Beyond Risk: Firm Financing and Interest Rates*

Miguel H. Ferreira † Nicholas Kozeniauskas ‡

First draft: June 25, 2025

This version: May 19, 2026

Abstract

Firm financing is a central component of modern macroeconomic theory, yet the empirical determinants of interest rates are not well understood. We study these using a novel loan-level dataset. Interest rates vary substantially across firms, even on observationally equivalent loans. Default risk—the core source of dispersion in standard models—can only account for 15% of variation. Bank heterogeneity and relationship banking have similarly limited explanatory power. Instead, most dispersion reflects persistent idiosyncratic firm-level heterogeneity—accounting for 27% of variation—that cannot be explained by observable firm characteristics. Our findings suggest that there are substantial frictions in credit markets outside existing frameworks.

Keywords: Firm financing, interest rates, financial frictions.

JEL Codes: E43; E44; G32.

*We thank Árpád Ábrahám, Isaac Baley, Diana Bonfim, Benjamin Born, Fernando Broner, Andrea Caggesse, Vasco M. Carvalho, Rui Castro, Diego Daruich, Chris Edmond, Miguel Faria-e-Castro, Manuel García-Santana, Mike Golosov, Basile Grassi, Danial Lashkari, Simone Lenzu, Chen Lian, Ellen McGratten, Davide Melcangi, Joao Monteiro, Alex Monge, Alessandra Peter, Tommaso Porzio, Hugo Reichardt, Filipe Saffie, Edouard Schaal, and Harold Uhlig, as well as seminar and conference participants at the PEJ annual meetings '24, Transpyrenean Macro Workshop '25, Midwest Macro spring meetings '25, SED '25, Vienna Macro Workshop '25, LuBraMacro '25, Lucca Macro Workshop '26, U. Helsinki, Bank of Portugal, QMUL, U. Barcelona, U. Bonn, UPF/CREi, Central Bank of Ireland, Nova SBE, UAB and U. Vienna for comments and suggestions. Thanks to FCT—Portuguese Foundation for Science and Technology for financially supporting this project through grants UID/00407/2025 and 2023.18151.ICDT. The views expressed in this article are the authors' and do not necessarily represent those of the Bank of Portugal or the Eurosystem.

†Queen Mary University of London, migueltchferreira@gmail.com.

‡Bank of Portugal and Universidade Católica Portuguesa, Católica-Lisbon School of Business and Economics, nicjkoz@gmail.com.

1 Introduction

Firm financing is a central component of modern macroeconomic theory. It is a key ingredient in theories of business cycles, monetary policy transmission, economic development, misallocation, firm dynamics, inequality, tax policy and macroeconomic aspects of international trade, among other things.[1] There is broad agreement that frictions in firm financing shape the allocation of capital in the economy and are quantitatively relevant for many macroeconomic outcomes. In these theories, interest rates are a key price. They determine firms borrowing decisions, affecting investment choices, firm-level outcomes and the aggregate economy. Despite their importance, the empirical determinants of interest rates are still not well understood.

This paper seeks to fill this gap and understand the sources of interest rate dispersion across firms in the context of bank loans. Our first contribution is to quantify how much dispersion can be accounted for by the leading mechanisms in existing frameworks. We begin with the dominant risk-based view of firm financing and show that default risk can only account for 15% of the observed variation in borrowing costs. We then turn to other prominent channels, namely heterogeneity across banks and firm-bank relationships. Both matter, but each accounts for only about 10% of overall dispersion. This leaves the majority of the variation originating from sources outside of the leading frameworks.

Our second contribution is to show what accounts for the remaining variation. We document that the largest source of interest rate dispersion, accounting for over one-quarter of the total variance, is persistent idiosyncratic firm-level heterogeneity that is unrelated to observable firm characteristics. The persistent differences in rates across firms are not absorbed by rich controls for firm size, age, industry, location, financial

[1] Some examples are as follows. Business cycles: Jermann & Quadrini (2012), Khan & Thomas (2013) and Buera & Moll (2015). Monetary policy transmission: Kiyotaki & Moore (1997), Bernanke et al. (1999) and Ottonello & Winberry (2020). Economic development: Buera et al. (2011), Moll (2014) and Itskhoki & Moll (2019). Misallocation: Midrigan & Xu (2014) and Gopinath et al. (2017). Firm dynamics: Cooley & Quadrini (2001), Albuquerque & Hopenhayn (2004) and Arellano et al. (2012). Inequality and tax policy: Boar & Midrigan (2023) and Guvenen et al. (2023). International trade: Antras & Caballero (2009) and Leibovici (2021).

position, credit history, productivity, production input mix, and many other observables, and they persist even when firms obtain new loans and switch banks. Our third contribution is to use this evidence to narrow the set of plausible explanations for that latent, persistent firm-level heterogeneity. We find little support for explanations based on local bank market power, the joint pricing of multiple loans to the same firm, interactions between entrepreneurs' personal and business finances, or unobservable demand shocks. Instead, the evidence points to search-related frictions as the most plausible source of the variation across firms.

The key to our analysis is a detailed new credit registry which we merge with comprehensive firm balance sheet and income statement data, additional risk measurement from the banking regulator, and geographic information on the location of bank branches. There are a number of important features of the data that make the analysis possible. It provides detailed measures of default risk that match up with the theoretical moments needed for pricing. It contains over half a million loans, with many observations at the bank level, the firm level, and even within firm-bank relationships. This facilitates the estimation of bank, firm and relationship effects on rates. Finally, with detailed information on firm characteristics, we can evaluate the factors behind differences in rates at the firm level.

The analysis starts by taking the core theories for frictional firm financing from the macroeconomic literature as a framework. While these theories are not designed with the full distribution of rates in mind, we are interested in quantifying how far the data departs from them. Do the theories merely omit second-order factors and, if not, what are the main sources of rate variation that are ignored? These core theories are based on the classic view of interest rates, formalized by Merton (1974), which looks at interest rates from the perspective of default risk. Most models either have lending without risk, so that all firms pay the same risk free rate (e.g. Buera et al., 2011; Moll, 2014),[2] or there is default risk and loans are priced so that lenders earn the risk free

[2] Various modeling techniques are used to achieve this. The most common approach is to impose a collateral constraint, assuming that a firm's existing assets constrain how much it can borrow, and that either it cannot default, or the environment is such that it will not choose to default in equilibrium (e.g.

rate in expectation (e.g. Bernanke et al., 1999). Under the second approach, interest rates differ across firms depending on their risk.

The first stage of the empirical analysis shows that the data departs substantially from this risk-based perspective. Risk free lending does not fit the data well because there is a lot of heterogeneity in interest rates after controlling for a rich array of loan characteristics. This is true for both collateralized and uncollateralized loans. To quantify the importance of default risk, we take the relevant pricing equation directly to the data using our empirical measures of default probabilities and loss given default. These characteristics of loans can only account for 15% of the variation in interest rates that we observe. Overall, once we control for loan characteristics and default risk, approximately 65% of the variation in interest rates remains unaccounted for. This corresponds to an unexplained standard deviation of rates of 1.46 percentage points.

The second stage of the analysis investigates potential explanations for the additional dispersion in interest rates. We first consider the role of banks, as they have received the most attention as an additional source of rate variation in the literature. This research has studied how bank characteristics like market power, cost heterogeneity and portfolio composition can affect loan prices (e.g. Klein, 1971; Monti, 1972; Ho & Saunders, 1981; Diamond, 1984), and there is a long history of research on the effect of firm-bank relationships on credit conditions (e.g. Sharpe, 1990; Petersen & Rajan, 1995).[3] These ideas have been incorporated into some macroeconomic models.[4] We evaluate the quantitative relevance of these theories for interest rate variance. They are quantitatively relevant, each accounting for approximately 10% of the variation in rates. However, nearly half of the variation remains unexplained, implying that there are substantial frictions in financial markets outside of the leading frameworks.

We show that over 30 percentage points of this residual variation can be attributed

Moll, 2014). A related approach assumes that loans are uncollateralized, but that there are endogenous borrowing limits so that firms never default (e.g. Buera et al., 2011).

[3]For additional empirical evidence of these effects see, for example, Petersen & Rajan (1994), Berger & Mester (1997) and Darmouni (2020)

[4]For examples of macroeconomic models with bank heterogeneity and relationship banking, see Corbae & D'Erasmo (2021) and Dempsey & Faria-e Castro (2025), respectively.

to persistent idiosyncratic firm-level heterogeneity above and beyond default risk, and that most of this (27 percentage points) is due to latent firm fixed effects. These latent fixed effects are the largest source of rate variation in the data. This result implies that if two observationally equivalent firms approach the same bank for the same loan, they obtain different rates, and these differences persist across subsequent loans, even when firms switch banks. We include rich controls for a wide variety of time-varying and immutable firm characteristics, so these effects are not due to heterogeneity in firm age, size, industry, location, financial position, credit history, productivity, production input mix and many other observables.

We consider a range of potential explanations for the large latent firm fixed effects in interest rates. The empirical results provide clear requirements: (1) the explanation must be based on a firm-specific source; (2) it must be able to generate highly persistent rate differences across firms; and (3) it must have a low correlation with the broad range of observable firm characteristics that we control for. There are a number of potential explanations based on bank behavior that relate to market power, geographic differences in markets, pricing multiple loans to a firm collectively rather than individually, or interactions between the pricing of loans to firms and personal loans to their owners or managers. We provide evidence against these channels. Variation in demand for loans due to unobservable shocks to investment opportunities, costs or liquidity is another possibility, but we also provide evidence against this. Instead, we argue that the market for bank loans to firms is an institutional setting that naturally gives rise to search frictions and that this type of theory has the necessary ingredients to fit the data well. Within a search setting, rate variation could come from differences in latent search ability, effects at the loan officer level, or differentiation based on non-economic factors such as socio-demographic characteristics or personal relationships.

Disentangling these search-related sources of rate variation is an important question for future research. Differences in search ability would typically be considered a technological feature of the economy, while effects at the loan officer level and differentiation on non-economic factors are more likely to be wedges generating misallocation. This is particularly relevant since we document persistent differences in rates that typically have a larger impact on investment than transitory differences. If the latent firm fixed effects in interest rates are pure wedges, then in the Hsieh & Klenow (2009) framework the implied aggregate output loss is 0.43%.

Contribution to the literature There is a long history of theoretical work on the determinants of interest rates and their heterogeneity across firms. Leading strands of this research have focused on the relevance of default risk, bank heterogeneity, and relationship banking. In macroeconomic models, default risk has been the most common source of rate heterogeneity (e.g. Bernanke et al., 1999), however there is also work incorporating bank effects and relationship banking (e.g. Corbae & D'Erasmo, 2021; Dempsey & Faria-e Castro, 2025). Our contribution is to quantify how much variation in interest rates can be attributed to each of these channels, show that a large share of variation comes from elsewhere, and provide evidence on the source of this variation.

There is, of course, a large existing empirical literature studying the determinants of interest rates. This literature has, in particular, focused on how rates vary as a result of default risk (e.g. Durrani et al., 2022), bank heterogeneity (e.g. Santos, 2011) and relationship banking (e.g. Petersen & Rajan, 1994). Compared to this literature, we focus on a different question. Rather than evaluating how a particular factor generates variation in rates, we quantify the overall contribution of each channel and provide guidance on where the substantial additional variation in rates comes from.

The most closely related paper is Amiti et al. (2026), which also studies the overall variation in interest rates on bank loans to firms. Both papers show that substantial dispersion in borrowing costs remains after accounting for loan characteristics, default risk, bank characteristics, and relationship variables. Our paper additionally quantifies how much of overall interest rate dispersion can be attributed to each of these sources, and uses a theoretically-based approach to measuring the role of default risk. The papers also differ in how they analyze the residual variation in rates. Amiti et al. (2026) interpret it through the lens of the Burdett & Judd (1983) model, structurally estimating the search costs required to match the data, and providing empirical evidence to support this interpretation. We instead show that the dominant source of the remaining dispersion is persistent idiosyncratic firm-level heterogeneity that is uncorrelated with observables, rule out several prominent potential explanations for what drives this heterogeneity, and argue that search-related frictions are the most plausible interpretation.

Faria-e-Castro et al. (2024a) studies rate variation on bank loans in the US. The analysis differs from ours by focusing on comparing the cost of bank loans to corporate bonds, and within-firm variation in rates. Cavalcanti et al. (2023) also studies rate variation, but in the context of a developing country (Brazil) in which interest rates look very different from those in our setting.[5] That paper also differs by focusing on how rates vary with observable firm characteristics, which is not a quantitatively important feature of our data. Consistent with our interpretation of the source of rate variation, Cerqueiro et al. (2011) and Mazet-Sonilhac (2025) also argue that discretion in loan pricing and search frictions, respectively, are important sources of interest rate variation. While the message of Cerqueiro et al. (2011) is consistent with our conclusions, it does not have sufficiently detailed data to study and quantify the breadth of channels that we do. Mazet-Sonilhac (2025) is focused specifically on documenting the existence of search frictions when firms lend from banks, rather than evaluating overall rate variation.

The remainder of the paper is structured as follows. Section 2 describes the data. Section 3 evaluates how much variation in interest rates can be accounted for by loan characteristics and default risk. Section 4 explores the importance of banks and firm-bank relationships for dispersion in interest rates. Section 5 evaluates the relevance of firm characteristics for interest rate differences. Section 6 investigates the source of fixed latent variation in interest rates across firms. Section 7 concludes.

[5] Interest rates are approximately an order of magnitude higher in the Brazilian data, and their standard deviation is nearly two orders of magnitude larger.

2 Data

The analysis uses two primary datasets, a credit registry covering the universe of credit instruments issued by banks to firms in Portugal, and a dataset with balance sheet and income statement information for the universe of incorporated firms.[6]

The credit registry has a monthly frequency, covering January 2019 to December 2022. For each month it provides detailed information about the characteristics of all outstanding credit instruments issued by banks to firms.[7] The set of credit instruments is broad, encompassing all types of contracts through which firms borrow from banks including traditional business loans, a range of credit lines, credit cards, mortgages and auto loans. The data has information about the current characteristics of each instrument including the amount outstanding, the interest rate, the duration of the instrument, whether the interest rate is fixed or variable and the time horizon for changes to it, and information about any collateral, among other things. Importantly, the data also includes estimates for the probability of default of firms and loss given default for loans. These are provided by the banks issuing credit instruments and will be a central component of our analysis. After an instrument is initiated, the registry records information about the repayments every month, and whether the instrument is in arrears or default. The information about credit instruments extends to their characteristics at origination and the time (month) of origination. This backward-looking feature means that despite the dataset starting in 2019, it has considerable information about loans issued before this date.

The other dataset provides balance sheet and income statement information for the universe of incorporated firms at an annual frequency, excluding financial corporations and general government entities. Incorporated firms account for 33% of firms and 72% of employment in the economy.[8] The data comes from annual mandatory

[6] The names of the datasets are *Central de Responsabilidades de Crédito do Banco de Portugal* and *Informação Empresarial Simplificada*.

[7] There is a small number of loans to public sector firms in the data. We drop these, restricting the sample to loans to private sector firms.

[8] Only sole proprietorships and non-employer firms without a separate legal status from their

reporting of financial statements to the Ministry of Finance. It covers 2010 to 2022, so it overlaps with the origination date of nearly all instruments in the credit registry.[9]

To construct our dataset we merge the credit registry with the firm data, dropping instruments that cannot be matched to the firm data in order to ensure a consistent sample across our analysis.[10] Only 32% of total credit in the data is excluded from the matched sample as a result of this, since the firm data only omits very small firms. The main types of instruments are traditional bank loans (35% of total credit), factoring and confirming (49%), credit cards (7%), auto loans (5%) and mortgages (2%). For our primary analysis, we restrict the sample to traditional bank loans, as these are analogous to firm financing in models and this ensures a more homogeneous sample.[11] We will refer to these as "firm financing loans" or simply "loans" from now on, and all analysis will be for this sample, unless stated otherwise. Nevertheless, we show in the Appendix that our main results are robust to using all credit instruments.

The market for firm financing loans consists of five main banks, with each accounting for 12% to 18% of the total value of loans in our sample. There are 157 smaller banks that collectively provide the remaining 29%.[12] The five main banks are the primary providers of both firm and consumer credit. They are large, national, general-purpose banks with similar operations to each other. To illustrate this, in the Appendix we show that these banks' loan distributions are similar across regions, industries and firm size.

Our sample has 150,700 firms. Relative to the population of firms in the industries covered by the data, the sample is skewed towards larger firms. Nearly all non-employer firms are excluded because they are not incorporated and therefore not cov-

owner are excluded. See the Appendix for details of the coverage of our sample.

[9]The share of loans in the credit registry with an origination date prior to 2010 is 2.3%.

[10]Instruments are dropped because either they were initiated before 2010, when the firm data starts, or were issued to unincorporated firms outside the firm data.

[11]We are excluding the small share of renegotiated and renewed loans in the sample to avoid history dependence in interest rates.

[12]The importance of large banks in other Euro Area 19 countries is similar. For example, the shares of banking sector assets held by the largest five banks in Portugal from 2010 to 2021, and the average for the Euro Area 19 countries, were 84–88% and 92–95%, respectively (World Bank Global Financial Development Indicators).

ered by the firm data. Larger firms are also more likely to have bank credit. Overall, our sample covers 37% of employer firms, 52% of total employment and 55% of aggregate output in 2019.[13] The average number of employees of our firms is 11.3, compared to 9.4 for all employer firms.

In terms of overall credit for firms, bank financing is the most important source. From 2010 to 2022 it represented 86% of total credit to non-financial corporations in Portugal. This degree of dependence on banks is typical across Europe. Over the same period, the average share of credit to non-financial corporations that came from banks in the Euro Area 19 countries was 89%.[14] While a substantial amount of firm financing comes from public debt markets in the U.S., these play a much smaller role in Europe.

To further illustrate the characteristics of the firms in the sample, we present summary statistics in Table 1. The statistics are within firm averages over the years that each firm is in the sample. The balance sheet and income statement variables reflect the characteristics of the population of firms. For the credit registry variables, the median firm has €119,830 in loans. The median number of loans is four and the median share of a firm's loan value that is collateralized is 78%. The majority of firms have loans from only one bank, but many firms have more bank relationships. At the 90th percentile, a firm averages three banking relationships over the years that it is in the sample.

Figure 1 presents summary statistics for all loans in the sample at their initiation date. Throughout the paper, we focus on evaluating the determinants of interest rates at loan initiation.[15] There are 598,028 loans in total. The bulk of these (78.5%) were originated from 2018 onward, with 19.2% originated from 2014–17, and the rest distributed over earlier years.[16] The average interest rate is 3.01% with a standard devia-

[13]Additional details of the comparison of our sample to the population of firms are provided in the Appendix. We also benchmark our sample against the size distribution of employer firms in the US.

[14]Data from the ECB. Details are in the Appendix.

[15]For some loans, the rate varies over time. We do not use this variation because it results in a less homogeneous sample that raises additional empirical complications.

[16]Recall that the data starts in September 2018, with loans originated earlier entering the sample if they were still active at this date. This is why origination dates are skewed towards recent years.

