated in this subsection consists of both SA and IRB banks. If higher PDs are mechanically driving the provisions, we would expect that the positive effect of specialization is mainly found in the sample of IRB banks, because SA banks are assumed to use more standardized credit risk measures. We investigate this in detail in Table 5. In columns (1) and (2), we make a rather simplified distinction between IRB and SA, respectively, by classifying a bank-firm-quarter observation as IRB if a PD is reported, and as SA otherwise. As hypothesized, we only find that specialization leads to higher pre-default provisioning in the IRB subsample (around 0.7 percentage points four quarters before the default), while we see no effect at all in the SA subsample. This split gives a first approximation of the difference between IRB and SA banks, but it is likely to be imperfect: the unavailability of PDs in AnaCredit for some observations might also be due to other reasons (e.g., data quality or coverage issues), not necessarily because the bank uses the SA for credit risk assessment. Therefore, in columns (3) and (4), we make a similar sample split, but distinguish between banks which report a PD for at least one observation in that quarter (IRB banks) and banks for which this is not the case (SA banks). Finally, in columns (5) and (6), we use additional supervisory data reported by banks in COREP to distinguish between IRB and SA banks.[21] Very similar to the earlier findings, specialization is reflected in higher provisioning towards later-defaulting firms for IRB banks, while we do not find significant results for SA banks. As a second, more direct way of

[21] More specifically, we assign banks as IRB banks if they report at least some IRB exposure in their corporate credit portfolio in COREP.

Table 4: Effect of sectoral specialization on provisioning

Dependent variable:	Provisions as percentage of outstanding loan amount							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
-4Q × $D_{90}^{Specialization}$	0.517***	0.496***	0.489***	0.459***	0.460***	0.454***	0.452***	0.448***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
-3Q × $D_{90}^{Specialization}$	0.489***	0.469***	0.428***	0.380***	0.378***	0.374***	0.374***	0.370***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
-2Q × $D_{90}^{Specialization}$	0.423***	0.402***	0.371***	0.379***	0.381***	0.372***	0.374***	0.370***
	(0.001)	(0.001)	(0.002)	(0.003)	(0.002)	(0.004)	(0.003)	(0.003)
-1Q × $D_{90}^{Specialization}$	0.230	0.245*	0.210	0.212	0.216	0.205	0.208	0.206
	(0.104)	(0.099)	(0.208)	(0.209)	(0.191)	(0.229)	(0.217)	(0.219)
0Q × $D_{90}^{Specialization}$	-0.319	-0.203	-0.221	-0.227	-0.245	-0.235	-0.232	-0.235
	(0.309)	(0.524)	(0.500)	(0.479)	(0.440)	(0.465)	(0.469)	(0.461)
Bank ROA		-0.053						
		(0.841)						
Bank size		1.283						
		(0.600)						
Bank CET1 buffer		0.071						
		(0.437)						
Bank loans/assets		0.019						
		(0.666)						
Bank NPL/loans		-0.425						
		(0.183)						
Maturity				0.098	0.045	0.098	0.100	0.103
				(0.168)	(0.563)	(0.168)	(0.155)	(0.146)
Exposure				1.643***	1.742***	1.624***	1.612***	1.617***
				(0.004)	(0.003)	(0.004)	(0.004)	(0.004)
Interest rate				0.621***	0.648***	0.621***	0.619***	0.619***
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$D^{Protection}$				-2.009***	-1.893***	-2.010***	-2.035***	-2.064***
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$D^{Same\ country}$				0.301	0.390	0.302	0.281	0.259
				(0.604)	(0.537)	(0.603)	(0.627)	(0.652)
$D^{Past\ due\ date}$				4.169***	4.152***	4.169***	4.164***	4.156***
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$D^{First\ default}$				3.776***	3.798***	3.775***	3.769***	3.766***
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$D^{Market\ power}$						0.099		
						(0.611)		
$D_{2018}^{Relationship}$							0.207**	
							(0.016)	
$D_{2019}^{Relationship}$								0.361***
								(0.000)
Bank FE	Yes	Yes	No	No	No	No	No	No
Bank x Quarter FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Firm x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loan type x Quarter FE	No	No	No	No	Yes	No	No	No
Rate type x Quarter FE	No	No	No	No	Yes	No	No	No
N obs	254,869	249,067	251,842	233,242	232,869	233,242	233,242	233,242
R^2	0.5533	0.5539	0.5724	0.6075	0.6080	0.6075	0.6075	0.6076

This table shows the results of the estimation of Equation (2), using provisions as percentage of outstanding loans as dependent variable. Standard errors are clustered at bank and firm level. The numbers in parentheses are p-values. *, **, and *** indicate significance at 10%, 5%, and 1% respectively.

Table 5: Effect of sectoral specialization on provisioning - IRB versus SA

Dependent variable:	Provisions as percentage of outstanding loan amount					
	(1)	(2)	(3)	(4)	(5)	(6)
-4Q × $D_{90}^{Specialization}$	0.699***	-0.104	0.641***	0.147	0.646***	0.159
	(0.001)	(0.557)	(0.000)	(0.435)	(0.000)	(0.422)
-3Q × $D_{90}^{Specialization}$	0.655***	-0.074	0.643***	0.091	0.655***	-0.014
	(0.000)	(0.678)	(0.000)	(0.668)	(0.000)	(0.948)
-2Q × $D_{90}^{Specialization}$	0.503**	0.174	0.512***	0.165	0.504**	0.085
	(0.018)	(0.432)	(0.009)	(0.497)	(0.011)	(0.712)
-1Q × $D_{90}^{Specialization}$	0.351	-0.011	0.275	0.387	0.288	0.165
	(0.142)	(0.967)	(0.264)	(0.219)	(0.242)	(0.598)
0Q × $D_{90}^{Specialization}$	0.426	0.298	0.661	0.175	0.685*	0.381
	(0.338)	(0.466)	(0.107)	(0.726)	(0.096)	(0.470)
Controls	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ
Bank x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
N obs	117,608	53,507	146,105	34,066	144,981	31,602
R^2	0.6369	0.6601	0.6290	0.6777	0.6298	0.6802
Sample	IRB	SA	IRB	SA	IRB	SA
Criterion	PD (obs)	PD (obs)	PD (bank)	PD (bank)	COREP	COREP

This table shows the results of the estimation of Equation (2), using provisions as percentage of outstanding loans as dependent variable. We make a distinction between IRB and SA. In columns (1) and (2), this distinction is made depending on whether the observation has a PD. In columns (3) and (4), this distinction is made depending on whether the bank reports at least one PD. In columns (5) and (6), this distinction is made depending on whether the bank reports non-zero IRB exposure in COREP. All specifications include bank-firm-quarter (BFQ) control variables. Standard errors are clustered at bank and firm level. The numbers in parentheses are p-values. *, **, and *** indicate significance at 10%, 5%, and 1% respectively.

investigating to what extent higher provisioning by specialized banks is driven by higher PDs, we add the PD as an additional explanatory variable to the main specification. As can be seen in Table A5 in the Appendix, a 1 percentage point higher PD leads to provisions which are around 0.16 percentage points higher. However, the significantly positive effects of specialization on provisions almost entirely disappear once controlling for the level of the PD[22], implying that specialized banks mechanically translate higher PDs towards later-defaulting firms into higher provisioning.

[22] Except for the coefficient one quarter before the default which is positive in some specifications.

4.4 Discussion

We now provide a short discussion on the economic relevance of the informational advantages of banks' loan portfolio specialization, focusing on our estimates four quarters in advance of the default. In the main analysis, we have established that the PD assigned to a later-defaulting firm is approximately 3.8 percentage points higher if the bank is specialized in the firm's sector (relative to a non-specialized bank). To get an idea of the economic relevance, we compare this effect with the interquartile range of PDs four quarters before the default (7.77%). This shows that the economic magnitude of the informational advantages linked to loan portfolio specialization is certainly not negligible, since it represents approximately 50% of the interquartile range. A similar back-of-the-envelope calculation regarding the results for provisioning shows somewhat more subdued effects. The increase in provisions by specialized banks of approximately 0.5 percentage points corresponds to slightly more than 25% of the interquartile range of provisions (1.77%), four quarters before the default. The somewhat lower relevance for provisioning is not surprising, given that PDs are only one input parameter in determining provisioning, next to, e.g., the availability of collateral.

In this context, we acknowledge that the focus of this study is limited to assessing the informational advantages of sectoral specialization for the credit risk assessment of banks' existing corporate loan portfolio. Since the AnaCredit dataset only captures outstanding corporate loans, not loan applications and/or rejections, we cannot investigate to what extent sectoral specialization may already help banks in better distinguishing between safe and likely-to-default borrowers during the loan application process, potentially leading to more loan application rejections for likely-to-default borrowers.

5 Robustness

In this section, we establish the robustness of our results to alternative specialization measures, subsamples, default definitions, and time periods. We also show that the effect is not driven by, among other things, individual countries or individual sectors.

In columns (1) to (5) of Table 6, we estimate Equation (2) using alternative specialization measures.[23] First, in column (1), we use the continuous specialization measure as defined in Equation (1) instead of the high specialization dummy. In column (2), we use an alternative continuous specialization measure which does not correct for the total lending by all banks in the same country to that sector (cf. the discussion in Section 2.2). Both alternatives produce similar results, with PDs around 0.3 percentage points higher four quarters before the default for every 1 percentage point increase in sectoral specialization. Similar to the baseline specification, the relative informational advantages of specialization decrease closer to the default date, although there is still a small positive effect one quarter before the default. In column (3), we use a continuous measure of specialization defined using less granular NACE1 sectors instead of NACE2 sectors. We find similar positive coefficients for the first quarters, again decreasing in magnitude when approaching the default date. In previous estimations, sectoral specialization coefficients in the quarter of default often showed a negative sign, albeit insignificant. Using NACE1 sectors, this negative coefficient in the quarter of default becomes significant, pointing towards less specialized banks assigning higher PDs in the quarter of the default than specialized banks, potentially overcompensating for their earlier underestimation. In column (4), we use the original relative specialization measure as defined by Paravisini et al. (2023), in which we correct for the size of the sector by dividing the pure exposure measure by the percentage exposure of all banks in the same country towards that industry (instead of subtracting). While the measure is scaled very differently compared to the other continuous measures, we still observe a significantly positive coefficient four quarters before the default, decreasing in magnitude and becoming insignificant closer to the default date. In terms of economic magnitude, a one standard deviation increase in this relative specialization measure corresponds to PDs which are 1 percentage point higher four quarters before the default. As a final alternative specialization measure, column (5) shows the results using a dummy which is 1 if the specialization measure is in the highest quartile (instead of the highest decile). Again, PDs by specialized banks are significantly higher four quarters before the default (around 3.9 percentage points) and decrease closer to the quarter of default.

As argued by Behn et al. (2022) and Plosser & Santos (2018), banks with low capital buffers

[23] The correlations between the specialization measures are presented in Table A6 in the Appendix.

