Beyond Risk: Firm Financing and Interest Rates*

Miguel H. Ferreira † Nicholas Kozeniauskas ‡

February 2026

Abstract

Understanding the determinants of interest rates is a central question in macroeconomics and finance. In the context of firm financing from banks, how do the core theories stack up against the data? We address this question using a novel loan-level dataset. There is a lot of dispersion in rates paid by firms, even on observationally equivalent loans, and risk—the core theory for dispersion—can account for less than 15% of this. Alternative theories for dispersion, based on banks and relationship banking, also have quantitatively low explanatory power. Instead, the data dictates that the major source of variation is persistent idiosyncratic firm characteristics that are uncorrelated with a wide range of observables. The results suggest that there are substantial frictions in financial markets outside existing frameworks.

Keywords: Financial frictions; firm financing; interest rates.
JEL Codes: E43; E44; G32.

*The views expressed in this article are the authors' and do not necessarily represent those of Banco de 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

Financial frictions are a prominent feature of modern macroeconomics. Although they attracted some attention before the financial crisis of 2007–08, their incorporation into a wide range of macroeconomic theories has grown significantly since then. They are used to understand 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 they shape the allocation of capital in the economy and are quantitatively relevant for many macroeconomic outcomes.

In these applications, interest rates are a key price. They determine the rate at which firms can borrow, affecting investment choices, firm-level outcomes and the aggregate economy. This paper provides new evidence on the determinants of interest rates and their dispersion across firms in the context of bank loans. We show that, on both counts, the dominant approaches in the literature miss important features of the data. Interest rates vary substantially across firms, and most of this variation does not come from the source that is the typical focus—default risk. It is also not due to bank heterogeneity or relationship banking, two other leading branches of the literature on interest rate dispersion. Instead, the main source of variation is persistent idiosyncratic firm characteristics, that are not closely connected to firm observables.

In investigating the determinants of interest rates, we start from the simplest theory and build up, focusing on the most widely adopted approaches in the literature. These theories serve as a useful conceptual framework, and it is also instructive to understand how far they fall from the data. The objective is to assess the quantitative relevance of existing theories for interest rates on bank loans and, to the extent that they do not match

[1] Some examples are as follows. Business cycles: Cooley et al. (2004), Jermann & Quadrini (2012), Khan & Thomas (2013), Gilchrist et al. (2014), Buera & Moll (2015) and Arellano et al. (2019). Monetary policy transmission: Kiyotaki & Moore (1997), Bernanke & Gertler (1989), 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), Manova (2013), Leibovici (2021) and Spencer (2022).

the data, to investigate what they are missing. This analysis is made possible by a detailed new credit registry, matched with comprehensive firm balance sheet and income statement data.

On the theory side, we start from a classic view of interest rates under which they depend on the characteristics of a loan (e.g. issuance date and maturity) and the risk that the bank is not repaid in full. We call this default risk. This approach was formalized by Merton (1974) and underlies modern macroeconomic models of financial frictions. The most common approach in this literature 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. 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). Both approaches imply that all firms pay the risk free rate determined by the characteristics of a loan. A third approach to financial frictions is to allow firms to default, and price loans so that once lenders account for default risk, they earn the risk free rate in expectation (e.g. Bernanke et al., 1999). Under this approach, interest rates differ across firms depending on their risk.

The first stage of the empirical analysis shows that the data departs substantially from these theories. For the theories based on risk free lending, there is a lot of heterogeneity in interest rates paid by firms after controlling for a rich array of loan characteristics, including maturity and issuance date effects to capture movements in the yield curve. This is true whether we focus on collateralized loans (which map to the collateralized lending model) or uncollateralized loans (which map to the theory of endogenous borrowing limits). We evaluate the theory that interest rates differ as a result of default risk using data on default probabilities and loss given default. Conveniently, these are precisely the moments of the data that the theory requires to adjust interest rates for default risk. These characteristics of a loan account for less than 10% of the variation in interest rates that we observe. Overall, once we control for the characteristics of a loan including the level of collateralization, default risk, the majority of the variation in interest rates remains unexplained. These results imply that, at best, there are substantial frictions in

financial markets that the classical theory of interest rates, and current macroeconomic models, ignore.

The second stage of the analysis investigates potential explanations for the additional dispersion in interest rates. We first consider the role of banks. There is existing research providing evidence that the characteristics of a bank can affect its loan prices (e.g. Santos & Winton, 2019), and a long history of theories under which firm-bank relationships affect credit conditions (e.g. Leland & Pyle, 1977; Campbell & Kracaw, 1980; Sharpe, 1990a; Petersen & Rajan, 1995a). We evaluate the quantitative relevance of these theories for the total variation in interest rates, and find that these channel can only account for a modest share of the variation in the data. Bank characteristics can account for 5–10% of the variation and firm-bank relationships have similar explanatory power. Jointly, these channels can account for no more than 15% of interest rate variation. This leaves the majority of the variation, 60% under our preferred estimate, still unexplained.

We leverage the richness of the data to investigate the source of this variation. The main result is that the majority is due to idiosyncratic and highly persistent firm characteristics that are *not* closely associated with observables in the data. These account for more than 35% of the total variation in interest rates.[2] This means that, regardless of the nature of the loan or the issuing bank, some firms pay more for loans than others, and these differences are highly persistent. Furthermore, these differences cannot be explained by the fact that some firms are larger, more profitable, have different levels of risk, are in different industries and geographic regions, are different ages, or have different financial positions—all of these factors, and more, and controlled for. The differences are due to firm characteristics *beyond* what can be observed in the data.

These results have several implications. They place quantitative constraints on how much interest rate variation models with existing mechanisms, such as default risk or relationship banking, should generate. They also suggest the existence of financial market frictions outside of those that are currently considered. The empirical results imply that the main source of interest rate variation must satisfy three properties: (1) it must be an

[2]The remaining variation (30% of the total) is primarily accounted for by variation in rates across sectors and over time (10–15%), and a final residual (10%).

idiosyncratic firm characteristic; (2) it must be highly persistent; and (3) it must be uncorrelated with the wide range of firm characteristics observable in the data, including the standard characteristics that define firms in models (e.g. productivity, capital and financial position). These constraints count against an explanation based geographic variation in banking markets, or implicit divisions in banking markets by industry, firm size, or any other observable characteristic. The requirement that the explanation generates persistent rate dispersion, is also inconsistent with explanations based on idiosyncratic shocks loan demand due to changes in investment opportunities or liquidity needs.

The final section of the paper discusses more promising explanations based on frictional markets and differences in firm ability when procuring loans, differentiation between firms or their owners based on non-economic characteristics, and unobservable connections between the finances of firms and their owners.

Contribution to the literature The paper's main contribution is to provide new empirical evidence on the sources of interest rate dispersion across loans. The results are useful for giving empirical structure to models of firm financing. They can be used to discipline models that focus on a single source of variation (e.g. bank heterogeneity should only generate 10–15% of total variation) and guide future theories that aim to explain the variation that existing theories do not account for. This has the potential to lead to a better understanding of the nature of financial frictions.

There are a number of existing papers that argue that various factors are associated with, or cause, differences in the interest rates paid by firms. For example, there are theories and evidence based on heterogeneity among banks (e.g. Bellifemine et al., 2022), expectations about the future performance of firms (Petersen & Rajan, 1995b), relationship lending between firms and banks (e.g. Berger & Udell, 1995; Ioannidou & Ongena, 2010; Bolton et al., 2016), firm-bank relationships resulting in bank market power (Dempsey & Faria-e-Castro, 2024), banks evergreening loans to keep firms alive (Faria-e-Castro et al., 2024b), and information frictions (Darmouni, 2020), among others. Relative to this literature we take a broader view of interest rate dispersion. Our focus is on understanding the total variation and evaluating how much variation different classes of theories can account for.

The two existing papers closest to our work are Cavalcanti et al. (2023) and Faria-e-Castro et al. (2024a). Cavalcanti et al. (2023) show that there is substantial variation in interest rates paid by firms in Brazil and quantify the macroeconomic costs of this. We show that there is a lot of dispersion in rates in a developed country as well, but our main contribution relative to this paper is to investigate the source of this variation. Faria-e-Castro et al. (2024a) also study the borrowing costs of firms. The key distinction of the present paper is that we are able to match our credit registry data with data on the characteristics of firms, including risk measures, so that we can investigate the source of interest rate variation across firms.

The remainder of the paper is structured as follows. Section 2 describes the data. Section 3 tests the dominant macroeconomic theories of financial frictions. Section 4 explores the importance of firms, banks and firm-bank relationships for dispersion in interest rates. Section 5 evaluates the relevance of firm characteristics for interest rate differences. Section 6 summarizes the evidence, and discusses interpretation and implications. Section 7 concludes.

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.[3]

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.[4] The set of credit instruments is broad, encompassing all types of contracts through which firms can lend 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

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

[4]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.

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. For credit lines there is information about their total size and the amount that has been drawn. The information about credit instruments extends to their past characteristics such as the time of origination and characteristics at 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 entities classified in general government. Incorporated firms account for 33% of firms and 72% of employment in the economy.[5] The data comes from mandatory reporting of financial statements to the Ministry of Finance each year. It covers 2010 to 2022, so it overlaps with nearly all instruments in the credit registry going back to their origination dates.[6]

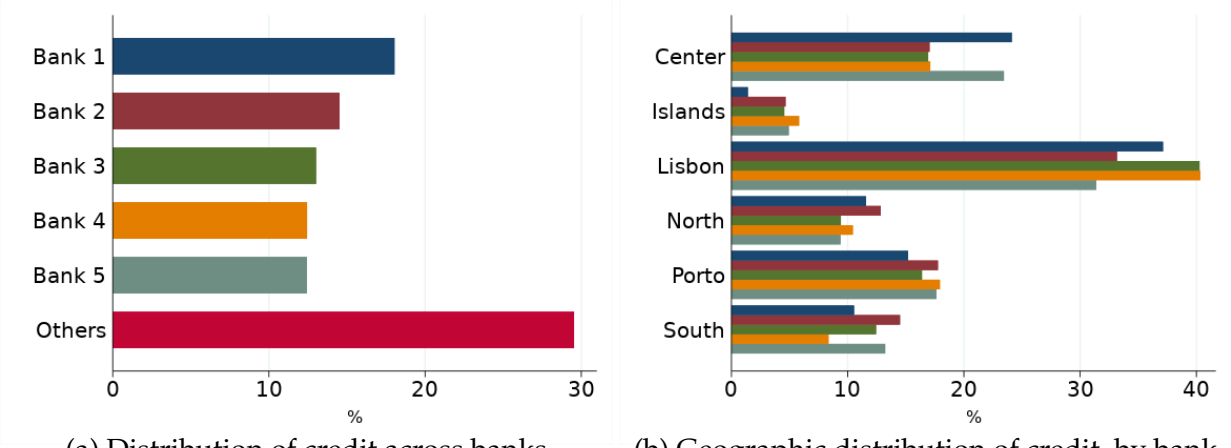
To construct our dataset we merge the credit registry with the firm data, dropping instruments that cannot be matched to the firm data in order to ensure a consistent sample across our analysis.[7] Only 32% of total credit in the data is lost as a result of this, since the firm data only omits very small 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.[8] We will refer to these as "firm financing loans" or sim-

[5]Only sole proprietorships and non-employer firms without a separate legal status from their owner are excluded. See the Appendix for details of the coverage of our sample.

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

[7]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.

[8]We are excluding the small share of renegotiated and renewed loans in the sample to avoid history



(a) Distribution of credit across banks (b) Geographic distribution of credit, by bank

Figure 1: **Distribution of firm financing credit across banks.** This figure presents moments of the distribution of firm financing loans by value in December 2019. Panel (a) shows the share of credit accounted for by each bank. Banks other than the largest five are grouped together. Panels (b) shows the distribution of credit across regions for each of the five main banks. The color for each bank is the same as in panel (a).

ply "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 (Figure 1(a)). There are 157 smaller banks that collectively provide the remaining 29%.[9] The five main banks are the main providers of both firm and consumer credit. They are large, national, general-purpose banks with similar operations to each other. To illustrate this, panel (b) show that their loans to firms have a similar geographic distribution, and in the Appendix we confirm that the distributions are similar across industry and firm size as well.

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 covered 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

dependence in interest rates.

[9]The market shares of the largest banks in other Euro Area 19 countries are similar. In the Appendix we show that share of banking sector assets held by the largest three and five banks in Portugal was similar to the average for the Euro Area 19 countries from 2010 to 2021.

2019.[10] 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 very important. 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%.[11] 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 3 banking relationships over the years that it is in the sample.

Figure 2 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.[12] There are 598,028 loans in total. The bulk of these (78.5%) were originated from 2018 onwards, with 19.2% originated from 2014–17, and the rest distributed over earlier years.[13] The average interest rate is 3.01% with a standard deviation of 2.25%.[14] The median loan is for €44,751. The right tail of the distribution is for much larger loans, with the 90^{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

[10] 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.

[11] See the Appendix for additional details.

[12] 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.

[13] 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 way origination dates are skewed towards recent years.

[14] Throughout the paper, interest rates are always stated on a per annum basis.

	p25	p50	p75	p90	St. dev.	# firms.
Value of loans (€'000s)	38.24	119.83	378.79	1387.72	(9848.76)	150,700
# of loans	3.00	4.00	8.00	14.00	(13.29)	
Av. loan value (€'000s)	11.35	24.93	58.46	183.84	(908.95)	
% collateralized	40	78	100	100	(35)	
# of banks	1.00	1.00	2.00	3.00	(1.14)	
Employees	1.50	3.75	8.75	21.00	(123.86)	
Total assets (€'000s)	101.10	261.73	778.33	2478.56	(65527.84)	
Revenue (€'000s)	87.60	228.28	649.41	1931.82	(40855.78)	
Leverage (%)	49	71	91	125	(255)	
Age	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 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.

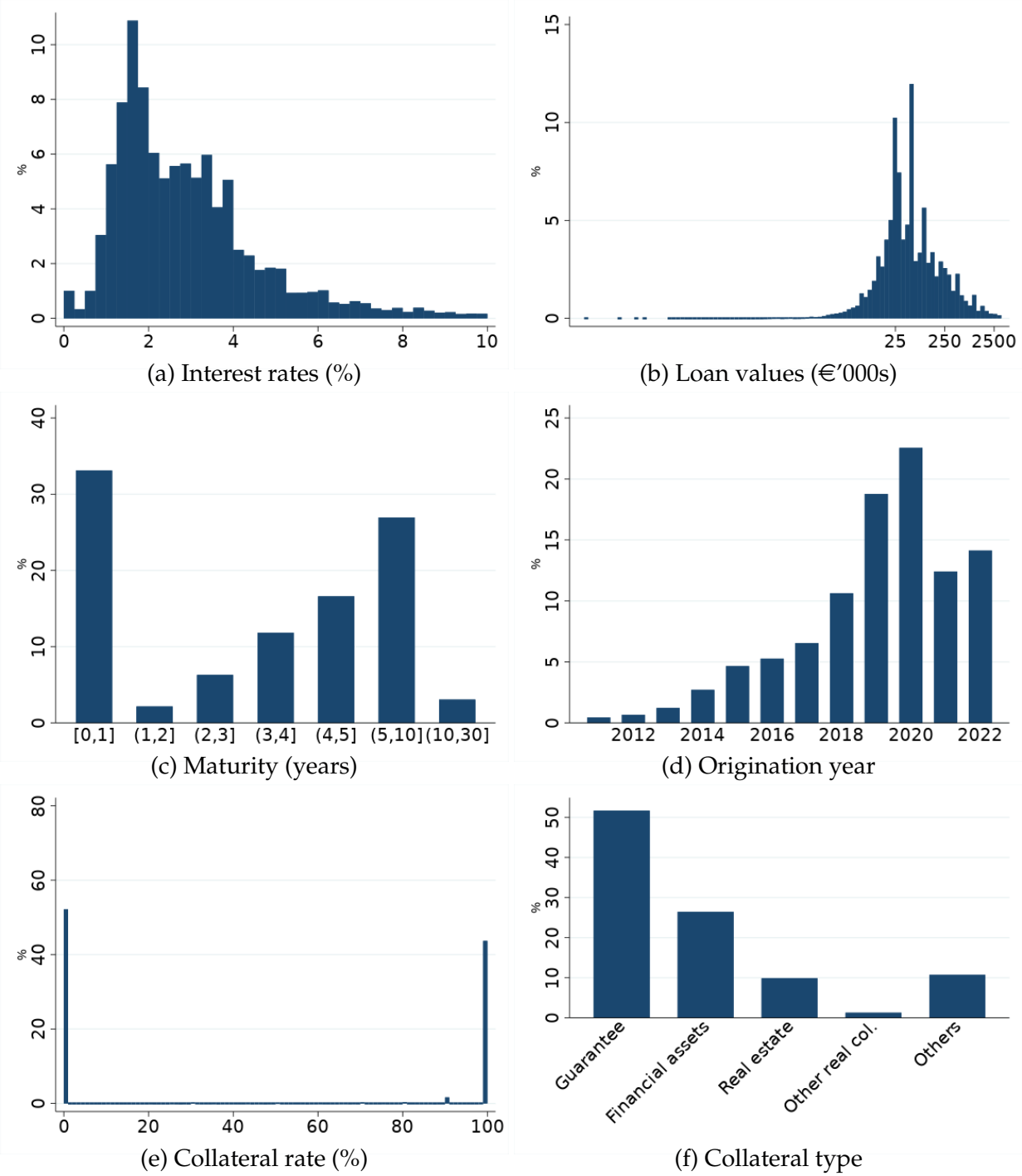
eight percent of loans are fully collateralized and virtually all other loans are uncollateralized. The most common types of collateral are guarantees and financial assets (51.7% and 26.4% by loan value, respectively), with real estate and other collateral also being used. Throughout the paper the analysis 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.[15] 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 banking regulator with probability of default estimates for all firms in the sample for 2010–2022.[16] We also use data on the postcodes of all bank branches for the same years.[17] When merged with the location of each firm, this data informs us about the banking market around each firm. Note that despite having the location of all branches, the credit

[15]See the Appendix for details on this.

[16]The methodology for estimating these probabilities of default is explained in Antunes et al. (2016). The dataset is called SIAC (*Sistema Interno de Avaliação de Crédito*).

[17]This data is a supplement to the balance sheet and income statement data from Banco de Portugal.



(a) Interest rates (%)

(b) Loan values (€'000s)

(c) Maturity (years)

(d) Origination year

(e) Collateral rate (%)

(f) Collateral type

Figure 2: **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 distribution in panel (f) is conditional on loans having some collateral.

registry does not record which branch each loan is from, only the bank. The postcodes are very granular, with seven digits. They correspond to an average of one postcode per

0.70 square kilometers for the country, and one per 0.09 square kilometers for the Lisbon metropolitan area (0.27 and 0.03 square miles, respectively).

3 Interest rates in macro models

The vast majority of macroeconomic models in which firms face external financing frictions fall into three categories. Under the most common approach it is assumed that a firm's existing assets constrain how much it can borrow and that either it cannot default, or that the environment is such that it will not choose to default in equilibrium (e.g. Moll, 2014). All firms therefore pay the risk free rate. An alternative approach is to assume that loans are uncollateralized, but that there are endogenous borrowing limits so that no firm defaults in equilibrium (e.g. Buera et al., 2011). This again implies that all firms pay the risk free rate. The third approach is to allow firms to default and price loans so that once lenders account for the probability of default and the loss conditional on default, they earn the risk free rate in expectation. Firms can differ in both their default probability and loss given default, and therefore pay different interest rates in equilibrium. In this section we evaluate these theories as general descriptions of firm financing.

3.1 Risk free lending

The first two theories both imply that firms borrow at the risk free rate. For the first theory a general form of the constraint that limits how much a firm can borrow is

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

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, and can be firm-specific and time-varying. This approach dates back to Kiyotaki & Moore (1997) and is particularly common in research studying business cycles, firm dynamics and capital misallocation.

A first observation is that it is common for loans to be collateralized, consistent with the first theory. Figure 11(e) in Section 2 shows that close to 50% of loans are fully collateralized, although most of the remaining loans have no collateral, so collateralized lending

is not a complete description of the data. In terms of interest rates, Figure 11(a) shows that there is a lot of dispersion. This is not necessarily inconsistent with these theories of risk free lending, as the dispersion could come from factors other than risk that are outside of typical models. For example, in the data loans are issued at different points in time when the prevailing risk free rates are different, they have different maturities and different interest rate structures (e.g. fixed versus variable). We have information on these factors, and more, in the data and can therefore strip out differences in interest rates associated with them.