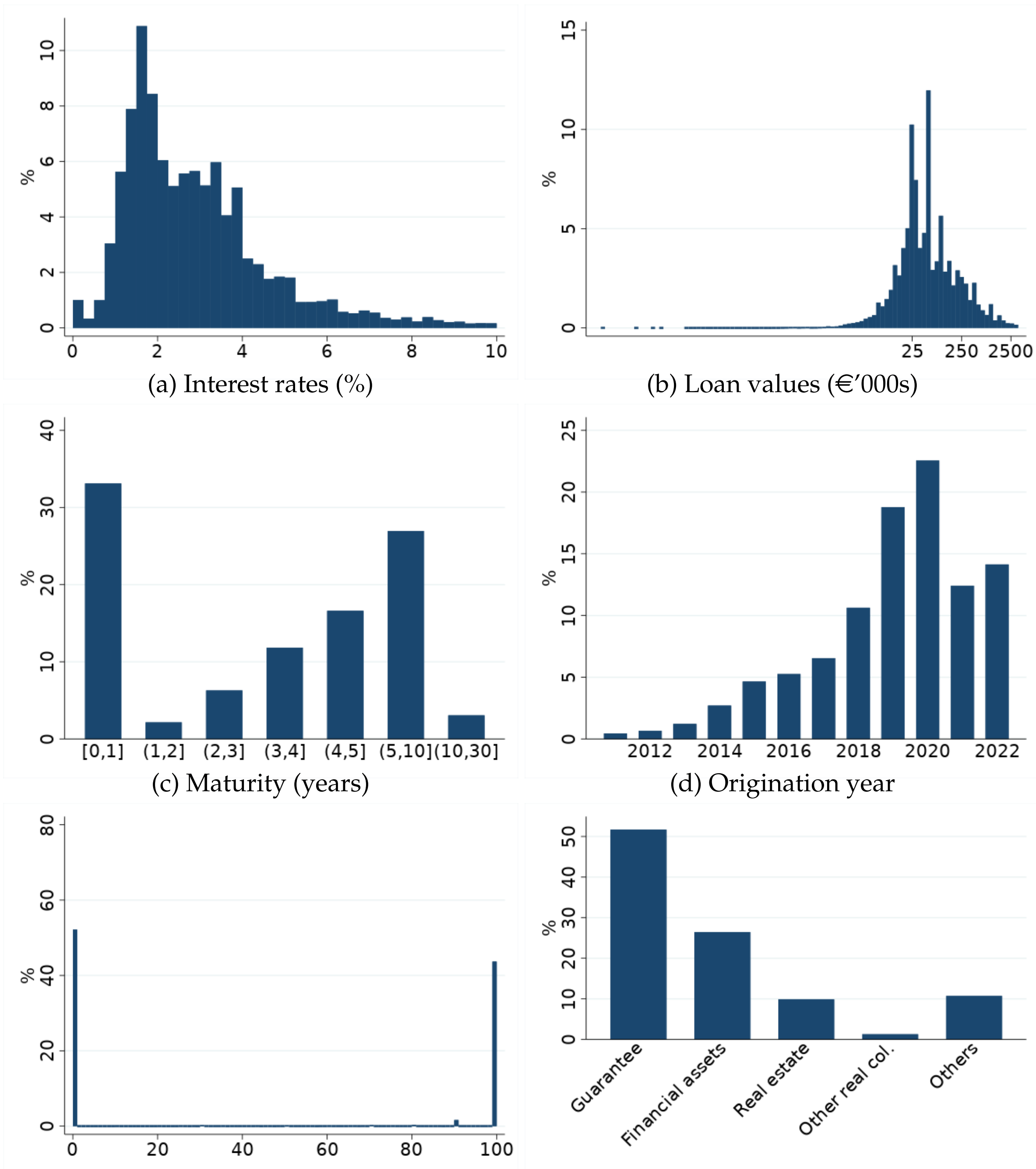
	Average	p25	p50	p75	p90	St. dev.	# firms
Loans value (€'000s)	1,063.31	38.24	119.83	378.79	1387.72	(9848.76)	
# of loans	7.07	3.00	4.00	8.00	14.00	(13.29)	
Av. loan (€'000s)	123.73	11.35	24.93	58.46	183.84	(908.95)	
% collateralized	67	40	78	100	100	(35)	
# of banks	1.63	1.00	1.00	2.00	3.00	(1.14)	150,700
Employees	12.53	1.50	3.75	8.75	21.00	(123.86)	
Total assets (€'000s)	2,277.66	101.10	261.73	778.33	2478.56	(65527.84)	
Revenue (€'000s)	1,657.86	87.60	228.28	649.41	1931.82	(40855.78)	
Leverage (%)	103	49	71	91	125	(255)	
Age	15.15	5.50	11.50	21.00	31.50	(13.10)	

Table 1: **Descriptive statistics for firms.** This table provides statistics for firms in the sample. For each variable, the average, four percentiles and the standard deviation are given. All moments are at the firm level and are averaged within firm over the years that each firm is in the sample. So, for example, the value of loans is the total value of a firm's loans, % collateralized is the share of a firm's loans that are collateralized, and # of banks is the number of banks that a firm has loans from. Value of loans, total assets and revenue are in thousands of 2016 Euros.

tion of 2.25%.[17] The median loan is for €44,751. The right tail of the distribution is for much larger loans, with the 90$^{\text{th}}$ percentile having a value of €338,872. The maturities of loans range one month to 30 years. Sixty-eight percent of loans have a maturity of less than 5 years and they account for 48% of total loan value. For collateral, the distribution is binary. Forty-eight percent of loans are fully collateralized and virtually all other loans are uncollateralized. The most common types of collateral are guarantees and financial assets (51.7% and 26.4% by loan value, respectively), with real estate and other collateral also being used. Throughout the paper the analysis focuses on interest rates rather than spreads, because spreads are not always reported, and the data shows that there is almost no variation in the implied financing costs of banks, conditional on maturity and loan origination date.[18] Since we control for time of origination and maturity in our analysis, using spreads instead of interest rates does not affect the results.

In addition to the main data, we use two supplementary datasets. As an alternative measure of the probability of default for firms, we use an annual dataset from the

[17] Throughout the paper, interest rates are always stated on a per annum basis.
[18] See the Appendix for details on this.



(a) Interest rates (%)

(b) Loan values (€'000s)

(c) Maturity (years)

(d) Origination year

(e) Collateral rate (%)

(f) Collateral type

Figure 1: **Distributions of loan characteristics at origination.** Panels (a), (b), (c), (d), (e) and (f) are the distributions of interest rates, loan values (in thousands of Euros), maturity (in years), origination years, collateral rates and collateral types, respectively. The scale for loan values in panel (b) is in log units. The distribution in panel (f) is conditional on loans having some collateral.

banking regulator with probability of default estimates for all firms in the sample for 2010–2022. We also use data on the seven-digit postcodes of all bank branches for the same years.[19] When merged with the location of each firm, this data provides information on the banking market around each firm. Note that despite having the location of all branches, the credit registry does not record which branch each loan is from, only the bank.

3 Risk free lending and default risk

The vast majority of macroeconomic models in which firms face external financing frictions think about interest rates from the perspective of risk. Either there is no risk, and all firms pay a common risk free rate, or there is heterogeneous default risk, and firms pay different rates to compensate banks for this. In this section we assess whether this approach provides a good first-order approximation to the data, or not.

3.1 Risk free lending

The first step is to assess how far the data is from risk free lending, as this is the simplest case and the most common assumption in macroeconomic models. In these models, risk free lending is typically achieved through one of two approaches. The most popular is to assume that there is a collateral constraint that limits how much a firm can borrow as a function of its assets. This takes the form

$$b_{i,t+1} \leq \lambda_{i,t} a_{i,t}$$

where $b_{i,t+1}$ is loan size of firm i from period t to $t+1$, $a_{i,t}$ is the period t assets of the firm and $\lambda_{i,t} > 0$ determines the tightness of the constraint. Lending is risk free either because the collateral can be claimed in the case of default, or because default is ruled out by assumption. This approach dates back to Kiyotaki & Moore (1997) and is particularly common in research studying business cycles, firm dynamics and capital misallocation. Under the second approach, it is assumed that there is an endogenous borrowing limit, with banks not lending more to a firm than it will be able to repay

[19] See Antunes et al. (2016) for the PD estimation methodology. The dataset is called SIAC (*Sistema Interno de Avaliação de Crédito*). This postcode data is a supplement to the balance sheet and income statement data. There is one 7-digit postcode per 0.70 square kilometers (0.27 square miles).

with certainty (e.g. Buera et al., 2011). Both approaches result in no default and firms paying a common risk free rate.

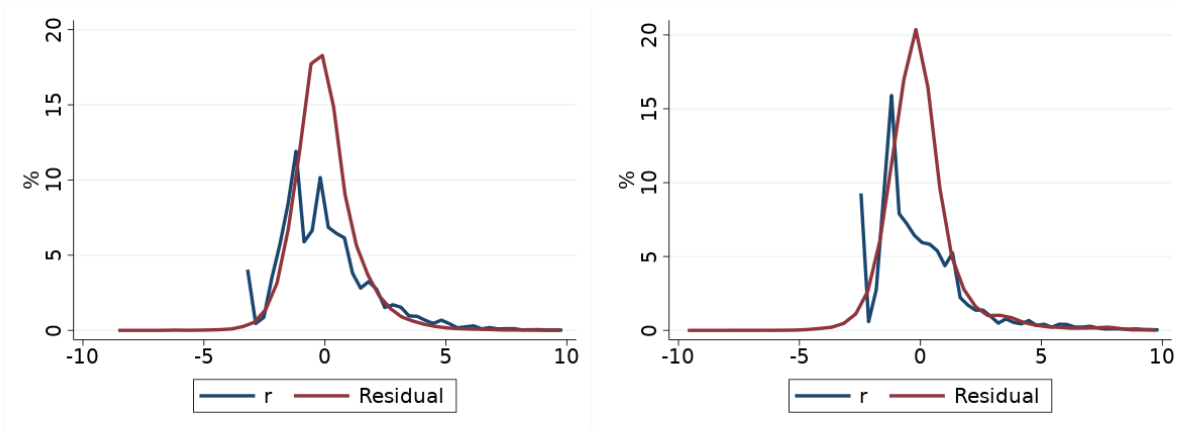
The data on collateralization rates aligns cleanly with these theories. Figure 1(e) in Section 2 shows that almost 50% of loans are fully collateralized, supporting the idea of no risk under the first theory. Nearly all of the remaining loans have no collateral. For these to be risk free, it would be necessary that loans are sufficiently small so that there is no realized default.

To assess whether lending is risk free in the spirit of these models, we are interested in whether two firms pay the same interest rate if they issue identical loans at the same time. In other words, can the dispersion in interest rates in Figure 1(a) be accounted for by loans having different characteristics such as the time at which they were issued, the maturity, the amount, the interest rate structure (e.g. fixed versus variable), et cetera? To implement this, we use the following regression:

$$r_{lt} = \beta \mathbf{L}_l + \mathcal{M}_{lt} + \varepsilon_{lt}, \tag{1}$$

where r_{lt} is the interest rate on loan l at its initiation date t, $\mathcal{M}_{lt}$ contains maturity-time fixed effects and $\mathbf{L}_l$ is a vector of other loan characteristics. Throughout the paper bold capital letters will be used to denote vectors of observable control variables, and calligraphic letters will denote vectors of fixed effects. For the maturity-time fixed effects, the time dimension is monthly and loans are grouped into maturities with the following durations (in years): (0, 0.25], (0.25, 0.5], (0.5, 1], and then in yearly intervals from (1, 2] up to (29, 30]. In this way, the maturity-time fixed effects control for differences in the risk free rate across the yield curve, and any changes in the curve at a monthly frequency. From here on, we will refer to the controls in $\mathbf{L}_l$ and $\mathcal{M}_{lt}$ as "loan characteristics."

The loan variables in $\mathbf{L}_l$ aim to control for all loan characteristics that are likely to affect interest rates and are available in the data: log of the loan value, frequency at which the interest rate can be adjusted, type of collateral, number of collateral assets, amortization structure, purpose of the loan, frequency of repayment, whether the



(a) Collateralized loans (b) Uncollateralized loans

Figure 2: **Interest rate distributions before and after controlling for loan characteristics.** Panels (a) and (b) present the distribution of interest rates for fully collateralized and uncollateralized loans, respectively. In both panels the blue line is the unconditional distribution, and the red line is the residuals after controlling for loan characteristics. All distributions are centered with their means at zero.

bank has the right to demand immediate repayment, whether the loan is securitized, and other special characteristics. All controls are categorical, except for the log of the loan value. The Appendix provides full details about these characteristics.

In taking equation (1) to the data, we split the data into loans that are fully collateralized and all others, which are mostly uncollateralized. This generates two subsamples that fit with the two theoretical approaches to risk free lending in the literature, which can be assessed separately. Figure 2 presents the results, comparing the distributions of interest rates with the distribution of residuals from equation (1) for each sample. The means of the two distributions are aligned at zero to make them visually comparable. This comparison tells us how much variation in interest rates remains after controlling for all of the loan, bank, maturity and time characteristics that we have discussed. If firms were lending at common risk free rates, then the distribution of residuals would be a mass point at zero.

For both collateralized and uncollateralized loans, we see that substantial variation in interest rates remains after controlling for loan characteristics, showing that firms are not lending at common risk free rates. Quantitatively, 52% of the variance in interest rates remains for collateralized loans, and for uncollateralized loans it is 60%. It is possible that some types of collateral could be more effective at insuring banks against risk than others, and that some of the residual interest rate variation is due to this. To assess this, in the Appendix we repeat the analysis separately for different classes of collateral: guarantees, real and financial assets, and other forms of collateral. The results show that the type of collateral is not materially important for the conclusion. More than 41% of the variation in rates remains unexplained in all cases.

3.2 Lending with risk

With the data departing substantially from risk free lending, we now ask whether the dominant theory of interest dispersion in macroeconomic models, default risk, can account for the additional variation.

In models adopting this theory, financial intermediaries earn the risk free rate on loans, plus a premium to compensate for the risk of default, and the loss when default occurs. Financial markets are assumed to be perfectly competitive and the spread is pinned down by intermediaries earning zero expected profit. While there is variation from paper to paper in the details of the assumptions that determine default probabilities and losses given default, the vast majority of papers use single period debt and the price of loan l to a firm at time t is determined by an equation with the form

$$Q_{lt} = \mathbb{E}_t \left[\Lambda_{t+1} \left(1 - \chi(s^{l,t+1}) LGD(s^{l,t+1}) \right) \right]. \tag{2}$$

Λ_{t+1} is the stochastic discount factor, $s^{l,t+1}$ is the history of states for the firm with loan l up to time $t+1$, χ is an indicator function for whether default occurs in a given state, and LGD (*loss given default*) is a function mapping the state for loan l to the share of the loan that is not recovered in the case of default. The expectation is taken over the set of possible states at time $t+1$, given the state in period t. For examples of this type of setup, see Cooley & Quadrini (2001) and Ottonello & Winberry (2020).[20]

[20] There are various microfoundations for the losses that arise from default. For example, Ottonello & Winberry (2020) assumes that the lender cannot recover the full value of the collateral. In Cooley & Quadrini (2001) there is asymmetric information and banks incur monitoring costs in the process of enforcing their right to collateral when default occurs.

Our objective is to take this equation to the data to evaluate whether default risk can explain the variation in interest rates that we observe. For our baseline approach, we follow the most common assumption in the theory, that banks are risk neutral. Later, we will investigate alternatives. Under this assumption, equation (2) can be expressed in terms of interest rates as

$$1 + r_{lt} = \frac{1 + \tilde{r}_t}{(1 - PD_{lt} LGD_{lt})}, \tag{3}$$

where $\tilde{r}_t$ is the one period risk free rate at time t. PD_{lt} is the overall probability that loan l issued at t defaults in the next period, and LGD_{lt} is a weighted average of loss given default across states next period, with the weights equal to share of total default risk for the loan in each state.[21]

In order to take this equation to the data we need measures of PDs and LGDs that are consistent with the theory.[22] The data provides two such measures of PDs. First, banks report their estimates of the PDs of firms each month in the credit registry. Second, the banking regulator estimates these PDs each year as well. The horizon at which the PDs are estimated is one year, and they are estimated at the firm rather than loan level, because this is the regulatory requirement. A firm is considered in default if, for three consecutive months, 2.5% or more of its total outstanding loans remain overdue. A loan is classified as overdue when there is an outstanding payment past its due date. This definition and the fact that default is measured at the firm rather than the loan level are fine from the perspective of the theory, as long as LGDs are estimated conditional on the same definition of default. Fortunately, this is precisely the case. For LGDs we have values reported by banks at the time that a loan is initiated.

From the perspective of the theory, the important thing for the PD and LGD estimates are that they are the banks' beliefs about these variables at the time they are issued, even if they are incorrect. Nevertheless, we verify the quality of the PD estimates by tracking all loans for one year after initiation, and checking that ex post

[21] Additional mathematical details are provided in the Appendix.

[22] Note that PD and LGD will be used as abbreviations for default probability and loss given default, and *PD* and *LGD* (italicized) are the variables measuring these in regressions.

default rates are in line with the ex ante estimates. There is some underestimation of PDs by both banks and the regulator, but for most of the sample default rates are approximately increasing one for one with PD values. See the Appendix for full details. For LGD values, we evaluate the potential effect of measurement error in robustness exercises below.

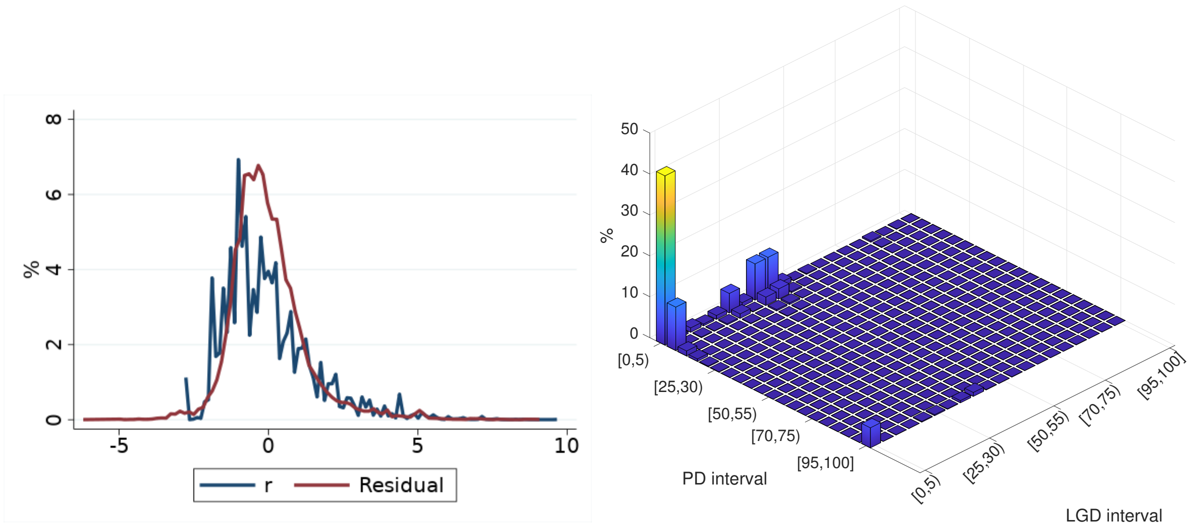
Since PDs are estimated at a one year horizon, we restrict the sample for the baseline analysis to loans with maturities up to one year to maintain a tight connection between the model and the data.[23] The PD and LGD values from banks are only available for loans issued from 2018, and do not have full coverage, whereas the regulator PDs are available for the full sample. We therefore use the regulator PDs for the baseline analysis, with the banks' LGDs, and in the Appendix we show that the results also hold for the bank PDs. Overall, the baseline sample size for this analysis is 89,896 loans. In a robustness exercise discussed below, we check that the results are not driven by any sample selection.

In taking equation (3) to the data, we also want to control for the effect of loan characteristics on rates, so that we can assess the total share of the variation that can be accounted for by default risk and loan characteristics. To do this, we take logs of both sides of the equation, rearrange, and then add the controls from before to the right hand side to give the regression equation:

$$r_{lt} = \beta_1 \log(1 - PD_{lt} LGD_{lt}) + \beta_2 \mathbf{L}_l + \mathcal{M}_{lt} + \varepsilon_{lt}. \tag{4}$$

Note that we allow the coefficient on the default risk term to be determined by the data to enable the best possible fit. Since the risk free rate is purely a function of maturity and time, it is absorbed by the maturity-time fixed effects and does not need to be included in the regression. For estimating this equation, we no longer divide the sample by whether loans are collateralized or not since this is not an important

[23] To perfectly map the theory to the data, we would want to use only loans with a maturity of exactly one year, because this is the horizon for the PD definition. The issue with this is that the sample size is small. We have therefore extended the sample to all loans with a maturity up to one year. This will result in an overestimate of price of default risk for these loans.



(a) Interest rate distribution & residuals

(b) Joint PD-LGD distribution

Figure 3: **Interest rates and risk.** Panel (a) presents the raw interest rate distribution (blue) and the residuals after controlling for loan characteristics and default risk (red). Both distributions are centered with their means at zero. Panel (b) presents the joint distribution of PDs and LGDs. PDs and LGDs are grouped into twenty bins: $[0\%, 5\%), [5\%, 10\%), \ldots, [95\%, 100\%]$.

dimension of the data for rate heterogeneity.

Figure 3(a) compares the actual interest rate distribution in blue, and the distribution after removing the effects of default risk and loan characteristics in red. The two distributions overlay each other, showing that risk is not a major factor in explaining the interest rate distribution. If risk accounted for the residual variation after the previous set of controls, then the red distribution would be concentrated on zero.