Table 6: Alternative specialization measures and subsamples

Dependent variable:	Probability of default								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
-4Q × specialization	0.307***	0.298***	0.319**	0.023**	3.894**	4.932***	5.683***	6.728***	4.951***
	(0.002)	(0.001)	(0.020)	(0.020)	(0.022)	(0.006)	(0.007)	(0.006)	(0.003)
-3Q × specialization	0.235***	0.224***	0.268**	0.020**	2.005**	4.603***	5.723***	5.437***	3.580***
	(0.005)	(0.006)	(0.013)	(0.013)	(0.018)	(0.007)	(0.004)	(0.004)	(0.006)
-2Q × specialization	0.118***	0.121***	0.153***	0.005	0.147	2.952**	4.025***	3.143***	2.774***
	(0.007)	(0.005)	(0.005)	(0.156)	(0.749)	(0.020)	(0.010)	(0.005)	(0.000)
-1Q × specialization	0.074*	0.071**	0.069**	-0.003	0.254	-0.383	-0.802	1.380	1.344**
	(0.061)	(0.049)	(0.043)	(0.580)	(0.633)	(0.530)	(0.250)	(0.241)	(0.039)
0Q × specialization	-0.122	-0.162	-0.268***	-0.011	-1.743	-0.199	1.131	-5.273**	-1.838
	(0.302)	(0.167)	(0.005)	(0.354)	(0.145)	(0.875)	(0.369)	(0.026)	(0.200)
Controls	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ
Bank x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N obs	117,608	117,608	117,608	117,608	117,608	99,961	75,027	24,190	69,489
R^2	0.6934	0.6934	0.6939	0.6935	0.6938	0.6917	0.6831	0.7071	0.7012
Specialization variable	Spec	Spec_{alt}	$\text{Spec}_{\text{NACE1}}$	Spec_{rel}	D_{75}^{Spec}	D_{90}^{Spec}	D_{90}^{Spec}	D_{90}^{Spec}	D_{75}^{Spec}
Specialization included	All	All	All	All	All	All	All	P90 & P10	P75 & P25
Banks	Baseline	Baseline	Baseline	Baseline	Baseline	Buffer>P10	Buffer>P25	Baseline	Baseline

This table shows the results of the estimation of Equation (2), using alternative specialization measures or subsamples (well-capitalized banks, and firms with loans from banks with strong specialization differences). All specifications include bank-firm-quarter (BFQ) control variables. Standard errors are clustered at bank and firm level. The numbers in parentheses are p-values. *, **, and *** indicate significance at 10%, 5%, and 1% respectively.

tend to systematically underreport their credit risk exposure. Hence, there is a concern that this subset of banks does not fully incorporate the potential informational advantages of specialization in their PDs. Therefore, following Faria-e Castro et al. (2024), columns (6) and (7) of Table 6 show the results of a robustness check in which banks with capital buffers (CET1 ratio minus overall capital requirements) in the lowest decile or quartile are removed, respectively. Indeed, our findings show that the results become even stronger (up to 4.9 or 5.7 percentage points four quarters before the default) when excluding banks with low capital buffers.

In the last columns of Table 6, we focus on a subsample of borrowers for which the banks differ more strongly in terms of their degree of specialization in the borrower's sector. More specifically, we only retain bank-firm pairs for which the bank is in the highest decile (quartile) or lowest decile (quartile) of the distribution, i.e., we remove moderately specialized banks. In combination with the inclusion of firm-quarter fixed effects, this implies that we are effectively comparing PDs assigned by highly specialized banks with PDs assigned by the least specialized banks to the same later-defaulting borrower. A priori, we expect that the informational advantages of specialization should be even larger in this subsample. Indeed, in column (8), we document that PDs assigned by banks in the highest decile of specialization are approximately 6.7 percentage points larger than PDs assigned by banks in the lowest decile. Column (9) shows that PDs by banks in the highest specialization quartile are almost 5 percentage points higher than PDs by banks in the lowest specialization quartile.

In Table 7, we report additional tests to ensure that our results are not driven by specialized banks causing firms to default on their outstanding loans (i.e., to ensure defaults are not endogenous to banks' specialization), and that the main findings are robust to alternative definitions of default. First, one might be concerned that the specialized bank could positively or negatively affect the likelihood of a (future) default of the firm by adjusting its credit conditions.[24] This is especially relevant if the specialized bank is the firm's most important credit supplier. In column (1), we therefore estimate the main specification on a restricted sample, only including firms which obtain more lending (in terms of total committed loan amount) from their non-specialized than from their specialized banks. In column (2), we impose the stricter requirement that the firm borrows at least twice as much from its non-specialized banks than from its specialized banks. In those subsamples with a large bulk of the firm's lending originated by non-specialized banks, it is less likely that any shift in lending or loan conditions of the specialized bank is affecting the future default. Both columns still validate our main findings, if anything even stronger in terms of magnitude and significance. In column (3) of Table 7, we explore this issue from a different angle by only retaining bank-firm combinations for which total committed amount did not change over the five quarters analyzed. In column (4), we only include observations where both committed amount and interest rate did not change. Again, we find qualitatively similar results to the baseline in both restricted samples, with specialized banks assigning higher PDs to later-defaulting firms several quarters before the default. Second, we also want to avoid that our results are driven by firms deciding to strategically default at their specialized banks rather than at their other lenders. In the baseline setup, we include all bank-firm pairs with an outstanding loan as soon as the firm defaults on a loan with one of its

[24] While it would be interesting to analyze whether specialized banks actively try to adjust their credit portfolio when they update the borrower's PD, e.g. to reduce their exposure to likely-to-default firms, this is outside the scope of this paper and we leave it for future research. We therefore focus in this robustness check on ruling out that this potential reaction by banks is actively causing defaults, thus biasing our findings.

banks. Hence, if firms systemically default first at their specialized banks, this may be an alternative explanation why specialized banks assign higher PDs. The main analysis already takes this concern into account by including a dummy variable capturing whether the first default occurs at this specific bank-firm pair or not. Nevertheless, in column (5) of Table 7, we go a step further and remove all bank-firm pairs from the sample if the firm did not default at that particular bank during 2023. We still find that specialized banks assign statistically higher PDs to defaulting firms in this robustness check, with an effect of around 1.7 percentage points four quarters before the default. Finally, in column (6), we show the results of our analysis using an alternative default definition. Following Becker et al. (2020), we classify borrowers as defaulting if they are at least 90 days past the due date on their loan. Results again confirm that sectoral specialization leads to informational advantages in advance of the default.

Table 7: Robustness - specialization and defaults

Dependent variable:	Probability of default					
	(1)	(2)	(3)	(4)	(5)	(6)
-4Q × $D_{90}^{Specialization}$	5.199**	5.202***	4.201**	5.816**	1.652**	1.666***
	(0.013)	(0.006)	(0.028)	(0.031)	(0.039)	(0.003)
-3Q × $D_{90}^{Specialization}$	5.492***	5.326***	4.880**	4.692*	1.734**	1.351***
	(0.009)	(0.006)	(0.012)	(0.061)	(0.038)	(0.005)
-2Q × $D_{90}^{Specialization}$	3.570***	3.334***	2.714*	3.439	0.387	-0.012
	(0.007)	(0.003)	(0.065)	(0.121)	(0.543)	(0.984)
-1Q × $D_{90}^{Specialization}$	1.212*	0.689	-0.259	-0.977	-0.550	-1.365
	(0.079)	(0.390)	(0.812)	(0.692)	(0.544)	(0.136)
0Q × $D_{90}^{Specialization}$	-0.347	-2.433	-0.896	-5.573	-1.768	0.205
	(0.852)	(0.243)	(0.654)	(0.132)	(0.373)	(0.847)
Controls	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ
Bank x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
N obs	24,609	16,088	38,253	15,989	29,858	95,671
R^2	0.6709	0.6587	0.7275	0.7492	0.8194	0.6975
Restriction	Non-spec. > spec.	Non-spec. > 2 x spec.	Constant amount	Constant rate & amount	Bank-firm has defaulted	90-day default definition

This table shows the results of the estimation of Equation (2) on various subsamples: only firms with more lending from non-specialized than from specialized banks in columns (1) and (2); only bank-firm observations with constant loan characteristics in columns (3) and (4); alternative default definitions in columns (5) and (6). All specifications include bank-firm-quarter (BFQ) control variables. Standard errors are clustered at bank and firm level. The numbers in parentheses are p-values. *, **, and *** indicate significance at 10%, 5%, and 1% respectively.

Up to this point, we focused exclusively on firms defaulting during 2023 in order to avoid distorting influences from the Covid-19 pandemic as much as possible, as explained in Section 2.3. Nevertheless, to show the robustness of our results to alternative time periods, we extend our sample period in Table 8 and also include defaults occurring in any quarter in 2022. Investigating PDs up to four quarters before the default implies that the earliest observations in the sample then correspond to 2021Q1. We again follow the same setup and show the results starting from a minimalist specification without any control variables and with only bank and firm-quarter fixed effects, up to the most saturated version with bank-firm-quarter controls, bank-quarter and firm-quarter fixed effects, as well as alternative versions including additional variables to control for banks' market share or relationship lending. In all specifications, ex-ante PDs to defaulting firms by specialized banks are again significantly higher than by non-specialized banks. Focusing on the baseline specification in column (4), the magnitude of this effect goes down from around 2.8 percentage points four quarters before the default to 2 percentage points two quarters before the default, before becoming insignificant.

Even though our analysis makes use of the full euro area credit register, we nevertheless have to check to what extent the results are potentially driven by observations from a single country only. This test is especially warranted since the majority of defaulting firms are headquartered in France, Spain and Italy (cf. Panel B in Figure 2). Therefore, we re-estimate Equation (2) on the sample of firms defaulting in 2023, but omit each of the countries one by one. As can be observed in Table A7 in the Appendix, there is not a single country fully driving the results, since the coefficients on the specialization variables remain significantly positive in the early quarters, before decreasing and becoming insignificant closer to the quarter of default as in the baseline, in each of the columns.[25] Similarly, Table A8 shows that the results do not change meaningfully when excluding each of the sectors (defined at NACE1 level) one by one.

Tables A9 to A11 in the Appendix provide the results of several additional robustness checks. All checks corroborate our main finding that specialized banks assign higher PDs to later-defaulting firms.[26] In Table A9, we first replace the baseline time-varying specialization measure by predetermined variables. Additionally, we follow De Jonghe et al. (2024) and construct dummy variables capturing whether the bank's specialization measure (uncorrected for the size of the sector) is in the highest decile or quartile of that specific sector's specialization distribution. We thus ensure an equal number of specialized banks in all sectors. In Table A10, we use

[25] The coefficient becomes smallest when excluding Spain, but remains highly statistically significant.
[26] Unreported placebo tests report no effects towards non-defaulting firms. Results available on request.