An additional possibility is that interest rates could vary because firms are lending from different banks. There are several reasons why different banks could charge different rates, which we will be agnostic about for now.[18] This is another factor that is typically outside of models of risk free lending. The data, however, is rich enough to control for this. Specifically we will use bank fixed effects.[19] We also control for maturity-time fixed effects to capture variation in the risk free rate across maturities and over time. The time dimensions is at the frequency of the data (monthly).

To assess how much variation in interest rates remains after controlling for loan characteristics and bank and maturity-time variation, we run the following regression:

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

where r_{lbt} is the interest rate on loan l from bank b at its initiation date t, $\mathbf{L}_{lt}$ is a vector of loan characteristics, and $\mathcal{M}_{lt}$ and $\mathcal{B}_b$ are maturity-time and bank fixed effects.[20] Throughout the paper, bold capital letters will denote vectors of observable control variables, and calligraphic letters will denote vectors of fixed effects. The time period for the analysis is monthly. For loan characteristics the data is very detailed and we aim to provide a

[18] See, for example, Ho & Saunders (1981); Angbazo (1997); Degryse & Ongena (2005); Freixas & Rochet (2008); Corbae & D'Erasmo (2021)

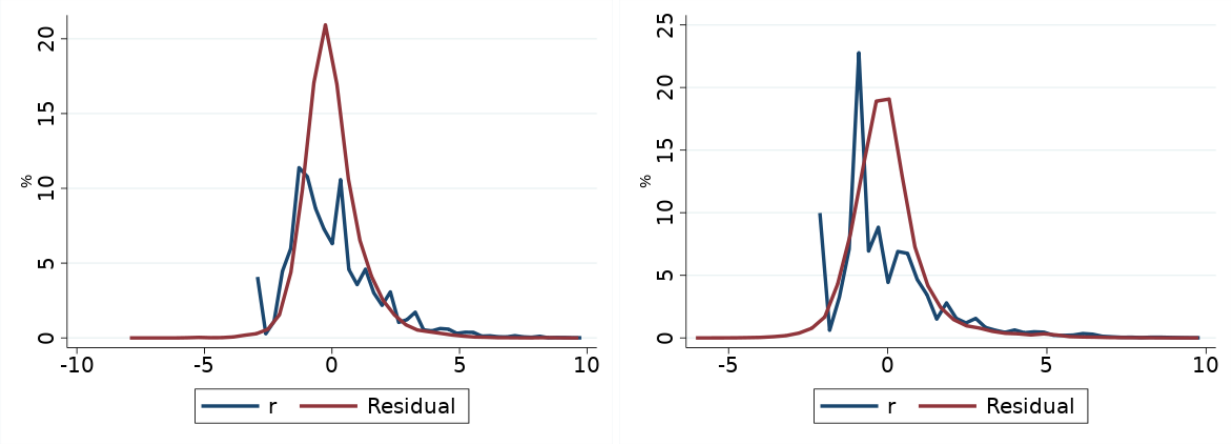
[19] The implicit assumption with this approach is that to the extent that two banks charge different interest rates for equivalent loans, the differences are constant across the loans. However, we allow these effects to differ at the level of firm-bank pairs later in the paper and the main results hold.

[20] We focus on interest rates as our main outcome variable instead of spreads, since we control for maturity–time fixed effects. These fixed effects absorb variation in the risk free rate both across time and across maturities, allowing us to isolate differences in lending rates besides shifts in the underlying term structure of interest rates. See the Appendix for additional details.

thorough description of the loans with these controls: amount, interest rate adjustability, maturity, type of collateral, amortization type/method, purpose of the loan, frequency of repayment, right to immediate repayment, if the loan is securitized or not, number of protections, and loan special characteristics. For the adjustability of the rate we have fixed effects for whether the loan has a fixed rate or can be adjusted at a daily, monthly, quarterly, biannual or annual frequency, or at the lender's discretion.[21] Loan maturities are categorized as follows (in years): (0, 0.25], (0.25, 0.5], (0.5, 1], and then in yearly intervals from (1, 2] up to (29, 30]. For the type of collateral we have no collateral, guarantees, financial assets, real estate, other collateral of real nature and others. The amortization method can be French (fixed installments, with interest and principal changing over time), German (fixed principal installment, with interest installment changing over time), fixed, unique payment of capital, increasing installments and others. The purpose of the loan can be for car purchases, liquidity management, exports and imports, equipment and other goods, real estate and land or without a specific purpose. The frequency of repayment is categorized as follows: monthly installments, quarterly, biannual, annual, unique payment of principal, unique payment of principal and interests and residual cases. Right to immediate repayment and the loan being securitized or not are dummy variables equal to one when the creditor can request immediate repayment of the loan and when the loan is securitized, respectively. The number of protections is zero for uncollateralized loans and for collateralized loans indicates the number of assets pledged as collateral. For special characteristics of the loan we have loans granted under Covid related policies/guarantees, consumer credit, no special characteristics or others. For more detailed information on the controls included, please see the Appendix.

To give the theory of collateralized lending the best chance of explaining the data we restrict the sample to fully collateralized loans. Figure 3(a) compares the distributions of interest rates in this sample and the distribution of residuals from equation (1), once it is estimated on the same sample. Comparing these distributions tells us how much varia-

[21] There are three additional frequencies of adjustment in the data, which we also include fixed effects for. They are other frequencies less than a year, frequencies greater than one year, and a category for all residual cases. These are categories are uncommon, representing 0.05%, 0.03% and 7.75% of the sample, respectively.



(a) Collateralized loans

(b) Uncollateralized loans

Figure 3: **Interest rate distributions before and after controlling for loan characteristics, banks and time.** 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 distribution after controlling for loan characteristics, bank and maturity-time fixed effects. All distributions are centered with their means at zero.

tion in interest rates remains after controlling for all of the loan, bank, maturity and time characteristics that we have discussed. The means of the two distributions are aligned at zero to make them visually comparable. We see that most of interest rate variation remains after the controls. The distribution is somewhat more concentrated around zero, but 49% of the variance of the distribution of rates remains.

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 distinguishing between different types of collateral is not materially important for the conclusion, with more than 43% of the variation in rates remaining unexplained in all cases.

The second main approach to modeling financial frictions departs from the assumption that loans are collateralized. Instead of collateral eliminating risk for banks, the theory posits that banks forecast the set of possible future outcomes for a firm and sets the terms of loans such that firms will always be able to fully repay. Since there is no risk, firms all pay the same risk free rate. To take this theory to the data we perform the analysis on the set of loans that are uncollateralized. The results for this sample are in Figure

3(b). The message is the same as for collateralized loans: more than 60% of the dispersion in interest rates remains after controlling for loan, bank, maturity and time characteristics.

The results from this section imply that there is substantial variation in interest rates beyond what can be reconciled with theories of risk free lending. Under the main specification, loan, bank, maturity and time characteristics can only explain around 50% of the observed variation in interest rates, even for fully collateralized loans.

3.2 Lending with risk

The dominant theory of interest dispersion in macroeconomic models is that interest rates vary across firms due to differences in risk. Can this theory explain the interest rate dispersion in the data?

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 for firm i 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 a 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. As examples of this type of setup, see Cooley & Quadrini (2001), Khan et al. (2021) and Ottonello & Winberry (2020).[22]

Our objective is to take this equation to the data to evaluate whether default risk can

[22]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. Kermani & Ma (2023) provided empirical evidence supporting this approach. 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.

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.[23]

In order to take this equation to the data we need measures of PDs and LGDs that are consistent with the theory.[24] 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.[25] The horizon at which the PDs are estimated is one year, and they are estimated at the firm rather than loan level, because these are the regulatory requirements. 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 its installment is not paid by the stipulated due date. From the perspective of the theory, the details of the definition and the fact that it is measured at the firm rather than the loan level is fine, as long as LGDs are estimated conditional on the same definition of default. Fortunately this is precisely the case. For LGDs we only have data reported by banks, which report their estimate when the loan is initiated.

Since PDs are estimated at a one year horizon, to maintain a tight connection between the model and the data we restrict the sample for the baseline analysis to loans with maturities up to one year.[26] For our baseline analysis we use the PD from the regulator since they have more coverage. However, in the Appendix we repeat the analysis using

[23] Full mathematical details are provided in the Appendix.

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

[25] The methodology for this is explained in Antunes et al. (2016).

[26] Several loans have various maturities that are less than a year. We think that, by attributing a one year PD to a loan with a maturity less than one year, we are therefore overestimating risk for some loans, which will add additional variation and work against our results.

the bank-reported risk variables and show that the results hold.[27]

In terms of coverage, the regulator has estimates of PDs at initiation date for all loans in our sample , and the banks estimates of PDs and LGD at initiation for loans granted after September 2018. Restricting our samples to those loans with an LGD value at initiation provides a sample of 89,896 loans. The coverage for these variables when reported by banks is less thorough.[28]

In taking equation (3) to the data we also want to control for the same factors that could account for variation in rates as we used in Section 3.1, so that we can assess the total share of the variation in rates that we can account for once we include risk. 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:

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

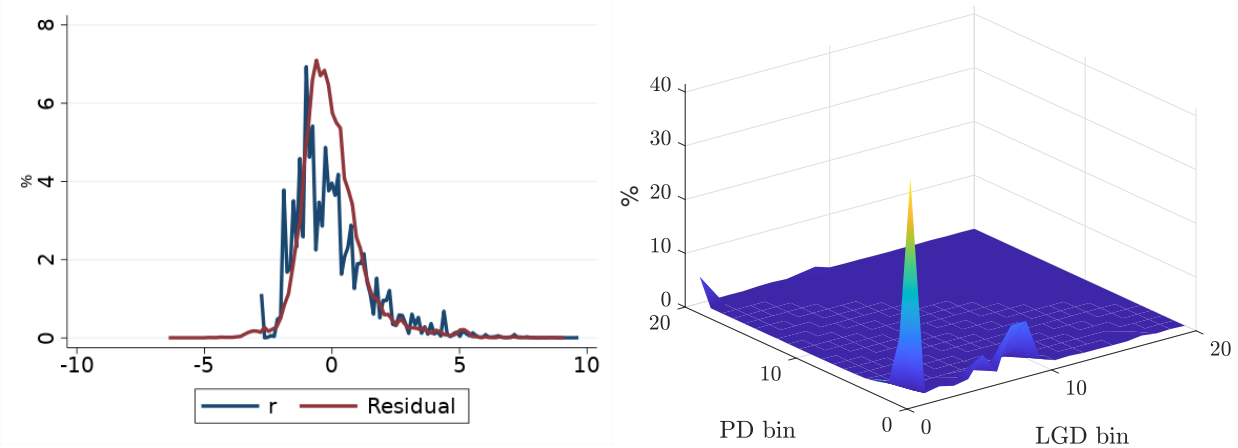
For this analysis we are no longer dividing the sample by whether loans are collateralized or not. We therefore add a variable for this to the loan controls $\mathbf{L}_{lt}$, and also add indicator variables for each of the possible collateral types. Note that since the risk free rate is purely a function of maturity and time, it is absorbed by the maturity-time fixed effects and does not need to be included in the regression.

Figure 4(a) compares the actual interest rate distribution in blue, and the distribution after controlling for risk, loan characteristics, bank and maturity-time effects 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. Quantitatively, risk can account for an additional 1.1% of the variance in rates beyond what is accounted for by the controls already in the analysis.[29] Risk is not very important because there is

[27] Another important detail for the analysis is the quality of the reported values. We want to be confident that PDs are measured well in the data. In order to do this we use the fact that we observe loans through to completion, so that we can ex post observe whether default occurs. We find a strong positive correlation between PDs and realized default. For more details, please see the Appendix.

[28] When using PDs from the banks the coverage is of 31,458 loans. The coverage is evenly distributed across all years in the sample.

[29] To calculate the explanatory power of risk we subtract the variance of the residuals from the equation (4) from the variance of the residuals of equation (1), estimated on the same sample, and then compare the



(a) Interest rate distribution & residuals

(b) Joint PD-LGD distirbution.

Figure 4: **Interest rates and risk.** Panel (a) presents the raw interest rate distribution (blue) and after controlling for loan characteristics, risk, bank and maturity-time effects (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%[,...,[95%,100%].

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%. Looking at the joint distribution of these variables in Figure 4(b), it is highly concentrated on values in the lower left quadrant where both PD and LGD are low.

We have performed a wide range of exercises to verify this result for the role of risk in determining interest rates. We now discuss these exercises and their results, with the full details provided in the Appendix.

The first robustness exercise we perform is to include all credit instruments in the regression sample. Given the sample restrictions we impose to guarantee a close mapping to the theory, plus the low coverage of the LGD variables, this restricts the sample in the baseline specification to 25,460 observations. When including all credit instruments the sample increases to 108,847, but results stay basically unchanged, with risk explaining less than 1%.

One potential concern is that there could be measurement error in the PDs and LGDs. We address this in two ways. First, we repeat the analysis using the PD estimates from

difference relative to the variance of the raw interest rate distribution in the same sample.

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 virtually the same.[30] Second, if there were measurement error in these reported values, then we evaluate how large this would have to be in order to explain the variation in rates that is observed. Specifically, we take the estimated coefficients for equation (4) 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, alternatively, what values LGDs would need to take holding PDs to their empirical values. The values are implausibly large. More than 20% of PDs and LGDs would need to be higher than 100%, with the median value of both distributions more than one order of magnitude larger compared to the reported values.

We also evaluate the effects of alternative stochastic discount factors for pricing loans. Under the baseline approach, we are assuming that the 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 lenders 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. The results show that the results are robust to these alternatives.

The fact that changes to the stochastic discount factor do not affect the results should

[30] In the Appendix we show that the PDs from banks and the regulator are very similar, which explains this result. Again, most loans have low PDs and low LGDs.

not be surprising, given the evidence already presented. The loans do not have a lot of risk to begin with, implying that their payoffs are similar across states. So changing the stochastic discount factor, which affects the weighting of states, is unlikely to affect the relative price of loans.

Equation (3) places a particular structure on the relationship between loan prices and risk. While this is the dominant approach in macroeconomic models, it is possible that this is misspecified and the true relationship takes another form. In the baseline regression we slightly deviate from this approach by allowing the coefficient on the risk term to be estimated and is not imposed to be -1. The estimated coefficient validates our approach, as it is -0.08 and statistically different from -1. Nonetheless, even if we impose the coefficient to be -1, which can amplify the potential effects of risk, the results are quantitatively similar to the baseline approach. 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.[31] Our results are robust to both of these approaches.

Additionally, to ensure that our findings are not specific to the subsample of loans for which LGD values are available, we replicate the semi-parametric analysis using only PDs. We proceed in two steps. First, we hold the sample fixed and re-estimate the semi-parametric specification excluding LGDs, to verify that their omission does not materially affect the results. Second, we extend the exercise to the full loan sample, for which PDs are available from the regulator. Results are robust across both exercises, indicating that our findings are not driven by the particular subsample of loans with LGD data.

A third approach that circumvents concerns about how risk is specified in the regression, as well reducing concerns about measurement that we discussed earlier, is to restrict the sample to loans with a low PD.[32] If there is little variation in rates after controlling for

[31] To allow the coefficients to vary as described we multiply each term by the relevant dummy variables for PD and LGD deciles.

[32] This approach just relies on firms with low PDs actually having low default risk. We verify this in

Robustness test	Risk explanatory power
Baseline	1.1%
B.1 All credit instruments	0.4%
B.2 Banks PDs	2.5%
B.5 Alternative stochastic discount factor	1.1%
B.6.2 Semi-parametric specification	9.5%
B.7 Semi-parametric PD only	4.3%, 3.0%
B.9 Extension to longer maturity loans	0.1%
B.10 Amortization types	2.0%, 0.6%, 0.1%
B.11 Exclusion of Covid period	0.1%
B.12.2 Size groups	0.1%, 0.1%, 1.1%, 1.1%, 1.2%
B.12.1 Age groups	0.8%, 1.0%, 0.1%

Table 2: **Robustness tests.** Notes: For the line "B.7 Semi-parametric PD only" results are presented for the samples for which LGD is reported and for the full sample. For the line "B.10 Amortization types" results are presented for unique payment of principal, french amortization and other types respectively. For the line "B.12.2 Size groups" results are presented for small, medium-small, medium, medium-large and large firms respectively. For the line "B.12.1 Age groups" results are presented for young, mature and old firms respectively.

risk and loan characteristics, then, if the sample is restricted to loans with very low risk and we run regression (1), there should be little residual variation in rates. To implement this we restrict the sample to loans with a PD below 0.29%, 0.55% and 1.14%, respectively the 5th, 10th and 25th percentile of the PD distribution. This changes the results little. For this subsamples 48%, 50% and 51% of the variance in rates remains unaccounted for by the controls.

To ensure as much consistency between the theory and data as possible, the baseline analysis restricts the sample to loans with a maturity of up to one year. In order to relax this restriction and extend the analysis to all loans, we generalize equation (3). To do this we need to make three assumptions. The PD provided in the data is for a one year horizon. We assume that this PD is constant over time for longer maturity loans, and that the realization of the default outcome is i.i.d. each year.[33] The LGD value 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

Appendix.

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

future for longer maturity loans. The third assumption is that the interest rate on a loan is expressed as the effective interest rate if the full repayment were made at maturity. In the Appendix we provide exercises to show that these assumptions are good approximations for the data. Under these assumptions, the multi-year analog of equation (2) implies that the interest rate for loan l, initiated at time t with a maturity of T periods can be expressed as:

$$1 + r_{ltT} = \frac{1 + \tilde{r}_{tT}}{(1 - PD_{lt} \times LGD_{lt})^{\frac{1}{T}}}, \tag{5}$$

where $\tilde{r}_{tT}$ is the risk free rate at time t for a maturity of T periods. We can take logs of this and take the equation to the data in the same way as for the single period case with equation (4). For this extended sample we find that the results are very similar to our baseline: the variation in interest rates explained by risk is 0.1%.

To ensure that our results are not driven by differences in loan repayment structures, we first restrict the sample to loans with a single payment of interest and principal, providing a closer correspondence to the theoretical framework. To further assess whether the main findings depend on the type of amortization, we also re-estimate the results separately for loans with French amortization (fixed total installments, with interest and principal components varying over time—20.9% of loans) and for loans with other amortization types (45% of loans). Results remain qualitatively similar across all repayment structures, confirming the robustness of our findings.

As a final robustness check, we verify that the results are not driven by the Covid period and potential guarantees offered by the government that would eliminate loan risk for banks. To do this we restrict the sample to loans initiated before 2020 and show that the results are very similar to our baseline. Additionally, to guarantee results are not driven by a specific group of firms, we verify that our results hold across groups of firms of different age and size.

Table 2 presents the explanatory power of risk in each of the different robustness exercises explained above. Not surprisingly, risk has the most explanatory power in the semi-parametric specification, when it accounts for 9.5% of the total variation in rates. Still, even in the semi-parametric approach, 60% of total variance is still unexplained.

For a detailed description of all the robustness exercises and the respective results, please check the Appendix.

4 Banks and relationship banking

The previous section showed that a substantial share of the variation in lending rates across loans persists even after controlling for observable loan characteristics and borrower risk. This section explores an alternative set of explanations for interest rate dispersion that originate from the supply side of credit markets. In particular, we draw on two strands of the banking literature that emphasize the role of (i) bank heterogeneity—differences in cost structures, risk profiles, and market power across lenders—and (ii) firm–bank relationships, whereby repeated interactions and private information shape loan pricing. This sections briefly review the dominant theories in these literatures and assess the extent to which each mechanism contributes to the observed dispersion in lending rates.

4.1 Banks and interest rate variation

There are several reasons why heterogeneity across banks can translate into differences in the interest rates they charge on loans. The banking 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).[34]

Our results so far already provide some guidance on this. Both regressions (1) and (4)

[34]Empirical evidence supporting these mechanisms includes Berger & Mester (1997); Berger & Hannan (1998); Wheelock & Wilson (2012, 2018); Darmouni (2020); Benetton (2021). Moreover, Corbae & D'Erasmo (2021) document systematic markup heterogeneity across banks of different sizes and develop a quantitative model that nests these three channels. More recently, Bellifemine et al. (2022) propose a heterogeneous-bank framework to quantify the role of banks in the propagation of macroeconomic shocks.

included bank fixed effects and therefore captured differences in rates across banks that are common across banks and over time. Those regressions still resulted in large residual variation in rates,[35] suggesting that banks may not be the major source of variation in rates. In this section we push this aspect of the analysis further, evaluating alternative specifications and quantifying the specific contribution of banks to rate variation.