To quantify the amount of variation in interest rates that can be attributed to risk, we adopt the Abowd et al. (1999) (AKM) method. For each observation, let $\hat{r}^c_{lt}$ denote the estimated value of component c from the right-hand side of equation (4). The three right-hand side components are the risk term, loan characteristics ($\beta_2 \mathbf{L}_l + \mathcal{M}_{lt}$) and the residual. The overall variance of interest rates is given by sum of the variances of each of these components plus the covariances between them:

$$\text{Var}(r_{lt}) = \sum_{c=1}^{N} \text{Var}(\hat{r}^c_{lt}) + \sum_{j=1}^{N} \sum_{c=1}^{N} \text{Cov}(\hat{r}^c_{lt}, \hat{r}^j_{lt}), \quad j \neq c, \tag{5}$$

where N is the number of right hand side components in the regression. The contribution of component c to the total variance is $\text{Var}(\hat{r}^c_{lt})/\text{Var}(r_{lt})$. The results imply that default risk accounts for only 1.2% of the variance in rates and 70% of the variance is unexplained.[24]

Risk is not very important for interest rates because there is simply not much risk in the data. A large fraction of firms have a PD close to zero. The median PD is 2.2% and the 90th percentile of the distribution is only 11%. The LGD distribution is equally concentrated around low values with its median being 0%. The joint distribution of these variables in Figure 3(b) highlights the low level of risk. There are some LGD values in the range of 30–50%, but the vast majority of PD and LGD value are less than 10%, and most are less than 5%.

We have performed a wide range of exercises to verify that default risk accounts for only a small share of the variation in interest rates. We now discuss these, with the additional details provided in the Appendix. For exercises that involve variations to the specification in equation (4), or estimating it on alternative samples, the results are summarized in Table 2.

One potential concern is that there could be measurement error in the PDs and LGDs. We address this in three ways. First, we repeat the analysis using the PD estimates from banks. Assuming that these are accurate reports of the banks' beliefs at the time when loans were initiated, they are the correct measure of risk for pricing loans, even if they are inaccurate ex post. When we use these measures the results are very similar, with default risk accounting for only 2.5% of interest rate variation.[25] Second, we consider the possibility that the reported PDs do not fully reflect the information of banks. If banks have better information and price loans accordingly, then the interest rate residuals in the regression would have predictive power for defaults. We have tested this by regressing a default indicator on these residuals and find that this is not the case. The R^2 for this analysis is 0.1%.[26] Third, we evaluate how large

[24]Loan characteristics account for 27% of the variance and the remaining 1% is due to the covariance terms.

[25]In the Appendix we show that the PDs from banks and the regulator are similar, which explains this result.

[26]For this analysis we use a more flexible semi-parametric specification for the relationship between

measurement error would need to be in order to explain the variation in rates that is observed. Specifically, we reestimate equation (4) imposing a coefficient of minus one on the risk term, following the theory, and ask what values the PDs would need to take to fit the interest rate observations holding LGDs at their values in the data and, we do the same for LGDs taking PDs from the data. The values are implausibly large. More than 20% of PDs and LGDs would need to be above 100%, and the medians would need to be an order of magnitude larger than the reported values.

We also evaluate the effects of alternative stochastic discount factors for pricing loans. Under the baseline approach, we are assuming that lenders are risk neutral and therefore discount at the risk free rate. In principle, lenders could discount the future at a different rate, or have a stochastic discount factor that varies across states so that if two loans have payments with different correlations with the stochastic discount factor, then they will be priced differently. For the first of these variations, note that when we take the pricing equation to the data, the stochastic discount factor is absorbed into the maturity-time fixed effects because its only variation is with respect to time. Our methodology is therefore robust to any common discount rate of lenders. To evaluate the second possibility, we care about the correlation between the likelihood that a firm defaults and the lender's valuation of the states in which default occurs. For a variable with information about the former, we choose the revenue of firms. For the value of states, we take guidance from consumption-based asset pricing theories and use the value of aggregate consumption. We take the correlation between these two variables and add it as a control to regression (4). We also allow it to interact with the risk term in this equation. This variation hardly affects the baseline results, with default risk now accounting for 1.2% of the interest rate variance. Given that loans do not have a lot of risk to begin with, it is not surprising that changing the weighting of different states does not affect the results.

Equation (3) places a particular structure on the relationship between loan prices and default risk. While this is the dominant approach in macroeconomic models, it

the risk variables and interest rates when extracting residuals, to allow the regression to capture the full information content of these variables. This specification is described below.

is possible that it is misspecified and that the true relationship takes another form. In the baseline regression, equation (4), we give the relationship additional freedom by allowing the coefficient on the risk term to be estimated rather than imposing a value of minus one. The estimated coefficient validates our approach, as it is -0.08 and statistically different from minus one. If we impose a coefficient of minus one, the explanatory power of default risk for interest rates decreases. To be more flexible with the relationship between interest rates and risk, we also use a semi-parametric specification. For this, we group PDs and LGDs into deciles, then take equation (4) and replace the risk term with six new terms: PD_{lt} and PD_{lt}^2, allowing the coefficients to differ for each decile of PDs; LGD_{lt} and LGD_{lt}^2, allowing the coefficient to differ for each decile of LGDs; $PD_{lt} \times LGD_{lt}$ and $PD_{lt}^2 \times LGD_{lt}^2$, allowing the coefficient to vary for each possible pair of PD and LGD deciles. This specification increases the share of rate variation due to default risk to 14.9%, indicating that the empirical relationship between default risk and interest rates deviates from the theory. Nevertheless, the majority of interest rate variance remains unaccounted for.

An alternative approach that circumvents concerns about how risk is specified in the regression, as well as reducing concerns about PD and LGD measurement, is to restrict the sample to loans with a low PD.[27] If variation in rates is primarily due to loan characteristics and default risk, then for low risk subsamples, loan characteristics should account for most of the variation in rates. To test this, we restrict the sample to loans with a PD below the 5th, 10th and 25th percentiles. This changes the results little, with 53%, 54% and 57% of the variance in rates remaining unaccounted for.

A limitation of the analysis is that the theory is based on a loan for which all principal and interest are repaid at maturity. This is not the case for all loans in the sample. To address this, we restrict the sample to loans for which all principal is paid at maturity, providing a closer correspondence to the theory. We also reestimate the results separately for loans with French amortization and for loans with other amortization

[27] Based on the theory, we would ideally like to focus on loans with the lowest values of $PD \times LGD$. We use just the PD instead, as this allows for a much larger sample.

	Risk explanatory power
Baseline	1.2%
Robustness test	
Bank PDs	2.7%
Alternative stochastic discount factor	1.2%
Semi-parametric specification	14.9%
Amortization types (all principal at end, French, other)	2.0%, 0.4%, 0.3%
Value of exposure	0.0%
Exclusion of Covid period	0.4%
Firm size groups (from small to large)	0.2%, 0.5%, 1.0%, 1.5%, 2.1%
Firm age groups (0–4, 5–15, more than 15 years)	0.7%, 0.6%, 1.2%
All credit instruments	0.7%
Semi-parametric: PDs only (baseline and full samples)	5.4%, 2.3%
Extension to longer maturity loans	0.7%

Table 2: **Robustness tests for default risk results.** "Semi-parametric using only PDs" refers to the semi-parametric specification in which PD values are used, but no LGD values. This analysis is done for the full sample, and the sample for which LGD values are available (i.e. the baseline sample). The amortization types are all principal paid at maturity, French amortization, and all other categories grouped together. In the exercise adjusting for the value of exposure, we multiply each LGD value by the expected exposure at default as a share of the loan value. The analysis excluding the Covid period uses loans issued before 2020 only. The five firm size groups are small (0–4 employees), medium-small (5–12 employees), medium (13–39 employees), medium-large (40–119 employees) and large (120 employees or more). The analysis for amortization types and firm size and age groups involves repeating the analysis for each of the specified subsamples.

structures. The results vary little from the baseline, with default risk accounting for 0.3–2.0% of rate variation for these subsamples. As as a second approach, we directly assess whether payments on loans prior to maturity affect the results by accounting for a bank's expected exposure at the time of default.[28] We multiply LGD by exposure (as a share of initial loan value) to allow for repayments before default to reduce a bank's potential losses. In principle, heterogeneity in exposure across loans could generate additional heterogeneity in rates. However, this does not increase the explanatory power of default risk for interest rates.

We verify that the results are not driven by sample selection due to incomplete reporting of LGDs, the Covid period, firms of particular ages or sizes, and that they hold when all credit instruments are included (see Table 2). For sample selection, the

[28]Exposure is the total outstanding amount on a loan (principal and interest) at a given point in time.

approach is to repeat the semi-parametric decomposition using only PDs, which are available for all loans, both within the LGD subsample and in the full loan sample. Omitting LGDs reduces the explanatory power of risk from 14.9% to 5.4%. Extending the analysis to the full sample produces a modest change to 2.3%, indicating that our findings are not driven by the particular subsample of loans with LGD data.

The final exercise extends the analysis to longer maturity loans. The baseline analysis restricts the sample to loans with a maturity of up to one year because that is the horizon for default probabilities. For loans with a maturity of $T \geq 1$ years, and all principal and interest paid at maturity, equation (3) can be generalized to

$$1 + r_{ltT} = \frac{1 + \tilde{r}_{tT}}{(1 - PD_{ltT} \times LGD_{ltT})^{\frac{1}{T}}}. \tag{6}$$

r_{ltT} is the per annum interest rate for loan l issued at time t with a maturity of T years, and $\tilde{r}_{tT}$ is the per annum risk free rate at time t for this maturity. PD_{ltT} is the overall probability of default over the T year horizon, and LGD_{ltT} is a weighted average of loss given default across states in T years, with the weights equal to the share of total default risk for the loan in each state.[29] As for the baseline analysis, to take this equation to the data we take logs of both sides and add controls to the right hand side for loan characteristics. To extend PD estimates beyond one year, we assume that they are constant over the maturity of a loan, and the realization of default is i.i.d. each year.[30] LGD in the data is the share of the loan value that is expected to be lost if there is default within the next year. We assume that this value is the same for a default occurring at any point in the future for longer maturity loans. With these assumptions, we can extend the analysis to loans of all maturities.[31] We find results to be very similar to our baseline, with risk explaining 0.7% of interest rate variation.

Our overall finding is that default risk has small explanatory power for the variance in interest rates. At best, under the semi-parametric approach, it accounts for

[29]See the Appendix for additional details for equation (6).

[30]This means that the default probability for loan l, initiated at time t, with a maturity of T years is $1 - (1 - PD_{lt})^T$.

[31]For loans with some principal or interest paid prior to maturity, or an adjustable rate, equation 6 should be treated as an approximation.

14.9% of the variance. After accounting for both default risk and loan characteristics, 62% of the total variance in rates is unaccounted for.

4 Banks and relationship banking

This section turns to leading alternative explanations for interest rate dispersion. The finance literature emphasizes the role of bank heterogeneity arising from differences in cost structures, risk profiles, and market power across lenders, and firm–bank relationships, whereby repeated interactions and private information shape loan pricing, as sources of interest rate dispersion. These ideas have been incorporated into macroeconomic theory too. This section quantifies the contribution of each of these sources to the observed dispersion in interest rates.

4.1 Banks and interest rate variation

Bank heterogeneity can translate into differences in the interest rates that they charge on loans in several ways. The finance literature identifies three main sources of such heterogeneity. First, market power, which directly affects lending markups, as originally formalized in the models of Monti (1972) and Klein (1971). Second, cost heterogeneity, since larger or more efficient banks typically face lower funding and monitoring costs, consistent with the intermediation and delegated monitoring frameworks of Diamond (1984). Third, risk composition, as banks with riskier asset portfolios demand higher spreads between lending and deposit rates to compensate for default and interest rate risk (Ho & Saunders, 1981; Angbazo, 1997).[32]

The objective of our analysis is to quantify the contribution of bank heterogeneity to interest rate variation, while being agnostic about the mechanism, by using rich fixed effects. Our baseline approach takes regression (4), and adds bank and firm fixed effects. We add both types of fixed effects to negate the risk that firms sort across banks in a way that would result in the bank fixed effects picking up what are actually

[32] Empirical evidence supporting these mechanisms includes Berger & Mester (1997) and Darmouni (2020) . Moreover, Corbae & D'Erasmo (2021) document systematic markup heterogeneity across banks of different sizes and develop a quantitative model that nests these three channels.

firm effects. The regression for loan l of firm f from bank b initiated at time t is:

$$r_{lfbt} = \beta_1 \mathbf{R}_{lt} + \beta_2 \mathbf{L}_l + \mathcal{M}_{lt} + \mathcal{B}_b + \mathcal{F}_f + \varepsilon_{lfbt}, \tag{7}$$

where $\mathbf{R}_{lt}$ contains controls for default risk, $\mathcal{B}_b$ and $\mathcal{F}_f$ denote bank and firm fixed effects, respectively, and the other controls are defined in the same way as in the previous section. From now on we control for risk using the semi-parametric method introduced in Section 3.2, including only PD terms. We omit LGD terms since including these reduces the sample size substantially, and we need a large sample in order to estimate bank and firm fixed effects. The robustness exercises in Section 3.2 showed that leaving out LGD controls decreases the explanatory power of risk by around 50%. We are interested in how much of the variation in rates can be accounted for by the bank fixed effects, $\mathcal{B}_b$. We emphasize that these fixed effects capture the effects of *all* relevant bank characteristics that are common over time and affect all firm financing loans in a bank's portfolio.

The bank and firm fixed effects in this setting are analogous to the firm and worker fixed effects, respectively, that are estimated in the labor literature on the determinants of wages. In that literature, identification of the fixed effects typically relies on variation arising from people changing jobs, which can generate mover bias.[33] Our identification does not rely on the same variation and is not subject to this issue. The banking side of the network is highly concentrated: a small number of large banks lend to many distinct firms, so bank fixed effects are estimated with large sample support and are well identified. At the same time, firms in the sample have multiple loans with different characteristics, even when they borrow from a single bank, providing additional within-firm variation that helps precisely estimate firm fixed effects conditional on well-identified bank effects. Put differently, identification of the two sets of effects is contingent on dense bank-to-many-firms links and substantial within-firm loan variation. The empirical network is therefore highly connected and we find no

[33] By mover bias we mean the bias that can arise when fixed effects are identified from people who move jobs and job mobility is non-random.

	Contribution
(1) Default risk	0.0049
(2) Loan characteristics	0.2555
(3) Bank fixed effects	0.0956
(4) Firm fixed effects	0.3265
(5) Residual	0.2766
(6) Covariances	0.0458
# Loans	521,856

Table 3: **Contributions of default risk, loan characteristics, and bank and firm fixed effects to interest rate variation.** Each row is the share of the total variance in interest rates due to each component of regression (7), calculated with equation (5).

evidence that only a small, unrepresentative subset of multi-bank firms drives our results. In the Appendix we report the distribution of the number of loans and banks per firm, and the distribution of the number of loans and firms per bank. Additionally, we later show that results do not change when restricting the sample to only firms with multiple bank relationships.

Table 3 reports the AKM decomposition of the variance in interest rates, presenting the contribution of each component of regression (7). The main result is in the third line: banks only account for 9.6% of the variance in interest rates and are therefore not a primary source of dispersion. Loan characteristics (which include maturity-time effects) remain important, accounting for 25.6% of variation, and the other standout contribution comes from firm fixed effects, which account for 32.7%. We will further evaluate firm effects in Section 5. An additional point to note is that the covariances between the components of the regression matter little for overall variance.[34]

As an extension of the analysis, we also consider a specification in which the bank and firm fixed effects, $\mathcal{B}_b$ and $\mathcal{F}_f$ respectively in equation (7), are replaced by bank-time and firm-time fixed effects. This specification allows the effects of individual banks on loans to vary over time. In order to estimate this specification, we need

[34]Neither the sum of the covariances nor the individual terms contribute substantially to the overall variance. For example, the largest covariance term—between bank fixed effects and loan characteristics—accounts for 6% of the total variance. The relatively low importance of the covariance terms indicate that there is little systematic sorting of specific firms (banks) into particular banks (firms) or loan characteristics.

observations for firms that originate multiple loans in the same period. There is a tradeoff between the sample size and frequency of variation that is permitted.[35] We use a one year time period, judging that this is sufficiently high frequency to capture changes in firm and bank characteristics that could affect interest rates, while preserving the sample size. This gives us a sample of 41,312 firms, relative to a sample of 84,535 under the baseline approach. The results are presented in the Appendix and show that allowing bank effects to vary with time does not change the contribution of banks to rate variation: under this specification it is 9.5%.

For additional verification of this result, we perform robustness exercises. One potential concern is that the firm fixed effects might be upwardly biased, and impact the bank fixed effect estimates, if there are not enough loans per firm. In the baseline analysis there are six loans per firm on average. We perform two exercises to address this. First, we restrict the sample to firm with at least 10 loans each and, second, we use all instruments instead of just firm financing loans. These modifications increase the average number of loans per firm to 36 and 12, respectively. For both specifications the results are similar to the baseline. Bank fixed effects account for less than 10% of rate variation, and firm fixed effects account for 42% and 37%, respectively.

Similar to our analysis in Section 3.2, we test the results by amortization type, by firm size and age groups, and exclude the Covid period. Overall, results are robust across all of these subsamples, with bank effects never explaining more than 15% of the variation in rates. Full details of these exercises are in the Appendix.

4.2 Firm-bank relationships and interest rate variation

While bank heterogeneity does not play a large quantitative role in accounting for interest rate variation, there could be additional important effects at the level of bank-firm relationships. Relationship banking theory holds that repeated interactions between firms and banks can affect credit terms. One rationale for this is informational: by learning about a borrower over time, banks can overcome adverse selection and

[35] As the frequency of the time dimension of the fixed effects increases, the number of firms that originate multiple loans in a single period falls, decreasing the sample size.

moral hazard problems (e.g. Diamond, 1984). A second rationale is contractual: long-term relationships enable intertemporal flexibility and implicit risk sharing through rate smoothing and covenant adjustment, as modeled by Boot et al. (1993) and Berlin & Mester (1998).[36] Relationships can also affect loan prices by generating monopoly power for banks (e.g. Sharpe, 1990) and leading banks to refinance weak firms (e.g. Bolton & Scharfstein, 1996; Faria-e-Castro et al., 2024b).

The objective of this section is to quantify the relevance of factors that are specific to individual relationships between firms and banks for overall interest rate variation, while being agnostic about the specific mechanisms. We again take a fixed effects approach. In doing so, it is important to separate factors that pertain to the relationship itself from those that are idiosyncratic to each firm or bank. The analysis makes use of regression (7), as well as the following variation of it:

$$r_{lfbt} = \beta_1 \mathbf{R}_{lt} + \beta_2 \mathbf{L}_l + \mathcal{M}_{lt} + \mathcal{FB}_{fb} + \varepsilon_{lfbt}, \tag{8}$$

where $\mathcal{FB}_{fb}$ is a vector of firm-bank fixed effects and the other variables have the same definitions as they have had throughout the paper. The difference with regression (7) is that the firm and bank fixed effects are now interacted. In order to isolate the effect of the relationships between firms and banks on interest rates, there are three steps. First, we estimate regression (8) and compute the contribution of firm-bank fixed effects to total interest rate variation using the AKM method. This provides a measure of the total contribution of firm factors, bank factors and relationship factors jointly. Second, we take the same sample of loans, reestimate equation (7), and use the AKM method to measure the separate bank and firm contributions to interest rate variation. Third, the difference between the sum of these values and the total contribution of firm-bank fixed effects in step one, is the contribution of the pure relationship component. For this analysis, firms need to have multiple loans from multiple banks. This restriction results in a sample of 463,624 loans to 73,858 firms.

[36] Empirically, Petersen & Rajan (1994) and Petersen & Rajan (1995) show that stronger, more exclusive relationships lower borrowing costs, reduce collateral requirements, and ease credit access, consistent with theories of intertemporal rent-sharing.

	Contribution
(1) Bank effects	0.095
(2) Firm effects	0.361
(3) Firm-bank relationship effects	0.115
# Loans	463,624

Table 4: **Contributions of firm, bank and relationship effects to interest rate variation.** The values are computed by estimating equation (8), and then equation (7) on the same sample. For each estimation we decompose the interest rate variance using equation (5). The contributions of the bank effects and the firm effects are the estimated contributions of the relevant fixed effects from equation (7). The firm-bank relationship effects contribution is the difference between the contribution of firm-bank fixed effects from equation (8), and the total contribution of firm and bank fixed effects from equation (7).