Table 8: Effect of sectoral specialization on PDs (including defaults in 2022)

Dependent variable:	Probability of default							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
-4Q × $D_{90}^{Specialization}$	2.311**	2.124**	2.398**	2.833***	2.799***	2.835***	2.783***	2.800***
	(0.022)	(0.029)	(0.018)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)
-3Q × $D_{90}^{Specialization}$	2.373**	2.234**	2.414**	2.698***	2.644***	2.699***	2.662***	2.677***
	(0.013)	(0.015)	(0.016)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)
-2Q × $D_{90}^{Specialization}$	1.602*	1.603*	1.791**	1.997***	1.959**	1.999**	1.964**	1.972**
	(0.065)	(0.069)	(0.046)	(0.010)	(0.011)	(0.010)	(0.011)	(0.011)
-1Q × $D_{90}^{Specialization}$	0.016	0.046	0.077	-0.151	-0.190	-0.149	-0.179	-0.166
	(0.974)	(0.927)	(0.876)	(0.778)	(0.729)	(0.780)	(0.736)	(0.755)
0Q × $D_{90}^{Specialization}$	-0.639	-0.634	-0.875	0.248	0.196	0.251	0.217	0.230
	(0.676)	(0.687)	(0.528)	(0.835)	(0.869)	(0.833)	(0.856)	(0.848)
Bank ROA		-3.900*						
		(0.083)						
Bank size		-0.491						
		(0.956)						
Bank CET1 buffer		-0.204						
		(0.644)						
Bank loans/assets		0.135						
		(0.313)						
Bank NPL/loans		-0.775						
		(0.262)						
Maturity				0.624	0.256	0.624	0.666	0.669
				(0.200)	(0.458)	(0.200)	(0.174)	(0.171)
Exposure				4.882	7.615	4.915	4.267	4.151
				(0.288)	(0.125)	(0.287)	(0.351)	(0.364)
Interest rate				0.485***	0.545***	0.485***	0.453***	0.460***
				(0.004)	(0.002)	(0.004)	(0.004)	(0.004)
$D^{Protection}$				-0.746	-0.408	-0.745	-0.973*	-0.996*
				(0.142)	(0.380)	(0.142)	(0.074)	(0.070)
$D^{Same\ country}$				-1.656	-1.384	-1.655	-1.905	-1.869
				(0.164)	(0.224)	(0.164)	(0.109)	(0.117)
$D^{Past\ due\ date}$				14.727***	14.549***	14.728***	14.676***	14.646***
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$D^{First\ default}$				17.001***	16.996***	17.001***	16.943***	16.926***
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$D^{Market\ power}$						-0.292		
						(0.875)		
$D_{2018}^{Relationship}$							1.507**	
							(0.011)	
$D_{2019}^{Relationship}$								1.647***
								(0.007)
Bank FE	Yes	Yes	No	No	No	No	No	No
Bank x Quarter FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Firm x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loan type x Quarter FE	No	No	No	No	Yes	No	No	No
Rate type x Quarter FE	No	No	No	No	Yes	No	No	No
N obs	235,747	235,186	235,636	216,818	216,720	216,818	216,818	216,818
R^2	0.6252	0.6258	0.6353	0.6926	0.6935	0.6926	0.6929	0.6929

This table shows the results of the estimation of Equation (2), including firms defaulting in 2022 and 2023. Standard errors are clustered at bank and firm level. The numbers in parentheses are p-values. *, **, and *** indicate significance at 10%, 5%, and 1% respectively.

variations of our main set of control variables: we include alternative market share measures and add a variable indicating the borrower's main bank, which is a relationship variable proposed by Bonfim et al. (2023). We also replace the baseline measure capturing the bank's exposure to the firm by other indicators such as the total outstanding loan amount, total committed loan amount, or number of outstanding loans (all in natural logarithms). Results are qualitatively similar when using these alternative specialization or control variables. In Table A11, we first focus on the distinction between significant institutions (SIs) and less-significant institutions (LSIs). Given that only IRB banks report PDs in AnaCredit, the vast majority of loans in our final sample are issued by SIs. However, a small number of loans (less than 2% of the observations in the baseline regression) are issued by LSIs, which are potentially different in terms of business model or other characteristics. We confirm that our baseline results hold when removing all loans by LSIs from the sample. Second, while most observations consist of IFRS 9 loans, some of the loans in the sample follow other accounting standards (e.g., GAAP). Since Behn & Couaillier (2023) argue that IFRS 9 offers greater discretion in terms of, e.g., provisioning behavior than GAAP, we also test our results for the subsample of borrowers with IFRS 9 loans only (around 90% of the observations), which does not impact the results. Column (3) and (4) of Table A11 show that the results are robust to only including bank-firm pairs with data available in all five quarters of the analysis, and only including "unexpected defaults", i.e., observations with zero days past due date four quarters before the default. Columns (5) and (6) show that the results are virtually unchanged when removing the top decile or top quartile of loans in terms of loan size. Among other things, this mitigates potential concerns that a few large loan exposures would influence the specialization measures. In the last columns of Table A11, we also assess the impact of alternative clustering choices, by clustering standard errors at the bank and sector level or at the bank-sector level (De Jonghe et al., 2024) and document that our baseline results remain highly statistically significant.

6 Extensions

6.1 Heterogeneity

In a first extension, we exploit the fact that our dataset captures the quasi-universe of corporate lending by euro area banks, including lending to SMEs, to investigate the heterogeneity of our results with respect to the size of the borrower. A priori, one would expect that specialization provides informational advantages (and hence earlier prediction of the default) especially for lending to smaller firms, since these are informationally more opaque and their creditworthiness is therefore more difficult to assess (Bharath et al., 2011; Blickle et al., 2024). In columns (1) to (4) of Table 9, we split the sample in four quartiles of increasing firm size. Column (1) shows that, four quarters before the default, banks assign PDs which are approximately 5.9 percentage points higher to the smallest defaulting firms if they are specialized in the firm's sector. For firms in the second and third quartile, this effect is around 4.1 to 4.3 percentage points. However, for the subsample of largest firms we do not find significantly higher PDs by specialized compared to non-specialized banks four quarters before the default, as documented in column (4). For the largest firms, we only find significance in later quarters and, except for the last quarter before the default, always lower in magnitude than for the small firms.

Subsequently, we examine heterogeneity with respect to relationship lending to investigate whether loan portfolio specialization and relationship lending act as complements or rather substitutes. In earlier specifications, we already included a dummy variable capturing whether or not the bank-firm pair had a longer-term lending relationship, and documented that relationship lending also leads to improved credit risk assessment (higher PDs towards later-defaulting firms). Now, we investigate whether the positive effect of loan portfolio specialization is heterogeneous depending on the presence or absence of a longer-term lending relationship. More specifically, we split the sample between bank-firm pairs with and without such a longer-term relationship and re-estimate the baseline specification. In columns (5) and (6), we split the sample based on the existence of a bank-firm relationships in 2018Q4, while we use a 2019Q4 indicator in columns (7) and (8). We find significantly higher ex-ante PDs for specialized banks in all subsamples, indicating that specialization leads to informational advantages for both banks with and without longer-term relationship with their borrowers. However, comparing columns (5)

Table 9: Heterogeneous effects of loan portfolio specialization

Dependent variable:	Probability of default							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
-4Q × $D_{90}^{\text{Specialization}}$	5.907**	4.129*	4.334*	2.756	4.468***	2.744*	4.081***	3.114**
	(0.011)	(0.054)	(0.060)	(0.148)	(0.001)	(0.064)	(0.002)	(0.036)
-3Q × $D_{90}^{\text{Specialization}}$	4.634**	4.378**	4.152*	3.743**	4.153***	2.958**	3.958***	3.594**
	(0.027)	(0.031)	(0.089)	(0.039)	(0.001)	(0.046)	(0.008)	(0.018)
-2Q × $D_{90}^{\text{Specialization}}$	5.145**	4.117**	1.986	2.489**	2.891**	1.812*	2.114**	2.767**
	(0.020)	(0.041)	(0.143)	(0.029)	(0.010)	(0.082)	(0.047)	(0.014)
-1Q × $D_{90}^{\text{Specialization}}$	-0.696	-1.685	0.083	1.725*	0.093	-0.868	-0.222	-0.027
	(0.671)	(0.195)	(0.952)	(0.059)	(0.924)	(0.387)	(0.861)	(0.975)
0Q × $D_{90}^{\text{Specialization}}$	3.747	1.917	-1.483	0.680	-0.426	-0.841	0.269	-1.805
	(0.166)	(0.362)	(0.445)	(0.742)	(0.803)	(0.633)	(0.894)	(0.239)
Controls	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ
Bank x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N obs	23,561	23,202	23,569	23,433	49,211	33,769	33,755	47,686
R^2	0.7131	0.6911	0.6627	0.6599	0.7123	0.7163	0.7176	0.7136
Firm size	Quartile 1	Quartile 2	Quartile 3	Quartile 4	All	All	All	All
Relationship lending	Yes & no	Yes & no	Yes & no	Yes & no	No	Yes	No	Yes
Relationship indicator	None	None	None	None	end-2018	end-2018	end-2019	end-2019

This table shows the results of the estimation of Equation (2), for subsamples of firms or bank-firm pairs. All specifications include bank-firm-quarter (BFQ) control variables. Standard errors are clustered at bank and firm level. The numbers in parentheses are p-values. *, **, and *** indicate significance at 10%, 5%, and 1% respectively.

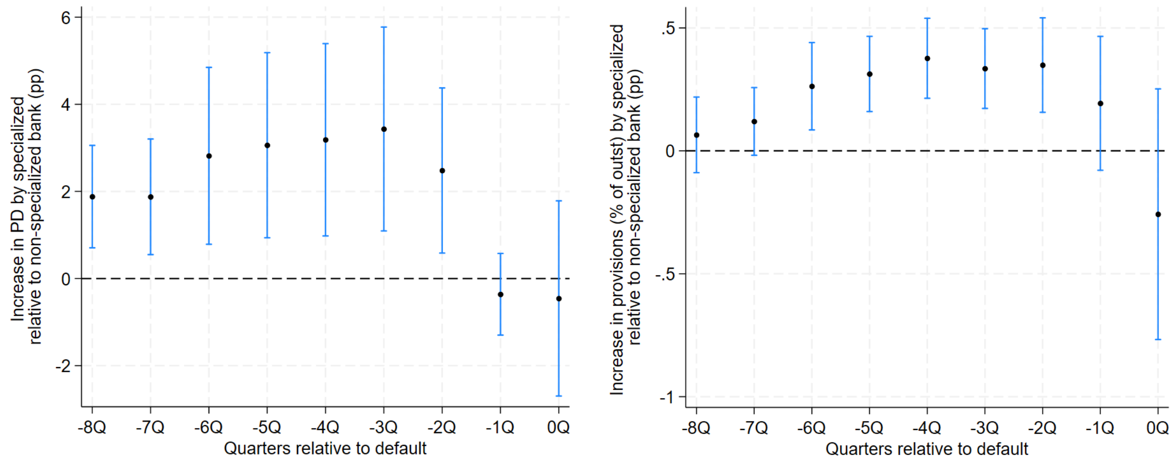
and (7) - no relationship lending - to columns (6) and (8) - relationship lending - reveals that the positive effect of specialization is larger in both magnitude and significance for the bank-firm pairs without longer-term relationship. These findings indicate that, while loan portfolio specialization and relationship lending both provide informational advantages for credit risk assessment, they are imperfect substitutes.

6.2 Time horizon of the specialization effect

In the main analysis, we provide evidence that PDs and provisions assigned by specialized banks to defaulting borrowers are significantly higher four quarters before the default. In this section, we explore three extensions to shed more light on the timing of the informational advantages of specialization.