Our methodology focuses on using bank fixed effects, since this allows us to capture the effects of banks on interest rates, while being agnostic about the mechanism through which these effects happen. Our baseline approach builds on regression (4), which contains controls for risk, loan characteristics and bank fixed effects. To better isolate the effects of banks, we take this regression and we add firm fixed effects. We do this to avoid the risk that firms sort across banks in way that would result in the bank fixed effects picking up what are actually firm effects. The regression for loan l of firm f from bank b initiated at time t is now:

$$r_{lfbt} = \beta_1 \mathbf{R}_{lt} + \beta_2 \mathbf{L}_{lt} + \mathcal{M}_{lt} + \mathcal{B}_b + \mathcal{F}_f + \varepsilon_{lfbt}, \tag{6}$$

where $\mathbf{R}_{lt}$ contains a linear and a quadratic term of the PD interacted with dummies for the ten deciles of the PD distribution and dummies for the loan maturity, and $\mathcal{F}_f$ contains the firm fixed effects.[36] We here include only PDs on the risk term to increase the number of observations in the sample. As shown in Section 3.2, leaving the LGD out 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 loans in a bank's portfolio.

Our identification of bank and firm fixed effects does not rely on the same mover variation that commonly generates mover bias in labor literature. 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 of different characteristics, even

[35] The results are in Figures 3 and 4(a).

[36] The other control variables in the regression are exactly the same as for regression (4). For the PD variable we follow our baseline approach of using the value from the regulator to maximize sample size.

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

Once we have estimated regression (6), we quantify the contribution of each component on the right-hand side to the total variance of interest rates following the Abowd et al. (1999) (AKM) method. For each loan-firm-bank-time observation, let $\hat{r}^c_{lfbt}$ denote the value of component c from the right-hand side of the estimated regression. The overall variance of interest rates is given by sum of the variances of each of the N components on the right-hand side plus the covariances between the different components. This is,

$$var(r_{lfbt}) = \sum_{c=1}^{N} var(\hat{r}^c_{lfbt}) + \sum_{j=1}^{N}\sum_{c=1}^{N} cov(\hat{r}^c_{lfbt}, \hat{r}^j_{lfbt}), \quad j \neq c. \tag{7}$$

Table 3 reports the results. The main results comes from the third line: under the baseline specification, banks only account for 9.6% of the variance in interest rates. They are not a primary source of dispersion. This exercise also provides a useful breakdown of the how much dispersion is due to the factors considered so far. Variation in rates over time and due to differences in maturity plus loan characteristics, where risk is included accounts for 28.6% of total variation. Interestingly, the time of issuance and maturity are by far the most important loan characteristics for determining rates. It is also striking that fixed firm characteristics account for a substantial amount of variation (32.7%). We evaluate this further in Section 5. Notice equally that the covariance terms between the different factors matter little for the overall interest rate variance.[37]

[37]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 relative low importance of the covariance terms indicate that there is little systematic sorting of specific firms (banks) into particular banks (firms) or loan characteristics.

	Contribution
(1) Risk + Loan char. + Matur-Time Component	0.2863
(2) Bank Component	0.0960
(3) Firm Component	0.3273
(4) Residual	0.2751
(5) Covariances	0.0153
# obs.	521,856

Table 3: **Contribution of loan characteristics, firm, bank and maturity-time fixed effects to interest rate variation.** The different lines represent the different components included in equation (6). Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (7).

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 (6), 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 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.[38] 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. This gives us a sample of 41,312 firms, relative to a sample of 84,535 under the baseline approach. The results are also presented in the Appendix and show that allowing bank effects to vary with time does not substantially change the results for the contribution of banks to rate variation. Note that the main change is that firm characteristics are now even more important, accounting for 62% of variation.

For additional verification of this result, we perform robustness exercises. One potential concern is that the firm fixed effects might be biased as the sample is not large enough. We have on average six loans per firm in the sample. The sample can be extended to include all instruments in the dataset, rather than just firm financing loans, which increases the number of loans per firm to twelve, guaranteeing a better fixed effect identification. Results are robust to this extension, with bank fixed effects still accounting for less than 10% of the total variation and the firm fixed component having the larger explanatory

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

power at 37%.

Similar to what we did in Section 3.2, we test the results by amortization type, when excluding the Covid period, and by firms of different sizes and ages. Overall, results are robust across all specifications, with the firm component always the most prevalent and the bank component 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 banks do not play a large quantitative role in accounting for interest rate variation, their specific relationships with firms can substantially affect borrowing terms. Relationship banking theory holds that repeated interactions allow banks to accumulate proprietary information and offer credit on better terms. The first rationale is informational: by learning about a borrower over time, banks can overcome adverse selection and moral hazard problems. This view originates in the information-based models of intermediation by Diamond (1984) and Ramakrishnan & Thakor (1984) which extend the asymmetric-information framework of Stiglitz & Weiss (1981), and is further developed in Von Thadden (1995), who shows that relationships are especially valuable when borrower quality is opaque.

The 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), Boot & Thakor (1994), and Berlin & Mester (1992, 1998). Empirically, Petersen & Rajan (1994, 1995c) and Berger & Udell (1995) show that stronger, more exclusive relationships lower borrowing costs, reduce collateral requirements, and ease credit access, consistent with theories of intertemporal rent-sharing. At the same time, relationships can create inefficiencies: Sharpe (1990b) and Rajan (1992) highlight information monopolies lead lenders to exploit the monopoly power to charge higher loan rates or impose less favorable terms, while Bolton & Scharfstein (1996) and Faria-e-Castro et al. (2024b) show that close ties can lead banks to refinance weak firms inefficiently.[39]

[39]For a comprehensive review of the origins of the firm-bank relationship theories, see Boot (2000). For recent quantitative evidence on the aggregate importance of relationship lending includes Bolton et al. (2016); Dempsey & Faria-e-Castro (2024).

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. In doing so, it is important to distinguish these factors from factors that are specific to each firm and bank, but separate from the relationship between them. The analysis makes use of regression (6), as well as the following variation of it:

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

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 (6) is that the firm and bank fixed effects are now interacted. In order to isolate the contribution of the relationship between firms and banks to 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 equation (7). This provides a measure of the total contribution of firm factors, bank factors and relationship factors.[40] Second, we take the same sample of loans and re-estimate equation (6). Using equation (7) we get a measure of bank factors and firm factors on their own 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.

Relative to the identification hypothesis for equation (6), to be able to estimate the firm-bank fixed effect, we need to additionally restrict our sample to firms that have multiple loans from multiple banks.

The results are presented in Table 4. Firm-bank relationships account for 11.4% of the overall variation in spreads and, once this component is removed the contribution of pure bank effects is 9.6%. So, in total, banks and relationship banking account for 21% of variation. While a meaningful amount, this only goes a fraction of the way towards accounting for the data. Interestingly, the contribution of firm fixed effects does not diminish at all, indicating that the characteristics of firms are quantitatively very important, and that the relevant characteristics are not associated with their banking relationships. More impor-

[40] Note that the sample is smaller than for the previous analysis because firms need to have received loans from multiple banks to be in the sample.

	Contribution
(1) Bank Component	0.096
(2) Firm Component	0.362
(3) Relationship Component	0.114
# obs.	463,624

Table 4: **Contribution of firm, bank and firm-bank fixed effects to interest rate variation.** The different lines represent the different components. Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (7). The Relationship Component is presented as the difference between the estimated firm-bank fixed effect and the separate firm and bank fixed effects estimated on the same sample.

tantly, restricting the sample to firms with multiple bank relationships does not alter the relative importance of the firm fixed effects, reinforcing the validity of our fixed effects identification.

To provide additional support for these results we perform additional exercises that we report in the Appendix. We extend the sample to include all credit instruments in the data and relationship banking accounts for 17.7% of variation. 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.[41] The frequency for the time dimension is one year. We find that this has little effect on the results, with firm-bank relationships still accounting for less than 15% of the total variation in interest rates.

5 Firm characteristics and interest rates

The results in the previous section indicated that firm characteristics are important for accounting for interest rate variation across loans. In this section we investigate what characteristics are relevant. We decompose the contribution of various firm characteristics by extending the analysis developed up to this point, and complement this with a statistical method that separates the effects of permanent and transitory firm characteristics on interest rates.

[41] This means that equation (8) is estimated with firm-bank-time fixed effects and equation (6) is estimated with firm-time and bank-time fixed effects.

5.1 Decomposition by firm characteristics

To understand what firm characteristics are important for determining interest rates we take a two-step approach. In the first step we take regression (6) 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 used as regressors. We then take the estimated fixed effects and evaluate how much of these can be accounted for by observable permanent firm characteristics.[42]

The regression for the first step is:

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

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 of the 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 that requires firm-bank fixed effects to economize on the number of fixed effects to estimate.[43] 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. 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.

The full list of controls for time-varying firm characteristics is provided in Table 5(a). We control for standard factors of production—total factor productivity (TFP),[44] capital, labor (employment and the wage bill) and intermediates—as they are commonly state

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

[43]Specifications with firm-bank fixed effects at the most demanding specifications 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.

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

(a) Time-varying characteristics		
Non-financial variables		
Production inputs	*Revenue & profit*	Investment rate
Capital stock	Revenue	Age
Number of employees	Revenue growth rate	
Wage bill	Profit share	
Intermediate inputs	TFP	
Financial variables		
Assets & liabilities	*Credit history*	*Bank composition*
Tangible assets	Past default indicator	Number banks
Financial assets	Past late repayment indicator	HHI bank credit
Total assets	*Trade credit*	
Total liabilities	Trade payables/total assets	
	Trade receivables/total assets	
(b) Time-invariant characteristics		
Industry	Home country	*Volatility*
Location	Legal form	Std. $\Delta \log(\text{revenue})$
Number of local banks	R&D indicator	Skewness $\Delta \log(\text{revenue})$
Top 5 banks	*Foreign operations*	Std. $\Delta \log(\text{TFP})$
Other banks	Exporter indicator	Skewness $\Delta \log(\text{TFP})$
Number of establishments	Number of foreign estabs.	

Table 5: **Controls for firm characteristics.** HHI stands for the Herfindahl–Hirschman index. All variables that are not rates or indicator variables are entered into the regressions in log units. Productivity is measured using the method from Ackerberg et al. (2015). For credit history, the indicators record whether a firm has ever defaulted on a loan or been late with a repayment. A payment is classified as late if firm is 30 days or more late on a repayment. Location is measured at the 4-digit postcode level. Industries are 5-digit, implying 833 industries in total. Legal forms of businesses are private limited liability company (87% of firms), public limited company (6%) and other types. The number of local banks is counted within the 7-digit postcode of a firm's headquarters. Number of establishments and number of foreign establishments are the maximum numbers observed for a firm in the data. Full details for the measurement of all variables are in the Appendix.

variable for a firm or otherwise are informative about production technologies. The investment rate contains information about the demand for financing, and revenue and profit are informative about the scope of the firm, its cash flow, and its ability to continue operating. Revenue growth contains information about the future trajectory of a firm, and age accounts for potential lifecycle patterns in demand for external financing and in firm outcomes (e.g. Haltiwanger et al., 2013; Hsieh & Klenow, 2014; Kochen, 2025). On the financial side, we control for the values of tangible, financial and total assets, as well

as total liabilities, to capture the general balance sheet position of firms. Credit history variables are included for any effects on loan pricing beyond default risk. The value of trade payables and receivables relative to the size of the firm provides information on cash flows. The composition of a firm's lending from banks, measured by its number of banks and the Herfindahl–Hirschman index (HHI) of its credit across these banks, is a measure of a firm's potential to generate competition between lenders. All variables are specified in log units, unless they that are rates, shares or indicator variables. For additional details about the variables and their measurement, see the notes to Table 5 and the Appendix.

To control for firm-bank relationships, we use five variables to capture the duration, depth and nature of the relationship. For duration, we use the time since the firm received its first loan from the bank. To capture the depth and nature of the present relationship we use four moments: the number of loans issued by the bank to the firm over the past five year; the log total value of these loans; and the number of loans and the value of these as a share of the number of loans and value of loans that the firm as received from all banks over the past five years.

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. The variables included in $\mathbf{F}_f$ are summarized in Table 5(b). We control for industry 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 postcode level.[45] 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

[45]These correspond to 131,000 geographic units, with an average area of 0.70 square kilometers (0.27 miles).

	Contribution
(1) Firm FE	0.762
(1.1) Firm FE: obs component	0.108
(1.2) Firm FE: unobs component	0.654
# Firms	59,528

Table 6: **Contribution of firm fixed effect and firm fixed observable characteristics.** The different lines represent the different components. Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (7).

branches located within the more granular 7-digit postcodes. 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 a 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 whether a firm engages in R&D with an indicator for whether it has R&D employees. Controls for whether a firm has ever exported and its number of foreign establishments address effects of international trade and multinational production. Finally, we want controls for the distributions of supply and demand shocks that firms face. We proxy for these with the standard deviations and skewness of the observed paths of TFP and revenue for each firm.

Table 6 presents the contribution of the firm fixed effect and the part of the firm fixed effect that is explained by the fixed firm observable characteristics. Relative to the baseline analysis, we lose 25,007 firms, as not all additional control variables are reported by all firms. Overall the firm fixed effect explains 76% of the total variance, a value twice as high as in our baseline results presented in the previous sections. This is explained by two factors. First, sample selection. The variables used tend to be reported mostly by the larger firms, causing the sample to be bias towards the larger firms, which we show in the Appendix have a higher firm fixed effect. Second, there is a strong negative covariance between firm characteristics and loan characteristics, which boosts the explanatory power of all the remaining components. In addition, Table 6 illustrates that the fixed

observable characteristics explain only a small fraction of the overall variance of interest rates (10%). This indicates that the strong firm fixed component is uncorrelated with both time-varying and invariant firm characteristics.

Robustness exercises, reported in the Appendix, extend the analysis to the sample of all credit instruments and add additional controls for permanent firm characteristics. Using all instruments results in an estimate for the contribution of firm fixed effects to total interest rate variation of 49%, similar to the baseline results, with time-invariant firm observables only explaining 4% of total variation. For the permanent characteristics of firms, Sterk et al. (2021) show that the initial conditions of a firm are a quantitatively important determinant of future outcomes. In this spirit we supplement the controls for the second stage of the analysis with additional measures of the initial state of a firm. Specifically, we use the log values of capital, number of employees, wage bill, intermediate inputs, TFP, revenue, and total liabilities at age one.[46] With these additional controls, we can explain approximately one quarter of the firm fixed effects or 21% of total interest rate variation.

5.2 Permanent vs. transitory characteristics

So far, we have established that firm fixed characteristics are the major driver of interest rate variation. Besides the loan, bank and firm-bank characteristics and the maturity-time fixed effects, we have considered several characteristics of the firm, including the risk component, and shown that this cannot explain much of the variation in interest rates. In this section we offer a more structural way of decomposing firm transitory and fixed components.

Autocorrelation and standard deviation If, as suggested by the evidence in Section 4.1, there is heterogeneous firm fixed effects, we should observe a relatively constant profile of standard deviation of interest rates and relatively high autocorrelation between interest rates within firm over different loans. Figure 5 plots these two statistics in our data, with panel (a) depicting the standard deviation and panel (b) depicting the autocorrelation

[46] We use values for a firm's second year in the data instead of the first (age one instead of age zero), so that measurement is more consistent. In the first year some firms will operate for larger share of the year than others.

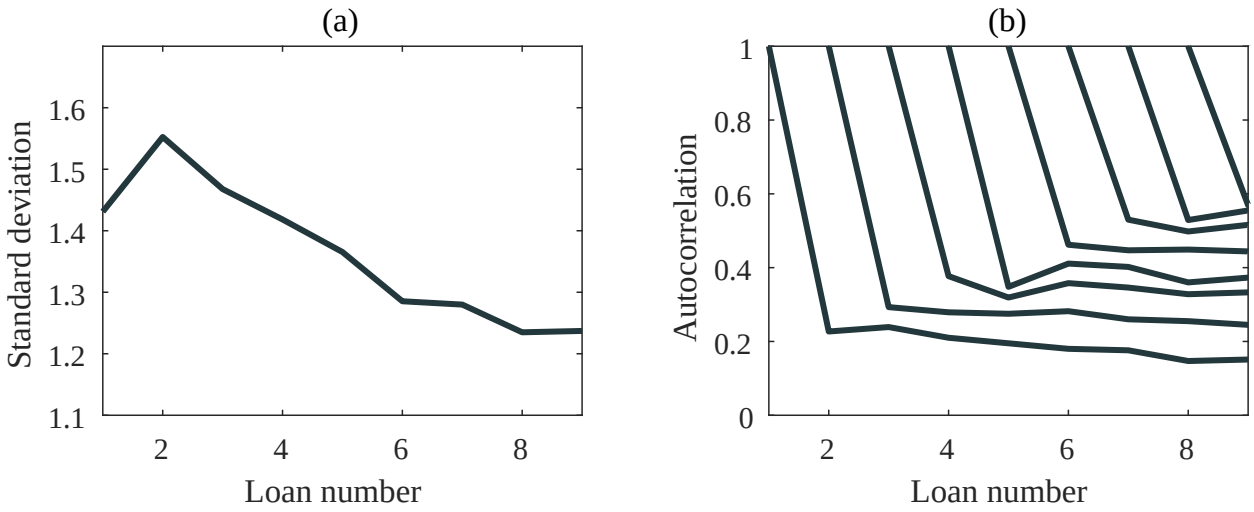


Figure 5: **Standard deviation and autocorrelation of interest rates across loans, within firm.** The left panel presents the standard deviation of interest rates by loan number, after controlling for loan characteristics, bank, maturity-time and sector fixed effects. The right panel presents the autocorrelation of spreads by loan number l and loan number $h \leq l$. Across lines h changes, while l changes along the lines.

profile of interest rates within firm over different loans.[47] First, the standard deviation is decreasing only mildly over loans, decreasing from 1.2 percentage points for the first loans firms get to 1.1 percentage points for the ninth loan. Second, the autocorrelation is relatively high between loan numbers, suggesting that initial conditions are also very important for interest rates.

Statistical model To gain a more structural understanding beyond descriptive statistics of the importance of ex-ante and ex-post heterogeneity for spreads over the firm's life-cycle, we follow Sterk et al. (2021) and adopt their statistical model.[48] This model uses the information provided by the autocovariance structure of interest rates to capture the importance of both types of heterogeneity.

Consider the following decomposition for interest rate r by firm i at loan number l:

$$\underbrace{r_{i,l}}_{\text{Interest rate}} = \underbrace{u_{i,l} + v_{i,l}}_{\text{Ex-ante component}} + \underbrace{w_{i,l} + z_{i,l}}_{\text{Ex-post component}}, \tag{11}$$

[47] To prevent differences across banks, loan characteristics, sectors and business cycle conditions from explaining the majority of the standard deviation and autocorrelation, we first control for loan characteristics, bank, maturity-time and sector fixed effects and then use the interest rate residuals.

[48] Note Sterk et al. (2021) apply this statistical model to employment, while we do it for interest rates.

where

$$
\begin{aligned}
u_{i,l} = \quad & \rho_u u_{i,l-1} + \theta_i, \quad && u_{i,-1} \sim iid\left(\mu_{\tilde{u}}, \sigma^2_{\tilde{u}}\right), \quad && \theta_i \sim iid\left(\mu_\theta, \sigma^2_\theta\right), \quad && |\rho_u| \leq 1 \\
v_{i,l} = \quad & \rho_v v_{i,l-1}, && v_{i,-1} \sim iid\left(\mu_{\tilde{v}}, \sigma^2_{\tilde{v}}\right), && && |\rho_v| \leq 1 \\
w_{i,l} = \quad & \rho_w w_{i,l-1} + \varepsilon_{i,l}, \quad && w_{i,-1} = 0, && \varepsilon_{i,l} \sim iid\left(0, \sigma^2_\varepsilon\right), \quad && |\rho_w| \leq 1 \\
z_{i,a} \sim \quad & iid\left(0, \sigma^2_z\right)
\end{aligned}
$$

In this interest rate process, the terms $u_{i,l}$ and $v_{i,l}$ capture the ex-ante profile while $w_{i,l}$ and $z_{i,l}$ capture the ex-post one. The ex-ante component is determined by three shocks that are drawn just prior to the time in which the firm gets its first loan, at $l = -1$. The shocks $v_{i,-1}$ and $u_{i,-1}$ represent the initial conditions of the firm, which allow for rich heterogeneity even across the first loans the firms get. θ_i is the permanent component, which will accumulate over the different loans at speed ρ_u. In particular, with $\rho_u < 1$, the long-run steady state level of interest rates will be given by $\frac{\theta_i}{1-\rho_u}$. Further, this specification allows for rich heterogeneity not only in terms of long term interest rates the firms will face, depending on the distribution of θ_i, but also in terms of the speed at which firms reach the long term spreads. As firms start at different points depending on $u_{i,-1}$ and $v_{i,-1}$ and each shock has its own persistence parameter, the path from initial to steady state spread will highly differ across firms.

The ex-post component is formed of two different shocks, one i.i.d. shock with expected value of zero, and a persistent one that follows an AR(1) process with i.i.d. innovations $\epsilon_{i,a}$ and persistence ρ_w. To abstract the ex-post component from affecting the ex-ante one, we set the initial conditions of the persistent shock to $w_{i,-1} = 0$.