The results are presented in Table 4. Firm-bank relationships account for 11.5% of the overall variation in rates and, once this component is removed, the contribution of pure bank effects is still 9.5%. So, in total, banks and relationship banking account for 21% of variation. While this is a meaningful amount, we are still only accounting for 53% of the variance in rates after including variation due to loan characteristics, default risk, bank effects, and relationship banking. For the remaining variation, firm fixed effects are still the standout factor, responsible for 36.1% of variation. As well as being quantitatively important, the relevant characteristics of firms are unrelated to their banking relationships. This analysis also shows that when the sample is restricted to firms with multiple bank relationships, it does not alter the importance of the firm fixed effects, reinforcing the validity of our fixed effects identification.

To provide further support for these results we perform additional exercises that we report in the Appendix. Since firm-bank relationships can change over time, our baseline estimate may underestimate their contribution. To evaluate this, we modify the regressions by adding a time dimension to the firm, bank and firm-bank fixed effects.[37] The frequency for the time dimension is one year. We use the sample of all credit instruments for this analysis, to increase the number of observations per firm-bank-year. We find that this has a modest effect on the results, with firm-bank relationships accounting for 16.4% of the total variation in interest rates.

[37]This means that equation (8) is estimated with firm-bank-time fixed effects and equation (7) is estimated with firm-time and bank-time fixed effects.

5 Firm characteristics and interest rates

The results in the previous section indicated that firm characteristics are important for accounting for interest rate variation. In this section we investigate what characteristics are important. We evaluate a broad range of observable characteristics alongside unobservable factors, and complement this with a statistical method that determines whether differences in rates across firms originate before or after their first loans.

To understand what firm characteristics are important for determining interest rates, we take a two-step approach. In the first step we take regression (7) and add controls for relevant time-varying firm characteristics, keeping firm fixed effects to account for permanent heterogeneity in interest rates not accounted for by the observable characteristics. We then take the estimated fixed effects and evaluate how much of these can be accounted for by observable permanent firm characteristics.[38]

The regression for the first step is:

$$r_{lfbt} = \beta_1 \mathbf{R}_{lt} + \beta_2 \mathbf{L}_l + \mathcal{M}_{lt} + \mathcal{B}_b + \beta_3 \mathbf{F}_{ft} + \mathcal{F}_f + \beta_4 \mathbf{BF}_{bft} + \varepsilon_{lfbt}, \tag{9}$$

where $\mathbf{F}_{ft}$ is a vector of control variables for firm f at time t, $\mathbf{BF}_{bft}$ is a vector of controls for characteristics of the relationship between firm f and bank b at time t, and all other variables maintain their existing definitions. We are adding controls for the characteristics of firm-bank relationships rather than using the methodology from Section 4.2, which requires firm-bank fixed effects, to maintain a broad sample and economize on the number of fixed effects to estimate.[39] In selecting the characteristics of firms to use in $\mathbf{F}_{ft}$, we aim to capture the state of the firm at the time the loan is granted by including everything that would be a state variable in relevant models of firms, as well as anything else in the data that could plausibly affect the interest rate charged by a bank.

[38]The second step is necessary because the effect of time-invariant observables cannot be separately identified from firm fixed effects in the first step.

[39]Specifications with firm-bank fixed effects are the most demanding that we run. Substituting them with measurable characteristics of firm-bank relationships creates the risk of overestimating the role of firm characteristics and firm fixed effects. However, this does not appear to be a major concern in our setting since the estimated role of fixed firm characteristics does not change substantially when firm-bank effects are considered—see Section 4.2.

Note that we are already controlling for default risk, loan characteristics, bank effects, and banking relationships, so we are assessing the relevance of firm characteristics for interest rates beyond these factors.

For time-varying characteristics, we control for standard factors of production—total factor productivity (TFP),[40] capital and labor (employment and the wage bill)—as they are commonly state variables for a firm or otherwise are informative about production technologies. The investment rate contains information about the demand for financing, and inventories, revenue and the profit share are informative about the scope of the firm, its cash flow, and its ability to continue operating.[41] Revenue growth contains information about the future trajectory of a firm. On the financial side, we control for the values of equity, financial and total assets, as well as total liabilities, to capture the general balance sheet positions of firms. Past default and late repayment indicators are included to capture any effect that credit history may have on loan pricing beyond default risk.[42] The values of trade payables and receivables relative to total assets provide information on cash flows. Continuous variables, and those that take a wide range of integer values, are mostly entered in the regression in log units.[43] For additional details about the variables and their measurement, see the Appendix.

To control for firm-bank relationships, we include the number of banks that a firm has borrowed from, the Herfindahl-Hirschman index for the distribution of a firm's credit across banks,[44] and five variables to capture the duration, depth and nature of the firm's relationship with the bank issuing the loan in question. For duration, we use the time since the firm received its first loan from the bank. To capture the depth and nature of the relationship we use four moments: the number of loans issued by the bank to the firm over the past five years; the log total value of these loans; and the

[40] Measured using the method from Ackerberg et al. (2015).

[41] For inventories, there is an indicator variable for values of zero, and otherwise it is in log units.

[42] A payment is classified as late if it is 30 days or more overdue.

[43] Specifically, we use log units for capital stock, number of employees, wage bill, revenue, TFP, financial assets, total assets, and total liabilities. For inventories, there is an indicator variable for values of zero, and otherwise it is in log units. Equity is specified in levels because of negative values.

[44] The number of banks is the total number of banks that a firm has obtained firm financing loans from prior to the month of loan initiation, and the Herfindahl–Hirschman index is computed using all of a firm's outstanding firm financing loans in the month prior to loan initiation.

number of loans and the value of these as a share of the number of loans and value of loans that the firm has received from all banks over the past five years.[45]

In the second step of the analysis, we take the estimated firm fixed effects from regression (9), and evaluate how much of the variation in these can be accounted for by observable time-invariant firm characteristics. For this, we use the regression:

$$\mathcal{F}_f = \beta \mathbf{F}_f + \varepsilon_f, \tag{10}$$

where $\mathbf{F}_f$ is a vector of time-invariant firm characteristics. We control for industry at the 5-digit level to capture industry specific demand and supply factors that may not be captured by firm-level variables. For location we use controls at the 4-digit post-code level.[46] These address any location-specific demand or supply factors affecting the firm, as well as geographic variation in banking markets that could affect interest rates. To further address the potential for local banking market effects, we control for the number of top five banks and the number of other banks with branches located within the more granular 5-digit and 7-digit postcodes (four variables).[47] To capture the scope of a firm's domestic operations, we use the maximum number of establishments that the firm attains in the data. A firm's country of origin may influence its outcomes by changing its familiarity with the local market, the distance of firm headquarters from local operations (Giroud, 2013), or the ease with which banks can enforce contracts. The legal form of businesses has been shown to correlate with firm outcomes, perhaps as an indicator of the owner's ambitions for the firm (Levine & Rubinstein, 2017). We control for R&D intensity with a firm's maximum number of R&D employees over time. Controls for whether a firm has ever exported and its number of foreign establishments address effects of international trade and multinational production.[48] Indicator variables for birth year address cohort effects, such as

[45]For all the firm-bank relationship variables, we only consider firm financing loans.

[46]Four-digit postcodes have an average area of 135 square km (18 square km in Lisbon).

[47]Five-digit postcodes have an average areas of 24 square km (3.9 square km in Lisbon).

[48]Some of the variables used for time-invariant firm controls have a small amount of variation within firms in the data: number of local banks, number of establishments, whether the firm has R&D employees, whether a firm exports, and its number of foreign establishments. However, since the amount of

	Contribution
(1) Default risk	0.0040
(2) Loan characteristics	0.2240
(3) Bank fixed effects	0.0931
(4) Time-varying firm characteristics	0.0093
(5) Relationship banking controls	0.0010
(6) Firm fixed effects	0.3050
(6a) Fixed firm characteristics	0.039
(6b) Latent firm fixed effects	0.267
(7) Residual	0.2871
(8) Covariances	0.0904
# Loans	412,199

Table 5: **Variance decomposition including firm controls.** Each row is the share of the total variance in interest rates due to each component of regressions (9) and (10), calculated with equation (5).

the interest rate level when a firm was born. Finally, we want controls for the distributions of supply and demand shocks that firms face. We proxy for these with the standard deviation and skewness of TFP and revenue for each firm.[49]

Table 5 presents the contribution of each component of the right hand sides of the first and second stage regressions to the overall variance of interest rates. We are particularly focused on the contributions of firm characteristics and latent firm fixed effects in rows (4), (6a) and (6b). Firm fixed effects explain 31% of the total variance, in line with the results from Section 4. Thus, the addition of the time-varying firm observables does not change the quantitative importance of firm fixed effects. Second, nearly all of the firm fixed effects are due to latent factors. Observable characteristics account for 3.9 percentage points of these fixed effects, with the remaining 26.7 percentage points being latent. Overall the strong firm fixed component in interest rates has a low correlation with both time-varying and invariant firm characteristics.

As a robustness check, reported in the Appendix, we extend the analysis to the

variation within firm for these variables is very small, we use time-invariant versions. Otherwise they would partially capture firm fixed effects in the first stage.

[49]The sample has 109,657 fewer loans and 25,503 fewer firms than in Section 4.1, due to missing observations for some variables and because some controls cannot be calculated for age zero firms (e.g. growth rates). The sample is now 412,199 loans and 59,032 firms.

sample of all credit instruments. This results in an estimate for the contribution of firm fixed effects to total interest rate variation of 48.8%, with time-invariant firm observables only accounting for 3.6 percentage points of this.

We conduct an additional exercise to assess whether the persistent heterogeneity in borrowing costs is already present when firms first enter the credit market, or instead emerges through shocks realized later in their borrowing histories. To do this, we apply the statistical framework of Sterk et al. (2021) to decompose firm-level sequences of interest rates on loans into an ex ante component and an ex post component.[50] We find that the ex ante component is quantitatively important, indicating that heterogeneity present prior to first loan issuance is a significant source of persistent rate dispersion. Quantitatively, ex ante factors account for roughly 15 percentage points of total interest rate variance, compared with 31 percentage points attributed to the firm fixed effect in Table 5. Further details are provided in the Appendix.

6 Interpretation of firm fixed effects

The analysis has shown that a substantial share of the variation in interest rates across firms is due to latent, highly persistent firm-level heterogeneity. This accounts for 27% of the total variation in interest rates, and is the single largest source of this variation. To quantify this in a different way, the standard deviation in rates due to this persistent heterogeneity is 60 basis points.[51] In this final part of the paper, we discuss potential sources of this variation. Understanding the source is an important step towards pinpointing the underlying frictions in financial markets, and evaluating their broader economic implications.

Constraints imposed by the empirics Our starting point is to specify the characteristics that an explanation must have to be consistent with the empirical analysis. There are three necessary properties. First, the explanation must be at the firm level. It needs to account for idiosyncratic variation in interest rates across firms, and it can-

[50] Sterk et al. (2021) apply the methodology to employment dynamics. We adapt it to interest rates.
[51] This is the share of variation from latent firm fixed effects times the standard deviation of rates.

not be a result of differences in default risk, differences in the types of loans that firms issue, selection of firms across banks that charge different rates, or firms attaining idiosyncratic rates because of individual relationships with banks. These possibilities are ruled out by our controls for default risk, loan characteristics, bank and bank-time fixed effects, and our analysis of relationship banking.[52] Thus the focus should be on factors that are idiosyncratic to firms.

Second, it must be able to explain highly persistent differences in interest rates across firms. The empirical analysis says that it is firm *fixed effects* that are the main source of variation. If the source of variation was idiosyncratic to firms but time-varying, then the fixed effects estimates would be smaller, and more variation would be captured by the residuals of the regressions. Explanations that generate transitory differences in rates across firms are therefore not enough. The explanation needs to be based on factors that are idiosyncratic to firms and highly persistent or permanent.

Third, the explanation must have a low correlation with the broad range of observable characteristics that we have controlled for. Without recapping the breadth of these controls—discussed in the previous section—let us highlight that we have controlled for differences across firm size, productivity level, a range of properties of productivity processes, capital, industry and an array of measures of the financial positions of firms. Given the extent of these controls, the results indicate that the explanation for the variation in interest rates lies outside of the firm characteristics that are typically considered.

Bank-related explanations The analysis has already considered the role of bank heterogeneity and relationship banking for interest rate variation in Section 4. Here we discuss alternative channels through which banks could generate firm fixed effects in interest rates, and explain why these explanations are challenged by the data.

If banks have market power, then this could generate persistent differences in interest rates across firms. The difficulty with this, is that the controls in the analysis

[52]For default risk, recall that we performed a number of robustness exercises to rule out there being a plausible amount of default risk that is priced by banks, but unobservable to the econometrician, that can account for the unexplained variation in rates.

address many types of market power. Bank controls address the ability of banks to charge consistently different prices across their loans, and bank-time controls make this more specific, addressing time variation in differential pricing of this type. If there is geographic variation in banking markets, then rate variation could show up across local markets rather than across firms (e.g. Bordeu et al., 2025). However, the firm-level analysis included firm location controls at the 4-digit postcode level that would capture this. We have also run this analysis (results in the Appendix) with richer controls at the 5-digit postcode level and still find small effects. To address the possibility that market power shows up at the bank-location level, we also test interacting bank fixed effects with 4-digit postcode dummy variables. This does not decrease the relevance of firm fixed effects for rate variation.

Alternatively, banks could generate firm fixed effects in interest rates through pricing a collection of loans jointly, rather than pricing each loan independently (e.g. Degryse & Van Cayseele, 2000).[53] If this were to account for firm fixed effects in rates, then the effect on a firm's interest rates would need to be common across the banks that it has loans with. It would not be enough that a firm concentrates most of its loans with one bank, and gets a different price because of this. That would show up empirically as a firm-bank effect. We have already run a specification that evaluates the relevance of the scope of a firm's borrowing from a bank when we controlled for the number and value of loans in a firm-bank relationship in the previous five years in Section 5. These controls accounted for little of the variation in rates. We push this analysis further with two additional exercises. First, we take residualized interest rates after controlling for loan characteristics, default risk and bank effects, and assess whether these differ across banks where a firm has many loans and banks where it has few. We also split the sample into firms whose financing is concentrated in one bank compared to firms with very dispersed financing. In both exercises we find small differences, suggesting that the joint pricing of loans does not have a significant impact

[53] Degryse & Van Cayseele (2000) provides evidence that firms with more products from a bank, and greater turnover on their account, pay lower rates.

on price variation.

Another way that this channel could affect interest rates is if the owner or manager of a firm has a personal financial relationship with a bank that affects the pricing of the firm's loans. For example, a firm could get a better rate at a bank where the owner has a mortgage as well. This channel also faces the issue that the effect would need to show up across the banks that a firm borrows from, and not just at an individual bank. To assess this explanation, we run two tests in the data. We evaluate the relevance of firm fixed effects for rate variation across the size distribution of firms.[54] The assumption is that the personal finances of the owner or manager will be smaller relative to the loans of a larger firm, and therefore less relevant to the pricing of its loans. We also split the sample into public and private firms, as the personal connection between a firm and its founder or manager is weaker for public firms. The results of both exercises count against the link between personal finance and business finance being quantitatively relevant for our results: the relevance of firm fixed effects for interest rate variation hardly diminishes with firm size, and is no smaller for public firms than private firms.

Unobservable variation in demand A second class of explanations that warrants discussion is unobservable idiosyncratic shocks to firms that affect their demand for loans. For example, firms may suffer cost shocks or shocks to their investment demand (e.g. an unobservable investment opportunity) that change their willingness to pay for a loan. These shocks are firm specific, and could be uncorrelated with firm observables. However, it is not enough for such shocks to cause temporary changes in interest rates paid by firms. The second requirement is that they generate highly persistent differences in rates. So, it would be necessary that some firms persistently have different investment demand, or have a persistently different liquidity position so that they routinely have more urgent financing needs. However, if one firm has persistently different investment demand to another, then over time this will result in

[54]For this, we do two things: reestimate the results on subsamples by firm size; take the estimated fixed effects for the full sample, and compute the total variance of these within each firm size group. The second exercise looks at total variance rather than shares of variance.

a different capital stock, which is controlled for. If the source of variation is persistent differences in liquidity needs, then it is hard to square this with the low correlation between the financial positions of firms and their interest rates. In addition, we estimate credit demand and supply shocks, following Amiti & Weinstein (2018). We then average the demand shocks over time within firms and include them as a control in regression (10). Results in the Appendix show that average demand shocks have no explanatory power for the firm fixed effects.

Search frictions Considering the institutional setting, search frictions are a natural consideration as a source of interest rate variation. The typical process for a firm to obtain a loan is that it applies to a bank, providing a range of information, the bank processes the application, and then the firm may receive a loan offer. Firms can go through this process at multiple banks, but it is costly.[55] Survey and empirical evidence support the existence of search frictions in this setting (Federal Reserve Bank of New York, 2014; Mazet-Sonilhac, 2025), and it is well understood from search theory that these can result in price dispersion (e.g. Burdett & Judd, 1983).

Within the framework of a search model, we would need two ingredients to generate firm fixed effects in interest rates: a source of rate differences that is uncorrelated with contemporaneous firm observables, and a source of persistence for interest rates.

Focus on the first of these ingredients and consider an environment in which firms have different abilities for searching for loans, and that ability affects the distribution of rates that a firm draws from once it matches with a bank. Firms can differ in ability due to different efficiencies in applying for loans, differences in the information that firms have about the rates that banks offer, or differences in bargaining ability.[56] Since these differences in ability are unobservable, they will generate differences in interest rates, at least for initial loans. Conditional on ability, rates can still vary across firms if there are idiosyncratic effects within or across loan officers, or if rates are affected by

[55]The types of costs that we have in mind are the time and effort of the firm's staff who are making the application, the time taken for the bank to process the application, as well as any costs incurred in paying lawyers, accountants or other bodies to prepare the relevant documents.

[56]For evidence of heterogeneity in search efficiency (or cost), information, and bargaining ability in the context of consumer finance, see Allen et al. (2019), Agarwal et al. (2024) and Berwart et al. (2025).

non-economic factors like personal relationships between bank employees and firm owners, or the socio-demographic characteristics of the owner.[57] Thus, there are two potential sources of unobservable rate variation at the firm level in this setting.

For thinking about persistence, consider rate variation due to unobservable ability. If differences in ability are persistent, for example because ability is determined by owner characteristics, then this will be enough to generate persistent rate differences. If, instead, the source of variation is loan officer effects or non-economic reasons, then rates would not necessarily be persistent across loans or banks. However, the search setting provides natural reasons for persistence. If a firm that is lending from one bank can search "on-the-job" (to borrow the language of the labor search literature) and switch if it finds a lower rate, then the current rate will affect future rates.[58] Another plausible source of persistence is that new banks observe rates on previous loans as part of the application process, and this affects the negotiated rate.

A challenge for search-based explanations is that the correlation between a firm's interest rates and its observable characteristics needs to remain low over time. It is clear that the mechanisms outlined can result in rate differences that are uncorrelated with firm observables for the first loans that firms receive. However, if one firm persistently has a lower rate than another, then under most models of investment this will result in the lower rate firm accumulating a larger capital stock. The question is whether this mechanism is weak enough in a search model for bank loans, combined with a standard investment model, to quantitatively match the latent firm fixed effects in interest rates. In fact, this is a challenge for any theory of persistent interest rate differences once endogenous investment choice is considered.

In terms of implications, understanding the relative strengths of the proposed explanations is important. If there are fundamental reasons for the market for bank-

[57] There is an extensive literature documenting idiosyncratic variation in the decision making of loan officers. See Bertrand & Burietz (2023) for a summary of this. For particular evidence of discrimination on socio-demographic characteristics of business owners, see Brock & De Haas (2023).

[58] For on-the-job search to be relevant, firms would need to have the option of continuing to lend with their existing bank on similar terms. In support of persistent firm-bank relationships, 91.5% of loans in the sample are received from banks that the relevant firm received their previous loan from, or which the firm has another active loan from at the time of issuance.

firm loans to be decentralized and require search, and firms inherently have different search abilities, then the observed equilibrium could be efficient. If, instead, rate differences are a result of idiosyncratic behavior by loan officers or differentiation for non-economic reasons, then they are wedges that can cause misallocation, and a more standardized approach to rate setting could yield benefits. It is also worth reiterating that we have documented persistent, rather than transitory, differences in rates. These would generally have a greater impact on firm investment decisions, increasing the importance of these questions.