First, we extend the time dimension of the analysis. While the baseline setup allows to investigate relative differences in PD and provisioning between specialized and non-specialized banks up to four quarters before the default, this extension includes data up to eight quarters before the default. Figure 5, Panel A presents the effect of specialization on PDs, while Panel B shows the effect on provisioning. While PDs by specialized banks are already somewhat higher (around 2 percentage points) eight quarters before the default, we find that the difference in provisioning between specialized and non-specialized banks is not statistically different from zero yet at that point. For both variables, we observe a similar subsequent evolution over time: an inverted U-shape indicates that the informational advantages of specialization first increase, towards a peak around three to four quarters before the default, before disappearing again in the last quarters (as described in detail before). This pattern suggests that the benefits of specialization described in this analysis relate mostly to monitoring advantages over the lifetime of the loan, with specialized banks having an advantage in detecting early warning signals in the immediate antecedence of the default, which is translated into an earlier adjustment of PDs and provisioning compared to non-specialized banks.

Figure 5: Effect of sectoral specialization - extended time period



A: Probability of default

B: Provisioning

Figure showing the point estimates and 90% confidence intervals of the β_q coefficients estimated in Equation (2), including up to eight quarters before the default, using PDs (Panel A) and provisions as percentage of outstanding loans (Panel B) as dependent variables.

In the second extension, we investigate more formally whether specialized banks indeed significantly increase PDs and provisioning more in advance of the default. More specifically, we examine whether the change in PDs and provisions between eight and four quarters before the default is significantly higher for specialized banks, by estimating Equation (3).

$$\Delta Y_{b,f,s} = \beta \cdot Specialization_{b,s} + \gamma \cdot Z_{b,f} + \eta_f + \alpha_b + \epsilon_{b,f,s} \tag{3}$$

The dependent variable captures the change in PD (or provisioning), assigned by bank b to firm f which is active in sector s, between eight and four quarters before the default. The main explanatory variable is a (dummy or continuous) measure indicating to what extent the bank is specialized in the sector of the firm. We include bank-firm control variables, as well as bank and firm fixed effects, which has similar implications as including bank-quarter and firm-quarter fixed effects in Equation (2). More specifically, the bank fixed effects allow to account for broader bank-specific changes in PD or provisioning towards all firms, while the firm fixed effects ensure that we are comparing changes in PDs and provisions by two or more banks to the same firm. All explanatory variables are measured eight quarters before the quarter of default. The results are presented in Table 10. In columns (1) and (2), we observe that specialized banks increase PDs of defaulting firms more between eight and four quarters before the default: using the high specialization dummy variable, we find that highly specialized banks increase PDs by almost 1 percentage point more than non-specialized banks, while the continuous measure indicates that a 1 percentage point higher specialization leads to a 0.13 percentage points larger increase in PD. In columns (3) and (4), we show that specialized banks also increase provisions to defaulting firms more: 0.17 percentage points for the high specialization dummy variable, and 0.03 percentage points for every percentage point increase in the continuous specialization measure. In columns (5) to (8), we employ similar placebo tests as before (focusing on the sample matched using PSM) and document that we do not find similar positive effects towards non-defaulting borrowers.

Summarizing, our analyses indicate that specialization allows banks to predict future defaults in their corporate loan portfolio better, as specialized banks assign higher PDs and provisions in advance of the default. Moreover, we find that an important part of the story is that these banks actively adjust PDs and provisions earlier than non-specialized banks. To cover this issue from all angles, our final extension examines whether specialized banks already have informational advantages (better future default predictions) at the moment when new loans are issued. For

Table 10: Effect of sectoral specialization on changes in PDs and provisions

Dependent variable:	Δ Probability of default		Δ Provisions		Δ Probability of default		Δ Provisions	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$D_{90}^{Specialization}$	0.935*		0.166**		-0.030		0.006	
	(0.068)		(0.041)		(0.809)		(0.743)	
Specialization		0.133**		0.025***		-0.019*		-0.000
		(0.030)		(0.002)		(0.062)		(0.956)
Maturity	-0.063	-0.054	0.014	0.015	0.000	-0.002	-0.001	-0.001
	(0.721)	(0.763)	(0.744)	(0.725)	(0.999)	(0.969)	(0.958)	(0.959)
Exposure	5.064	4.180	-0.433	-0.494	1.146	1.574*	-0.135	-0.130
	(0.371)	(0.461)	(0.306)	(0.239)	(0.231)	(0.097)	(0.459)	(0.498)
Interest rate	-0.520**	-0.525**	0.091**	0.091**	-0.155***	-0.156***	-0.029***	-0.029***
	(0.019)	(0.018)	(0.023)	(0.024)	(0.000)	(0.000)	(0.002)	(0.003)
$D^{Protection}$	0.169	0.134	-0.149	-0.151	0.111	0.119	0.046	0.047
	(0.764)	(0.811)	(0.258)	(0.253)	(0.287)	(0.250)	(0.135)	(0.134)
$D^{Same\ country}$	-4.637**	-4.362**	0.079	0.105	-0.424*	-0.431*	-0.247***	-0.248***
	(0.028)	(0.037)	(0.727)	(0.644)	(0.080)	(0.076)	(0.006)	(0.006)
$D^{Past\ due\ date}$	-0.412	-0.465	0.236	0.230	0.499	0.503	-0.023	-0.023
	(0.736)	(0.704)	(0.405)	(0.415)	(0.181)	(0.178)	(0.874)	(0.874)
$D^{First\ default}$	2.655**	2.671**	0.448***	0.450***				
	(0.026)	(0.025)	(0.001)	(0.001)				
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N obs	18,302	18,302	33,274	33,274	74,495	74,495	136,174	136,174
R^2	0.5677	0.5678	0.4897	0.4898	0.5374	0.5374	0.4471	0.4471
Sample	Default	Default	Default	Default	Non-default	Non-default	Non-default	Non-default

This table shows the results of the estimation of Equation (3). Columns (1), (2), (5) and (6) use the change in PD as dependent variable, while columns (3), (4), (7) and (8) use the change in provisions (as percentage of outstanding loans) as dependent variable. Changes are between eight and four quarters before the default. Columns (1) to (4) show the results for defaulting firms, while columns (5) to (8) show the results for a sample of non-defaulting firms, determined using PSM. Standard errors are clustered at bank and firm level. The numbers in parentheses are p-values. *, **, and *** indicate significance at 10%, 5%, and 1% respectively.

this analysis, we focus on all loans originated between 2019 and 2022 to firms which default for the first time in 2023. Formally, we estimate a slightly adapted version of Equation (2), in which the main specialization variable is interacted with an indicator for the year of origination of the loan. This specification allows to investigate whether new loans, issued in every year between 2019 and 2022 to a later-defaulting firm, have significantly higher PDs or provisions at the time of origination if they are issued by a specialized bank. In contrast to the previous analyses, which all compared specialized and non-specialized banks' credit risk assessment regarding outstanding loans, this analysis only includes new loan originations. Given that we still want to compare lending by two banks to the same firm in the same time period, we collapse the dataset in this part of the analysis to the bank-firm-year level (i.e., all new loan issuances by bank b to firm f in year t) instead of the bank-firm-quarter level, to ensure sufficient observations for a meaningful analysis.

The main findings of this final extension are shown in Table 11. Based on columns (1) and (2), we observe that the coefficient on the specialization variables, measured as a dummy or continuously, consistently shows a positive sign in all years. However, the coefficient is rarely significant (only at 10% using the continuous specialization measure, for loans originated in 2020). Thus, we cannot confidently state that PDs by specialized banks are already significantly higher than PDs by non-specialized banks at the moment when a new loan is issued to a later-defaulting borrower. In columns (3) and (4), we do not observe significant differences in provisioning between specialized and non-specialized banks either. For completion, placebo tests for non-defaulting firms determined using PSM are included in columns (5) to (8) of Table 11.

Table 11: Effect of sectoral specialization on PDs and provisioning at origination

	Probability of default		Provisions		Probability of default		Provisions	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2019 × $D_{90}^{Specialization}$	0.445		-0.035		0.478		-0.022	
	(0.391)		(0.565)		(0.217)		(0.327)	
2020 × $D_{90}^{Specialization}$	0.645		-0.069		0.039		-0.005	
	(0.219)		(0.111)		(0.792)		(0.753)	
2021 × $D_{90}^{Specialization}$	0.745		-0.023		0.088		0.001	
	(0.176)		(0.779)		(0.620)		(0.963)	
2022 × $D_{90}^{Specialization}$	0.636		-0.037		-0.172		-0.027	
	(0.489)		(0.821)		(0.145)		(0.288)	
2019 × Specialization		0.027		0.003		0.001		-0.001
		(0.548)		(0.671)		(0.964)		(0.716)
2020 × Specialization		0.103*		0.004		0.034		0.005*
		(0.060)		(0.684)		(0.104)		(0.081)
2021 × Specialization		0.094		0.009		0.017		0.003
		(0.110)		(0.377)		(0.403)		(0.131)
2022 × Specialization		0.043		0.004		-0.022		0.004
		(0.746)		(0.736)		(0.206)		(0.199)
Controls	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ
Bank x Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm x Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N obs	54,874	54,874	97,040	97,040	138,976	138,976	237,288	237,288
R^2	0.5961	0.5961	0.5477	0.5477	0.6578	0.6578	0.5304	0.5304
Sample	Default	Default	Default	Default	Non-default	Non-default	Non-default	Non-default

This table shows the results of the estimation of an adapted version of Equation (2), in which we focus on new loan issuances (at the bank-firm-year level) instead of outstanding loans. Columns (1), (2), (5) and (6) use the PD as dependent variable, while columns (3), (4), (7) and (8) use provisions (as percentage of outstanding loans) as dependent variable. Columns (1) to (4) show the results for defaulting firms, while columns (5) to (8) show the results for a sample of non-defaulting firms, determined using PSM. Standard errors are clustered at bank and firm level. The numbers in parentheses are p-values. *, **, and *** indicate significance at 10%, 5%, and 1% respectively.

These results confirm that the informational advantages of specialization described in this paper are driven by better monitoring, and by a quicker adjustment of credit risk assessment by specialized banks over the lifetime of the loan, not necessarily by better screening at origination. However, it is again important to draw the reader's attention to the caveat discussed in Section 4.4: the unavailability of information regarding loan applications and rejections in the credit register implies that a full analysis of the potential screening benefits of specialization is beyond the scope of this paper, as our current analysis does not investigate whether specialized banks' screening abilities allow them to reject more likely-to-default borrowers' loan applications as well.

7 Conclusion

The presence of informational asymmetries between lenders and borrowers is fundamental to understanding the role of banks as financial intermediaries. While the literature has argued that specialization of the loan portfolio should enable banks to obtain better information about their borrowers, leading to more accurate credit risk monitoring, the empirical investigation of these informational advantages has been largely indirect given the private nature of banks' internal risk assessments.

In this paper, we provide a direct empirical test of the informational advantages related to sectoral specialization. Using granular credit register data for euro area corporate loans, we find that specialization indeed leads to informational advantages for credit risk assessment: ex-ante, banks assign PDs which are 3.8 percentage points higher to later-defaulting firms, if they are highly specialized in the sector of that firm. We demonstrate that these higher PDs are mostly driven by specialized banks actively raising PDs more (and especially earlier) than non-specialized banks, not by higher PDs at origination of the loan, indicating that specialization contributes to early warning signals which banks actively integrate into their credit risk evaluations. We show that the effects are not driven by higher risk aversion towards all firms in specialized sectors, and that the effect is strongest for lending towards smaller firms. While we document positive effects of loan portfolio specialization for lending to borrowers with and without longer-term relationship, the effect is stronger for the latter, suggesting that sectoral specialization and relationship lending can act as substitutes, but only to some extent. Additionally, we find that the informational advantages also feed through into higher provisioning towards later-defaulting firms.