We calibrate the model by minimizing the sum of squared differences between the model and empirical autocovariance. As the model only captures residual components and firm permanent and persistent components, we first clean the raw interest rates from loan characteristics, and bank and maturity-time fixed effects, as neither of these components is considered in the model. We then calculate the autocovariance of cleaned residuals. Table 7 presents the parameters resulting from the calibration strategy. As we have 45 moments to match with only 8 parameters, we repeat the process 1000 times and

ρ_u	ρ_v	ρ_w	σ_θ	σ_u	σ_v	σ_ϵ	σ_z
0.473	0.695	0.534	0.367	0.668	1.374	0.229	0.253
(0.102)	(0.207)	(0.204)	(0.080)	(0.607)	(1.613)	(0.543)	(0.906)

Table 7: Equally-weighted minimum distance estimates of the statistical model. Standard deviation in parenthesis.

report the standard deviation of the parameters. The model fit to the data can be found in panel (a) of Figure 6, which plots the interest rate autocovariance generated by the model in blue and the empirical counterpart in orange.

Panel (a) of Figure 6 plots the model fit to the data when calibrated to all firms in the Portuguese economy with at least nine loans. Two key parameters of the model are ρ_u and σ_θ, as, together, they imply that permanent heterogeneity exists. The point estimates imply that ex-ante conditions matter, as both ρ_u and σ_θ are statistically different from zero. The point estimates indicate a long-run standard deviation of spreads of 0.70 percentage points, while in the data, after nine loans, the standard deviation is 1.24 percentage points. This again demonstrates that there seems to be differences across firms that originate from ex-ante conditions.

To test if the model would predict a similar importance of the firm fixed effect to the empirical estimates, we simulate the model for the same number of firms we have in the baseline regression, 84,535 firms, and we include a probability of firm exit to match the average number of loans per firm, 6.17. We then regress the interest rate r_{lf} on loan l for firm f on a firm fixed effect $\mathcal{F}_f$ in the simulated data

$$r_{lf} = \mathcal{F}_f + \varepsilon_{lfbt}. \tag{12}$$

To account for the variance of the interest rates explained by the firm fixed component we again use the AKM decomposition, as in Section 4. Table 8 illustrates that the firm fixed effect accounts for 70% of the variance in interest rates in the simulation. The model was calibrated to match the interest rate variance after cleaning for loan characteristics, bank and maturity-time fixed effects, which is 55% of total raw interest rate variance in the data. Scaling the 70% for the total variance accounted for by the model, one gets that

	Contribution
(1) Firm component	0.7017
(2) Residual	0.2983
obs.	520,059

Table 8: **Contribution of firm fixed effect to interest rate variation.** The different lines represent the different components. Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (7).

the firm fixed component accounts for 39% of the total interest rate variance present in the data, a number remarkably close to the baseline results presented in Section 5.

Finally, to more clearly identify the ex-post and ex-ante contributions, one can also derive the formula for the model autocovariance, enabling a clear identification of the contribution of both components. The autocovariance formula is given by

$$Cov[r_{i,a}, r_{i,a-j}] = \underbrace{\left(\sum_{k=0}^{a} \rho_u^k\right)\left(\sum_{k=0}^{a-j} \rho_u^k\right)\sigma_\theta^2 + \rho_u^{2(a+1)-j}\sigma_{\hat{u}}^2 + \rho_v^{2(a+1)-j}\sigma_{\hat{v}}^2}_{\text{Ex-ante component}}$$
$$+ \underbrace{\sigma_\epsilon^2 \rho_w^j \sum_{k=0}^{a-j} \rho_w^{2k} + \sigma_z^2 \mathbf{1}_{j=0}}_{\text{Ex-post component}}$$

and its derivation can be found in Sterk et al. (2021). The autocovariance is a function of variance and persistence parameters of both ex-ante and ex-post shocks, as described above. Panel (b) of Figure 6 illustrates the importance of the ex-ante component for the variance as a function of loan number. The ex-ante component contribution is close to 40% at the first loan, and around 26% of the variance of interest rates after the ninth loan, or when scaling down by the variance accounted for the model is 15% of total interest rate variance. When comparing to the robustness exercise in Section 5 in which we include age one control variables, we find those to explain around 10% of the total interest rate variance, very close to the 15% we find from the model that are accounted for ex-ante conditions.

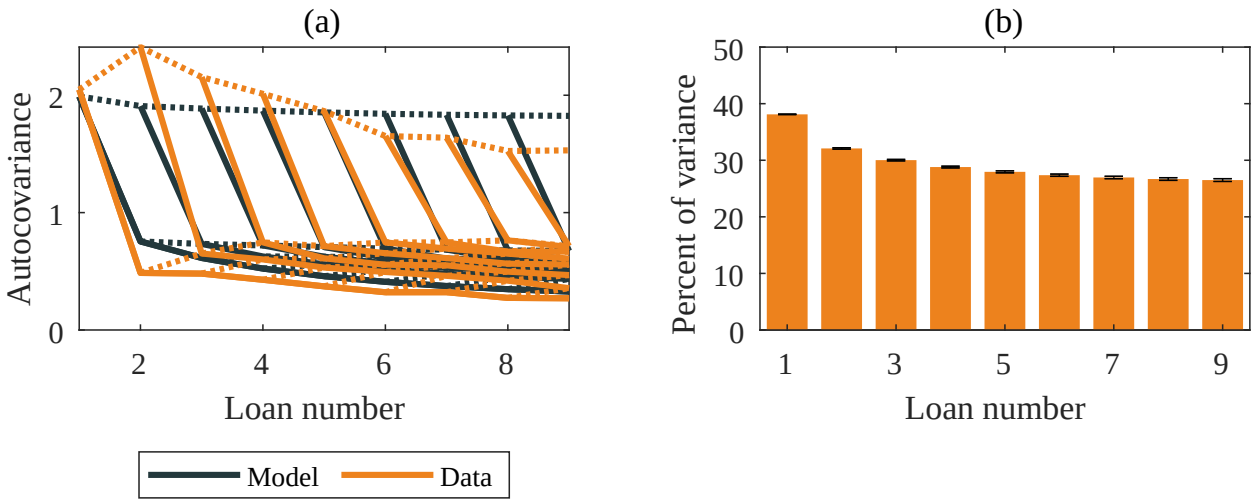


Figure 6: **Model fit to the data and ex-ante contribution.** Panel (a) plots both the autocovariance structure of the spread in the calibrated statistical model (blue) and in the data (orange). Panel (b) presents the percent of variance explained by ex-ante factors as a function of the loan number, according to the statistical model.

6 Interpretation of firm fixed effects

The analysis so far has shown that a substantial share of the variation in interest rates across firms is due to latent firm factors that are highly persistent over time. These account for approximately 30% of the total variation in interest rates, and are the single largest source of this variation. 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.

The starting point for this discussion is to understand the characteristics that an explanation must have in order to be consistent with the empirical analysis. In some detail, we will discuss three necessary properties for an explanation.

First, it must be firm level. The explanation needs to account for idiosyncratic variation in interest rates across firms, and it cannot 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 char-

acteristics, bank and bank-time fixed effects, and our analysis of relationship banking.[49] Thus the focus should be on the characteristics of the firm itself.

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 have been smaller, and more variation would have been captured by the residuals of the regressions. Explanations that generate transitory differences in rates across firms are therefore not enough. An explanation does not have to imply permanent differences in rates across firms, but the differences need to be persistent enough to generate the large fixed effects that we see. This suggests a focus on highly persistent characteristics of firms.

Third, the explanation must be uncorrelated 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 again that we have controlled for differences across firm size, firm age, 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 we typically consider.

One specific class of explanation that the firm controls rule out, is variation due to differences in rates across local banking markets. Since firms typically don't move location once they are established, there could be persistent variation in interest rates across firms if competition in local banking markets varies, and this effects equilibrium interest rates. However, our firm controls include the geographic location of firms up to 4-digit postcodes. Therefore market-specific effects at this geographic level are controlled for.

A second class of explanation that warrants additional 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. Explanations of this

[49] 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.

nature are difficult to square with the data. It is not enough for such shocks to cause temporary changes in interest rates paid by firms. The variation in the data is highly persistent within firms, 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 a different capital stock, which is controlled for. If the source of variation is persistent differences in liquidity needs, then this likely to show up in higher liabilities or trade payables, which we control for.

We now turn to discussing three explanations that could plausibly fit the three necessary requirements. The first explanation would require a combination of two factors: frictions in the market for loans between banks and firms that allow for price differences to be sustained in equilibrium, and an idiosyncratic firm characteristic causing these price differences to persistently benefit some firms relative to others. It is not hard to imagine frictions in the market for bank loans that could sustain price differences. To obtain a loan, firms typically need to apply to a bank, wait for the application to be processed, and then, if the bank is willing to lend, the firm receives a loan offer with specific terms. The time and effort involved in doing this makes it costly to compare options across banks. Furthermore, with a small number of banks dominating the market (recall Figure 1), competitive pressures to reduce price variation may be limited. Within this environment, there are several characteristics of firms that are unobservable and could generate persistent differences in rates. Differences in efficiency in applying for loan offers will determine a firm's capacity to compare offers across banks and obtain a better rate. Firms can differ in their information about the rates that banks are willing to lend at, affecting their decisions about how much to search and what rates to accept. They can also differ in their ability to bargain with banks. While this type of heterogeneity has received more attention in consumer finance than firm finance, the idiosyncrasies of individuals could transfer to the firms that they own or operate more than is typically assumed.

This explanation clearly satisfies the first two requirements. Its main challenge is the third requirement. Suppose that you have two firms that are observationally equivalent

which are obtaining new bank loans. If one of them is able to obtain a better rate due to one of the unobservable characteristics just discussed, then in a typical model of firm investment, the firm with a better rate will invest more and grow larger over time. This implies a correlation between the rate and observable firm characteristics. The question is whether this correlation is quantitatively strong enough that the difference in interest rates is empirically accounted for by firm observables.

In a context in which the market for loans is frictional, there could alternatively be variation in rates for non-economic reasons. One such reason is differentiation between firms based on personal relationships between the loan officer and the representative of the firm (e.g. owner or manager), or socio-demographic characteristics of this person. Random variation in rates is also possible. This could be due to variation across loan officers within a bank, or variation over time within loan officers. A challenge for this class of explanations is that differences in rates across firms need to be persistent, but this is not insurmountable. The initial interest rate that a firm obtains could act a reference for future loans with the same bank. This will then also influence rates obtained from other banks in the future, because the existing rate at the first bank will act as a reference point. This explanation faces the same challenge as the first one: will the observable characteristics of firms end up being too highly correlated with interest rates?

The third explanation that we consider is that the financial position of the owners of a firm affect its lending terms. This effect would need to be above and beyond the ability of the owners to provide a guarantee that the loan is repaid, because we have already controlled for default risk and the effect of collateral. However, as an example, a bank could be willing to lend at a lower rate to owners who also have deposits, investments or loans from the bank. In other words, rates could depend on the bundle of services provided by a bank to a firm and its owners. At a broad level, this is a firm-specific source of variation, it is persistent since firm ownership is very stable, and it is unobservable in the data. Since we see that differences in rates across firms persistent even when they change banks, this explanation would require that owners have private financial relationships with the multiple banks that their firms do business with, or that loan pricing is influenced by the financial position of firms owners, even if they don't have a personal

financial relationship with the bank.

To provide guidance on the plausibility of this theory, we look at how the role of firm fixed effects in interest rate variation changes with firm size. The assumption for this analysis is that, on average, owners of firms are less likely to have a direct role in procuring financing the larger the firm is. We use the framework outlined in Section 5.1 and run the analysis for subsamples of firms grouped according to size. The results show that the estimated contribution of firm fixed effects to total variation in interest rates does not diminish with firm size. We interpret this as as suggestive evidence against the personal finances of owners being an important determinant of firm interest rates.

7 Conclusion

Understanding the determinants of interest rates has been a long standing question in macroeconomics and finance. In recent decades its importance has grown as firm financing has become an increasingly important component of macroeconomic theory. This paper studies this question in the context of bank lending to firms, evaluating how core theories of interest rates quantitatively perform in the data, and what they are missing. This analysis is possible due to a detailed new credit registry, matched with firm balance sheet and income statement data.

We find that there is large variation in interest rates 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 characteristics, that are not closely connected to a broad range of firm observables, including industry, size, productivity and financial position.

The results imply constraints on how much interest rate variation should be generated by existing theories, and suggest that there may be substantial additional frictions in the market for firm financing. In terms of the nature of these frictions, the data dictates that

the main source of interest rate variation must satisfy three properties: (1) it must be an idiosyncratic firm characteristic; (2) it must be highly persistent; and (3) it must be uncorrelated with the broad range of firm characteristics observable in the data, including the standard characteristics that define firms in models (e.g. productivity, capital and financial position).

We discuss a set of plausible explanations that satisfy these requirements based on frictional markets and differences in firm ability to procure loans, differentiation between firms based on non-economics characteristics of owners or managers, and unobservable connections between the finances of firms and their owners. Important directions for future research are to determine the cause of rate dispersion, understand what this implies for the nature of financial frictions, and evaluate the broader implications for the economy.

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.

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

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., & Kehoe, P. J. (2019). Financial frictions and fluctuations in volatility. *Journal of Political Economy*, 127(5), 2049–2103.

Arellano, C., Bai, Y., & Zhang, J. (2012). Firm dynamics and financial development. *Journal of Monetary Economics*, 59(6), 533–549.

Bellifemine, M., Jamilov, R., & Monacelli, T. (2022). Hbank: Monetary policy with heterogeneous banks. Working paper.

Benetton, M. (2021). Leverage regulation and market structure: A structural model of the uk mortgage market. *The Journal of Finance*, 76(6), 2997–3053.

Berger, A. N. & Hannan, T. H. (1998). The efficiency cost of market power in the banking industry: A test of the "quiet life" and related hypotheses. *Review of economics and statistics*, 80(3), 454–465.

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.

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

Berlin, M. & Mester, L. J. (1992). Debt covenants and renegotiation. *Journal of Financial Intermediation*, 2(2), 95–133.

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. (1989). Agency costs, net worth, and business fluctuations. *American Economic Review*, 79(1), 14–31.

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).

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

Bolton, P., Freixas, X., Gambacorta, L., & Mistrulli, P. E. (2016). Relationship and transaction lending in a crisis. *Review of Financial Studies*, 29(10), 2643–2676.

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. *The American economic review*, (pp. 1165–1183).

Boot, A. W. & Thakor, A. V. (1994). Moral hazard and secured lending in an infinitely repeated credit market game. *International economic review*, (pp. 899–920).

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.

Campbell, T. S. & Kracaw, W. A. (1980). Information production, market signalling, and the theory of financial intermediation. *Journal of Finance*, 35(4), 863–882.

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

Clymo, A. & Rozsypal, F. (2023). *Firm cyclicality and financial frictions*. Technical report,

Danmarks Nationalbank Working Papers.

Cooley, T., Marimon, R., & Quadrini, V. (2004). Aggregate consequences of limited contract enforceability. *Journal of Political Economy*, 112(4), 817–847.

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.

Degryse, H. & Ongena, S. (2005). Distance, lending relationships, and competition. *The Journal of Finance*, 60(1), 231–266.

Dempsey, K. & Faria-e-Castro, M. (2024). A quantitative analysis of bank lending relationships. Working paper.

Diamond, D. W. (1984). Financial intermediation and delegated monitoring. *The review of economic studies*, 51(3), 393–414.

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.

Freixas, X. & Rochet, J.-C. (2008). *Microeconomics of banking*. MIT press.

Gilchrist, S., Sim, J. W., & Zakrajšek, E. (2014). Uncertainty, financial frictions, and investment dynamics. Working paper.

Giroud, X. (2013). Proximity and investment: Evidence from plant-level data. *The 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.

Haltiwanger, J., Jarmin, R. S., & Miranda, J. (2013). Who creates jobs? small versus large versus young. *Review of Economics and Statistics*, 95(2), 347–361.

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. (2014). The life cycle of plants in india and mexico *. *Quarterly Journal of Economics*, 129(3), 1035–1084.

Ioannidou, V. & Ongena, S. (2010). "time for a change": loan conditions and bank behav-

ior when firms switch banks. *Journal of Finance*, 65(5), 1847–1877.

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.

Kermani, A. & Ma, Y. (2023). Asset specificity of nonfinancial firms. *Quarterly Journal of Economics*, 138(1), 205–264.

Khan, A., Senga, T., & Thomas, J. K. (2021). Default risk and aggregate fluctuations in an economy with production heterogeneity. Working paper.

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.

Kochen, F. (2025). Finance over the life cycle of firms. Working paper.

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

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

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.

Manova, K. (2013). Credit constraints, heterogeneous firms, and international trade. *Review of Economic Studies*, 80(2), 711–744.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The 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. *The journal of finance*, 49(1), 3–37.

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

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

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

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

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

Santos, J. A. & Winton, A. (2019). Bank capital, borrower power, and loan rates. *Review of Financial Studies*, 32(11), 4501–4541.

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

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

Spencer, A. H. (2022). Policy effects of international taxation on firm dynamics and capital structure. *Review of Economic Studies*, 89(4), 2149–2200.

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

Stiglitz, J. E. & Weiss, A. (1981). Credit rationing in markets with imperfect information. *The American economic review*, 71(3), 393–410.

Von Thadden, E.-L. (1995). Long-term contracts, short-term investment and monitoring. *The Review of Economic Studies*, 62(4), 557–575.

Wheelock, D. C. & Wilson, P. W. (2012). Do large banks have lower costs? new estimates of returns to scale for us banks. *Journal of Money, Credit and Banking*, 44(1), 171–199.

Wheelock, D. C. & Wilson, P. W. (2018). The evolution of scale economies in us banking. *Journal of Applied Econometrics*, 33(1), 16–28.

Appendix

Contents

A Additional data details **2**
- A.1 Sample construction and composition . . . 2
- A.2 Banking market and firm financing sources: comparison with Europe . . . 4
- A.3 Interest rates vs. spreads . . . 5
- A.4 Controls for loan characteristics . . . 6
- A.5 Descriptive statistics for sample of all instruments . . . 8

B Robustness exercises and extensions for Section 3.2 **10**
- B.1 All credit instruments . . . 10
- B.2 Banks PDs instead of regulator PDs . . . 11
- B.3 Validation of PD distributions . . . 11
- B.4 PD and LGD distributions that would be need to explain rates . . . 12
- B.5 Alternative stochastic discount factors . . . 14
- B.6 More flexible specifications for risk . . . 16
  - B.6.1 Imposing risk coefficient to be minus one . . . 16
  - B.6.2 Semi-parametric estimation . . . 17
- B.7 Validation that selective LGD reporting does not drive results . . . 17
- B.8 Low risk loans only . . . 18
- B.9 Extension to longer maturity loans . . . 19
- B.10 Amortization type . . . 19
- B.11 Exclusion of Covid period . . . 19
- B.12 Firm groups . . . 20
  - B.12.1 Age groups . . . 21
  - B.12.2 Size groups . . . 22

C Robustness exercises and extensions for Section 4 **22**
- C.1 Distribution of loans across banks and firms . . . 22
- C.2 Bank-time and firm-time fixed effect specification . . . 22
- C.3 All instruments . . . 25
- C.4 Amortization type . . . 27
- C.5 Exclusion of Covid period . . . 27
- C.6 Firm groups . . . 27
  - C.6.1 Age groups . . . 28
  - C.6.2 Size groups . . . 29

D Robustness exercises and extensions for Section 4.2 **29**
- D.1 All credit instruments . . . 29
- D.2 Extension of firm-bank to firm-bank-time fixed effects . . . 30

E Additional details for Section 5.1 **31**
E.1 All credit instruments 31
E.2 Age one controls in the second stage 32
E.3 Excluding firm observables in the first stage 32

F Loan pricing details **33**
F.1 PD and LGD definitions for one period loans 33
F.2 Extension to longer maturity loans 34
F.3 Interest rates 36

A Additional data details

A.1 Sample construction and composition

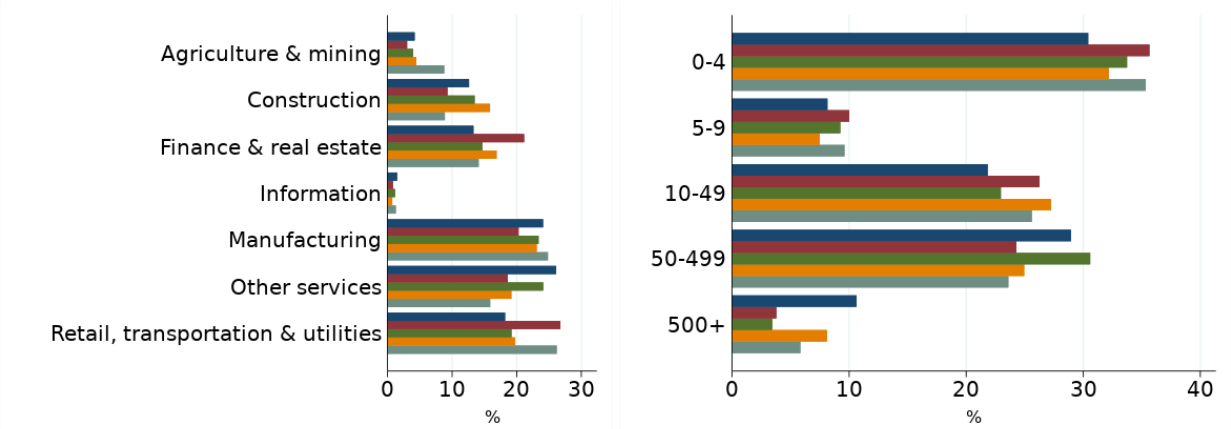
This section provides additional details about the distribution of loans issued by each of the five major banks, the comparison of our sample to the universe of Portuguese firms, how we compute the share of aggregate moments covered by our sample, the share of credit from banks covered by our sample, and a comparison of the size distributions of our firms and employer firms in the US.