To give a sense of the aggregate importance of this distinction, we evaluate the cost of misallocation for the case in which the latent firm fixed effects in interest rates are purely wedges using the Hsieh & Klenow (2009) framework. We use these estimated fixed effects as measures of the interest rate distortions for firms. This type of distortion is in line with evidence in David & Venkateswaran (2019) that permanent firm level wedges in the cost of the capital are one of the primary sources of capital misallocation. Holding total input supplies fixed, so that losses are purely from misallocation and not any distortion to the total capital stock, equilibrium output relative to the efficient level is

$$\frac{Y}{Y^*} = \prod_{s=1}^{S} \left(\sum_{i=1}^{N_s} \left(\frac{Y_{si}^*}{Y_s^*} \right)^{\frac{\sigma-1}{\sigma}} \left[\left(\frac{K_{si}}{K_{si}^*} \right)^{\alpha_s} \left(\frac{L_{si}}{L_{si}^*} \right)^{1-\alpha_s} \right]^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\theta_s \sigma}{\sigma-1}} . \tag{11}$$

Y, Y_s and Y_{si} are aggregate, sector s, and firm i in sector s output levels, respectively, K_{si} and L_{si} are the capital and labor choices at the firm level, and stars denote values when there are no wedges in interest rates. There are S sectors and N_s firms in sector s. θ_s is the share of expenditure in sector s, α_s is the capital share in that sector, and σ is the elasticity of substitution between goods within sectors.[59] Losses depend on the efficient output shares of firms within sectors, and how much each firm's capital and labor choices are distorted. The Appendix details that, given our direct measure of interest rate wedges, the misallocation loss can be measured using only firm-level

[59] Within sectors, goods are aggregated by CES. Across sectors, the aggregator is Cobb-Douglas.

productivity and values for the parameters in equation (11). Under this approach, we are not inferring distortions from differences in marginal revenue products, and thereby avoiding the identification challenges associated with this.[60] Instead, we are directly using a measure of interest rate distortions and evaluating their implied losses in isolation. We measure losses using 2019 data, since we have data on all loans for this year, and it is pre-pandemic. The aggregate output loss is 0.43%.[61]

7 Conclusion

Firm financing is an increasingly important component of modern macroeconomic theory and interest rates are key prices. However, we still have a limited understanding of the empirical determinants of these. This paper addresses this question with analysis made possible by a detailed new credit registry, matched with firm balance sheet and income statement data.

We show that interest rates vary substantially even after controlling for a rich set of loan characteristics such as maturity, issuance date, and the amount and type of collateral. Furthermore, only a small share of this variation is accounted for by the core theory of rate dispersion, default risk. Theories for dispersion based on bank heterogeneity and relationship banking also have quantitatively low explanatory power. Instead, the main source of variation is persistent idiosyncratic firm-level heterogeneity that is uncorrelated with a broad range of firm observables, including firm age, size, industry, location, financial position, credit history, productivity, production input mix, and many others.

These findings suggest that there are substantial additional frictions in credit markets, outside of existing frameworks. The evidence implies that the main source of interest rate variation must satisfy three properties: it must operate at the firm level; generate highly persistent differences in rates across firms; and be only weakly correlated with the broad set of firm characteristics observed in the data, including the

[60] For research addressing these challenges see Asker et al. (2014) and Bils et al. (2021), for example.
[61] Full details of the model and quantification are provided in the Appendix.

standard characteristics that define firms in macroeconomic models.

We evaluate several candidate explanations that could satisfy these requirements. The evidence provides little support for explanations based on local bank market power, the joint pricing of multiple loans to the same firm, interactions between entrepreneurs' personal and business finances, or unobservable demand shocks. Instead, it points most strongly toward search-related frictions in the market for bank loans. An important task for future research is to distinguish among the specific channels through which these frictions operate, such as differences in firms' ability to search for credit, heterogeneity at the loan-officer level, or differentiation based on non-economic characteristics, and to understand their macroeconomic implications.

References

Abowd, J. M., Kramarz, F., & Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2), 251–333.

Ackerberg, D. A., Caves, K., & Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83(6), 2411–2451.

Agarwal, S., Grigsby, J., Hortaçsu, A., Matvos, G., Seru, A., & Yao, V. (2024). Searching for approval. *Econometrica*, 92(4), 1195–1231.

Albuquerque, R. & Hopenhayn, H. A. (2004). Optimal lending contracts and firm dynamics. *Review of Economic Studies*, 71(2), 285–315.

Allen, J., Clark, R., & Houde, J.-F. (2019). Search frictions and market power in negotiated-price markets. *Journal of Political Economy*, 127(4), 1550–1598.

Amiti, M., Kashyap, A. K., Kovner, A., & Weinstein, D. (2026). Why do firms pay different interest rates on their bank loans? Working paper.

Amiti, M. & Weinstein, D. E. (2018). How much do idiosyncratic bank shocks affect investment? evidence from matched bank-firm loan data. *Journal of Political Economy*, 126(2), 525–587.

Angbazo, L. (1997). Commercial bank net interest margins, default risk, interest-rate risk, and off-balance sheet banking. *Journal of Banking & Finance*, 21(1), 55–87.

Antras, P. & Caballero, R. J. (2009). Trade and capital flows: A financial frictions perspective. *Journal of Political Economy*, 117(4), 701–744.

Antunes, A., Gonçalves, H., & Prego, P. (2016). Firm default probabilities revisited. *Economic Bulletin and Financial Stability Report Articles*, (pp. 21–45).

Arellano, C., Bai, Y., & Zhang, J. (2012). Firm dynamics and financial development.

Journal of Monetary Economics, 59(6), 533–549.

Asker, J., Collard-Wexler, A., & De Loecker, J. (2014). Dynamic inputs and resource (mis) allocation. *Journal of Political Economy*, 122(5), 1013–1063.

Berger, A. N. & Mester, L. J. (1997). Inside the black box: What explains differences in the efficiencies of financial institutions? *Journal of Banking & Finance*, 21(7), 895–947.

Berlin, M. & Mester, L. J. (1998). On the profitability and cost of relationship lending. *Journal of Banking & Finance*, 22(6-8), 873–897.

Bernanke, B., Gertler, M., & Gilchrist, S. (1999). The financial accelerator in a quantitative business cycle framework. In J. B. Taylor & M. Woodford (Eds.), *Handbook of Macroeconomics, Volume 1, Part C* (pp. 1341–193). Elsevier.

Bertrand, J. & Burietz, A. (2023). (Loan) price and (loan officer) prejudice. *Journal of Economic Behavior & Organization*, 210, 26–42.

Berwart, E., Higgins, S., Kulkarni, S., & Truffa, S. (2025). Search and negotiation with biased beliefs in consumer credit markets. Working paper.

Bils, M., Klenow, P. J., & Ruane, C. (2021). Misallocation or mismeasurement? *Journal of Monetary Economics*, 124, S39–S56.

Boar, C. & Midrigan, V. (2023). Should we tax capital income or wealth? *American Economic Review: Insights*, 5(2), 259–274.

Bolton, P. & Scharfstein, D. S. (1996). Optimal debt structure and the number of creditors. *Journal of Political Economy*, 104(1), 1–25.

Boot, A. W., Greenbaum, S. I., & Thakor, A. V. (1993). Reputation and discretion in financial contracting. *American Economic Review*, (pp. 1165–1183).

Bordeu, O., González, G., & Sorá, M. (2025). Bank competition and investment costs across space. Working paper.

Brock, J. M. & De Haas, R. (2023). Discriminatory lending: Evidence from bankers in the lab. *American Economic Journal: Applied Economics*, 15(2), 31–68.

Buera, F. J., Kaboski, J. P., & Shin, Y. (2011). Finance and development: A tale of two sectors. *American Economic Review*, 101(5), 1964–2002.

Buera, F. J. & Moll, B. (2015). Aggregate implications of a credit crunch: The importance of heterogeneity. *American Economic Journal: Macroeconomics*, 7(3), 1–42.

Burdett, K. & Judd, K. L. (1983). Equilibrium price dispersion. *Econometrica*, 51(4), 955–969.

Cavalcanti, T., Kaboski, J. P., Martins, B., & Santos, C. (2023). Financing costs and development. Working Paper.

Cerqueiro, G., Degryse, H., & Ongena, S. (2011). Rules versus discretion in loan rate setting. *Journal of Financial Intermediation*, 20(4), 503–529.

Cooley, T. F. & Quadrini, V. (2001). Financial markets and firm dynamics. *American Economic Review*, 91(5), 1286–1310.

Corbae, D. & D'Erasmo, P. (2021). Capital buffers in a quantitative model of banking industry dynamics. *Econometrica*, 89(6), 2975–3023.

Darmouni, O. (2020). Informational frictions and the credit crunch. *Journal of Finance*, 75(4), 2055–2094.

David, J. M. & Venkateswaran, V. (2019). The sources of capital misallocation. *American Economic Review*, 109(7), 2531–2567.

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.

Dempsey, K. & Faria-e Castro, M. (2025). A quantitative analysis of bank lending relationships. *Journal of Financial Economics*, 170, 104083.

Diamond, D. W. (1984). Financial intermediation and delegated monitoring. *Review of Economic Studies*, 51(3), 393–414.

Durrani, A., Metzler, J., Michail, N., & Werner, J.-G. (2022). Bank lending rates and the remuneration for risk: evidence from portfolio and loan level data. Working paper.

Faria-e-Castro, M., Jordan-Wood, S., & Kozlowski, J. (2024a). An empirical analysis of the cost of borrowing. Working paper.

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

Federal Reserve Bank of New York (2014). *Small Business Credit Survey, Spring 2014*. Federal Reserve Bank of New York.

Giroud, X. (2013). Proximity and investment: Evidence from plant-level data. *Quarterly Journal of Economics*, 128(2), 861–915.

Gopinath, G., Kalemli-Özcan, Ş., Karabarbounis, L., & Villegas-Sanchez, C. (2017). Capital allocation and productivity in South Europe. *Quarterly Journal of Economics*, 132(4), 1915–1967.

Guvenen, F., Kambourov, G., Kuruscu, B., Ocampo, S., & Chen, D. (2023). Use it or lose it: Efficiency and redistributional effects of wealth taxation. *Quarterly Journal of Economics*, 138(2), 835–894.

Ho, T. S. & Saunders, A. (1981). The determinants of bank interest margins: Theory and empirical evidence. *Journal of Financial and Quantitative Analysis*, 16(4), 581–600.

Hsieh, C.-T. & Klenow, P. J. (2009). Misallocation and manufacturing tfp in china and india. *Quarterly Journal of Economics*, 124(4), 1403–1448.

Itskhoki, O. & Moll, B. (2019). Optimal development policies with financial frictions. *Econometrica*, 87(1), 139–173.

Jermann, U. & Quadrini, V. (2012). Macroeconomic effects of financial shocks. *American Economic Review*, 102(1), 238–271.

Khan, A. & Thomas, J. K. (2013). Credit shocks and aggregate fluctuations in an economy with production heterogeneity. *Journal of Political Economy*, 121(6), 1055–1107.

Kiyotaki, N. & Moore, J. (1997). Credit cycles. *Journal of Political Economy*, 105(2), 211–248.

Klein, M. A. (1971). A theory of the banking firm. *Journal of Money, Credit and Banking*, 3(2), 205–218.

Leibovici, F. (2021). Financial development and international trade. *Journal of Political Economy*, 129(12), 3405–3446.

Levine, R. & Rubinstein, Y. (2017). Smart and illicit: Who becomes an entrepreneur and do they earn more? *Quarterly Journal of Economics*, 132(2), 963–1018.

Mazet-Sonilhac, C. (2025). Search frictions in credit markets: Evidence from broadband expansion in france. Working paper.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, 29(2), 449–470.

Midrigan, V. & Xu, D. Y. (2014). Finance and misallocation: Evidence from plant-level data. *American Economic Review*, 104(2), 422–458.

Moll, B. (2014). Productivity losses from financial frictions: Can self-financing undo capital misallocation? *American Economic Review*, 104(10), 3186–3221.

Monti, M. (1972). *Deposit, credit and interest rate determination under alternative bank objective function*. North-Holland/American Elsevier.

Ottonello, P. & Winberry, T. (2020). Financial heterogeneity and the investment channel of monetary policy. *Econometrica*, 88(6), 2473–2502.

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

Petersen, M. A. & Rajan, R. G. (1995). The effect of credit market competition on lending relationships. *Quarterly Journal of Economics*, 110(2), 407–443.

Santos, J. A. (2011). Bank corporate loan pricing following the subprime crisis. *The Review of Financial Studies*, 24(6), 1916–1943.

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

Sterk, V., Sedláček, P., & Pugsley, B. (2021). The nature of firm growth. *American Economic Review*, 111(2), 547–579.

Online Appendix

Beyond Risk: Firm Financing and Interest Rates

Contents

A Additional data details **1**
- A.1 Sample construction and composition . . . 1
- A.2 Interest rates vs. spreads . . . 2
- A.3 Descriptive statistics for sample of all instruments . . . 3

B Additional details for Section 3 **3**
- B.1 Controls for loan characteristics . . . 3
- B.2 Analysis by collateral type for Section 3.1 . . . 6
- B.3 Additional details for Equation (3) . . . 6
- B.4 Validation of PDs . . . 7
- B.5 PD and LGD distributions that would be needed to explain rates . . . 7
- B.6 Extension to longer maturity loans . . . 9
- B.7 Additional details for other robustness exercises in Section 3.2 . . . 10

C Additional details for Section 4 **11**
- C.1 Distribution of loans across banks and firms . . . 11
- C.2 Robustness exercises for Section 4.1 . . . 13
- C.3 Robustness exercises for Section 4.2 . . . 13

D Additional details for Section 5 **14**
- D.1 Definitions of firm control variables . . . 14
- D.2 All credit instruments . . . 16
- D.3 Ex ante vs. ex post heterogeneity . . . 17

E Additional details for Section 6 **20**
- E.1 Firm location controls at the 5-digit postcode level . . . 20
- E.2 Bank-location fixed effects . . . 20
- E.3 Pricing loan portfolios . . . 20
- E.4 Firm size and public vs. private firm tests . . . 21
- E.5 Demand shocks . . . 23
- E.6 Misallocation exercise . . . 24

A Additional data details

A.1 Sample construction and composition

Figure A.1 presents within-bank distributions of loans across geographic areas, industries and firm size groups for the five largest banks. This evidence shows that these banks have similar operations, and are not specialized or dividing the market.

The left panel of Table A.1 compares the size distribution of our sample of firms in 2019 with the size distribution of the universe of Portuguese firms in the same year.[1] The results show that firms with loans (our sample) are selected from larger firms in the population. To compute the shares of employer firms, total employment and aggregate output that are attributable to firms in our sample (presented in the main text), we use aggregate values from Eurostat.

When we merge the credit registry with the firm data, some credit instruments are dropped because they are issued to firms outside our data. These account for 12% of firm financing loans and 11% of the value of these loans. For the sample of all credit instruments, the share that is lost is 22% (32% by value).

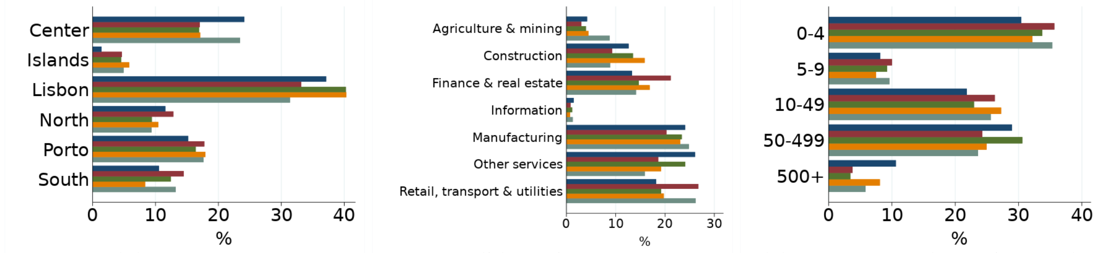
We also compare the firm size distribution of our sample with that of employer firms in the U.S. in 2019 (right panel of Table A.1).[2] The distributions for the number of firms in each size category are similar. However, the firms in the biggest firms in the US are larger than in Portugal, with firms more than 500 employees accounting for 52.9% of employment compared to 22.4% in our sample. The average firm size is 11.3 employees in our sample and 24.7 employees for U.S. employer firms.

For the average share of banking sector assets held by the largest five banks for Euro Area 19 countries, and the average share of credit to non-financial corporations that came from banks in the Euro Area 19, the data sources are the World Bank's Global Financial Development Indicators and the ECB, respectively.[3]

[1] A firm is included in this exercise if it is in our sample and it has at least one active firm financing loan at some point in 2019.

[2] The U.S. data is from the Business Dynamics Statistics of the Census Bureau.

[3] The World Bank data series is called "five-bank asset concentration," code GFDD.OI.06. From the ECB, we use: loans to non-financial corporations (codes QSA.Q.N.PT.W0.S11.S1.N.L.LE.F4.T.Z.XDC.



(a) Regions (b) Industry (c) Firm size (employees)

Figure A.1: **Distribution of firm financing loans across regions, industries and firm size, by bank.** This figure presents moments of the distribution of firm financing loans by value in December 2019 for each of the five major banks. Panels (a), (b) and (c) show the distributions across regions, industries and firm size categories. Each color is a different bank. Firm size is measured with the number of employees.

Size	Sample	Portugal	Size	Sample	USA
0 employees	12.3%	69.8%	1 – 4 employees	57.5%	58.6%
1 – 9 employees	68.8%	26.2%	5 – 9 employees	20.9%	18.0%
10 – 49 employees	16.0%	3.4%	10 – 19 employees	11.5%	11.4%
50 – 249 employees	2.6%	0.5%	20 – 99 employees	8.6%	10.0%
>250 employees	0.4%	0.1%	100 – 499 employees	1.3%	1.7%
			>500 employees	0.2%	0.4%

Table A.1: **Firm size distributions: sample, Portugal and U.S.** Distribution of firms across firm size categories in 2019. The sample is all firms with firm financing loans in 2019. The left side comparison is to the universe of firms in Portugal (data from Statistics Portugal, Integrated business accounts system). The right side comparison is between employer firms in the sample and the universe of employer firms in the US (data are from Business Dynamics Statistics).

A.2 Interest rates vs. spreads

We analyze interest rates instead of spreads, since spreads are not reported for all loans. This could be important if it means that we neglect heterogeneity in banks' cost of funds. We assess whether this omission is material using the 78% of loans for which spreads are reported. For each loan l extended by bank b at time t, we define the risk free component of the interest rate as $\tilde{r}_{lbt} \equiv r_{lbt} - s_{lbt}$, where r_{lbt} denotes the contractual interest rate and s_{lbt} is the reported spread. To purge these of variation due to maturity and issuance date, we regress them on maturity-month-year fixed effects,

T.S.V.N.T for Portugal and QSA.Q.N.I8.W0.S11.S1.N.L.LE.F4.T.Z.XDC.T.S.V.N.T for the Euro Area); and debt securities and loans of non-financial corporations (codes QSA.Q.N.PT.W0.S11.S1.N.L.LE. F3T4.T. Z.XDC.T.S.V.N.T for Portugal and QSA.Q.N.I8.W0.S11.S1.N.L.LE.F3T4.T. Z.XDC.T.S.V.N.T for the Euro Area).

denoted by $\mathcal{M}_{lt}$, and study the residuals. These are tightly concentrated around zero. Eighty percent of loans have a residual of approximately zero, and 90% of loans have a residual of less than or equal to one basis point. Thus, cross-bank heterogeneity in funding costs is quantitatively small and should not affect the analysis.

A.3 Descriptive statistics for sample of all instruments

Section 2 of the main text outlines the share of credit due to each type of credit instrument in the data. Figure A.2 replicates Figure 1 from the main text, now including all credit instruments.[4] The main changes are that these other types of credit are for lower values on average (panel b) and have a shorter maturity (panel c). This is not surprising as credit lines and credit cards are mostly short term loans with smaller amounts. Summary statistics for the sample of firms when all credit instruments are considered are in Table A.2.[5] This sample has approximately 100,000 more firms than the baseline, and on average firms are now slightly smaller and have more loans.