Our results highlight that sectoral specialization can be beneficial for the safety and soundness of individual banks by enhancing banks' credit risk assessment. By revealing key details of the channel through which these benefits materialize, our analysis sheds new light on possible avenues to optimally combine the conventional financial stability benefits of diversification with the informational advantages from sectoral expertise.

References

Acharya, V. V., Hasan, I., & Saunders, A. (2006). Should banks be diversified? Evidence from individual bank loan portfolios. *The Journal of Business*, 79(3), 1355–1412.

Baele, L., De Jonghe, O., & Vander Vennet, R. (2007). Does the stock market value bank diversification? *Journal of Banking & Finance*, 31(7), 1999–2023.

Bao, Y. (2022). Peer information in loan pricing. *Journal of Corporate Finance*, 76, 102248.

Beck, T., De Jonghe, O., & Mulier, K. (2022). Bank sectoral concentration and risk: Evidence from a worldwide sample of banks. *Journal of Money, Credit and Banking*, 54(6), 1705–1739.

Becker, B., Bos, M., & Roszbach, K. (2020). Bad times, good credit. *Journal of Money, Credit and Banking*, 52(S1), 107–142.

Behn, M. & Couaillier, C. (2023). Same same but different: credit risk provisioning under IFRS 9. Working paper 2841, ECB.

Behn, M., Haselmann, R., & Vig, V. (2022). The limits of model-based regulation. *The Journal of Finance*, 77(3), 1635–1684.

Berger, A. N. & Udell, G. F. (1995). Relationship lending and lines of credit in small firm finance. *Journal of Business*, 68(3), 351–381.

Berger, P. G. & Ofek, E. (1996). Bustup takeovers of value-destroying diversified firms. *The Journal of Finance*, 51(4), 1175–1200.

Beyhaghi, M., Howes, C., & Weitzner, G. (2024). The information advantage of banks: Evidence from their private credit assessments. Working paper.

Bharath, S. T., Dahiya, S., Saunders, A., & Srinivasan, A. (2011). Lending relationships and loan contract terms. *The Review of Financial Studies*, 24(4), 1141–1203.

Blickle, K., Parlatore, C., & Saunders, A. (2024). Specialization in banking. *The Journal of Finance*, forthcoming.

Bonfim, D., Cerqueiro, G., Degryse, H., & Ongena, S. (2023). On-site inspecting zombie lending. *Management Science*, 69(5), 2547–2567.

Boot, A. W. (2000). Relationship banking: What do we know? *Journal of Financial Intermediation*, 9(1), 7–25.

Böve, R., Düllmann, K., & Pfingsten, A. (2010). Do specialization benefits outweigh concentration risks in credit portfolios of German banks? Discussion Paper 2010-10, Bundesbank.

Boyd, J. H. & Prescott, E. C. (1986). Financial intermediary-coalitions. *Journal of Economic Theory*, 38(2), 211–232.

Bruno, B., Marino, I., & Nocera, G. (2023). Internal ratings and bank opacity: Evidence from analysts' forecasts. *Journal of Financial Intermediation*, 56, 101062.

Chu, Y., Deng, S., & Xia, C. (2019). Bank geographic diversification and systemic risk. *The Review of Financial Studies*, 33(10), 4811–4838.

De Jonghe, O. (2010). Back to the basics in banking? A micro-analysis of banking system stability. *Journal of Financial Intermediation*, 19(3), 387–417.

De Jonghe, O., Dewachter, H., Mulier, K., Ongena, S., & Schepens, G. (2020). Some borrowers are more equal than others: Bank funding shocks and credit reallocation. *Review of Finance*, 24(1), 1–3.

De Jonghe, O., Mulier, K., & Samarin, I. (2024). Bank specialization and zombie lending. *Management Science*.

Degryse, H. & Van Cayseele, P. (2000). Relationship lending within a bank-based system: Evidence from European small business data. *Journal of Financial Intermediation*, 9(1), 90–109.

Diamond, D. W. (1984). Financial intermediation and delegated monitoring. *The Review of Economic Studies*, 51(3), 393.

Faria-e Castro, M., Paul, P., & Sánchez, J. M. (2024). Evergreening. *Journal of Financial Economics*, 153, 103778.

Ferreira, M. A. & Matos, P. (2012). Universal banks and corporate control: Evidence from the global syndicated loan market. *The Review of Financial Studies*, 25(9), 2703–2744.

Fraisse, H., Lé, M., & Thesmar, D. (2020). The real effects of bank capital requirements. *Management Science*, 66(1), 5–23.

Giannetti, M. & Saidi, F. (2019). Shock propagation and banking structure. *The Review of Financial Studies*, 32(7), 2499–2540.

Giometti, M. & Pietrosanti, S. (2022). Bank specialization and the design of loan contracts. Working Paper 2022-14, FDIC Center for Financial Research.

Goedde-Menke, M. & Ingermann, P.-H. (2024). Loan officer specialization and credit defaults. *Journal of Banking & Finance*, 161, 107077.

Goetz, M. R., Laeven, L., & Levine, R. (2013). Identifying the valuation effects and agency costs of corporate diversification: Evidence from the geographic diversification of U.S. banks. *The Review of Financial Studies*, 26(7), 1787–1823.

Goetz, M. R., Laeven, L., & Levine, R. (2016). Does the geographic expansion of banks reduce risk? *Journal of Financial Economics*, 120(2), 346–362.

Hayden, E., Porath, D., & Westernhagen, N. V. (2007). Does diversification improve the performance of German banks? Evidence from individual bank loan portfolios. *Journal of Financial Services Research*, 32(3), 123–140.

Jensen, M. (1986). Agency costs of free cash flow, corporate finance and takeovers. *American Economic Review*, 76(2).

Jiang, S. & Li, J. Y. (2022). He who lends knows. *Journal of Banking & Finance*, 138, 106412.

Kysucky, V. & Norden, L. (2016). The benefits of relationship lending in a cross-country context: A meta-analysis. *Management Science*, 62(1), 90–110.

Leland, H. E. & Pyle, D. H. (1977). Informational asymmetries, financial structure, and financial intermediation. *The Journal of Finance*, 32(2), 371–387.

López-Espinosa, G., Mayordomo, S., & Moreno, A. (2017). When does relationship lending start to pay? *Journal of Financial Intermediation*, 31, 16–29.

Nyola, A. P., Sauviat, A., Tarazi, A., & Danisman, G. O. (2021). How organizational and geographic complexity influence performance: Evidence from European banks. *Journal of Financial Stability*, 55, 100894.

Paravisini, D., Rappoport, V., & Schnabl, P. (2023). Specialization in bank lending: Evidence from exporting firms. *The Journal of Finance*, 78(4), 2049–2085.

Petersen, M. A. & Rajan, R. G. (1994). The benefits of lending relationships: Evidence from small business data. *The Journal of Finance*, 49(1), 3–37.

Plosser, M. C. & Santos, J. A. (2018). Banks' incentives and inconsistent risk models. *The Review of Financial Studies*, 31(6), 2080–2112.

Rajan, R. G. (1992). Insiders and outsiders: The choice between informed and arm's-length debt. *The Journal of Finance*, 47(4), 1367–1400.

Ramakrishnan, R. T. & Thakor, A. V. (1984). Information reliability and a theory of financial intermediation. *The Review of Economic Studies*, 51(3), 415–432.

Rossi, S. P., Schwaiger, M. S., & Winkler, G. (2009). How loan portfolio diversification affects risk, efficiency and capitalization: A managerial behavior model for Austrian banks. *Journal of Banking & Finance*, 33(12), 2218–2226.

Sharpe, S. A. (1990). Asymmetric information, bank lending, and implicit contracts: A stylized model of customer relationships. *The Journal of Finance*, 45(4), 1069–1087.

Shim, J. (2019). Loan portfolio diversification, market structure and bank stability. *Journal of Banking & Finance*, 104, 103–115.

Stiroh, K. J. (2004). Diversification in banking: Is noninterest income the answer? *Journal of Money, Credit and Banking*, 36(5), 853–882.

Stiroh, K. J. (2006). A portfolio view of banking with interest and noninterest activities. *Journal of Money, Credit and Banking*, 38(5), 1351–1361.

Tabak, B. M., Fazio, D. M., & Cajueiro, D. O. (2011). The effects of loan portfolio concentration on Brazilian banks' return and risk. *Journal of Banking & Finance*, 35(11), 3065–3076.

Winton, A. (1999). Don't put all your eggs in one basket? Diversification and specialization in lending. Working paper, SSRN.

Yildirim, A. (2020). The effect of relationship banking on firm efficiency and default risk. *Journal of Corporate Finance*, 65, 101500.

Appendix

Additional figures

In Section 2.2, we present descriptives regarding the main specialization measure, defined based on Equation (1). These descriptives are reported for the full sample of euro area banks with loan level data available in 2022. In this Appendix, we split the sample of banks and show a comparison of the 77 IRB banks included in the baseline analysis with all other (non-included) banks. While Figure A1 shows that the full distribution is rather similar, with the vast majority of observations again situated between -1% and 1%, we observe in Figure A2 that the average bank in our sample of 77 banks is highly specialized in fewer sectors (mean of 6 sectors, versus 9 sectors for the other banks).