Figure 1(a) in the main text presents the geographic distribution of loans for each of the five main banks in 2019. In Figure 7 we replicate this exercise for the distribution across industries and firm size to illustrate that the major banks have similar operations, and are not specialized or dividing the market.

Table 9 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.[50] Relative to Portugal, the sample substantially underrepresents zero-employee firms and overrepresents larger firms, indicating an upward size bias in employment relative to 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.

Finally, we compare the firm size distribution of our sample with that of employer

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



(a) Industry distribution of credit, by bank

(b) Firm size distribution of credit, by bank

Figure 7: **Distribution of firm financing loans across 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) and (b) show the distributions across industries and firm size categories. The color for each bank is the same as in Figure 1(a) in the main text. Firm size is measured with the number of employees.

Size indicator	Number of firms		Employment Share	
	Sample	Portugal	Sample	Portugal
0 employees	12.3%	69.8%	-	-
1 - 9 employees	68.8%	26.2%	20.2%	44.3%
10 - 49 employees	16.0%	3.4%	27.6%	193%
50 - 249 employees	2.6%	0.5%	22.3%	15.2%
$>$ 250 employees	0.4%	0.1%	30.0%	21.2%

Table 9: **Firm size distributions in our sample and in Portugal.** Distribution of firms and employment shares by firm size category in 2019. Data for Portugal correspond to the sample used in this study; Data for Portugal are from the Statistics Portugal, Integrated business accounts system.

firms in the U.S. in 2019. The U.S. data comes from the Business Dynamics Statistics of the Census Bureau. Table 10 reports the distribution of firms across employment size categories, as well as the corresponding employment shares. In both countries, the firm size distribution is skewed toward small firms, but more so in Portugal. Portugal has more weight on all size categories below 500 employees, and less weight above this threshold. Overall, the average firm size is 11.3 employees in our sample, compared to 24.7 employees for U.S. employer firms.

Size indicator	Number of firms		Employment Share	
	Sample	USA	Sample	USA
1 - 4 employees	49.8%	57.5%	9.5%	5.2%
5 - 9 employees	12.2%	20.9%	10.7%	5.1%
10 - 19 employees	6.0%	11.5%	11.9%	6.5%
20 - 99 employees	4.2%	8.6%	25.9%	16.3%
100 - 499 employees	0.6%	1.3%	19.6%	14.0%
$>$ 500 employees	0.03%	0.2%	22.4%	52.9%

Table 10: **Firm size distributions in our sample and the U.S.** Distribution of firms and employment shares by firm size category. Data for Portugal correspond to 2019 data for firms in our sample with active firm financing loans in 2019; U.S. data are from Business Dynamics Statistics for 2019.

A.2 Banking market and firm financing sources: comparison with Europe

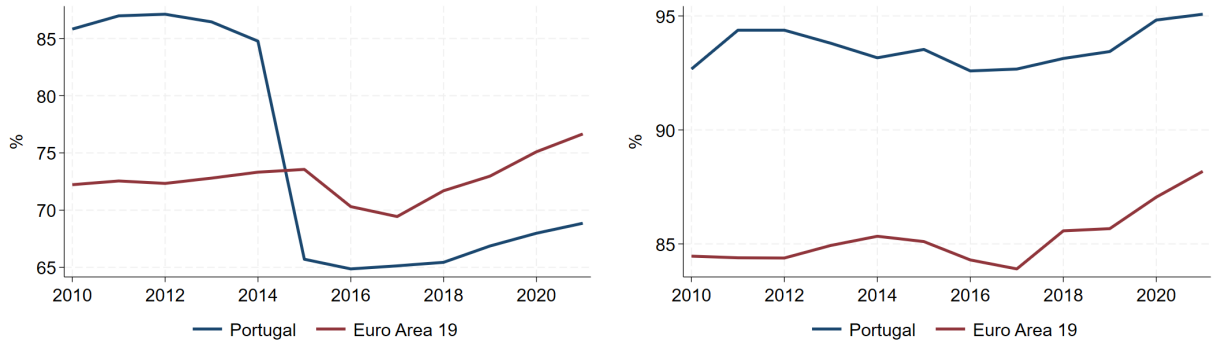
In this section, we report additional details for two data moments reported in Section 2: the market shares of the largest banks in the Euro Area 19 countries, and the degree of dependence on bank financing of firms in these countries.

For banking market shares, we compare the share of total banking sector assets held by the three and five largest banks between 2010 and 2021 in Portugal, with the average for the Euro Area 19 countries. The data are from the World Bank's Global Financial Development Indicators.[51] In both Portugal and the Euro Area, large banks have a similar presence. They account for 84–88% of assets in Euro Area countries, and 92–95% in Portugal. The shares for the largest three banks are also similar. The sharp decline in this share in 2015 in Portugal was due the resolution of Banco Espírito Santo, one of the country's largest banks at the time.

The second moment to explain is the share of external firm financing accounted for by credit from banks in Portugal and the Euro Area 19 countries. Quarterly data covering 2010:Q1 to 2022:Q4 are drawn from the European Central Bank.[52] The measure of total

[51]The respective series are: assets of the three largest commercial banks as a share of total commercial banking assets ("bank concentration," code GFDD.OI.01); and assets of the five largest commercial banks as a share of total commercial banking assets ("five-bank asset concentration," code GFDD.OI.06). Total assets include total earning assets, cash and due from banks, foreclosed real estate, fixed assets, goodwill, other intangibles, current and deferred tax assets, discontinued operations, and other assets.

[52]The series used are: Loans to non-financial corporations (codes QSA.Q.N.PT.W0.S11.S1.N.L.LE.F4.T._Z.XDC._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



(a) Three largest banks

(b) Five largest banks

Figure 8: **Share of banking sector assets held by largest banks.** Panels (a) and (b) show the share of total banking sector assets held by the three and five largest banks, respectively, between 2010 and 2021. The blue line corresponds to Portugal, and the red line is the average across the 19 Euro Area countries. Source: World Bank, Global Financial Development Indicators.

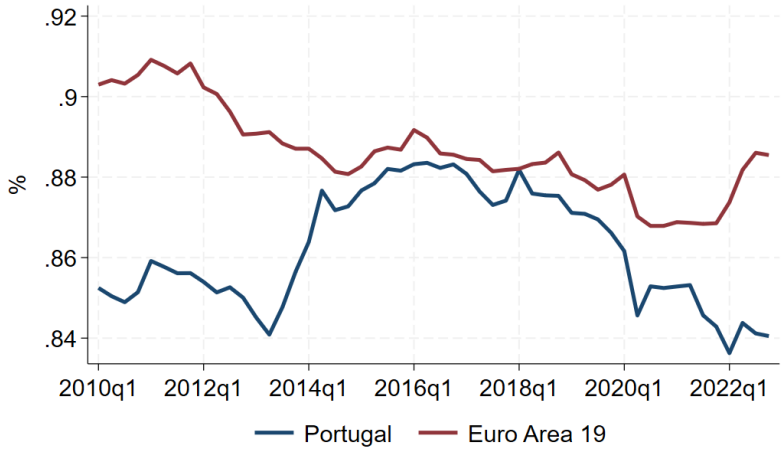


Figure 9: **Share of bank loans in total firm credit.** Share of loans in total loans and debt securities issued by non-financial corporations, 2010–2022. The blue line represents Portugal, and the red line is the average across the 19 Euro Area countries. Source: European Central Bank.

external financing is the sum of credit from banks and debt securities. Figure 9 shows that firms in both Portugal and the Euro Area rely heavily on bank credit. On average, it accounts for 86.2% of total external financing for Portuguese firms and 88.6% for firms across the Euro Area.

A.3 Interest rates vs. spreads

As outlined in Section 2, we analyze interest rates instead of spreads since spreads are not reported for all loans in the data. This could be important if it means that we neglect

the Euro Area).

heterogeneity in banks' cost of funds. We assess whether this omission is material using the set of loans for which spreads are reported. These loans make up 78% of the loans in our main sample. We begin by constructing a loan-level proxy for the risk-free rate. For each loan l extended by bank b at time t, we define the risk-free component of the interest rate as

$$rf_{lbt} = r_{lbt} - s_{lbt},$$

where r_{lbt} denotes the contractual interest rate and s_{lbt} is the reported spread.

Banks' effective cost of financing varies systematically with the maturity of loans and the issuance date. To purge such variation, we regress the constructed risk-free rate on maturity–month–year fixed effects, denoted by $\mathcal{M}_{lt}$:

$$rf_{lbt} = \mathcal{M}_{lt} + \varepsilon_{lbt}.$$

This specification absorbs all variation in benchmark rates associated with the term structure and aggregate time effects, isolating residual differences in banks' cost of funds.

Figure 10 shows that the residual component of banks' financing costs is 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 1 basis points. Thus, cross-bank heterogeneity in funding costs—after controlling for maturity and time fixed effects—is quantitatively small, and treating interest rates, rather than spreads, as the relevant object of analysis is unlikely to confound firm-level interest rate dispersion with variation in banks' own cost of funds.

A.4 Controls for loan characteristics

The full details of the controls for loan characteristics that are used throughout the analysis are as follows:

- **Interest rate adjustability**: The loan 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%) frequency, or at the lender's discretion (0.1%). There are three additional

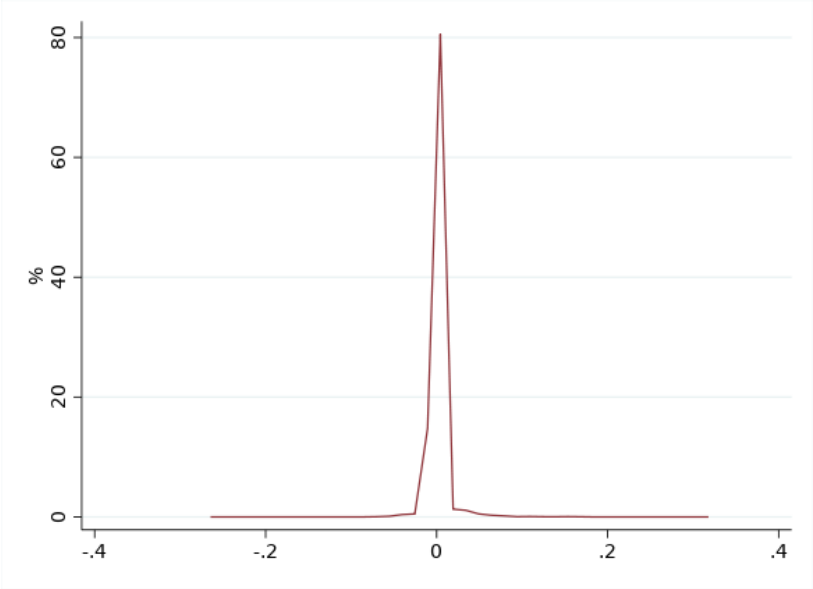


Figure 10: **Variation in bank financing costs.** This figure presents the distribution of risk-free interest rates after controlling for maturity-month-year fixed effects.

frequencies of adjustment in the data, other frequencies less than a year (0.03%), frequencies greater than one year (0.02%), and a category for all residual cases (9.2%).

- **Frequency of repayment**: The loan can be repaid with monthly installments (42.5% of loans), quarterly (5.0%), biannual (0.4%), annual (0.5%), unique payment of principal (3.9%), unique payment of principal and interests (16.3%), and a category for all residual cases (31.3%).
- **Type of collateral**: No collateral (52% of loans), 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%).
- **Right to immediate repayment**: For 16.2% of loans, the creditor has the right to demand immediate repayment of the loan.
- **Amortization type**: The amortization method can be French (fixed installments, with interest and principal changing over time - 23.9% of loans), German (fixed principal installment, with interest installment changing over time - 0.2%), fixed (6.6%), unique payment of capital (25.5%), increasing installments (0.01%) and category for all residual cases (43.4%).
- **Purpose of the loan**: the loan 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 without a specific purpose (63.2%).

- **Number of protections**: More than one asset can be pledged as collateral. 75.7% of loans only have one asset pledged, 14.6% have two assets, 4.8% have three assets, 2.2% have 4 assets and 2.8% have 5 assets or more pledged.
- **Securitized loan**: Banks can also securitize loans. 97.7% of loans are securitized.
- **Loan special characteristics**: Loans can have no special characteristics (98.2% of loans), or one special characteristic or more. The special characteristics can be loans granted under Covid related policies or guarantees (0.6%), consumer credit loans (1.4%) and a category for all residual cases (0.3%).
- **Maturity**: Loan maturities are categorized as follows (in years): (0, 0.25], (0.25, 0.5], (0.5, 1], and then in yearly intervals from (1, 2] up to (29, 30].

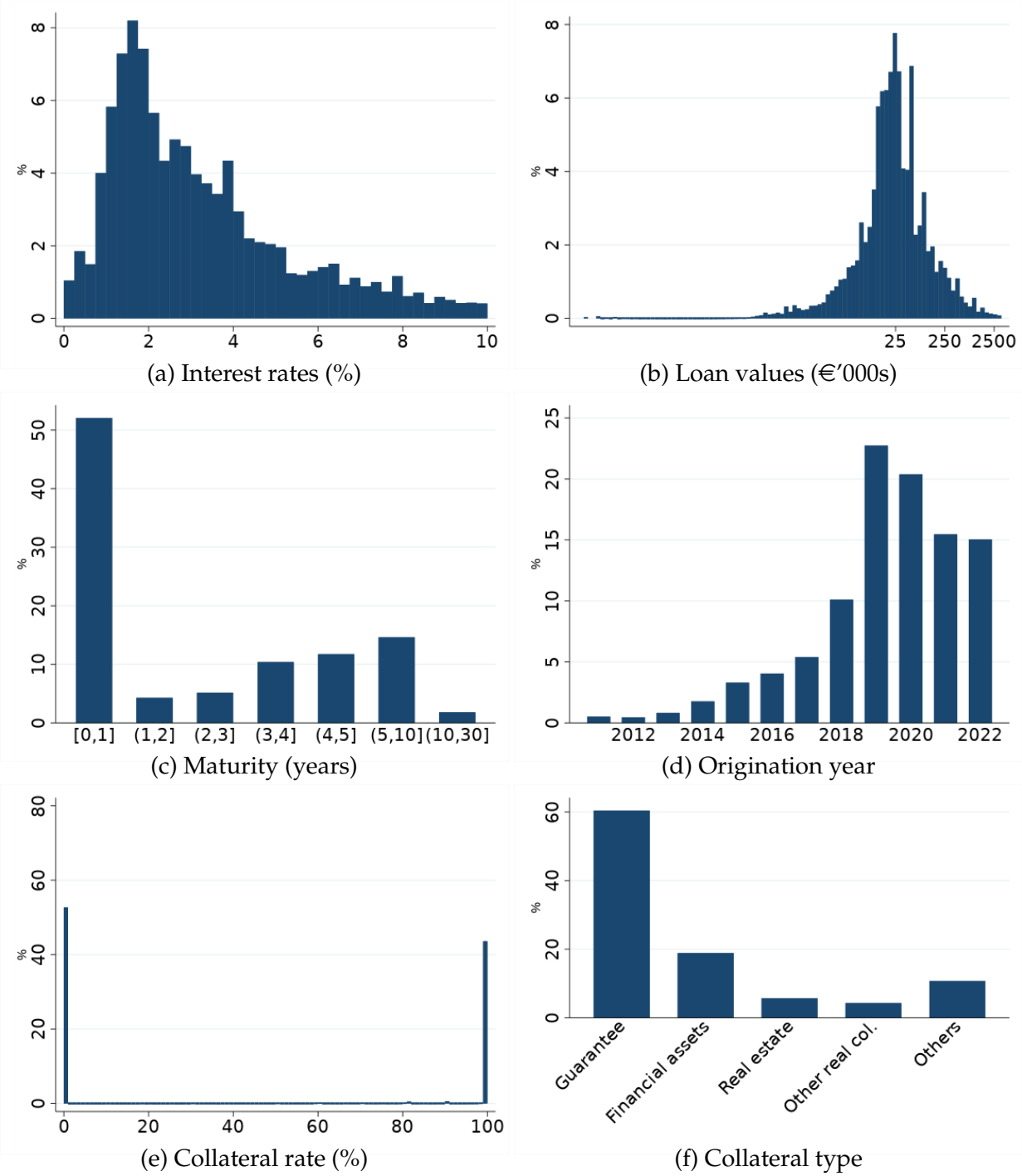
A.5 Descriptive statistics for sample of all instruments

For all of the main results of in the paper, we verify that they hold when we use the full sample of all credit instruments. In this section we introduce the additional instruments and provide descriptive statistics about them.

35% of total credit issued by banks to firms in the sample is what we refer to as "firm financing loans" or simply "loans,". The other main classes of credit are factoring and confirming—credit lines received to finance trade credit received or granted (49%)—, credit cards (7%), auto loans—loans conceded for car purchases (5%)— and mortgages (2%).

Figure 11 replicates Figure 2 from the main text, but now includes all types of credit given by banks to firms. The interest rate and collateral rate distribution hardly change (panels a and d). These other types of loans 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.

To illustrate the characteristics of the firms in the sample when considering all credit instruments we present summary statistics from their loan portfolios, balance sheets and



(a) Interest rates (%)

(b) Loan values (€'000s)

(c) Maturity (years)

(d) Origination year

(e) Collateral rate (%)

(f) Collateral type

Figure 11: **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 distribution in panel (f) is conditional on loans having some collateral.

income statements in Table 11.[53] The statistics are within firm averages over the years

[53]Notice that there are firms that have no firm financing loans which have other credit instruments. This

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

Table 11: **Descriptive statistics for firms.** 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.

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 €63,400 in loans. The median number of loans is 6 and the median share of a firm's loan value that is collateralized is 50% The majority of firms have loans from only one bank, but many firms have more bank relationships. At the 90th percentile a firm averages 3 banking relationships over the years that it is in the sample. Overall, firms in this sample are slightly smaller and younger than firms use in the baseline sample, which considers only firm financing loans.

B Robustness exercises and extensions for Section 3.2

B.1 All credit instruments

To increase the number of observations we extend the analysis to include all credit instruments. This increases the number of observations in the baseline specification from 25,460 to 108,847. Figure 12 shows that results are unchanged by extending the sample to

results in the sample of firms across the two the samples to differ.

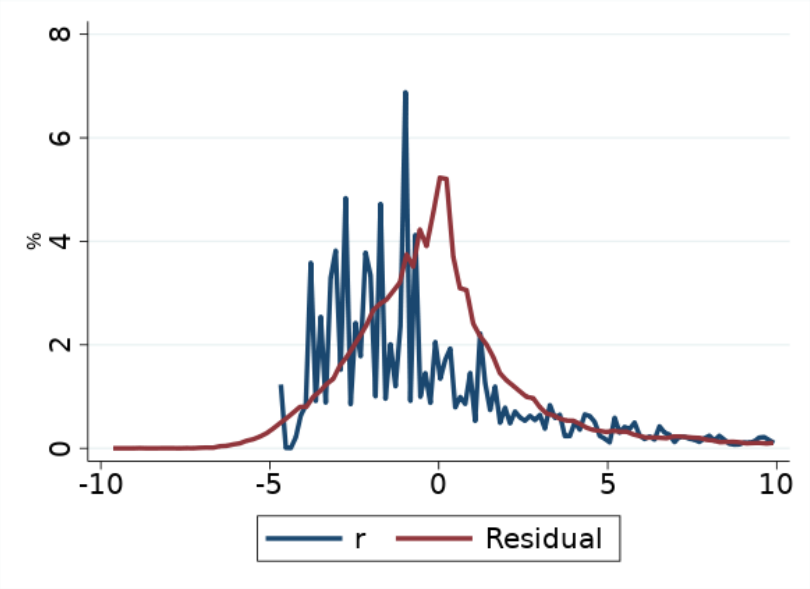


Figure 12: **Interest rates and risk - All credit instruments.** The raw interest rate distribution (blue) and after controlling for loan characteristics, risk, bank and maturity-time effects (red). Both distributions are centered with their means at 0.

include all credit instruments. In this case, risk explain 0.4% of the total variation.