B Additional details for Section 3

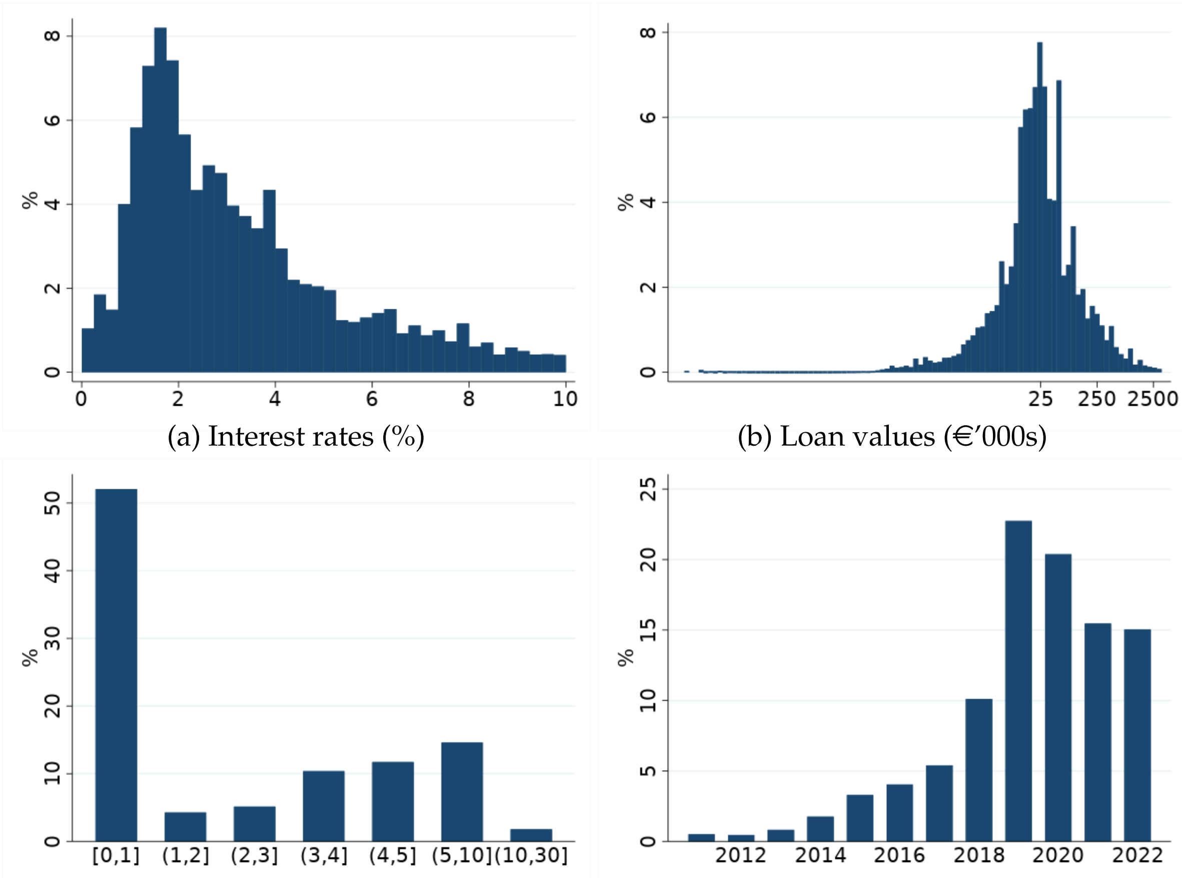
B.1 Controls for loan characteristics

The loan characteristics controls are comprised of maturity-time fixed effects (described in the main text) and additional characteristics in $\mathbf{L}_l$. The details of the elements of $\mathbf{L}_l$ are as follows. With the exception of loan amount, all variables are categorical and are controlled for with a set of dummy variables. We provide the distribution of the baseline sample across the categories of each characteristic.

- *Loan value.* Log of the value of the loan in Euros.
- *Frequency of rate adjustment.* Loans can be fixed rate (52.4% of loans) or can be adjusted at a daily (1.6%), monthly (13.4%), quarterly (6.8%), biannual (7.6%) or annual (8.8%) frequencies, at other frequencies less than a year (0.03%), at frequencies

[4]The distributions of collateral rates and collateral types are very similar to in Figure 1, so these are omitted here, but available on request.

[5]There are firms without firm financing loans that have other credit instruments. Thus the sets of firms differ across the two samples.



(a) Interest rates (%) (b) Loan values (€'000s)

(c) Maturity (years) (d) Origination year

Figure A.2: **Distributions of loan characteristics at origination: all credit instruments** Panels (a), (b), (c) and (d) are the distributions of interest rates, loan values (in thousands of Euros), maturity (in years), and origination years, respectively. The scale for loan values in panel (b) is in log units.

greater than one year (0.02%), or at the lender's discretion (0.1%). There is also a category for all residual cases (9.2%).

- *Type of collateral.* Collateral are classified as guarantees (29.4%), financial assets (7.5%), real estate (2.7%), other collateral of real nature (2.4%) and a category for all residual cases (5.8%). Fifty-two per cent of loans have no collateral.

- *Number of collateral assets.* Among collateralized loans, 75.7% have one asset pledged, 14.6% have two, 4.8% have three, 2.2% have four and 2.8% have five or more.[6]

- *Amortization type.* The amortization method can be French (fixed installments with interest paid at the end of each period—23.9% of loans), German (fixed installments with interest paid in advance of each period—0.2%), fixed principal pay-

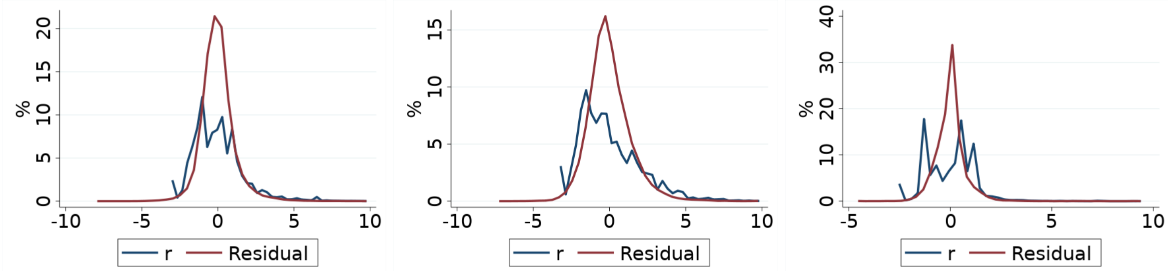
[6]We use the term "assets" broadly here to refer to all types of collateral.

	Average	p25	p50	p75	p90	St. dev.	# firms
Loans value (€'000s)	888.85	8.26	63.40	273.67	1092.77	(9017.03)	
# of loans	12.91	3.00	6.00	12.00	26.00	(87.48)	
Av. loan (€'000s)	56.56	1.96	10.45	25.58	70.76	(603.59)	
% collateralized	49	12	50	84	100	(37)	
# of banks	1.50	1.00	1.00	2.00	3.00	(0.99)	252,956
Employees	10.83	1.00	3.00	7.00	16.50	(114.21)	
Total assets (€'000s)	2,022.60	65.63	186.41	573.86	1896.84	(55506.16)	
Revenue (€'000s)	1,442.11	60.65	165.83	485.69	1499.35	(29007.92)	
Leverage (%)	130	44	70	95	164	(355)	
Age	14.21	4.50	10.50	20.00	31.00	(13.00)	

Table A.2: **Descriptive statistics for firms: all credit instruments.** This table provides statistics for firms in the sample. For each moment, four percentiles and the standard deviation are given. All moments are at the firm level and are averaged within firm over the years that each firm is in the sample. So, for example, the value of loans is the total value of a firm's loans, % collateralized is the share of a firm's loans that are collateralized, and # of banks is the number of banks that a firm has loans from. Value of loans, total assets and revenue are in thousands of 2016 Euros. All credit instruments are included

ments (6.6%), all principal paid at maturity (25.5%), increasing installments (0.01%), and a category for all residual cases (43.4%).

- *Purpose of the loan.* Loans can be for car purchases (14.5% of loans), liquidity management (13.8%), exports and imports (4.2%), equipment and other goods purchases (1.8%), real estate and land (2.6%) or no specific purpose (63.2%).
- *Frequency of repayment.* Loans can be repaid with monthly (42.5% of loans), quarterly (5.0%), biannual (0.4%) or annual (0.5%) installments, with a single principal payment (3.9%), with a single principal and interest payment (16.3%), and there is a category for all residual cases (31.3%).
- *Right to immediate repayment.* For 16.2% of loans, the bank has the right to demand immediate repayment of the loan.
- *Securitized loan.* 97.7% of loans are securitized by banks.
- *Loan special characteristics.* The controls include dummy variables for whether loans were granted under Covid-related policies including guarantees (0.6%), are covered by consumer credit law (1.4%), or are a syndicated loans, related to real estate, or are involved in a default process (0.3%). The last three categories are grouped together as they include small numbers of loans. Loans can have more than one of these characteristics, but 98.2% of loans have none of them.



(a) Real and financial assets (b) Guarantees (c) Other types of collateral

Figure B.1: **Interest rate distributions before and after controlling for loan characteristics, by collateral type.** Panel (a), (b) and (c) present the distribution of interest rates for fully collateralized loans backed with real and financial assets, guarantees and other types of collateral, respectively. In both panels the blue line is the unconditional distribution, and the red line is the distribution after controlling for loan characteristics (i.e. the residuals from equation 1). All distributions are centered with their means at zero.

B.2 Analysis by collateral type for Section 3.1

For the robustness exercises discussed at the end of Section 3.1, Figure B.1 plots the distribution of interest rates and of the residuals from equation (1) for each class of collateral. For all cases there is substantial residual dispersion, with more than 41% of the variation in rates remaining unexplained in each case.

B.3 Additional details for Equation (3)

Equation (2) in the main text can be written as

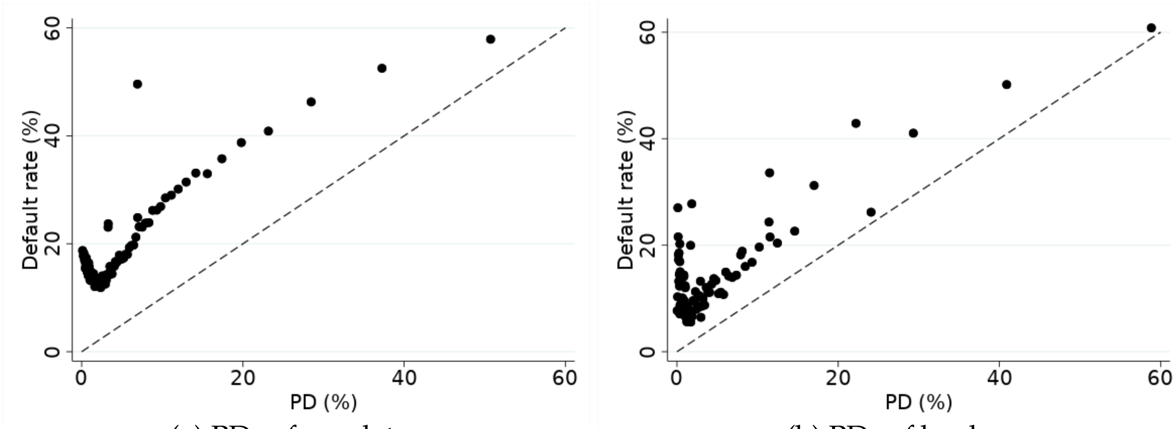
$$Q_{lt} = \sum_{s_{l,t+1} \in \mathcal{S}_{l,t+1}} p(s_{l,t+1}) \Lambda_{t+1}(s_{l,t+1}) \left(1 - \chi(s_{l,t+1}) LGD(s_{l,t+1})\right),$$

where $\mathcal{S}_{l,t+1}$ is the set of possible states in the next period, and $p(s_{l,t+1})$ is the probability of a state $s_{l,t+1} \in \mathcal{S}_{l,t+1}$.[7] For the case of a risk free lender, which is used as the baseline, this can be expressed as equation (3), with

$$PD_{lt} \equiv \sum_{s_{l,t+1} \in \mathcal{S}_{l,t+1}} p(s_{l,t+1}) \chi(s_{l,t+1}),$$

$$LGD_{lt} \equiv \frac{\sum_{s_{l,t+1} \in \mathcal{S}_t} p(s_{l,t+1}) \chi(s_{l,t+1}) LGD(s_{l,t+1})}{PD_{lt}}.$$

[7]Note that the stochastic discount factor is common across loans, so it only depends on the aggregate elements of the state $s_{l,t+1}$, and not on those that are idiosyncratic to loan l.



(a) PDs of regulator (b) PDs of banks

Figure B.2: **Estimated vs. realized PDs.** In panel (a) the PDs from the regulator are ranked and sorted into 100 equally sized bins, with each bin assigned the average value of its elements. The vertical axis provides the share of loans in each bin that default. Panel (b) repeats the exercise for PDs reported by the banks. The dashed lines are 45° lines along which PDs equal ex post default rates.

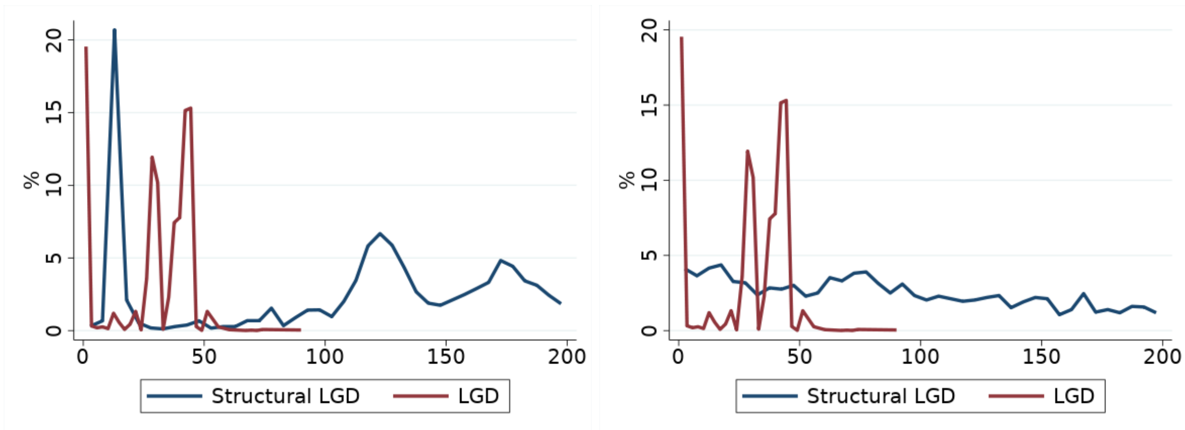
B.4 Validation of PDs

To check the quality of PD measurement, we use the fact that we observe loans through to completion, so that we can observe defaults. We take the PD values from the data, rank them from lowest to highest, group them into 100 evenly sized bins based on this ranking, and then compare the values of the average PDs and ex post realized default rates within each bin. The results are presented in Figure B.2 and show that the PDs are high quality. Realized default rates are monotonically increasing in the ex ante estimates and quantitatively they closely track each other.

B.5 PD and LGD distributions that would be needed to explain rates

To evaluate what the LGD and PD distributions would need to be for risk to fully explain the interest rate distribution, we start by estimating equation (4), imposing that the coefficient on the risk term equals minus one, in accordance with the theory. Given the estimated coefficients, we then ask what values the LGDs would need to take, taking the PDs as given, to fit observed interest rates perfectly. That is,

$$-\log(1 - PD_{lt}LGD_{lt}) = \hat{\varepsilon}_{lt},$$



(a) PDs from the regulator.

(b) PDs from banks.

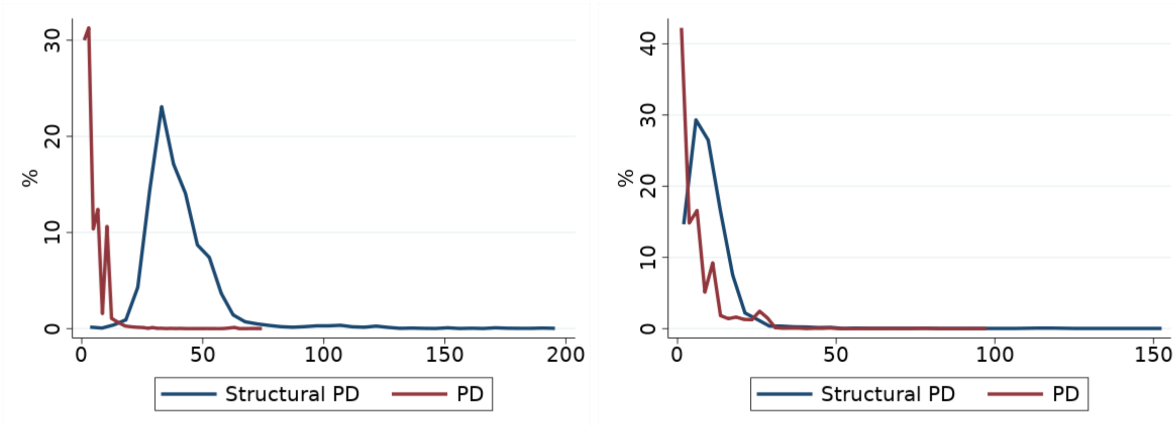
Figure B.3: **LGDs: actual vs. required to match interest rate distribution.** Probability density functions of the LGDs required to match the interest rate distribution (blue), and the LGDs reported in the data (red). In panels (a) and (b), the PDs used for the estimation are from the regulator and the banks, respectively.

where $\hat{\varepsilon}_{lt}$ is the estimated value of ϵ_{lt}.[8] We repeat this for PDs, taking LGDs as given.

Figure B.3 presents results for LGDs. The left panel is the LGD distribution when we use the PDs from the regulator, and the right panel uses the PDs from the banks. The distributions (in blue) are compared to the LGD distribution from the data (in red). In both panels there is a large mass of LGDs that would need to be above 100%, meaning that a bank looses more than the value of the loan when default occurs. In fact, the median LGD values are 545% and 129% in panels (a) and (b), respectively (in both cases the x-axis is capped at 200% for clarity).

For PDs, the results are presented with the same format in Figure B.4. Panel (a) has the estimated and observed PDs for all loans with LGD estimates, and panel (b) has the estimated and observed PDs for all loans with LGDs and PDs reported by banks. Thirty to forty percent of PDs in the data are close to 0% (depending on the measure used), but the median PD needed to match interest rates is 40% in panel (a). Additionally, the estimated PDs have an average over 100%. In panel (b), the average PD needed to fit rates is an order of magnitude larger than in the data.

[8]One potential problem with this approach is that, by construction, ϵ_{lt} is centered around zero, which could force LGD and PD into negative values as well. To go around this, we add to ϵ_{lt} the constant estimated from equation (4).



(a) PDs from the regulator.

(b) PDs from banks.

Figure B.4: **PDs: actual vs. required to match interest rate distribution.** Probability density functions of the PDs required to match the interest rate distribution (blue), and the PDs reported in the data (red). Panel (a) uses all loans for which LGDs are reported in the data and PDs from the regulator. Panel (b) uses all loans for which PDs and LGDs are reported by banks, and the PDs from the banks.

B.6 Extension to longer maturity loans

To extend the analysis in Section 3.2 to loans with maturity greater than one year, we use equation (6). To derive this, consider a T period loan l, issued at time t, for which all principal and interest are paid in the final period. Let a history of states from time t to $t+T$ be denoted s_{lt}^T, the set of possible such histories be $\mathcal{S}_{lt}^T$, r_{ltT} be the per period interest rate on the loan and $\tilde{r_{tT}}$ be the per period risk free rate. The lender is risk neutral. The interest rate on the loan satisfies

$$1 = \sum_{s_{lt}^T \in \mathcal{S}_{lt}^T} p(s_{lt}^T) \frac{(1+r_{ltT})^T}{(1+\tilde{r_{tT}})^T} (1 - \chi(s_{lt}^T) LGD(s_{lt}^T)),$$

where $p(s_{lt}^T) \in [0,1]$ is the probability of observing the history of states $s_{lt}^T \in \mathcal{S}_{lt}^T$, and $\chi(s_{lt}^T)$ and $LGD(s_{lt}^T)$ are a default indicator and the loss given default, respectively, for this history of states. Equation (6) in the main text follows from this, with PD_{ltT} and LGD_{ltT} defined as:

$$LGD_{ltT} \equiv \frac{\sum_{s_{lt}^T \in \mathcal{S}_{lt}^T} p(s^T) \chi(s_{lt}^T) LGD(s_{lt}^T)}{\sum_{s_{lt}^T \in \mathcal{S}_{lt}^T} p(s_{lt}^T) \chi(s_{lt}^T)}, \quad PD_{ltT} \equiv \sum_{s_{lt}^T \in \mathcal{S}_{lt}^T} p(s_{lt}^T) \chi(s_{lt}^T).$$

The regression for taking equation (6) to the data is

$$r_{ltT} = \beta_1 \frac{1}{T} \log(1 - PD_{ltT} LGD_{ltT}) + \beta_2 \mathbf{L}_l + \mathcal{M}_{ltT} + \varepsilon_{ltT}.$$

	Default indicator
Residuals	0.558***
	(0.153)
R^2	0.001
# Obs.	25,460

Table B.1: **Explanatory power of interest rate residuals for default.** Standard errors are in parentheses. *** denotes $p < 0.01$, ** denotes $p < 0.05$, and * denotes $p < 0.1$.