Figure A1: Specialization measure - full distribution

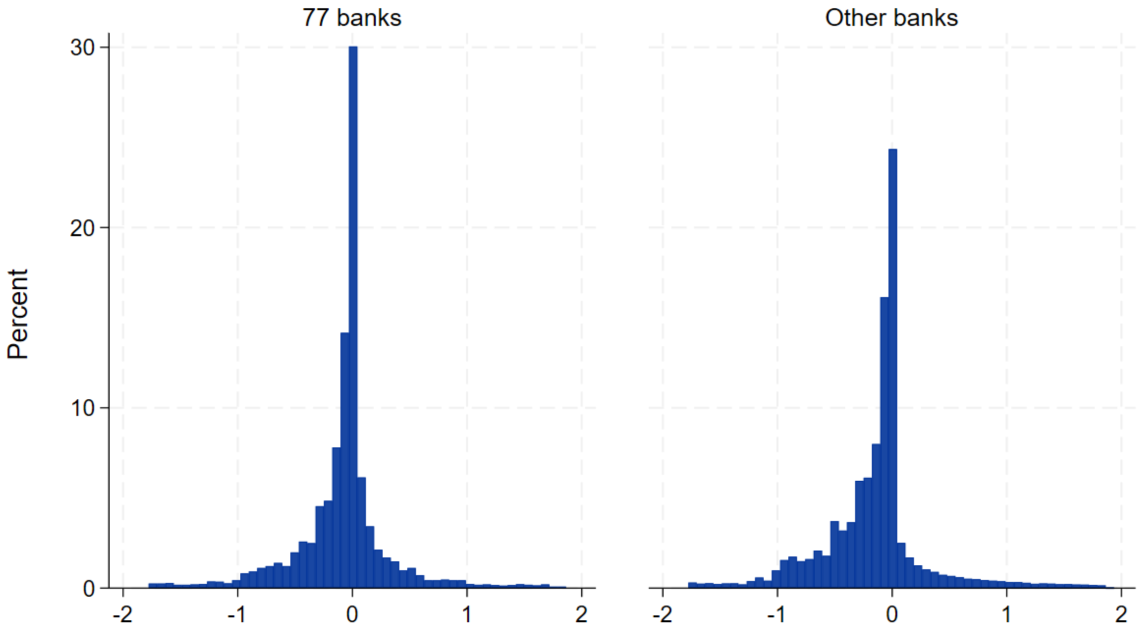
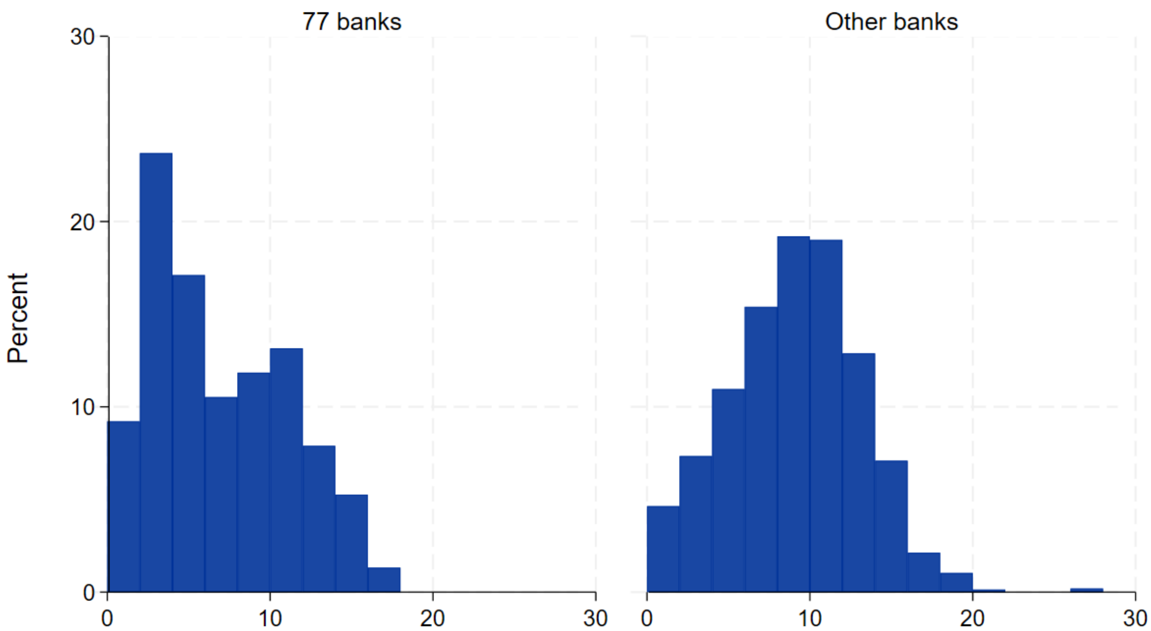


Figure A2: Specialization measure - number of highly specialized sectors per bank



Additional tables

In Table A1 in this Appendix, we include a list of the most important variables, with their source and definition. Detailed descriptive statistics for the (non-defaulting) placebo sample determined using PSM are presented in Table A2. Table A3 shows descriptive statistics for the broader sample of defaulting firms used in the provisioning analyses, while the results for the placebo tests regarding provisioning are included in Table A4. Table A5 documents that most positive effects of sectoral specialization on provisioning disappear after controlling for PDs, implying that the findings for provisioning are driven by a rather mechanical relation with PDs. In Table A6, we present the correlations between the main specialization measures used in the baseline analysis.

The remaining tables provide the results of multiple robustness checks, which are discussed in Section 5. First, in Table A7 and Table A8, we re-estimate the baseline Equation (2), but on a restricted sample in which (one by one) each of the countries or sectors is omitted, respectively. This shows that our main results are not driven by a single country or sector. In Table A9 and Table A10, we include time-invariant and within-sector specialization variables, and alternative control variables, respectively. Finally, in Table A11, we present findings when including only loans by SIs, loans under IFRS accounting standards, bank-firm pairs for which an observation is available in all five quarters (balanced), bank-firm pairs with zero days past due date four quarters before the default, as well as when removing the largest loans. We also check the robustness of our results to alternative clustering of the standard errors. Overall, our main results remain qualitatively similar in all tables and are in line with the hypothesis that sectoral specialization leads to informational advantages: banks assign higher PDs to later-defaulting borrowers if they are specialized in the borrower's sector, compared to less specialized banks. For all robustness checks, placebo tests are available on request.

Table A1: Variable definitions

Variable	Source	Granularity	Definition
Dependent variables:			
PD	AnaCredit	Bank-firm-quarter	Probability of default over one-year horizon (IRB banks only)
Provisions (% of outstanding)	AnaCredit	Bank-firm-quarter	Accumulated impairments as percentage of outstanding loan amount
Specialization variables:			
Specialization	AnaCredit	Bank-sector-quarter	Continuous sectoral (NACE2) specialization measure, as defined in Equation (1)
$D_{90}^{\text{Specialization}}$	AnaCredit	Bank-sector-quarter	Dummy which is 1 if continuous (NACE2) sectoral specialization measure is in highest decile of distribution (0 otherwise)
$D_{75}^{\text{Specialization}}$	AnaCredit	Bank-sector-quarter	Dummy which is 1 if continuous (NACE2) sectoral specialization measure is in highest quartile of distribution (0 otherwise)
$\text{Specialization}_{\text{alt}}$	AnaCredit	Bank-sector-quarter	Alternative continuous sectoral (NACE2) specialization measure, without correction term for lending directed to that sector by all banks in the country
$\text{Specialization}_{\text{NACE1}}$	AnaCredit	Bank-sector-quarter	Continuous sectoral (NACE1) specialization measure, as defined in Equation (1)
$\text{Specialization}_{\text{rel}}$	AnaCredit	Bank-sector-quarter	Relative continuous sectoral specialization measure, constructed by dividing the bank's percentage exposure to the sector by the percentage exposure of all banks in the same country to that sector, as originally defined in Paravisini et al. (2023)
Control variables:			
Bank ROA	COREP / FINREP	Bank-quarter	Return on assets: profits divided by total assets
Bank size	COREP / FINREP	Bank-quarter	Natural logarithm of total assets
Bank CET1 buffer	COREP / ECB calc.	Bank-quarter	CET1 capital ratio minus overall capital requirements (including P2G)
Bank loans/assets	COREP / FINREP	Bank-quarter	Loans and advances as percentage of total assets
Bank NPL/loans	COREP / FINREP	Bank-quarter	Non-performing loans and advances as percentage of total loans and advances
Maturity	AnaCredit	Bank-firm-quarter	Natural logarithm of average residual maturity (in days)
Exposure	AnaCredit	Bank-firm-quarter	Total outstanding loan amount to firm as percentage of total AnaCredit loan portfolio
Interest rate	AnaCredit	Bank-firm-quarter	Weighted average interest rate
$D^{\text{Protection}}$	AnaCredit	Bank-firm-quarter	Dummy which is 1 if at least one of the borrower's loans with that bank has collateral (0 otherwise)
$D^{\text{Same country}}$	AnaCredit	Bank-firm-quarter	Dummy which is 1 if bank and firm are headquartered in the same country (0 otherwise)
$D^{\text{Past due date}}$	AnaCredit	Bank-firm-quarter	Dummy which is 1 if firm is past due date on at least one loan with that bank (0 otherwise)
$D^{\text{First default}}$	AnaCredit	Bank-firm	Dummy which is 1 if first default for firm occurs at that bank-firm pair (0 otherwise)
$D_{90}^{\text{Market share}}$	AnaCredit	Bank-sector-quarter	Dummy which is 1 if concentration measure (bank's exposure to that sector as percentage of all banks' exposures to that sector) is in highest decile of distribution (0 otherwise)
$D_{2018}^{\text{Relationship}}$	AnaCredit	Bank-firm	Dummy which is 1 if bank-firm pair had loan in 2018Q4 (0 otherwise)
$D_{2019}^{\text{Relationship}}$	AnaCredit	Bank-firm	Dummy which is 1 if bank-firm pair had loan in 2019Q4 (0 otherwise)

Table A2: Descriptive statistics - PSM analysis

Variable	**N. obs**	**Mean**	**Stdev**	**Min**	**P25**	**P50**	**P75**	**Max**
Dependent variables:								
PD	471,248	2.99	7.24	0.00	0.50	1.15	2.68	100.00
Provisions (% outstanding)	454,644	0.74	1.86	0.00	0.05	0.16	0.51	14.03
Specialization variables:								
Specialization	471,248	0.23	3.29	-40.32	-0.26	0.00	0.32	96.18
$D_{90}^{Specialization}$	471,248	0.18	0.38	0.00	0.00	0.00	0.00	1.00
$D_{75}^{Specialization}$	471,248	0.50	0.50	0.00	0.00	0.00	1.00	1.00
$Specialization_{alt}$	471,248	4.51	7.38	0.00	0.95	2.06	5.37	100.00
$Specialization_{NACE1}$	471,248	0.69	5.16	-40.32	-1.42	0.03	1.31	87.50
$Specialization_{rel}$	471,248	106.48	43.60	24.08	86.13	99.83	117.69	621.15
Control variables:								
Bank ROA	470,826	0.50	0.27	-0.21	0.34	0.46	0.65	1.85
Bank size	470,830	13.23	1.10	7.36	12.43	13.52	14.19	14.79
Bank CET1 buffer	470,280	2.75	1.73	-1.13	1.75	2.65	3.48	14.42
Bank loans/assets	470,830	61.65	6.52	43.35	57.92	61.66	65.90	79.17
Bank NPL/loans	470,830	2.97	1.00	0.82	2.27	3.08	3.52	7.91
Maturity	466,465	6.67	1.02	3.25	6.28	6.89	7.29	8.73
Exposure	471,248	0.00	0.02	0.00	0.00	0.00	0.00	0.44
Interest rate	443,310	2.76	1.83	0.00	1.41	2.32	3.96	8.58
$D^{Protection}$	471,248	0.88	0.33	0.00	1.00	1.00	1.00	1.00
$D^{Same\ country}$	471,248	0.91	0.29	0.00	1.00	1.00	1.00	1.00
$D^{Past\ due\ date}$	471,238	0.01	0.10	0.00	0.00	0.00	0.00	1.00
$D^{First\ default}$	471,248	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$D^{Market\ power}$	471,248	1.00	0.07	0.00	1.00	1.00	1.00	1.00
$D_{2018}^{Relationship}$	471,248	0.52	0.50	0.00	0.00	1.00	1.00	1.00
$D_{2019}^{Relationship}$	471,248	0.63	0.48	0.00	0.00	1.00	1.00	1.00

This table shows the number of observations, mean, standard deviation, minimum, 25^{th} percentile, median, 75^{th} percentile and maximum for the main variables in the PSM analysis. Except for bank size (natural logarithm), maturity (natural logarithm) and the dummy variables, all variables are expressed in percentage.