B.2 Banks PDs instead of regulator PDs

The regulator-provided PDs are estimated using balance sheet and income statement data and may not fully capture the information available to banks or their subjective assessments of a firm's probability of default. To ensure that our results are not driven by the use of regulator PDs, we re-estimate equation (4) using the PDs reported by banks themselves. Figure 14 shows that relying on banks' own PDs does not increase the explanatory power of risk, confirming the robustness of our results.

B.3 Validation of PD distributions

An important detail for the analysis is the quality of the reported PDs. We want to be confident that PDs are measured well in the data. In order to do this we use the fact that we observe loans through to completion, so that we can ex post observe whether default occurs. To evaluate the quality of the PD estimates we take their values from the data, rank them from lowest to highest, group them into 100 bins based on this ranking, and then compare the average values of the estimated with the ex post realized rates

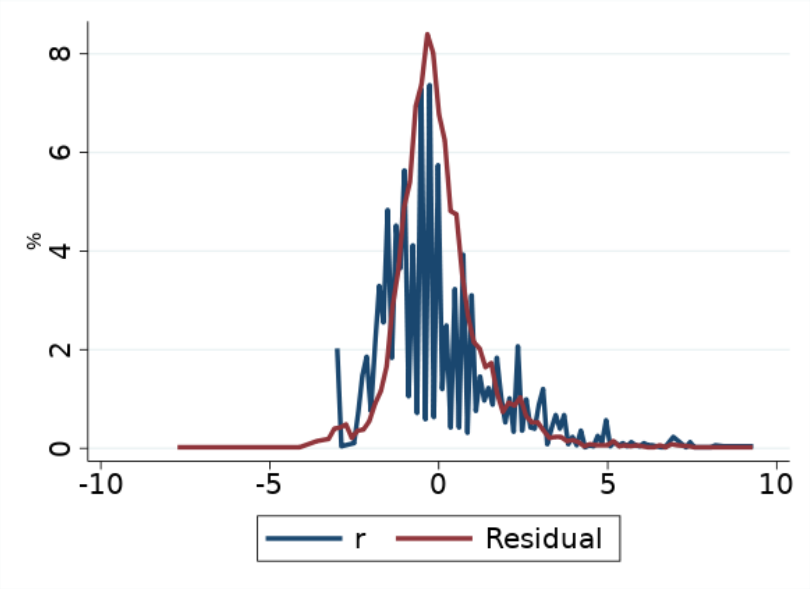


Figure 13: **Interest rates and risk - Banks PDs.** The raw interest rate distribution (blue) and after controlling for loan characteristics, risk, bank and maturity-time effects (red). Both distributions are centered with their means at 0.

of default. The results are presented in Figure 14 and show that the variables are high quality. Realized default rates are monotonically increasing in the ex ante estimates and quantitatively they closely track each other.

B.4 PD and LGD distributions that would be need to explain rates

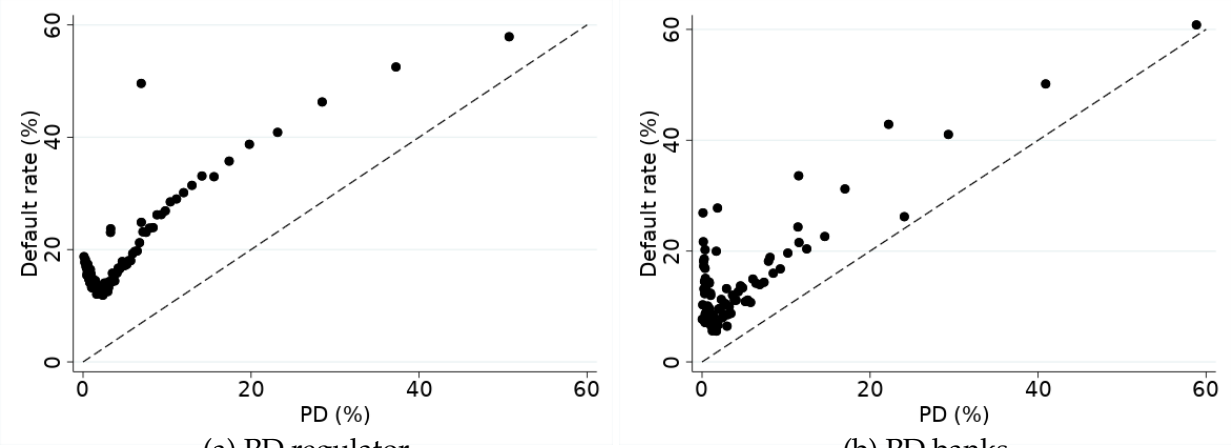
To have a better idea of what the LGD and PD distributions would need for risk to fully explain the interest rate distribution, we invert the problem and given the interest rate and LGD (PD) distribution, we calculate the PD (LGD) distribution needed to generate an interest rate distribution similar to the one observed in the data.

Continuing on the methodology proposed above, we take equation (4) and compute the interest rate on loan l from bank b at time t cleaned of controls other than risk as the residual ϵ_{lbt}. Then use the equation

$$-\log(1 - PD_{lt}LGD_{lt}) = \epsilon_{lbt}, \tag{13}$$

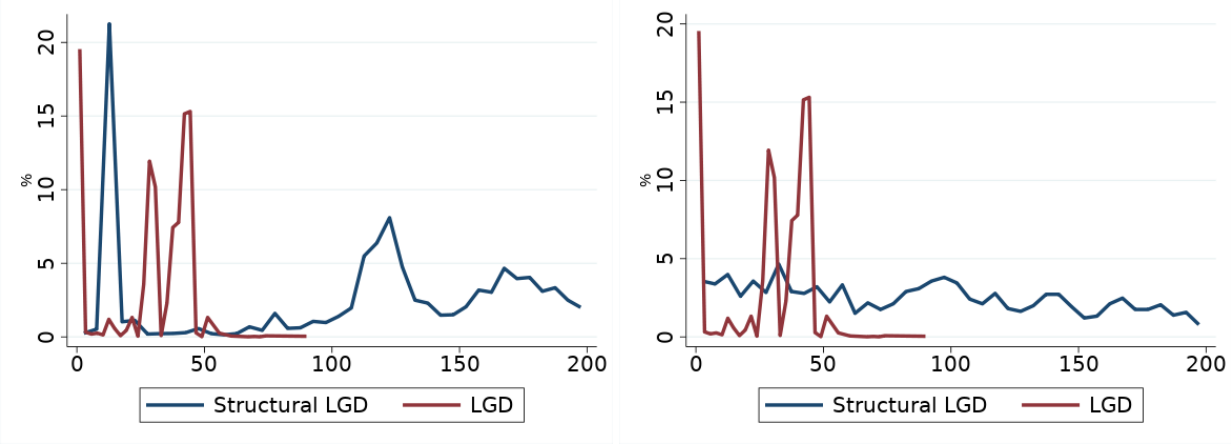
to determine what values of PD_{lt} and LGD_{lt} would be needed to fit the data.[54]

[54]One potential problem with this approach is that, by construction, ϵ_{lbt} is centered around zero, which could force LGD and PD into negative values as well. To go around this, we add to ϵ_{lbt} the constant



(a) PD regulator (b) PD banks

Figure 14: **Estimated vs. realized PDs.** In panel (a) the PDs from the regulator are ranked and sorted into 20 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.



(a) PDs estimated by the Bank of Portugal. (b) PDs reported by the commercial banks.

Figure 15: **LGDs: actual vs. required to match interest rate distribution.** Probability density function of the required LGDs to match the interest rate distribution following equation (15) in blue, and of the actual LGD distribution in red.

LGD We start by calculating the LGD needed to match the interest rate distribution, given the PD distribution. The LGD is given by

$$LGD_{lt} = \frac{1 - \exp(\epsilon_{lbt})}{PD_{lt}}. \tag{14}$$

Figure 15 plots on the left hand panel the probability density function (PDF) of estimated LDG using the PDs estimated by the Bank of Portugal and on the right hand panel

estimated from equation (4).

the same object using the PDs reported by the commercial banks. The reported LGD distribution, in red in both panels, is concentrated below 50%, which translates into banks expected actual loss for loans to be below 50% of the loan amount. In contrast, the estimated LGD distributions generate a large fraction of values above 100%, which would mean the bank would need to lose more money than the loan amount to justify the interest rate on some of the loans. In fact, we limit the x-axis on both panels of Figure 15 to 200, but the median is 573 and 253 for panel (a) and (b) respectively. So, the LGD distributions would need to be infeasibly high for risk to explain the interest rate distribution, given the PD distribution estimated both by the Bank of Portugal and commercial banks.

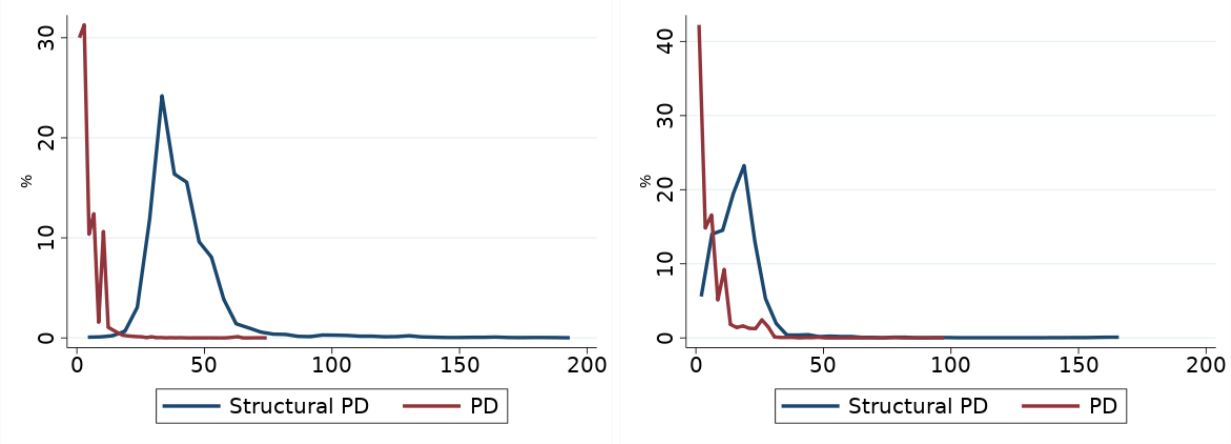
PD We next calculate, given the observed LGD, what would the PD distribution need to be to generate a interest rate distribution in line with the observed one, given the LGD distribution. The PD is given by

$$PD_{lt} = \frac{1 - \exp(\epsilon_{lbt})}{LGD_{lt}}. \tag{15}$$

Figure 16 plots the PDFs for both the structural PD and the observed PD. On the panel (a) the estimated and observed PDs for the universe of loans for which we have LGD estimates and on panel (b) the estimated and observed PDs for the universe of loans for which we have both LGDs and PDs reported by commecial banks. It is possible to observe in the Figure that the PDs required to generate an interest rate distribution in line with the empirically observed one, there would need to be much more default risk in the economy. While between 30% and 40% of the reported PDs – depending on the measure used – are close to 0%, the estimated median PD is 42%, which is higher than the 90th percentile in the distribution reported by the Bank of Portugal. Additionally, the estimated PDs have an average above 100%. So, similar to the conclusions when estimating the required LGDs, the required PDs to match the observed interest rate distribution are infeasibly high.

B.5 Alternative stochastic discount factors

The main specification assumes that lenders are risk neutral and price the loans according to the risk-free rate. In principle, lenders could discount the future at a different



(a) PDs estimated by the Bank of Portugal. (b) PDs reported by the commercial banks.

Figure 16: **PDs: actual vs. required to match interest rate distribution.** Probability density function of the required PDs to match the interest rate distribution following equation (15) in blue, and of the actual LGD distribution in red.

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. To account for this we want to control for the correlation between the likelihood that a firm defaults in a given state and the valuation of the lender in the state when default occurs. For a variable with information about potential firm default we choose firm revenues. For the value of states by lender, we take guidance from consumption-based asset pricing theories and use the value of aggregate consumption. We then take the correlation between these two variables and add it to equation (4) interacted with the risk term. Therefore, we estimate the following specification

$$r_{lbt} = -\log(1 - PD_{lt}LGD_{lt}) + \beta_1 \Lambda_f \log(1 - PD_{lt}LGD_{lt}) + \beta_2 \mathbf{L}_{lt} + \mathcal{M}_{lt} + \mathcal{B}_b + \varepsilon_{lbt}, \quad (16)$$

where Λ_f is the correlation between firm f's revenues and aggregate consumption during the period the firm is in the sample. Results in Figure 17 show that the addition of the correlation coefficient Λ_f does not increase significantly the explanatory power of the risk component.

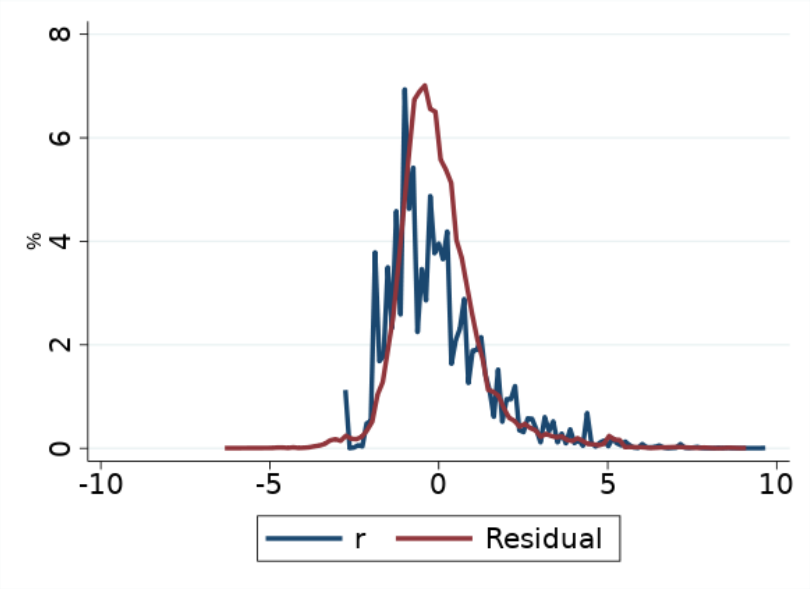
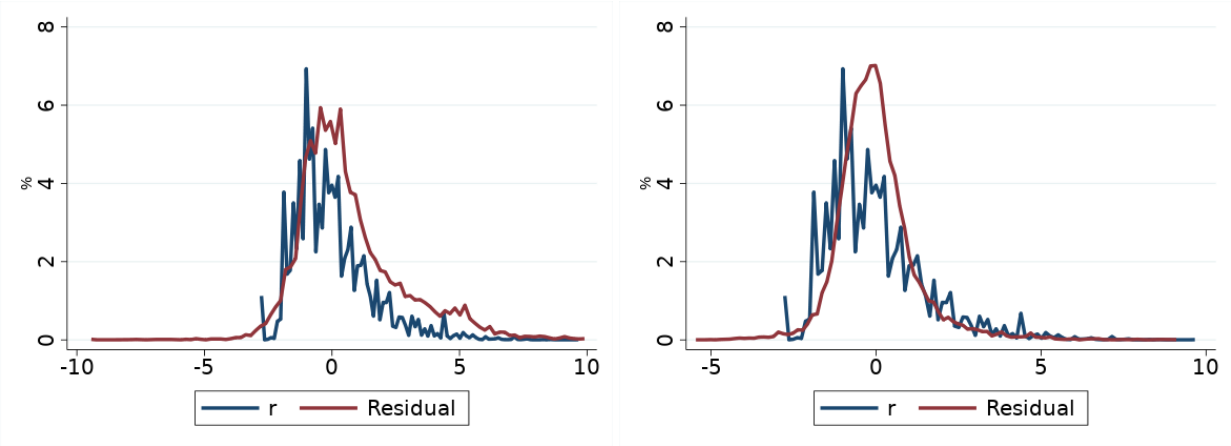


Figure 17: **Interest rates and risk - alternative SDF.** The raw interest rate distribution (blue) and after controlling for loan characteristics, risk, correlation coefficient between state of the firm and aggregate state, bank and maturity-time effects (red). Both distributions are centered with their means at 0.



(a) Imposing coefficient to -1

(b) Semi-parametric

Figure 18: **Interest rates and risk.** The raw interest rate distribution (blue) and after controlling for loan characteristics, risk, bank and maturity-time effects (red). Both distributions are centered with their means at 0. Panel (a) when estimating the coefficient on the risk term. Panel (b) for the semi-parametric estimation.

B.6 More flexible specifications for risk

B.6.1 Imposing risk coefficient to be minus one

In regression (4) we allow the coefficient on the risk term to be freely estimated, while the theory suggests the coefficient should be -1. While this is the dominant approach in

macroeconomic models, it is possible that this is misspecified and the true relationship takes another form. In fact the estimated coefficient is -0.08 and statistically different from zero. But even if we take the theory seriously and impose the coefficient to be -1 the results are not meaningfully different. To test this, we estimate the following specification

$$r_{lbt} = -\log(1 - PD_{lt}LGD_{lt}) + \beta_2 \mathbf{L}_{lt} + \mathcal{M}_{lt} + \mathcal{B}_b + \varepsilon_{lbt}. \tag{17}$$

Results in Figure 18(a) illustrates that imposing the coefficient on the risk term to be -1 does not improve the explanatory power of risk.

B.6.2 Semi-parametric estimation

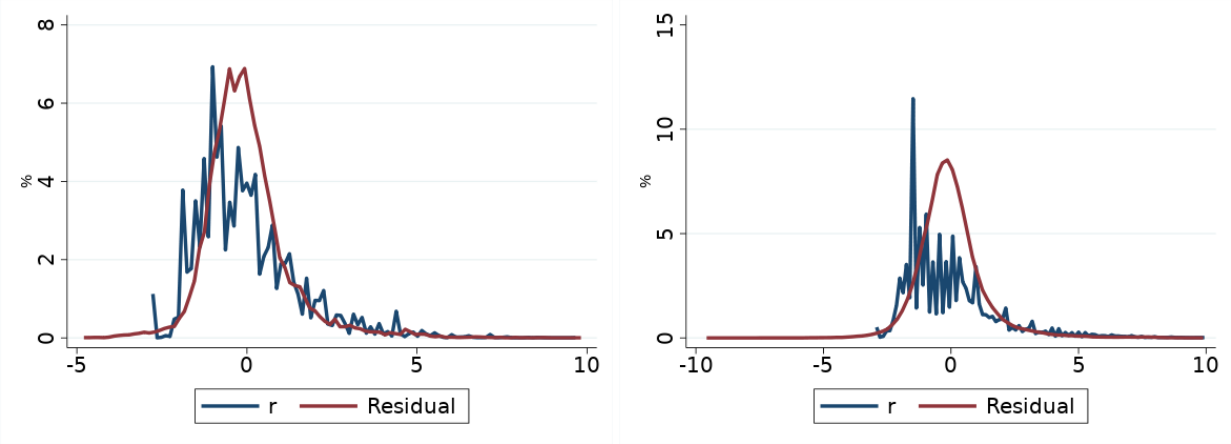
To be more flexible with the relationship between interest rates and risk, we also use a semi-parametric specification of the following form:

$$\begin{aligned} r_{lbt} = & \sum_{p \in \mathcal{PD}} \gamma_p PD_{lt} \mathbf{1}_{i \in \mathcal{S}^{(P)}} + \sum_{d \in \mathcal{LGD}} \theta_d LGD_{lt} \mathbf{1}_{i \in \mathcal{S}^{(D)}} + \sum_{d \in \mathcal{LGD}} \sum_{p \in \mathcal{PD}} \lambda_{dp} PD_{lt} LGD_{lt} \mathbf{1}_{i \in \mathcal{S}^{(P)}} \mathbf{1}_{i \in \mathcal{S}^{(D)}} \\ & + \sum_{p \in \mathcal{PD}} \sigma_p PD_{lt}^2 \mathbf{1}_{i \in \mathcal{S}^{(P)}} + \sum_{d \in \mathcal{LGD}} \alpha_d LGD_{lt}^2 \mathbf{1}_{i \in \mathcal{S}^{(D)}} + \sum_{d \in \mathcal{LGD}} \sum_{p \in \mathcal{PD}} \mu_{dp} PD_{lt}^2 LGD_{lt}^2 \mathbf{1}_{i \in \mathcal{S}^{(P)}} \mathbf{1}_{i \in \mathcal{S}^{(D)}} \\ & + \beta_1 \mathbf{L}_{lt} + \mathcal{M}_{lt} + \mathcal{B}_b + \varepsilon_{lbt}. \end{aligned} \tag{18}$$

The terms $\mathcal{S}^{(P)}$ and $\mathcal{S}^{(D)}$ represents the p^{th} PD and d^{th} LGD group, respectively. We partition the sample into ten deciles for each of the PD and LGD distribution 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. Results in Figure 18(b) illustrate that even with this more flexible specification, risk still doesn't explain much of the interest rate variation.