B.7 Additional details for other robustness exercises in Section 3.2

Bank PDs instead of regulator PDs Figure B.2 shows that the PDs from the regulator and the banks have similar relationships with ex post default, so it is not surprising the changing between these measures does not affect results. The sample size for this analysis with bank PDs is 31,458 loans because banks do not report default risk variables for all loans.[9]

Predictive power of interest rate residuals for default The results the regression of interest rate residuals on a default indicator, as described in the main text, are reported in Table B.1. While the coefficient of interest is positive and significant, indicating that banks are pricing some additional risk that is not captured by PDs and LGDs, quantitatively the effect is very small.

Alternative stochastic discount factor The specification for the analysis that controls for the correlation between the revenue of a firm and aggregate consumption is

$$r_{lt} = \beta_1 \log(1 - PD_{lt}LGD_{lt}) + \beta_2 \Lambda_f \log(1 - PD_{lt}LGD_{lt}) + \beta_3 \Lambda_f + \beta_4 \mathbf{L}_l + \mathcal{M}_{lt} + \varepsilon_{lt},$$

where Λ_f is the correlation between firm f's revenue and aggregate consumption during the period that the firm is in the sample.

Semi-parametric specification The semi-parametric specification that allows for a more flexible relationship between default risk and interest rates is:

$$r_{lt} = \sum_{p \in \mathcal{PD}} \gamma_p PD_{lt} \mathbf{1}_{i \in \mathcal{S}^{(P)}} + \sum_{d \in \mathcal{LGD}} \theta_d LGD_{lt} \mathbf{1}_{i \in \mathcal{S}^{(D)}} + \sum_{d \in \mathcal{LGD}} \sum_{p \in \mathcal{PD}} \lambda_{dp} PD_{lt} LGD_{lt} \mathbf{1}_{i \in \mathcal{S}^{(P)}} \mathbf{1}_{i \in \mathcal{S}^{(D)}}$$

[9]Banks report PDs for 89,896 loans in the sample, but not all of these observations have an LGD, and this analysis requires both.

$$+ \sum_{p \in \mathcal{PD}} \sigma_p PD_{lt}^2 \mathbf{1}_{i \in \mathcal{S}^{(P)}} + \sum_{d \in \mathcal{LGD}} \alpha_d LGD_{lt}^2 \mathbf{1}_{i \in \mathcal{S}^{(D)}} + \sum_{d \in \mathcal{LGD}} \sum_{p \in \mathcal{PD}} \mu_{dp} PD_{lt}^2 LGD_{lt}^2 \mathbf{1}_{i \in \mathcal{S}^{(P)}} \mathbf{1}_{i \in \mathcal{S}^{(D)}}$$
$$+ \beta_1 \mathbf{L}_l + \mathcal{M}_{lt} + \varepsilon_{lt}.$$

The terms $\mathcal{S}^{(P)}$ and $\mathcal{S}^{(D)}$ represent the p^{th} PD and d^{th} LGD group, respectively. We partition the sample into ten deciles for each of the PD and LGD distributions to capture potential nonlinear relationships between the risk terms and interest rates. We additionally include the squared terms of both PD and LGD variables to further account for potential nonlinear relations.

Effect of selective LGD reporting This analysis, explained in the main text, uses the following version of the non-parametric specification which includes terms for PDs, but not LGDs:

$$r_{lt} = \sum_{p \in \mathcal{PD}} \gamma_p PD_{lt} \mathbf{1}_{i \in \mathcal{S}^{(P)}} + \sum_{p \in \mathcal{PD}} \sigma_p PD_{lt}^2 \mathbf{1}_{i \in \mathcal{S}^{(P)}} + \beta_1 \mathbf{L}_l + \mathcal{M}_{lt} + \varepsilon_{lt}. \tag{B.1}$$

Amortization structure The main text outlines subsample analysis for the three most common amortization categories: all principal paid at maturity, French amortization, and the residual category with assorted amortization structures.[10]

Exposure at default For assessing the impact of variation in expected exposure at default on the results, we use the regression

$$r_{lt} = \beta_1 \log(1 - PD_{lt} LGD_{lt} VE_{lt}) + \beta_2 \mathbf{L}_l + \mathcal{M}_{lt} + \varepsilon_{lt},$$

where VE_{lt} is expected exposure at default, as a share of the initial loan value.

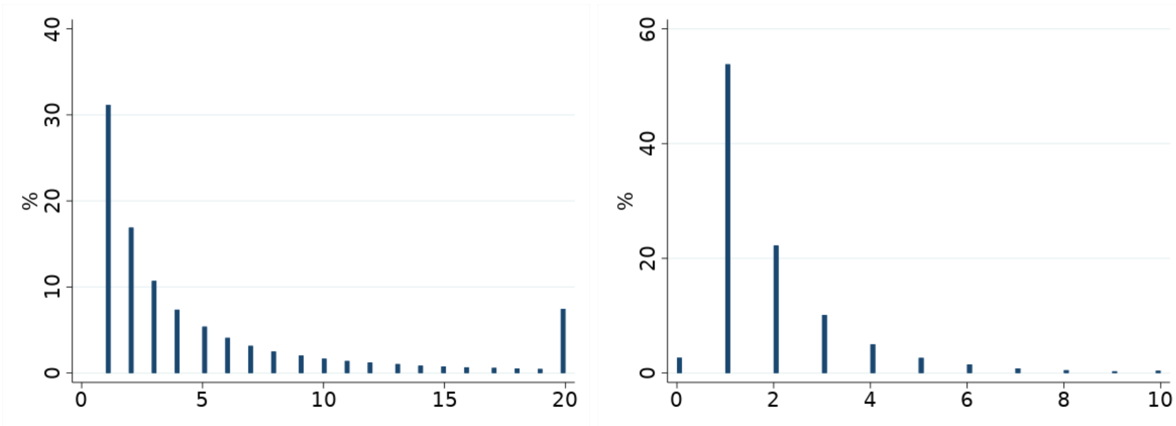
C Additional details for Section 4

C.1 Distribution of loans across banks and firms

Figure C.1 presents the distribution of number of loans and number of bank relationships that each firm has in the data. All loans and bank relationships over the duration of the data are included. The majority of firms have a small number of loans issued by a small number of banks. There is then a right tail, indicating that some firms

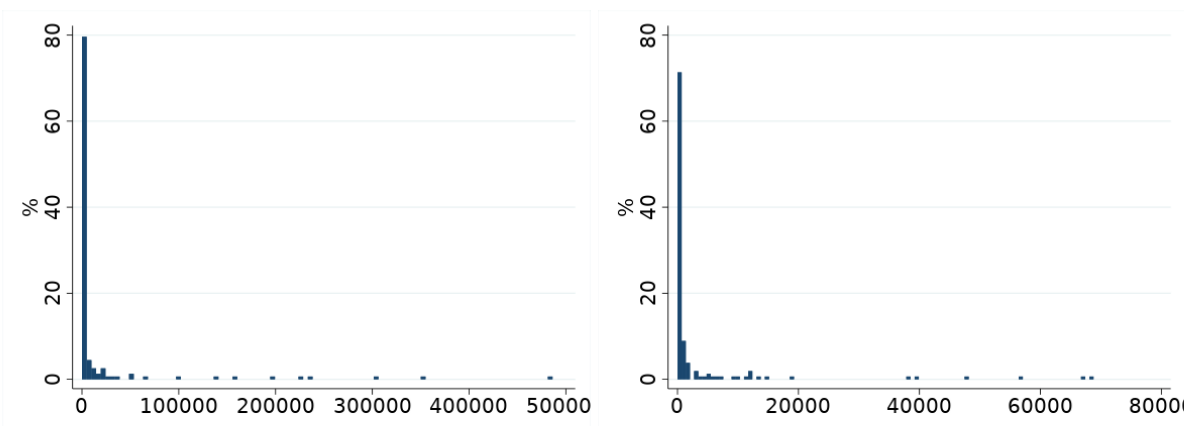
[10] Recall that these subsamples account for 28.6%, 20.9% and 45% of loans, respectively.

have many loans originating from multiple banks. Figure C.2 plots the distribution of number of loans and firm relationships for each bank. In contrast to the corresponding firm-level distributions, the majority of banks issue thousands of loans and have relationships with thousands of firms.



(a) Loans (b) Banks

Figure C.1: **Loan and bank relationship distributions across firms.** Panels (a) and (b) present the distributions of number of loans by firm and the number of bank relationships by firm. In panel (a), all firms with 20 loans or more are grouped in the last bar. In panel (b), all firms with 10 banks or more are grouped into the last bar.



(a) Loans (b) Firms

Figure C.2: **Loan and firm relationship distributions across banks.** On the left hand side the distribution of number of loans by bank over time. On the right hand side the distribution of number of firm relationships by bank over time.

C.2 Robustness exercises for Section 4.1

The main text discusses robustness exercises that extend the analysis to use bank-time and firm-time fixed effects instead of bank and firm fixed effects;[11] use the sample of all credit instruments;[12] and test whether the results hold for firms with at least 10 loans,[13] across amortization structures, when loans issued during the Covid period (from January 2020 onward) are excluded, and across firm age and size groups. The results are presented in Tables C.1 and C.2. The contributions of bank (or bank-time) fixed effects to interest rate variation range from 5.8% to 15.1% across these exercises, in line with the baseline results. Also, firm fixed effects (or firm-time fixed effects) are a large source of rate variation across specifications, with a minimum of 26.4% and a maximum of 61.0%. They are particularly important under the firm-time specification, and for large firms.

C.3 Robustness exercises for Section 4.2

To check the robustness of results for Section 4.2 we redo the analysis on the sample of all credit instruments, and also add a time dimension to the firm, bank and firm-bank fixed effects.[14] The frequency of the time dimension is one year. The results are presented in Table C.3. Relationship banking accounts of 17.6% of rate variation in this sample under the baseline specification, and adding a time dimension to the fixed effects barely changes this. Instead, it increases the importance of the firm effects.

[11] To estimate this specification, we need firms that originate multiple loans in the same time period. The number of firms in the sample decreases from 84,535 to 41,312.

[12] For this exercise we add instrument type as a fixed effect in the vector $\mathbf{L}_{lt}$. The number of firms in the sample increases from 84,535 to 135,350 and the number of observations from 521,856 to 1,598,466. This strengthens the identification of the fixed effects, as we now have almost twice as many observations per firm and more within firm variation.

[13] In total, 6,955 firms satisfy this restriction. This increases the number of loans per firm from six to 36.

[14] When using all instruments, instrument types are included as fixed effects in $\mathbf{L}_{lt}$. The number of firms in the sample increases from 73,858 to 115,673 and the number of observations goes from 463,624 to 1,469,225. With time added to the fixed effects, this sample decreases to 1,159,644.

	Contribution						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Default risk	0.000	0.003	0.002	0.006	0.002	0.001	0.002
(2) Loan characteristics	0.129	0.171	0.329	0.137	0.180	0.245	0.216
(3) Bank fixed effects	0.095	0.084	0.097	0.101	0.063	0.093	0.092
(4) Firm fixed effects	0.610	0.415	0.373	0.391	0.542	0.400	0.412
(5) Residual	0.171	0.283	0.253	0.258	0.214	0.166	0.142
(6) Covariances	−0.004	0.047	−0.051	0.114	0.001	0.096	0.138
# Obs.	336,098	251,847	1,598,466	248,490	89,488	56,823	163,153

Table C.1: **Contributions of risk, loan characteristics, firm and bank fixed effects to interest rate variation: robustness exercises.** Each row is the share of the total variance in interest rates due to each component of regression (7), calculated with equation (5). In sequential order, the columns contain results for: (1) changing the bank and firm fixed effects to bank-time and firm-time fixed effects; (2) firms with 10 loans or more; (3) all instruments; (4) pre-2020 loans; (5) loans with all principal paid at maturity; (6) loans with French amortization; and (7) loans with all other amortization structures.

	Contribution							
	Young	Mature	Old	S	S-M	M	M-L	L
(1) Default risk	0.002	0.007	0.005	0.004	0.006	0.005	0.004	0.017
(2) Loan characteristics	0.373	0.291	0.225	0.346	0.315	0.284	0.250	0.176
(3) Bank fixed effects	0.151	0.120	0.079	0.147	0.135	0.112	0.080	0.058
(4) Firm fixed effects	0.288	0.310	0.384	0.264	0.277	0.325	0.393	0.588
(5) Residual	0.204	0.256	0.283	0.228	0.271	0.296	0.230	0.266
(6) Covariances	-0.017	0.023	0.029	0.015	0.002	-0.015	0.047	-0.088
# Obs.	72,390	151,635	273,319	137,468	115,493	119,218	73,346	54,023

Table C.2: **Contributions of risk, loan characteristics, firm and bank fixed effects to interest rate variation: robustness exercises for firm age and size.** Each row is the share of the total variance in interest rates due to each component of regression (7), calculated with equation (5). Young, mature and old firms are less than 5, 5–15, and more than 15 years old, respectively. Small (S), small-medium (S-M), medium (M), medium-large (M-L) and large (L) firms have less than 5, 5–12, 13–40, 41–120, and more than 120 employees, respectively.

D Additional details for Section 5

D.1 Definitions of firm control variables

This section provides definitions of firm-level control variables used in the analysis. Variables measured in monetary units are deflated using the GDP deflator. The following variables are specified in log units in the regression: capital, number of employees,

	Contribution	
(1) Bank effects	0.079	0.087
(2) Firm effects	0.381	0.535
(3) Firm-bank relationship effects	0.176	0.164
Time variation in fixed effects	No	Yes
# Obs.	1,469,225	1,159,644

Table C.3: **Contributions of bank, firm, and relationship effects to interest rate variation: all credit instruments** The values are computed by estimating equation (**??**), and then estimating the same equation replacing firm-bank-time fixed effects with firm-time and bank-time fixed effects on the same sample. For each estimation we decompose the interest rate variance using equation (5). The contributions of the bank effects and the firm effects are the estimated contributions from the second stage. The firm-bank relationship effects contribution is the difference between the contribution of firm-bank fixed effects from the first stage, and the total contribution of firm and bank fixed effects from the second stage. The sample of all credit instruments is used. In the second column, the fixed effects are changed to bank-time, firm-time and firm-bank-time, with time variation at a yearly frequency.

revenue, wage bill, TFP, financial assets, total assets, total liabilities, and the number of R&D employees. For inventories, there is an indicator variable for values of zero, and otherwise it is in log units.

Time-varying characteristics

- *Capital stock.* Book value of fixed tangible assets.
- *Number of employees.* Total number of employees.
- *Wage bill.* Total labor costs.
- *Revenue.* Total operating revenue (turnover).
- *Revenue growth rate.* Log difference of revenue from previous year.
- *Profit share.* Operating profits as a share of total revenue.
- *Total factor productivity (TFP).* Firm-level productivity estimated with the Ackerberg et al. (2015) method, based on capital, intermediate inputs, and labor costs.
- *Investment rate.* Annual change in fixed tangible assets divided by the previous year's level of fixed tangible assets.
- *Inventories.* Value of inventories held.
- *Equity, financial assets, total assets and total liabilities.* Values reported in balance sheet.
- *Past default indicator.* Indicator equal to one if the firm has had overdue credit exceeding 2.5% of total credit for three consecutive months between September 2018 and the date at which the loan was issued.[15]
- *Past late repayment indicator.* Indicator equal to one if the firm has been at least 30 days late on a loan repayment between September 2018 and the date at which the

[15] The starting date is September 2018, because this is the starting date of the dataset. This applies to the past late repayment indicator as well.

loan was issued.
- *Trade payables over total assets.* Trade credit from suppliers relative to total assets.
- *Trade receivables over total assets.* Trade credit to customers relative to total assets.

Time-invariant characteristics

- *Industry.* Five-digit industry classification according to the Portuguese Classification of Economic Activities. There are 833 such industries.
- *Location.* Firm headquarter location measured with four-digit postcodes.
- *Number of local banks (top 5).* Number of the five largest banks that have branches located in a firm's five and seven-digit postcode areas (two variables).
- *Number of local banks (other banks).* Number of other banks with branches located in a firm's five and seven-digit postcode areas (two variables).
- *Number of establishments.* Maximum number of establishments operated by the firm during the sample period.
- *Home country.* Country of origin of the firm.
- *Legal form.* Legal structure of firm. The options are: private limited liability company (87% of firms), public limited company (6%) and other legal structures.[16]
- *Number of R&D employees.* Maximum number of employees engaged in research and development activities during the sample period.
- *Exporter indicator.* Indicator equal to one if the firm reports having exports in any year in the sample period.
- *Number of foreign establishments.* Maximum number of foreign establishments operated by the firm during the sample period.
- *Birth year indicators.* Indicator variable for the founding years of firms. Firms born before 1943 are grouped together because for these years there are less than 50 firms founded per year in our sample.
- *Revenue growth standard deviation.* Standard deviation of annual log revenue change.
- *Revenue growth skewness.* Skewness of annual log revenue change.
- *TFP growth standard deviation.* Standard deviation of the annual log TFP change.
- *TFP growth skewness.* Skewness of the annual log TFP change.

D.2 All credit instruments

The results for the Section 5 analysis on the sample of all credit instruments are reported in Table D.1.

[16] Aside from private limited liability and public limited companies, all other legal categories have only a small share of firms (a maximum of 0.35%), so we group them together along with firms whose legal structure is not reported (5.4%).

	Contribution
(1) Firm fixed effects	0.488
(1a) Firm fixed characteristics	0.036
(1b) Latent firm fixed effects	0.452
# Firms	89,766

Table D.1: **Contributions of firm fixed effects and time-invariant observable firm characteristics: all credit instruments.** Each row is the share of the total variance in interest rates due to due select components of regressions (9) and (10), calculated with equation (5).

D.3 Ex ante vs. ex post heterogeneity

We use the statistical methodology of Sterk et al. (2021) to assess whether the persistent heterogeneity in borrowing costs that we document in the data is already present when firms first enter the credit market, or instead emerges through shocks realized later in their borrowing histories.[17] These persistent differences arise after controlling for loan characteristics, default risk, bank heterogeneity and relationship banking, so we use residualized interest rates that remove the effects of these factors.[18] Consider a firm's sequence of interest rates, using $\check{r}_{f,n}$ to denote the residualized interest rate on the n^{th} loan of firm f. The Sterk et al. (2021) methodology decomposes such a sequence with

$$\check{r}_{f,n} = \underbrace{u_{f,n} + v_{f,n}}_{\text{ex ante component}} + \underbrace{w_{f,n} + z_{f,n}}_{\text{ex post component}} . \tag{D.1}$$

$u_{f,n}$ and $v_{f,n}$ are random variables that are realized before the first loan is issued, capturing the effect and any factor affecting a firm's interest rates that predated its first loan. $w_{f,n}$ and $z_{f,n}$ are AR(1) and i.i.d. stochastic processes, with new shocks realized when each loan is issued. These capture the effects of factors arising from the time of first loan issuance onward on rates. See equation (1) in Sterk et al. (2021) for the full details.

To estimate the interest rate process we follow the approach of Sterk et al. (2021)

[17] The original application is to employment, while here study interest rates.

[18] Specifically, we estimate a simplified version of equation (9) that omits the firm fixed effects and the controls for firm observables in $\mathbf{F}_{ft}$, to give us interest rates purged of variation due to loan characteristics, default risk, and bank and firm-bank relationship effects.