Table A3: Descriptive statistics - provisioning analysis

Variable	**N. obs**	**Mean**	**Stdev**	**Min**	**P25**	**P50**	**P75**	**Max**
Dependent variables:								
PD	156,355	22.72	33.81	0.00	1.99	6.24	23.43	100.00
Provisions (% outstanding)	254,931	5.15	10.83	0.00	0.19	0.89	4.24	62.71
Specialization variables:								
Specialization	254,931	0.74	5.18	-57.12	-0.29	0.02	0.60	98.28
$D_{90}^{Specialization}$	254,931	0.26	0.44	0.00	0.00	0.00	1.00	1.00
$D_{75}^{Specialization}$	254,931	0.53	0.50	0.00	0.00	1.00	1.00	1.00
$\text{Specialization}_{alt}$	254,931	6.09	11.03	0.00	0.95	2.12	6.02	100.00
$\text{Specialization}_{NACE1}$	254,931	1.37	7.31	-57.12	-1.40	0.10	2.07	93.75
$\text{Specialization}_{rel}$	254,931	127.26	111.47	16.97	84.29	102.17	126.13	857.02
Control variables:								
Bank ROA	251,169	0.55	0.39	-0.33	0.31	0.46	0.66	2.00
Bank size	251,253	12.36	1.98	6.63	11.48	13.22	13.75	14.79
Bank CET1 buffer	250,906	3.46	2.95	-0.87	1.78	2.90	4.08	15.21
Bank loans/assets	251,253	61.09	8.33	40.80	56.55	60.91	66.15	86.84
Bank NPL/loans	251,189	3.06	1.41	0.68	2.09	2.97	3.73	8.78
Maturity	251,284	6.87	1.00	3.33	6.54	7.06	7.44	8.98
Exposure	254,931	0.02	0.10	0.00	0.00	0.00	0.00	0.82
Interest rate	245,813	3.40	2.21	0.00	1.70	2.99	4.82	10.45
$D^{Protection}$	254,931	0.88	0.32	0.00	1.00	1.00	1.00	1.00
$D^{Same\ country}$	254,931	0.89	0.31	0.00	1.00	1.00	1.00	1.00
$D^{Past\ due\ date}$	254,841	0.17	0.38	0.00	0.00	0.00	0.00	1.00
$D^{First\ default}$	254,931	0.43	0.49	0.00	0.00	0.00	1.00	1.00
$D^{Market\ power}$	254,931	0.92	0.28	0.00	1.00	1.00	1.00	1.00
$D_{2018}^{Relationship}$	254,931	0.35	0.48	0.00	0.00	0.00	1.00	1.00
$D_{2019}^{Relationship}$	254,931	0.46	0.50	0.00	0.00	0.00	1.00	1.00

This table shows the number of observations, mean, standard deviation, minimum, 25th percentile, median, 75th percentile and maximum for the main variables in the provisioning analysis. Except for bank size (natural logarithm), maturity (natural logarithm) and the dummy variables, all variables are expressed in percentage.

Table A4: Placebo tests for provisioning

Dependent variable:	Provisions as percentage of outstanding loan amount					
	(1)	(2)	(3)	(4)	(5)	(6)
-4Q $\times$ $D_{90}^{\text{Specialization}}$	0.473***	-0.021	-0.015	-0.001	-0.000	0.010
	(0.003)	(0.274)	(0.496)	(0.971)	(0.996)	(0.565)
-3Q $\times$ $D_{90}^{\text{Specialization}}$	0.379***	-0.012	0.003	0.010	0.011	0.021
	(0.003)	(0.503)	(0.873)	(0.651)	(0.435)	(0.240)
-2Q $\times$ $D_{90}^{\text{Specialization}}$	0.276**	-0.029	0.007	-0.002	0.016	0.002
	(0.041)	(0.120)	(0.740)	(0.933)	(0.305)	(0.890)
-1Q $\times$ $D_{90}^{\text{Specialization}}$	0.018	-0.018	-0.011	-0.022	0.018	-0.000
	(0.921)	(0.332)	(0.626)	(0.269)	(0.306)	(0.978)
0Q $\times$ $D_{90}^{\text{Specialization}}$	-0.693*	-0.009	0.009	-0.005	0.013	-0.020
	(0.063)	(0.652)	(0.673)	(0.813)	(0.425)	(0.350)
Controls	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ
Bank x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
N obs	149,246	796,332	476,810	530,991	958,640	891,666
R^2	0.5881	0.5379	0.5275	0.5357	0.5409	0.5419
Sample	Default	Non-default	Non-default	Non-default	Non-default	Non-default
Matching	Has match	PSM	ROA	Leverage	Amount	Rate

This table shows the results of the estimation of Equation (2) for defaulting firms (column (1)) or (matched) non-defaulting firms (columns (2) to (6)), using provisions as percentage of outstanding loans as dependent variable. All specifications include bank-firm-quarter (BFQ) control variables. Standard errors are clustered at bank and firm level. The numbers in parentheses are p-values. *, **, and *** indicate significance at 10%, 5%, and 1% respectively.

Table A5: Effect of sectoral specialization on provisioning, controlling for PD

Dependent variable:	Provisions as percentage of outstanding loan amount							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
-4Q × $D_{90}^{Specialization}$	0.367	0.245	0.172	0.104	0.082	0.103	0.110	0.105
	(0.107)	(0.284)	(0.441)	(0.643)	(0.711)	(0.646)	(0.623)	(0.636)
-3Q × $D_{90}^{Specialization}$	0.252	0.176	0.145	0.067	0.029	0.067	0.071	0.068
	(0.213)	(0.414)	(0.534)	(0.749)	(0.887)	(0.752)	(0.737)	(0.746)
-2Q × $D_{90}^{Specialization}$	0.120	0.134	0.088	0.090	0.048	0.089	0.092	0.091
	(0.558)	(0.509)	(0.672)	(0.683)	(0.822)	(0.686)	(0.676)	(0.680)
-1Q × $D_{90}^{Specialization}$	0.333	0.408**	0.331*	0.390*	0.347*	0.389*	0.392*	0.391*
	(0.123)	(0.042)	(0.094)	(0.065)	(0.084)	(0.065)	(0.064)	(0.064)
0Q × $D_{90}^{Specialization}$	0.319	0.431	0.528	0.475	0.411	0.474	0.478	0.476
	(0.553)	(0.383)	(0.271)	(0.311)	(0.369)	(0.312)	(0.308)	(0.309)
PD	0.160***	0.161***	0.162***	0.157***	0.156***	0.157***	0.157***	0.157***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Bank ROA		-0.440						
		(0.377)						
Bank size		-7.832						
		(0.201)						
Bank CET1 buffer		-0.379*						
		(0.075)						
Bank loans/assets		0.046						
		(0.666)						
Bank NPL/loans		0.187						
		(0.775)						
Maturity				0.161	0.112	0.161	0.156	0.160
				(0.141)	(0.407)	(0.141)	(0.148)	(0.147)
Exposure				5.767**	6.626**	5.745**	5.858**	5.785**
				(0.032)	(0.012)	(0.032)	(0.030)	(0.031)
Interest rate				0.913***	1.000***	0.913***	0.915***	0.913***
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$D^{Protection}$				-2.024***	-1.851***	-2.024***	-1.992***	-2.013***
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$D^{Same\ country}$				0.214	0.047	0.214	0.244	0.221
				(0.739)	(0.951)	(0.737)	(0.708)	(0.734)
$D^{Past\ due\ date}$				1.944***	1.912***	1.944***	1.948***	1.945***
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$D^{First\ default}$				1.241***	1.269***	1.241***	1.248***	1.243***
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$D^{Market\ power}$						0.147		
						(0.885)		
$D_{2018}^{Relationship}$							-0.202	
							(0.160)	
$D_{2019}^{Relationship}$								-0.058
								(0.710)
Bank FE	Yes	Yes	No	No	No	No	No	No
Bank x Quarter FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Firm x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loan type x Quarter FE	No	No	No	No	Yes	No	No	No
Rate type x Quarter FE	No	No	No	No	Yes	No	No	No
N obs	127,682	127,318	127,609	117,608	117,550	117,608	117,608	117,608
R^2	0.6765	0.6775	0.6902	0.7078	0.7092	0.7078	0.7078	0.7078

This table shows the results of the estimation of Equation (2), using provisions as percentage of outstanding loans as dependent variable, and with the probability of default as an additional explanatory variable. Standard errors are clustered at bank and firm level. The numbers in parentheses are p-values. *, **, and *** indicate significance at 10%, 5%, and 1% respectively.

Table A6: Correlation table of specialization variables

Variable	(1)	(2)	(3)	(4)	(5)	(6)
(1) Specialization	1.000					
(2) $D_{90}^{Specialization}$	0.509	1.000				
(3) $D_{75}^{Specialization}$	0.368	0.474	1.000			
(4) $Specialization_{alt}$	0.463	0.442	0.120	1.000		
(5) $Specialization_{NACE1}$	0.698	0.404	0.352	0.312	1.000	
(6) $Specialization_{rel}$	0.570	0.501	0.558	0.184	0.477	1.000

Table A7: Effect of sectoral specialization on PDs - omitting individual countries

Dependent variable:	Probability of default												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
-4Q × $D_{90}^{Specialization}$	3.801**	3.861**	3.862**	1.171***	3.786**	4.943**	3.790**	4.702**	3.810**	3.789**	3.907**	3.795**	3.823**
	(0.015)	(0.015)	(0.016)	(0.001)	(0.016)	(0.017)	(0.015)	(0.014)	(0.015)	(0.015)	(0.014)	(0.015)	(0.015)
-3Q × $D_{90}^{Specialization}$	3.747**	3.857**	3.789**	1.011**	3.722**	4.892**	3.743**	4.768**	3.727**	3.744**	3.839**	3.748**	3.770**
	(0.013)	(0.012)	(0.015)	(0.012)	(0.014)	(0.013)	(0.013)	(0.010)	(0.014)	(0.013)	(0.013)	(0.013)	(0.013)
-2Q × $D_{90}^{Specialization}$	2.646**	2.676**	2.666**	0.819**	2.627**	3.518**	2.633**	3.213**	2.633**	2.621**	2.695**	2.633**	2.642**
	(0.022)	(0.022)	(0.025)	(0.038)	(0.024)	(0.018)	(0.022)	(0.026)	(0.023)	(0.023)	(0.023)	(0.022)	(0.023)
-1Q × $D_{90}^{Specialization}$	-0.246	-0.250	-0.438	0.665	-0.263	-0.482	-0.249	-0.408	-0.258	-0.298	-0.273	-0.251	-0.240
	(0.674)	(0.672)	(0.450)	(0.112)	(0.653)	(0.491)	(0.669)	(0.565)	(0.657)	(0.608)	(0.648)	(0.666)	(0.681)
0Q × $D_{90}^{Specialization}$	-0.351	-0.508	-0.048	-1.641	-0.330	1.452	-0.314	-0.697	-0.302	-0.279	-0.565	-0.314	-0.377
	(0.784)	(0.690)	(0.971)	(0.316)	(0.796)	(0.216)	(0.806)	(0.637)	(0.814)	(0.827)	(0.662)	(0.805)	(0.768)
Controls	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ
Bank x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N obs	117,478	115,587	115,422	69,230	117,000	80,295	117,582	94,941	117,480	117,467	114,074	117,598	117,095
R^2	0.6937	0.6941	0.6927	0.7225	0.6938	0.6855	0.6937	0.6875	0.6937	0.6937	0.6922	0.6937	0.6935
Without country	AT	BE	DE	ES	FI	FR	IE	IT	LU	NL	PT	SI	SK

This table shows the results of the estimation of Equation (2), for subsamples in which one of the countries has been removed (based on the location of the firm). All specifications include bank-firm-quarter (BFQ) control variables. Standard errors are clustered at bank and firm level. The numbers in parentheses are p-values. *, **, and *** indicate significance at 10%, 5%, and 1% respectively.