B.7 Validation that selective LGD reporting does not drive results

The LGD variable is only available for a subsample of loans, as banks don't report LGDs for the universe of loans. To guarantee our results are not driven by selection of the subsample we replicate the semi-parametric exercise but using only the regulator PDs,



(a) LGD sample (b) Full sample

Figure 19: **Interest rates and risk - PD only.** The raw interest rate distribution (blue) and after controlling for loan characteristics, risk, bank and maturity-time effects (red). Both distributions are centered with their means at 0. Panel (a) when estimating the semi-parametric on the sample for which we have LGD values. Panel (b) for the semi-parametric estimation on the full sample.

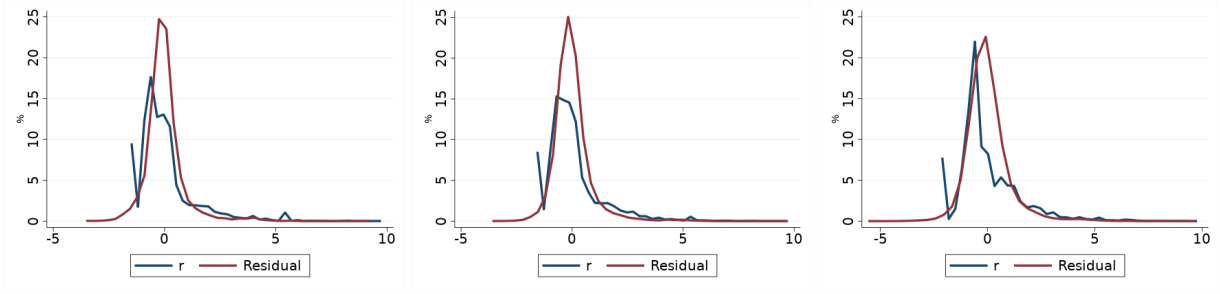
for which we have values for the full sample. We estimate the following specification

$$r_{lbt} = \sum_{p \in \mathcal{PD}} \gamma_p PD_{lt} \mathbf{1}_{i \in \mathcal{S}^{(P)}} + \sum_{p \in \mathcal{PD}} \sigma_p PD_{lt}^2 \mathbf{1}_{i \in \mathcal{S}^{(P)}} + \beta_1 \mathbf{L}_{lt} + \mathcal{M}_{lt} + \mathcal{B}_b + \varepsilon_{lbt}. \quad (19)$$

First, to guarantee our results in the full sample are not driven by the omission of the LGD variable, we estimate the above specification in the subsample for which we have LGD values. Figure 19(a) illustrates that results are similar to when equally including LGD. Then, we extend the analysis to the entire sample. Figure 19(b) illustrates that results are robust to including the full sample in the risk analysis.

B.8 Low risk loans only

To circumvent concerns about how risk is specified in the regression, as well reducing concerns about measurement that we discussed earlier, we restrict the sample to loans with a low PD. To implement this we restrict the sample to loans with a PD below 0.29%, 0.55% and 1.14%, respectively the 5th, 10th and 25th percentile of the PD distribution. Then we estimate equation (1) for the three different groups. Results for the three groups in Figures 20(a), 20(b) and 20(c) reflect than even for low risk loans, after controlling for loan characteristics, bank and maturity-time fixed effects, majority of variance in interest rates remains.



(a) PD 5^{th} percentile (b) PD 10^{th} percentile (c) PD 25^{th} percentile

Figure 20: **Interest rates and risk - low PDs.** The raw interest rate distribution (blue) and after controlling for loan characteristics, bank and maturity-time effects (red). Both distributions are centered with their means at 0. Panel (a) PDs in the bottom 5^{th} percentile. Panel (b) PDs in the bottom 10^{th} percentile. Panel (c) PDs in the bottom 25^{th} percentile

B.9 Extension to longer maturity loans

To ensure as much consistency between the theory and data as possible, the baseline analysis restricts the sample to loans with a maturity of up to one year. We here relax this assumption and re-estimate equation (4) for loans of all maturities.[55] Figure 21 reflects that results are robust to the inclusion of longer maturity loans.

B.10 Amortization type

The theory assumes that both the interest and the principal are only repaid when the loan matures. In the data, for 28.6% of loans, the principal is repaid in a unique repayment. To guarantee as much consistency with the theory as possible, we re-estimate (4) for loans with a unique repayment of principal. Additionally, to guarantee the results are not driven by a specific amortization type, we re-estimate (4) restricting to the other two largest amortization types, French (fixed install- ments, with interest and principal changing over time - 20.9% of loans) and category for all residual cases (45%). Figure 22 illustrates that results are robust across the different amortization types.

B.11 Exclusion of Covid period

To guarantee results are not driven by the Covid period and potential guarantees offered by the government offered during this period that would eliminate risk for banks, we

[55]For a detailed discussion of how to extend the risk equation to longer maturities, please see F.2.

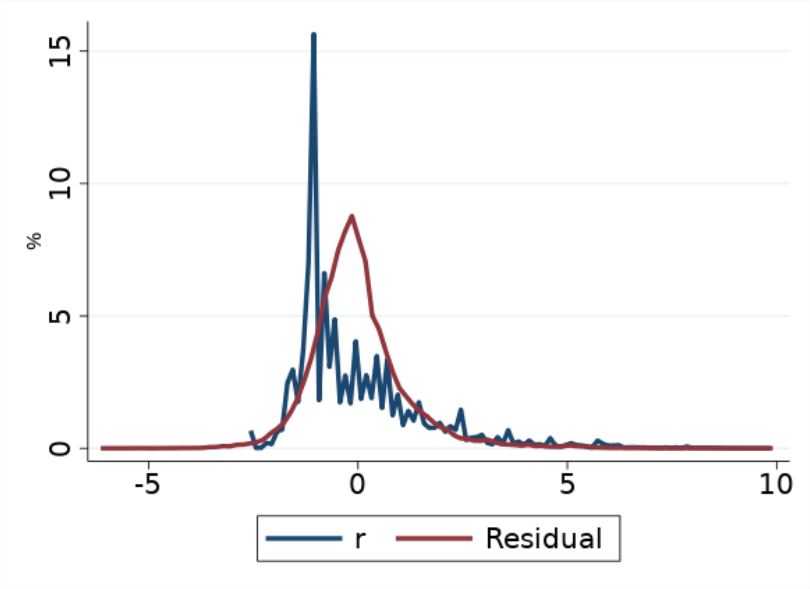
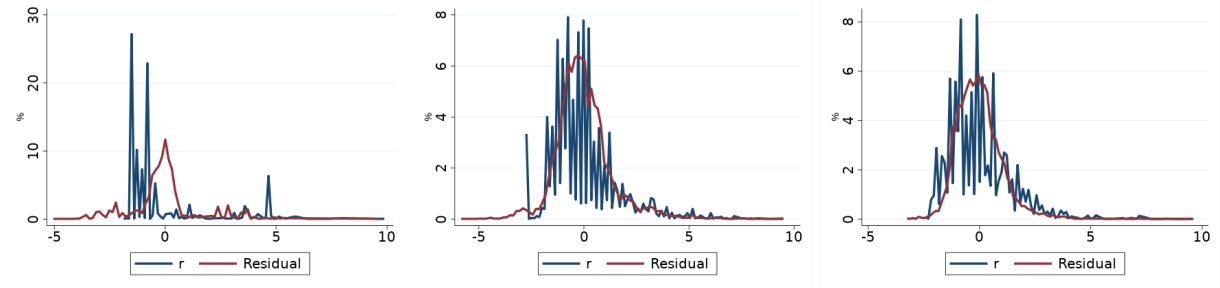


Figure 21: **Interest rates and risk - all maturities.** The raw interest rate distribution (blue) and after controlling for loan characteristics, risk, bank and maturity-time effects (red). Both distributions are centered with their means at 0.



(a) Unique payment (b) French amortization (c) Others

Figure 22: **Interest rates and risk - amortization type.** The raw interest rate distribution (blue) and after controlling for loan characteristics, bank and maturity-time effects (red). Both distributions are centered with their means at 0. Panel (a) for loans with a unique payment of capital. Panel (b) for loans with a French type amortization. Panel (c) for loans with other amortization types.

re-estimate equation (4) for loans initiated prior to 2020. Figure 23 illustrates results are robust to the exclusion of the Covid period.

B.12 Firm groups

To guarantee results are not driven by a specific group of firms, we repeat the analysis considering sub-groups of the firm distribution.

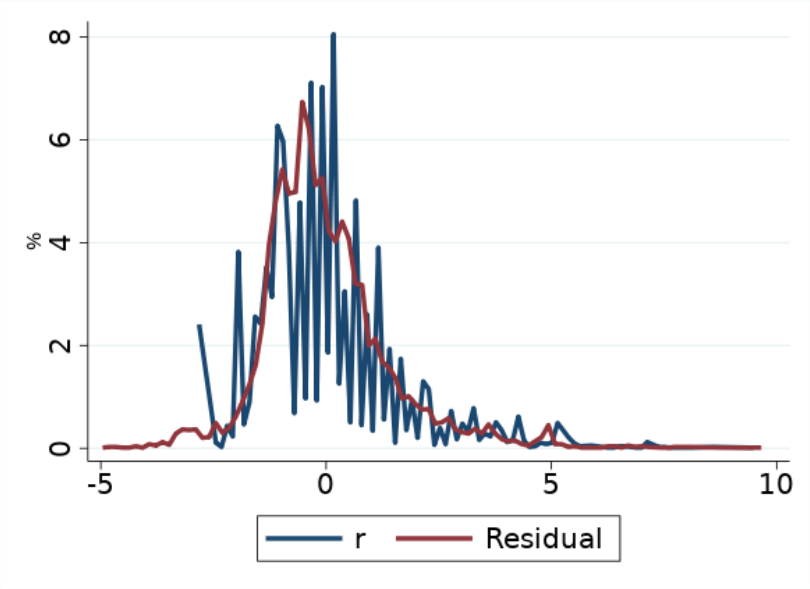
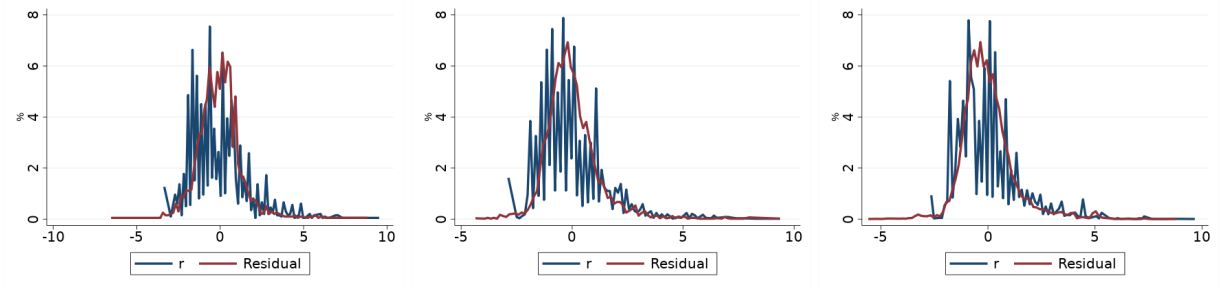


Figure 23: **Interest rates and risk - loans initiated pre-2020.** The raw interest rate distribution (blue) and after controlling for loan characteristics, risk, bank and maturity-time effects (red). Both distributions are centered with their means at 0.



(a) Young (b) Mature (c) Old

Figure 24: **Interest rates and risk - Age groups.** The raw interest rate distribution (blue) and after controlling for loan characteristics, bank and maturity-time effects (red). Both distributions are centered with their means at 0. Panel (a) for young firms. Panel (b) for mature firms. Panel (c) for old firms.

B.12.1 Age groups

We split firms into three different age groups, following Clymo & Rozsypal (2023). Young firms, below 5 years of age, mature firms, between 5 and 15 years of age, and old firms, older than 15 years. We then re-estimate equation (4) for the different groups of firms. Figure 24 illustrates that results are robust across firms of different ages.

B.12.2 Size groups

We split firms according to number of employees, following Clymo & Rozsypal (2023). Small firms are below 5 employees, small-medium firms between 5 and 12 employees, medium firms between 13 and 40 employees, medium-large firms between 40 and 120 employees and large firms above 120 employees. We then re-estimate equation (4) for the different groups of firms. Figure 25 illustrates results are robust across firms of different sizes.

C Robustness exercises and extensions for Section 4

C.1 Distribution of loans across banks and firms

Figure 26 presents the distribution of number of loans and number of bank relationships by firm over time. 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 have many loans originating from multiple banks. Figure 27 plots the distribution of number of loans and firm relationships by bank over time. Opposite to the same distributions by firm, majority of banks issue thousands of loans and have relationships with thousands of firms.

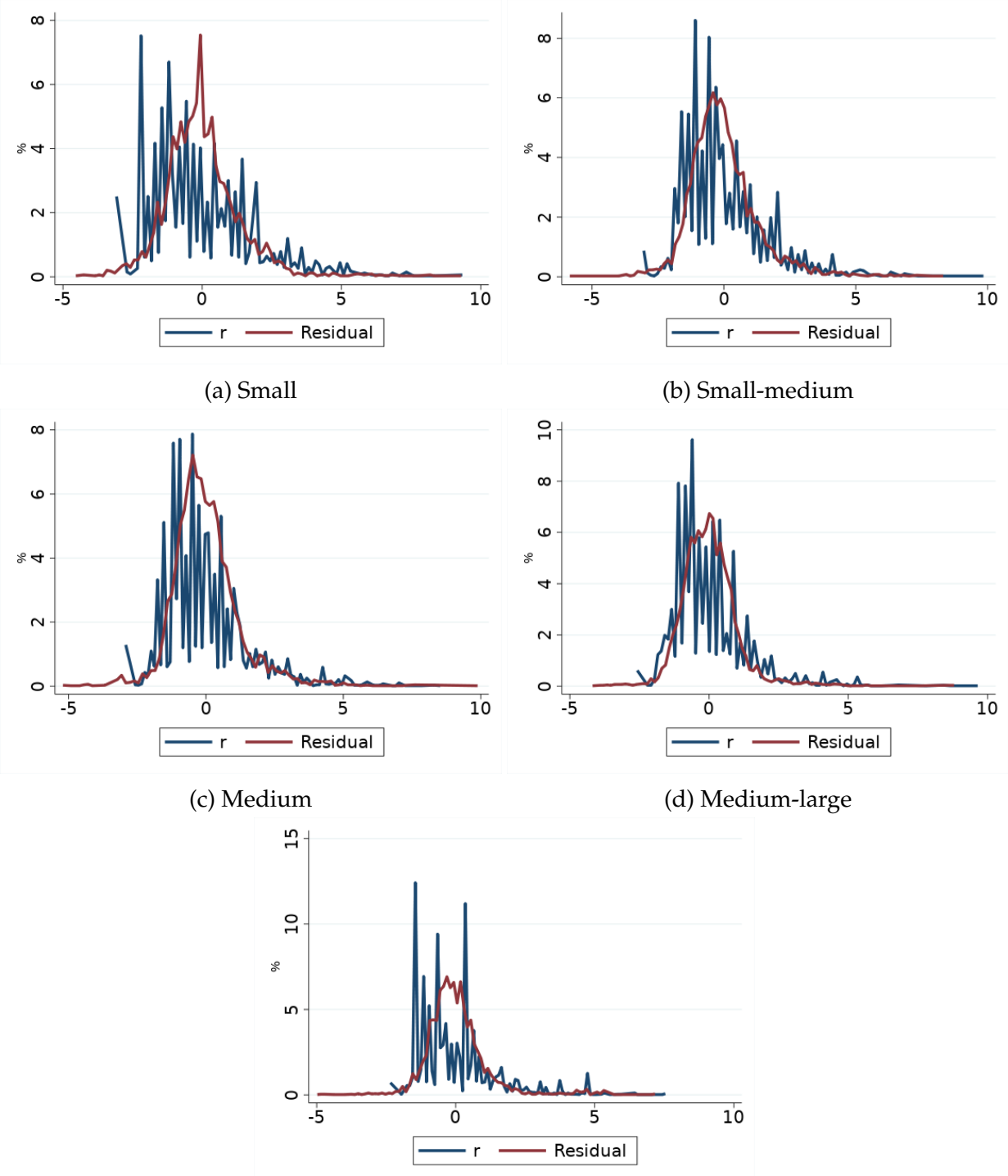
C.2 Bank-time and firm-time fixed effect specification

We extend the baseline analysis in Section 4.1 to include firm-time $\mathcal{F}_{ft}$ and bank-time $\mathcal{B}_{bt}$ fixed effects

$$r_{lfbt} = \beta_1 \mathbf{R}_{lt} + \beta_2 \mathbf{L}_{lt} + \mathcal{M}_{lt} + \mathcal{B}_{bt} + \mathcal{F}_{ft} + \varepsilon_{lfbt}. \tag{20}$$

To estimate the above specification, we need firms that originate multiple loans in the same time period. We use a one year time period, enough to capture changes in firm and time characteristics while preserving close to 50% of the original sample.[56] To estimate the importance of each component, we again use the Abowd et al. (1999) (AKM) method. For each loan-firm-bank-time observation, let $\hat{r}^c_{lfbt}$ denote the value of component c from

[56] Number of firms in the sample goes down from 84,535 to 41,312.



(a) Small

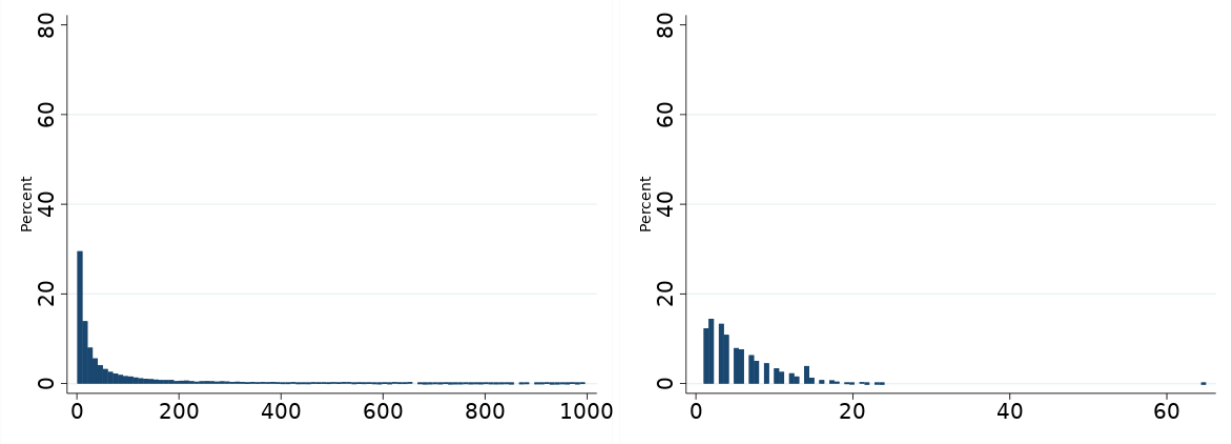
(b) Small-medium

(c) Medium

(d) Medium-large

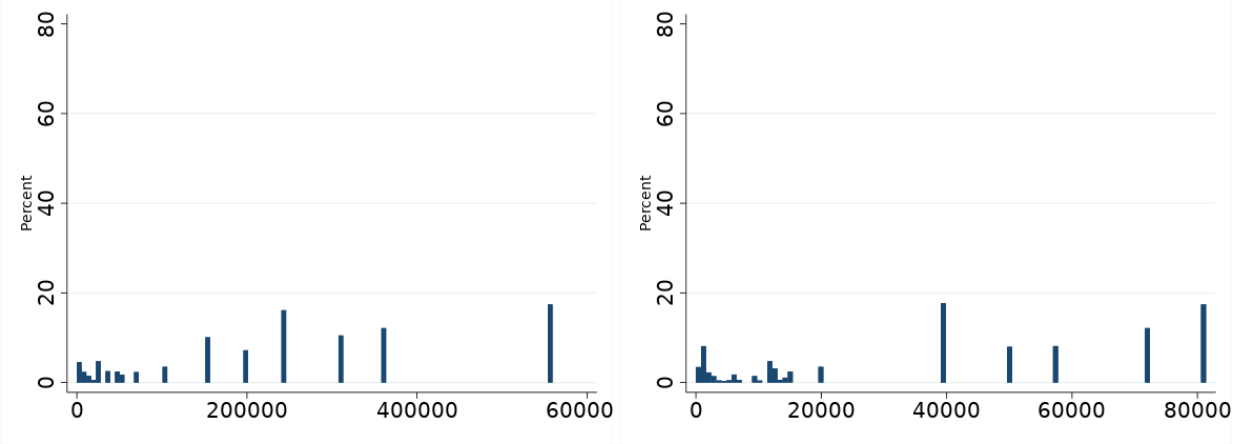
(e) Large

Figure 25: **Interest rates and risk - PD only.** The raw interest rate distribution (blue) and after controlling for loan characteristics, risk, bank and maturity-time effects (red). Both distributions are centered with their means at 0. Panel (a) for small firms. Panel (b) for small-medium firms. Panel (c) for medium firms. Panel (d) for medium-large firms. Panel (e) for large firms.