ρ_u	ρ_v	ρ_w	σ_θ	σ_u	σ_v	σ_ϵ	σ_z
0.413	0.140	0.317	0.405	1.321	3.927	0.581	0.753
(0.099)	(0.086)	(0.256)	(0.033)	(0.667)	(2.397)	(0.237)	(0.313)

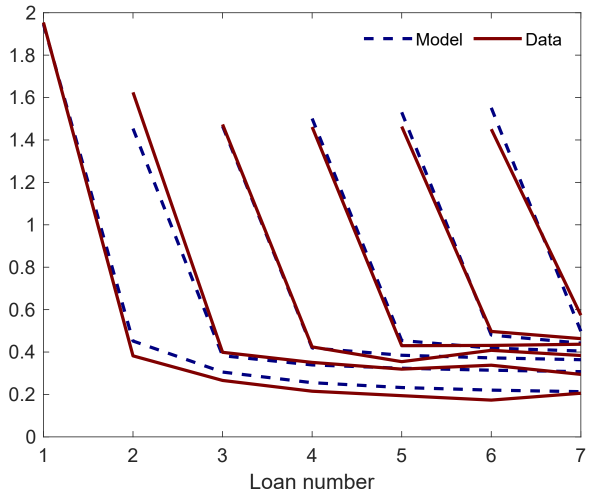
Table D.2: **Parameter values.** Parameter values for the interest rate process. See equation (1) in Sterk et al. (2021) for the full specification of the process.

by estimating the eight parameters so that simulated interest rates match the autocovariance matrix for interest rates paid by firms from the data. To construct this autocovariance matrix, we need to observe a sequence of loans for a large enough sample of firms from their first loan. In our data, we can be sure to observe all loans issued to firms born from September 2018 onward, however this sample is too small to generate reliable autocovariance estimates. We therefore extend the sample to include all firms born from 2010 onward and use the interest rates on the first seven loans observed for each firm.[19] Estimation of the parameters of the model is by simulated method of moments, minimizing the sum of squared deviations of the 28 moments of the autocovariance matrix.[20] Table D.2 presents the parameter values. We repeat the estimation process 1000 times and report the standard deviation of the parameters. The fit of the model to the data is presented in Figure D.1(a).

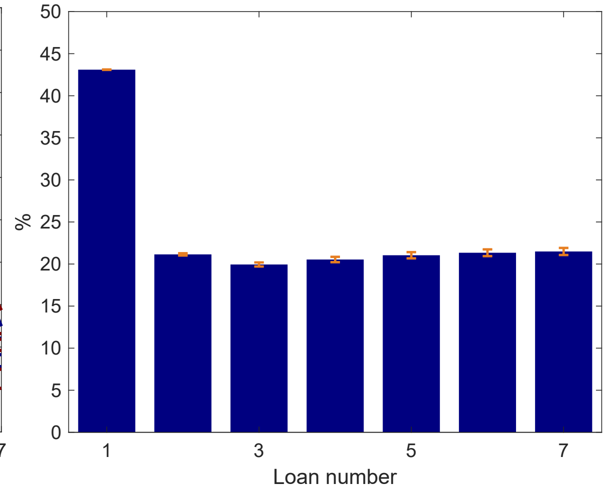
Two key parameters of the process are ρ_u and σ_θ, which are the persistence of the transitory component of $u_{f,n}$ and the standard deviation of the permanent component, respectively. Together, they determine how much long run heterogeneity in interest rates results from ex ante factors. The point estimates imply a long-run standard deviation of spreads of 0.55 percentage points. For comparison, our empirical estimate of the standard deviation in interest rates due to firm fixed effects is 1.01.

[19]While using firms born from 2010 onward gives us a larger sample, it comes with the limitation that we do not observe loans that matured before September 2018. So, for firms born before this date, we will observe loans that have a long enough maturity, but will miss shorter maturity loans. To verify that this does not substantially affect the results, we replicate the analysis restricting the sample to firms born in 2015 or later and find very similar results (these are available on request). For firms born from 2015 to August 2018, there are still some loans that we do not observe, but this issue is reduced relative to the baseline analysis. We use autocovarainces up to the seventh loan because beyond this the sample gets small for reliable estimates. We also use an unbalanced panel for the autocovariance estimation, allowing firms with fewer than seven loans to be included, in order to maximize the sample size.

[20]The autocovariance matrix is constructed using equation (2) from Sterk et al. (2021). To use this formula loans need to be numbered starting from zero instead of one.



(a) Autovariance of interest rate



(b) Ex ante contribution to rate variance

Figure D.1: **Autocovariance of interest rates and ex ante contribution to rate variance.** Panel (a) presents the autocovariances of interest rates from the model and the data. The line that starts at loan one presents the variance of loan one interest rates (the value in the loan one column) and the covariance of these rates with the rates for each future loan number. The line starting at loan two does the same for loan two, etc. Panel (b) presents the percentage of the variance of interest rates across firms that is accounted for by ex ante factors as a function of the loan number.

	Contribution
(6) Firm fixed effects	0.3006
(6a) Fixed firm characteristics	0.057
(6b) Latent firm fixed effect	0.243
# Loans	412,199

Table D.3: **Variance decomposition including firm controls: location controls at 5-digit postcode level.** Each row is the share of the total variance in interest rates due to select components of regressions (9) and (10), calculated with equation (5). In this version, firm location controls are at the 5-digit postcode level instead of the 4-digit level.

Figure D.1(b) presents the share of the variance of interest rates across firms that is due to the ex ante component, for loan numbers one to seven. The ex ante component is particularly important for determining first loan interest rates, accounting for approximately 43% of the variance. This contribution halves for the second loan, but then persists at around 21% for all loans thereafter. This indicates that firm heterogeneity prior to first loan issuance is an important source of persistent rate variation. To tie this to our earlier estimate of the contribution of firm fixed effects, this translates to ex ante factors accounting for approximately 15 of the 31 percentage points of total interest rate variation due to the fixed firm component in Table 5.[21]

E Additional details for Section 6

E.1 Firm location controls at the 5-digit postcode level

Table D.3 presents results for the exercise in which we replace the control for firm location at the 4-digit postcode level, with 5-digit postcodes.[22] We only report the components of the decomposition that are relevant for this exercise. Firm fixed characteristics can now account for 5.7 percentage points of the firm fixed effect for interest rates, a modest increase from 3.9 percentage points.

E.2 Bank-location fixed effects

To test the possibility that bank market power shows up at the bank-location level, we reestimate regression (9) replacing the bank fixed effects with bank-postcode fixed effects. The postcode is determined by the location of the firm receiving the loan, at the 4-digit level. Since the postcode is firm level variable, we present the results including and excluding firm fixed effects.[23] Table E.1 shows that bank-location fixed effects account for about 10 percentage points more variation in interest rates, but that firm fixed effects are not diminished.

E.3 Pricing loan portfolios

To test whether banks price loan portfolios jointly, and in a way that matters for firm-level rate variation, we compute residualized interest rates after controlling for loan characteristics, default risk, and bank fixed effects. We then compare residuals across

[21] Firm fixed effects account for 1.01 percentage points of interest rate variance in the earlier decomposition. The variance of the residualized interest rates that are used for the statistical model is 1.70 percentage points. Since the ex ante component accounts for approximately 21% of that residual variance, this corresponds to $0.21 \times 1.70 = 0.36$ percentage points. This is 15% of the total interest rate variance in our sample.

[22] There are 684 4-digit postcodes, and 3,851 5-digit postcodes, so they are substantially more granular.

[23] We do this for two reasons: there are a large number of fixed effects when both bank-location and firm fixed effects are included, which may cause poor identification; and the bank-location and firm fixed effects may be correlated.

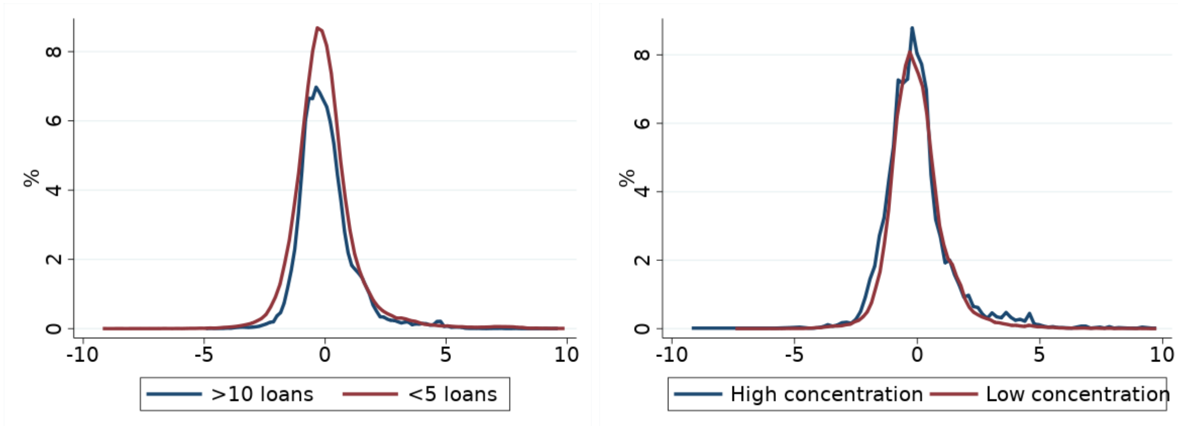
	Contribution	
(1) Default risk	0.0021	0.0001
(2) Loan characteristics	0.3228	0.2379
(3) Bank-location fixed effects	0.1974	0.2281
(4) Time-varying firm characteristics	0.0619	0.0116
(5) Relationship banking controls	0.0022	0.0006
(6) Firm fixed effects	-	0.3874
(7) Residual	0.4334	0.2687
(8) Covariances	0.0465	-0.1221
# Loans	452,692	410,329
Firm FE	No	Yes

Table E.1: **Variance decomposition including firm controls: with bank-postcode fixed effects.** Each row is the share of the total variance in interest rates due to each component of regression (9), calculated with equation (5). The bank fixed effects in regression (9) are replaced with bank-postcode fixed effects. Results are presented with an without the inclusion of firm fixed effects.

banks with which a firm has more than 10 loans and those with which it has fewer than 5 loans. If banks price loans jointly, residuals should vary within a firm across banks as a function of the intensity of the lending relationship. The left panel of Figure E.1 shows that residuals are slightly more dispersed for banks with which the firm has 10 or more loans—consistent with potential omitted variables related to joint pricing—but the magnitude is small and the two distributions largely overlap. In addition, we split the sample into firms with an HHI for borrowing across banks above the 70$^{\text{th}}$ percentile and below the 30$^{\text{th}}$ percentile, capturing highly concentrated and more dispersed borrowing structures, respectively. Figure E.1(b) shows that residuals are not quantitatively different across these groups, providing further evidence against meaningful differences in interest rates due to joint pricing of loans in a portfolio.

E.4 Firm size and public vs. private firm tests

To assess the relevance of personal financial relationships between the owner or manager of a firm and a bank, we compare firm fixed effects across firm size groups, and also between public and private firms. The underlying hypothesis is that such relationships should be less relevant for larger firms and for public firms. We have already shown that the share of interest rate variance due to firm fixed effects increases with



(a) Loans

(b) Concentration

Figure E.1: **Interest rate distribution: by number of loans from a bank and concentration of loans.** The interest rate distributions are for residuals after controlling for loan characteristics, default and bank fixed effects. The left panel splits the sample into firms with more than 10 loans from a bank (blue) and less than 5 loans from a bank (red). The right panel splits the sample into firms in the top 30% of loan concentration (blue) and bottom 30% of loan concentration (red). Loan concentration is measured with HHI of a firm's loans across banks.

	S	S-M	M	M-L	L
Std. dev. of firm fixed effects	1.05	0.96	0.83	0.83	0.96
# Firms	79,404	103,238	110,671	67,772	51,114

Table E.2: **Standard deviation of firm fixed effects by firm size group.** Each value is the standard deviation, in percentage points, of the firm fixed effect for firms in a specific size group. Small (S), small-medium (S-M), medium (M), medium-large (M-L) and large (L) firms have less than 5, 5–12, 13–40, 41–120, and more than 120 employees, respectively.

firm size (Table C.2), which challenges this. A potential concern is that this pattern could be driven by lower overall dispersion in interest rates among larger firms. To address this, we estimate the firm fixed effects from equation (9) and examine their standard deviation directly, rather than their contribution to total variance. Table E.2 shows that, although the standard deviation declines slightly with firm size, the differences are quantitatively small. For public vs. private firms, we estimate equation (9) separately for these groups and present the variance decomposition in Table E.3. Rate variation due to firm fixed effects is nearly twice as large for public firms as for private firms, contrary to the hypothesis. These exercises indicate that firm fixed effects are unlikely to be driven by personal financial relationships.

	Contribution	
	Private	Public
(1) Default risk	0.0001	0.0002
(2) Loan characteristics	0.2720	0.1496
(3) Bank fixed effects	0.1188	0.0540
(4) Time-varying firm characteristics	0.0103	0.0045
(5) Relationship banking controls	0.0029	0.0078
(6) Firm fixed effects	0.2619	0.5125
(7) Residual	0.2935	0.2501
(8) Covariances	0.0539	0.0337
# Loans	299,618	99,425

Table E.3: **Contributions of risk, loan, firm and firm-bank relationship characteristics, firm and bank fixed effects to interest rate variation: public vs. private firms.** Each row is the share of the total variance in interest rates due to each component of regression (9), calculated with equation (5). The sample is divided in public and private firms.

	Contribution	
(1) Firm fixed effects	0.301	0.301
(1a) Fixed firm characteristics	0.039	0.039
(1b) Latent firm fixed effects	0.262	0.262
Demand shocks	No	Yes
# Firms	27,875	27,875

Table E.4: **Variance decomposition including firm controls: with and without a control for demand shocks.** The first row is the share of the total variance in interest rates due to firm fixed effects in regression (7), calculated with equation (5). This contribution is decomposed using regression (10). In the left column demand shocks are not included in the fixed firm characteristics, and in the right column they are.

E.5 Demand shocks

To test whether persistent differences in credit demand explain the importance of latent firm fixed effects for the variance of interest rates, we estimate credit demand and supply shocks following Amiti & Weinstein (2018). We then average the demand shocks within firms over time, and add these as a control in the second stage regression of the Section 5 analysis. Table E.4 shows that these demand shocks provide no additional explanatory power for the latent firm fixed effects in rates.

E.6 Misallocation exercise

The model used for the misallocation calculations at the end of Section 6 is from Hsieh & Klenow (2009) (HK). We use an identical environment and maintain their notation.[24] The only change is we only have interest rate wedges and no output wedges (i.e. $\tau_{Ysi} = 0$ for all s and all i). Recall that s indexes sectors and i indexes firms within sectors. Equation (11) in the main text provides the ratio of equilibrium output to output in the case in which all capital wedges are zero ($\tau_{Ksi} = 0$ for all s and all i), denoted Y/Y^*. We assume that the total supplies of capital and labor are fixed, as in HK, so that we are measuring pure misallocation effects and not the effect of any distortion to the aggregate capital stock that results from the interest rate distortions. To compute Y/Y^*, the equations for the efficient within-sector firm output share, and the firm level distortions to capital and labor choices are:[25]

$$\frac{Y_{si}^*}{Y_s^*} = \frac{A_{si}^{\sigma}}{\left(\sum_{i=1}^{N_s} A_{si}^{\sigma-1}\right)^{\frac{\sigma}{\sigma-1}}}, \quad \frac{K_{si}}{K_{si}^*} = \frac{\left(\sum_{i=1}^{N_s} A_{si}^{\sigma-1}\right)\left(\frac{1}{1+\tau_{ksi}}\right)^{1+(\sigma-1)\alpha_s}}{\sum_{i=1}^{N_s} A_{si}^{\sigma-1}\left(\frac{1}{1+\tau_{ksi}}\right)^{1+(\sigma-1)\alpha_s}} \times \frac{K_s}{K_s^*}, \quad \frac{L_{si}}{L_{si}^*} = \frac{\left(\sum_{i=1}^{N_s} A_{si}^{\sigma-1}\right)\left(\frac{1}{1+\tau_{Ksi}}\right)^{\alpha_s(\sigma-1)}}{\sum_{i=1}^{N_s}\left(\frac{A_{si}}{(1+\tau_{Ksi})^{\alpha_s}}\right)^{\sigma-1}},$$

where

$$\frac{K_s}{K_s^*} = \frac{\left(\sum_{s=1}^{S} \theta_s\alpha_s\right)\left(\sum_{i=1}^{N_s}\left(A_{si}(1+\tau_{Ksi})^{-\frac{1}{\sigma-1}-\alpha_s}\right)^{\sigma-1}\right)\Big/\left(\sum_{i=1}^{N_s}\left(A_{si}(1+\tau_{Ksi})^{-\alpha_s}\right)^{\sigma-1}\right)}{\sum_{s=1}^{S}\left(\theta_s\alpha_s\left(\sum_{i=1}^{N_s}\left(A_{si}(1+\tau_{Ksi})^{-\frac{1}{\sigma-1}-\alpha_s}\right)^{\sigma-1}\right)\Big/\left(\sum_{i=1}^{N_s}\left(A_{si}(1+\tau_{Ksi})^{-\alpha_s}\right)^{\sigma-1}\right)\right)}.$$

To quantify Y/Y^* we need values for A_{si} and τ_{Ksi} for all firms, θ_s and α_s for each sector, and σ. We set $\sigma = 3$, following HK, and also adopt the HK methodology for computing values for θ_s, α_s, and A_{si}. θ_s maps to the value added share of sector s, which comes directly from the balance sheet and income statement data. The labor share is $1 - \alpha_s$, which we measure with the total cost of wages relative to value added

[24] One small notational change is using N_s for the number of firms in sector s instead of M_s.

[25] A sketch of the derivation is as follows. Equation (11) in the main text can be derived using the aggregator for the final good, the within-sector aggregator and the production function. To derive the four equations provided here, start by solving for K_{si}, L_{si}, Y_{si} and P_{si} in the standard way. Y_{si}^*/Y_s^* follows from this. For input $X \in \{K, L\}$, you can aggregate X_{si} to get X_s and X, and use:

$$\frac{X_{si}}{X_{si}^*} = \frac{X_{si}/X_s}{X_{si}^*/X_s^*} \times \frac{X_s/X}{X_s^*/X}.$$

This relies on the fact that the aggregate supplies of K and L are fixed.

by sector. We compute θ_s and α_s for each year from 2010 to 2022, and average the values over these years. Productivity can be expressed as

$$A_{si} = \kappa_s \frac{(P_{si}Y_{si})^{\frac{\sigma}{\sigma-1}}}{K_{si}^{\alpha_s} L_{si}^{1-\alpha_s}}.$$

where $\kappa_s \equiv (P_s^{\sigma} Y_s)^{-1/(\sigma-1)}$ is a sector level constant.[26] Since the equations for computing Y/Y^* are all invariant to the value of κ_s, we normalize $\kappa_s = 1$. We then measure productivity using the value of nominal output, $P_{si}Y_{si}$, the average of book value of capital and the beginning and end of a year for K_{si}, and the total cost of wages for L_{si}. We also trim the 1% tails of A_{si}. We use 2019 data for computing A_{si} because we have the universe of loans for this year and there was no major economic shock such as the pandemic or the sovereign debt crisis at this time. The real interest rate paid by firm i in the model is $(1+\tau_{Ksi})R$. For R, we use the average interest rate from our sample, weighted by loan value, minus the average annual inflation rate in Portugal from 2013 to 2021, which was 0.58%.[27] The inflation rate was much higher in 2010–12 and 2022, however we omit these years to be conservative with the estimate of the size of the interest rate wedges.[28] Using equation (10) from the main text, the latent firm fixed effect for the interest rates of firm i is $\check{\mathcal{F}}_i \equiv \mathcal{F}_i - \beta \mathbf{F}_i$, and we measure the distortion for the model as$\tau_{Ksi} = \check{\mathcal{F}}_i/R$. Finally, when computing Y/Y^*, we include all firms that operated in 2019 for which we have a fixed effect estimate.

Redoing the analysis using measures of A_{si} for each year from 2019 to 2022 results in estimated losses that range from 0.43% to 0.50%.[29]

[26] This can be derived using the demand function for firm i.

[27] Ideally we would adjust each nominal interest rate by the expected inflation rate over the period of each loan. Without such detailed inflation expectations data, we proxy for these expected inflation rates with the average realized one year inflation rate. For most of the sample period, inflation was low and stable, reducing the approximation error. The inflation rate is calculated as the average of the monthly year-on-year change in the consumer price index from January 2013 to December 2021, from the Portuguese Statistics Institute.

[28] A lower benchmark real interest rate, R, implies larger estimated distortions for all firms. Thus, a lower inflation rate reduces the estimated losses.

[29] We restrict attention to 2019 onward because for earlier years we do not have the universe of loans, so that our data is not fully representative of the economy.