Table A8: Effect of sectoral specialization on PDs - omitting individual sectors

Dependent variable:	Probability of default																	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
-4Q × $D_{90}^{Specialization}$	4.266***	3.798**	4.239**	3.888**	3.791**	3.417**	4.277**	3.930**	2.367**	3.786**	3.853**	3.940**	3.958**	3.795**	3.802**	3.810**	3.783**	3.798**
	(0.009)	(0.015)	(0.011)	(0.016)	(0.015)	(0.018)	(0.017)	(0.013)	(0.011)	(0.014)	(0.050)	(0.016)	(0.016)	(0.015)	(0.015)	(0.015)	(0.014)	(0.015)
-3Q × $D_{90}^{Specialization}$	4.225***	3.750**	4.049**	3.801**	3.750**	3.294**	4.763**	3.840**	2.344***	3.722**	4.032**	3.655**	3.865**	3.749**	3.753**	3.776**	3.735**	3.736**
	(0.008)	(0.013)	(0.012)	(0.014)	(0.013)	(0.018)	(0.013)	(0.012)	(0.009)	(0.012)	(0.028)	(0.019)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
-2Q × $D_{90}^{Specialization}$	3.002**	2.640**	2.946**	2.708**	2.646**	2.337**	4.194**	2.725**	1.208**	2.629**	2.520*	2.558**	2.683**	2.634**	2.637**	2.637**	2.630**	2.628**
	(0.012)	(0.022)	(0.025)	(0.022)	(0.022)	(0.029)	(0.025)	(0.022)	(0.024)	(0.021)	(0.052)	(0.025)	(0.024)	(0.022)	(0.023)	(0.023)	(0.022)	(0.023)
-1Q × $D_{90}^{Specialization}$	-0.046	-0.244	-0.113	-0.194	-0.238	-0.767	1.866**	-0.251	-0.494	-0.293	-1.001	-0.339	-0.294	-0.250	-0.235	-0.288	-0.272	-0.263
	(0.945)	(0.676)	(0.848)	(0.734)	(0.682)	(0.364)	(0.031)	(0.670)	(0.471)	(0.623)	(0.142)	(0.560)	(0.608)	(0.667)	(0.684)	(0.628)	(0.642)	(0.651)
0Q × $D_{90}^{Specialization}$	-0.784	-0.309	-0.020	-0.210	-0.363	0.180	-0.066	-0.413	-1.949	-0.080	0.379	-0.114	-0.328	-0.314	-0.294	-0.281	-0.308	-0.300
	(0.603)	(0.809)	(0.988)	(0.870)	(0.776)	(0.902)	(0.971)	(0.748)	(0.164)	(0.954)	(0.751)	(0.929)	(0.806)	(0.805)	(0.818)	(0.824)	(0.810)	(0.814)
Controls	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ
Bank x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
[0.1em] N obs	112,392	117,389	96,763	117,124	117,164	104,010	86,072	112,758	106,595	113,550	109,508	110,131	113,203	117,598	116,651	115,858	116,234	116,245
R^2	0.6953	0.6938	0.6931	0.6936	0.6936	0.6920	0.6969	0.6949	0.6916	0.6935	0.6947	0.6926	0.6949	0.6937	0.6940	0.6938	0.6940	0.6936
Without sector	A	B	C	D	E	F	G	H	I	J	L	M	N	O	P	Q	R	S

This table shows the results of the estimation of Equation (2), for subsamples in which one of the sectors (NACE1 level) has been removed. All specifications include bank-firm-quarter (BFQ) control variables. Standard errors are clustered at bank and firm level. The numbers in parentheses are p-values. *, **, and *** indicate significance at 10%, 5%, and 1% respectively.

Table A9: Robustness - time-invariant and within-sector specialization measures

Dependent variable:	Probability of default							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
-4Q × specialization	3.821**	2.935***	0.297***	0.327***	2.975*	1.546**	2.537	4.818**
	(0.015)	(0.005)	(0.002)	(0.004)	(0.055)	(0.018)	(0.167)	(0.011)
-3Q × specialization	3.735**	2.700***	0.236**	0.254**	2.602*	1.225**	4.409*	3.840**
	(0.022)	(0.008)	(0.012)	(0.014)	(0.070)	(0.033)	(0.056)	(0.020)
-2Q × specialization	2.137**	1.182**	0.129**	0.127**	0.440	0.199	1.191*	2.460***
	(0.035)	(0.037)	(0.029)	(0.024)	(0.499)	(0.698)	(0.051)	(0.001)
-1Q × specialization	0.600	-0.183	0.130**	0.106**	-1.535***	1.537*	-0.361	2.198***
	(0.452)	(0.724)	(0.015)	(0.026)	(0.007)	(0.096)	(0.696)	(0.003)
0Q × specialization	-0.025	-2.150	-0.079	-0.113	-0.327	-2.010	-3.105	-3.918***
	(0.985)	(0.110)	(0.499)	(0.362)	(0.801)	(0.248)	(0.133)	(0.009)
Controls	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ
Bank x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N obs	116,745	116,745	116,745	116,745	117,608	117,608	117,608	117,608
R^2	0.6938	0.6936	0.6935	0.6935	0.6934	0.6933	0.6935	0.6939
Specialization variable	D_{90}^{Spec}	D_{90}^{Spec}	Spec	Spec	$D_{90,\text{ within}}^{NACE2}$	$D_{75,\text{ within}}^{NACE2}$	$D_{90,\text{ within}}^{NACE1}$	$D_{75,\text{ within}}^{NACE1}$
Predetermined specialization	2021Q4	2020Q4	2021Q4	2020Q4	No	No	No	No

This table shows the results of the estimation of Equation (2), using specialization variables which are either time-invariant (predetermined) or constructed based on the within-sector distribution of specialization. All specifications include bank-firm-quarter (BFQ) control variables. Standard errors are clustered at bank and firm level. The numbers in parentheses are p-values. *, **, and *** indicate significance at 10%, 5%, and 1% respectively.

Table A10: Robustness - alternative control variables

Dependent variable:	Probability of default					
	(1)	(2)	(3)	(4)	(5)	(6)
-4Q × $D_{90}^{Specialization}$	4.183**	3.794**	3.811**	3.850**	3.801**	3.847**
	(0.010)	(0.015)	(0.014)	(0.013)	(0.015)	(0.012)
-3Q × $D_{90}^{Specialization}$	4.137***	3.748**	3.768**	3.793**	3.759**	3.795**
	(0.009)	(0.013)	(0.012)	(0.011)	(0.013)	(0.011)
-2Q × $D_{90}^{Specialization}$	3.027**	2.633**	2.654**	2.682**	2.634**	2.681**
	(0.014)	(0.022)	(0.020)	(0.019)	(0.022)	(0.018)
-1Q × $D_{90}^{Specialization}$	0.125	-0.251	-0.233	-0.209	-0.243	-0.194
	(0.843)	(0.666)	(0.692)	(0.723)	(0.677)	(0.746)
0Q × $D_{90}^{Specialization}$	0.094	-0.314	-0.299	-0.262	-0.305	-0.272
	(0.941)	(0.805)	(0.813)	(0.835)	(0.811)	(0.829)
Market share	-0.102**					
	(0.025)					
$D_{75}^{Market\ share}$		0.104				
		(0.982)				
$D^{Main\ bank}$			-0.467			
			(0.323)			
Outstanding amount (log)				-0.794***		
				(0.007)		
Committed amount (log)					-0.241***	
					(0.000)	
Number of loans (log)						-1.170*
						(0.067)
Controls	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ
Bank x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
N obs	117,608	117,608	117,608	117,608	117,608	117,608
R^2	0.6938	0.6937	0.6938	0.6940	0.6938	0.6939

This table shows the results of the estimation of Equation (2), using alternative control variables. All specifications include bank-firm-quarter (BFQ) control variables. Standard errors are clustered at bank and firm level. The numbers in parentheses are p-values. *, **, and *** indicate significance at 10%, 5%, and 1% respectively.

Table A11: Robustness - other

Dependent variable:	Probability of default							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
-4Q × $D_{90}^{Specialization}$	3.785**	3.916**	3.826**	3.538**	3.960***	3.978**	3.794**	3.794***
	(0.018)	(0.016)	(0.029)	(0.023)	(0.009)	(0.013)	(0.022)	(0.002)
-3Q × $D_{90}^{Specialization}$	3.712**	3.788**	3.935**	3.599**	3.772**	3.568**	3.748**	3.748***
	(0.017)	(0.017)	(0.019)	(0.020)	(0.012)	(0.015)	(0.028)	(0.003)
-2Q × $D_{90}^{Specialization}$	2.691**	2.731**	2.746**	2.710**	2.629**	2.543*	2.634*	2.634**
	(0.025)	(0.025)	(0.037)	(0.021)	(0.041)	(0.055)	(0.069)	(0.021)
-1Q × $D_{90}^{Specialization}$	-0.308	-0.421	-0.521	-0.513	-0.846	-1.130	-0.251	-0.251
	(0.605)	(0.480)	(0.499)	(0.402)	(0.162)	(0.123)	(0.859)	(0.882)
0Q × $D_{90}^{Specialization}$	-0.528	-0.060	-1.465	-0.970	0.098	-0.280	-0.314	-0.314
	(0.682)	(0.964)	(0.309)	(0.515)	(0.936)	(0.851)	(0.846)	(0.879)
Controls	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ	BFQ
Bank x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm x Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N obs	115,677	114,274	87,049	97,162	100,797	77,425	117,608	117,608
R^2	0.6939	0.6938	0.7036	0.6737	0.6970	0.7044	0.6937	0.6937
Restriction	SI	IFRS	Balanced	Unexpected	Loans<P90	Loans<P75	No	No
Bank and firm cluster	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Bank and NACE2 cluster	No	No	No	No	No	No	Yes	No
Bank x NACE2 cluster	No	No	No	No	No	No	No	Yes

This table shows the results of the estimation of Equation (2), using a variety of robustness checks: only SIs are included in column (1); only IFRS exposures are included in column (2); only bank-firm pairs with data in all five quarters are included in column (3); only bank-firm pairs with zero days past due date four quarters before default ("unexpected defaults") are included in column (4); large loans (upper decile or upper quartile) are removed in columns (5) and (6); alternative clustering is applied in columns (7) and (8). All specifications include bank-firm-quarter (BFQ) control variables. Standard errors are clustered at bank and firm level in columns (1) to (6), at bank and NACE2 level in column (7), and at bank-NACE2 level in column (8). The numbers in parentheses are p-values. *, **, and *** indicate significance at 10%, 5%, and 1% respectively.