(a) Loans (b) Banks

Figure 26: **Loan and bank relationship distribution across firms.** On the left hand side the distribution of number of number of loans by firm over time. On the right hand side the distribution of number of bank relationships by firm over time.



(a) Loans (b) Firms

Figure 27: **Loan and firm relationship distribution across banks.** On the left hand side the distribution of number of number of loans by bank over time. On the right hand side the distribution of number of firm relationships by bank over time.

the right-hand side of the estimated regression. The overall variance of interest rates is given by sum of the variances of each of the N components on the right-hand side plus the covariances between the different components. This is,

$$var(r_{lfbt}) = \sum_{c=1}^{N} var(\hat{r}^{c}_{lfbt}) + \sum_{j=1}^{N} \sum_{c=1}^{N} cov(\hat{r}^{c}_{lfbt}, \hat{r}^{j}_{lfbt}), \quad j \neq c. \tag{21}$$

Table 12 presents the results. Overall results are qualitatively in line with the results presented in Section 4.1, with the firm component still having the largest explanatory

	Contribution
(1) Risk + Loan char. + Matur-Time Component	0.1641
(2) Bank-Time Component	0.0957
(3) Firm-Time Component	0.6172
(4) Residual	0.1698
(5) Covariances	-0.0467
# Obs.	216,522

Table 12: **Contribution of loan characteristics, firm-time, bank-time and maturity-time fixed effects to interest rate variation.** The different lines represent the different components included in equation (22). Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (21).

power. Quantitatively, the main difference comes from the increased explanatory power of the firm-time fixed effect. While the firm component accounted for close to 33% of the overall variation in the baseline regression, the firm-time component accounts for 61% of the variation, absorbing part of the explanatory power of loan characteristics and maturity-time fixed effect contributions and diminishing the variance of the residual.

Additionally, as the firm-time fixed effect is the component restricting the sample, we estimate the contributions of each component when including the bank-time FE but only a firm fixed effect. As banks issue multiple loans every time period, the time component can now be month-year, similar to the time component on the maturity-time fixed effect. In this case, the number of observations in the sample is almost identical, decreasing from 521,856 to 518,739. We estimate the following specification

$$r_{lfbt} = \beta_1 \mathbf{R}_{lt} + \beta_2 \mathbf{L}_{lt} + \mathcal{M}_{lt} + \mathcal{B}_{bt} + \mathcal{F}_f + \varepsilon_{lfbt}. \tag{22}$$

Results presented in Table 13 illustrate that even when including the bank-month-year fixed effects, the most important component is still the firm fixed component.

C.3 All instruments

We extend the baseline analysis in Section 4.1 to include all instruments in the sample. We estimate the following specification

$$r_{lfbt} = \beta_1 \mathbf{R}_{lt} + \beta_2 \mathbf{L}_{lt} + \mathcal{M}_{lt} + \mathcal{B}_{b} + \mathcal{F}_f + \varepsilon_{lfbt}. \tag{23}$$

	Contribution
(1) Risk + Loan char. + Matur-Time Component	0.1104
(2) Bank-Time Component	0.2917
(3) Firm Component	0.3191
(4) Residual	0.2410
(5) Covariances	0.0378
# Obs.	518,739

Table 13: **Contribution of loan characteristics, firm, bank-time and maturity-time fixed effects to interest rate variation.** The different lines represent the different components included in equation (22). Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (21).

	Contribution
(1) Risk + Loan char. + Matur-Time Component	0.3794
(2) Bank Component	0.0971
(3) Firm Component	0.3686
(4) Residual	0.2518
(5) Covariances	-0.0969
# Obs.	1,598,466

Table 14: **Contribution of loan characteristics, firm, bank and maturity-time fixed effects to interest rate variation. All instruments included.** The different lines represent the different components included in equation (22). Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (21).

The instrument type is now additionally included as a fixed effect in the vector $\mathbf{L}_{lt}$ In this case the number of firms in the sample increase from 84,535 to 135,350 and the number of observations goes from 521,856 to 1,598,466. This strengthens even more the fixed effect identification, as we now have almost twice as many observations per firm and more within firm variation. Nonetheless, results are qualitatively and quantitatively very stable. The firm component still is the most important, with its importance almost identical to the baseline results—changing from 32.7% to 37.3%. The largest change came from the increase in importance of the loan characteristics, which can be explained by the inclusion of a instrument type fixed effect.

	Contribution		
	Unique repayment	French	Other
(1) Risk + Loan char. + Matur-Time Component	0.1948	0.3207	0.2221
(2) Bank Component	0.0647	0.0911	0.0910
(3) Firm Component	0.5406	0.4107	0.4132
(4) Residual	0.2124	0.1603	0.1409
(5) Covariances	-0.0125	0.0173	0.1328
# Obs.	89,488	56,823	163,153

Table 15: **Contribution of loan characteristics, firm, bank and maturity-time fixed effects to interest rate variation by amortization types.** The different lines represent the different components included in equation (22). The different columns represent the estimates for the three different amortization types: unique repayment of capital, French and residual cases. Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (21).

C.4 Amortization type

We here repeat the robustness exercise done for risk and re-estimate equation (6) for the same three different amortization types: unique repayment of principal, French amortization and all residual cases. Results are similar across the different amortization types, with the firm fixed component always being the most important and the bank component not a major driver of interest rate variation.

C.5 Exclusion of Covid period

To guarantee results are not driven by the Covid period and any potential compression of rates or changes in the banking market that may have been caused by government policies, we re-estimate equation (6) for loans initiated prior to 2020. Results in Table 16 indicate that the overall qualitative and quantitative conclusions do not change by excluding loans originated during the Covid period.

C.6 Firm groups

To guarantee results are not riven by specific groups of firms, we re-estimate equation (6) for firms of different sizes and ages.

	Contribution
(1) Risk + Loan char. + Matur-Time Component	0.1902
(2) Bank Component	0.1004
(3) Firm Component	0.3895
(4) Residual	0.2565
(5) Covariances	0.0633
# Obs.	248,490

Table 16: **Contribution of loan characteristics, firm, bank and maturity-time fixed effects to interest rate variation. Loans initiated pre-2020.** The different lines represent the different components included in equation (22). Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (21).

	Contribution		
	Young	Mature	Old
(1) Risk + Loan char. + Matur-Time Component	0.3899	0.3604	0.2437
(2) Bank Component	0.1495	0.1201	0.0792
(3) Firm Component	0.2894	0.3087	0.3859
(4) Residual	0.2026	0.2548	0.2811
(5) Covariances	-0.0315	-0.0441	0.0102
# Obs.	72,390	151,635	273,319

Table 17: **Contribution of loan characteristics, firm, bank and maturity-time fixed effects to interest rate variation by age groups.** The different lines represent the different components included in equation (22). The different columns represent the estimates for the three different age groups: young, mature and old. Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (21).

C.6.1 Age groups

We split firms into the same three groups as in Appendix B.12.1 and estimate equation (6) separately for each of the three different groups. Results in Table 17 indicate that the conclusions are robust across the three age groups. For young, mature and old firms, the firm fixed component is always the most important component. Results indicate that this firm fixed component is increasing on age, while the loan characteristics and bank component are more important for young firms.

	Contribution				
	Small	Medium-small	Medium	Medium-Large	Large
(1) Risk + Loan char. + Matur-Time Component	0.3723	0.3438	0.3090	0.3603	0.2239
(2) Bank Component	0.1462	0.1351	0.1120	0.0765	0.0533
(3) Firm Component	0.2640	0.2796	0.3252	0.4040	0.5803
(4) Residual	0.2276	0.2693	0.2939	0.2284	0.2634
(5) Covariances	-0.0100	-0.0278	-0.0400	-0.0691	-0.1209
# Obs.	137,468	115,493	119,218	73,346	54,023

Table 18: **Contribution of loan characteristics, firm, bank and maturity-time fixed effects to interest rate variation by size groups.** The different lines represent the different components included in equation (22). The different columns represent the estimates for the five different size groups: small, medium-small, medium, medium-large and large. Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (21).

C.6.2 Size groups

We split the firm into five different size groups, the same in Appendix B.12.2. We then estimate equation (6) for each of the five groups separately. Results in Table 18 illustrate the results are qualitatively similar across groups, with the firm fixed component always being the most important. Similar to the results by age groups, the firm component increases on firm size, accounting for 70% of the overall variation in rates for the largest firms. Additionally, while the covariance terms are relatively small and unimportant across the four groups of smaller firms, they become important for the largest firms. The increased importance of the covariance terms comes from a strong sorting of the largest firms into specific banks (-8.5%).

D Robustness exercises and extensions for Section 4.2

D.1 All credit instruments

We extend the baseline analysis in Section 4.1 to include all instruments in the sample. We estimate the following specification

$$r_{lfbt} = \beta_1 \mathbf{R}_{lt} + \beta_2 \mathbf{L}_{lt} + \mathcal{M}_{lt} + \mathcal{FB}_{fb} + \varepsilon_{lfbt}. \tag{24}$$

	Contribution
(1) Bank Component	0.079
(2) Firm Component	0.377
(3) Relationship Component	0.177
# Obs.	1,469,225

Table 19: **Contribution of firm, bank and firm-bank fixed effects to interest rate variation. All credit instruments.** The different lines represent the different components. Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (7). The Relationship Component is presented as the difference between the estimated firm-bank fixed effect and the separate firm and bank fixed effects estimated on the same sample.

The instrument type is now additionally included as a fixed effect in the vector $\mathbf{L}_{lt}$ In this case the number of firms in the sample increase from 84,535 to 135,350 and the number of observations goes from 463,624 to 1,469,225. This strengthens even more the fixed effect identification, as we now have almost twice as many observations per firm and more within firm variation. Results presented in Table 19 illustrate that results are robust to the inclusion of all credit instruments. The overall importance of the relationship component increases slightly compared to the baseline scenario, but the bank and relationship component account for 25.5% of the overall variation, still considerable less than the firm component.

D.2 Extension of firm-bank to firm-bank-time fixed effects

As the firm-bank relationship may evolve over time, our baseline specification may underestimate the contribution of the relationship component. To address this, we further test the results when including a firm-bank-time fixed effect $\mathcal{FB}_{fbt}$:

$$r_{lfbt} = \beta_1 \mathbf{R}_{lt} + \beta_2 \mathbf{L}_{lt} + \mathcal{M}_{lt} + \mathcal{FB}_{fbt} + \varepsilon_{lfbt}. \tag{25}$$

This places additional restrictions on the sample, as this requires firms to initiate multiple loans from different banks in the same time period. The frequency for the time is one year. To guarantee the validity of the results with the reduce sample, we estimate specification (25) for firm financing loans only and for all credit instruments. Results in Table 20 illustrate that the firm component is still the most important one, with the bank and relationship components still accounting for less of the overall variation than the firm

	Contribution	
	Firm financing	All instruments
(1) Bank-Time Component	0.096	0.086
(2) Firm-Time Component	0.708	0.522
(3) Relationship-Time Component	0.004	0.154
# Obs.	262,876	1,159,644

Table 20: **Contribution of firm-time, bank-time and firm-bank-time fixed effects to interest rate variation.** The different lines represent the different components. The different columns represent the results when estimating equation (25) on the sample of firm financing loans only and on the sample with all credit instruments. Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (7). The Relationship-Time fixed effect is presented as the difference between the estimated firm-bank-time fixed effect and the separate firm-time and bank-time fixed effects estimated on the same sample.

	Contribution
(1) Firm FE	0.488
(1.1) Firm FE: obs component	0.036
(1.2) Firm FE: unobs component	0.452
# Firms	89,766

Table 21: **Contribution of firm fixed effect and firm fixed observable characteristics. All credit instruments.** The different lines represent the different components. Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (7).

one.

E Additional details for Section 5.1

E.1 All credit instruments

Our baseline sample focuses on firm financing relationships. As a first robustness check, we expand the sample to include all loans in the data.

Table 21 reports the results. When including all loans, firm fixed effects account for 49% of the total variation in interest rates. Of this, 3.6 percentage points are explained by observable firm characteristics, while 45.2 percentage points remain attributable to the unobservable component.

Relative to the baseline specification, the overall contribution of firm fixed effects decreases. Importantly, however, the bulk of the explanatory power continues to stem from

	Contribution
(1) Firm FE	0.762
(1.1) Firm FE: obs component	0.207
(1.2) Firm FE: unobs component	0.556
# Firms	5,030

Table 22: **Contribution of firm fixed effect and firm fixed observable characteristics. Age 1 controls.** The different lines represent the different components. Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (7).

the firm fixed component, which cannot be explained with firm observable fixed characteristics.

E.2 Age one controls in the second stage

Our second robustness exercise augments the second-stage regression (equation (10) in the main text) with firm characteristics measured at age one. The purpose of this exercise is to test whether early-life firm characteristics account for a larger fraction of the firm fixed effect.

Table 22 presents the results. In this specification, firm fixed effects account for 76% of total interest rate variation. Of this, 20.7 percentage points are explained by observable firm characteristics, while 55.6 percentage points remain unobservable. Including age-one controls increases the fraction of the firm fixed effect that can be attributed to observables, consistent with the idea that persistent firm characteristics partly reflect early economic conditions. Nevertheless, even in this specification, the unobservable component remains quantitatively important.

E.3 Excluding firm observables in the first stage

Finally, we assess whether our results depend on the inclusion of firm observable controls in the first stage. In this robustness exercise, we do not include any additional firm-level observable controls in the first-stage regression. We then estimate equation (10) directly on the residuals from equation (6).

Table 23 reports the corresponding decomposition. Firm fixed effects account for 32.7% of total interest rate variation. Of this, 3.7 percentage points are explained by ob-

	Contribution
(1) Firm FE	0.327
(1.1) Firm FE: obs component	0.037
(1.2) Firm FE: unobs component	0.290
# Firms	62,772

Table 23: **Contribution of firm fixed effect and firm fixed observable characteristics. No firm time varying controls.** The different lines represent the different components. Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (7).

servables and 29.0 percentage points are attributed to the unobservable component. The relative importance of the unobservable component remains large. This indicates that our results are not driven by the specific set of firm-level controls included in the first-stage specification.

F Loan pricing details

This section provides mathematical details for the loan pricing equations used in Section 3.2 of the main text.

F.1 PD and LGD definitions for one period loans

In the main text, equation (2) provides an equation for pricing one period loans that 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 $p(s^{l,t+1})$ id the probability of a state $s^{l,t+1} \in \mathcal{S}_{l,t+1}$.[57] For the case of a risk free lender, which is used as the baseline, this can be expressed as equation (3), with

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

[57] 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.

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

F.2 Extension to longer maturity loans

Consider a loan issued at time $t = 0$ with a maturity of T periods. The state of the economy at time 0 is s_0 and this state evolves stochastically thereafter with s_t denoting the state at time t and $s^t \equiv s_o, s_1, \ldots, s_t$ denoting the history of states up to time t. Let $\mathcal{S}_t$ denote the set of possible values for s^t. Let $y(s^t)$ denote the payment of the firm promised under the loan contract after the history of states s^t, $p(s^t)$ denote the probability of history s^t at time t so that $\sum_{s^t \in \mathcal{S}_t} p(s^t) = 1$, $\chi(s^t) \in [0, 1]$ denote the probability of default at time t after history s^t, $LGD(s^t)$ denote loss given default in this case and $\Lambda(s^t)$ be the stochastic discount factor. Note that this setup nests that standard case in which loan repayment are only time dependent and non-repayment at some history s^t results in non-repayment at all future points in time too (i.e. default is once and for all).

The time 0 price of a loan is

$$\begin{aligned} Q_0 &= \sum_{t=1}^{T} \sum_{s^t \in \mathcal{S}_t} \Lambda(s^t) p(s^t) y(s^t) (1 - \chi(s^t) LGD(s^t)) \\ &= \tilde{Q}_0 - \sum_{t=1}^{T} \sum_{s^t \in \mathcal{S}_t} \Lambda(s^t) p(s^t) y(s^t) \chi(s^t) LGD(s^t), \end{aligned}$$

where

$$\tilde{Q}_0 \equiv \sum_{t=1}^{T} \sum_{s^t \in \mathcal{S}_t} \Lambda(s^t) p(s^t) y(s^t)$$

is what the price of the loan would if it were risk free. Define a period t measure of loss given default, LGD_t, as

$$LGD_t \equiv \frac{\sum_{s^t \in \mathcal{S}_t} \Lambda(s^t) p(s^t) y(s^t) \chi(s^t) LGD(s^t)}{\sum_{s^t \in \mathcal{S}_t} \Lambda(s^t) p(s^t) y(s^t) \chi(s^t)},$$

which is a weighted sum of the loss given default across states in period t, with the weights determined by the stochastic discount factor, the probability of each state, the value of the payment due in each state and the probability of default in each state. In short, the weights depend on how costsly default in each state would be from the per-

spective of the bank. Our pricing formula is now

$$Q_0 = \tilde{Q}_0 - \sum_{t=1}^{T} LGD_t \sum_{s^t \in \mathcal{S}_t} \Lambda(s^t) p(s^t) y(s^t) \chi(s^t).$$

Next define a period t measure of default probability as

$$\chi_t = \frac{\sum_{s^t \in \mathcal{S}_t} \Lambda(s^t) p(s^t) y(s^t) \chi(s^t)}{\sum_{s^t \in \mathcal{S}_t} \Lambda(s^t) p(s^t) y(s^t)}$$

so that the loan price can be expressed as

$$Q_0 = \tilde{Q}_0 - \sum_{t=1}^{T} LGD_t \chi_t y_t.$$

y_t is the present discounted value of period t loan repayments in the case of no default:

$$y_t \equiv \sum_{s^t \in \mathcal{S}_t} \Lambda(s^t) p(s^t) y(s^t).$$

χ_t is a weighted average of default probabilities across states in period t, with the weights determined by the stochastic discount factor, the probability of each state and the value of the payment due in each state. Again, the weights represent the importance of the state to the bank.

To aggregate the loss given default measures over time define

$$LGD \equiv \frac{\sum_{t=1}^{T} LGD_t \chi_t y_t}{\sum_{t=1}^{T} \chi_t y_t}.$$

It then follows that

$$\begin{aligned} Q_0 &= \tilde{Q}_0 - LGD \sum_{t=1}^{T} \chi_t y_t \\ \implies \frac{Q_0}{\tilde{Q}_0} &= 1 - \frac{LGD \sum_{t=1}^{T} \chi_t y_t}{\sum_{t=1}^{T} y_t}. \end{aligned}$$

The second line uses the fact that $\tilde{Q}_0 = \sum_{t=1}^{T} y_t$. Finally aggregate the probability of default measures over time with

$$\chi \equiv \frac{\sum_{t=1}^{T} \chi_t y_t}{\sum_{t=1}^{T} y_t},$$

and we have

$$\frac{Q_0}{\tilde{Q}_0} = 1 - LGD\chi.$$

F.3 Interest rates

An issue with the formula derived in the previous section is that we don't observe loan prices in the data conditional on future promised payments. Or at least they are not easy to observe. It would be better to have a formula directly in interest rates.

Consider a multi-period loan in which the only payment occurs in the final period. I think that you could argue that any loan can be represented in this way, but I will put this aside for now. Let a history of states from time 1 to T be s^T, the set of possible such histories be $\mathcal{S}_T$, r be the interest rate on a loan and $\tilde{r}$ be the risk free rate. The lender is risk neutral. The price of the loan is

$$\begin{aligned}
Q_0 &= \sum_{s^T \in \mathcal{S}_T} p(s^T) \frac{(1+r)^T Q_0}{(1+\tilde{r})^T} (1 - \chi(s^T) LGD(s^T)) \\
\implies \left(\frac{1+\tilde{r}}{1+r}\right)^T &= \sum_{s^T \in \mathcal{S}_T} p(s^T)(1 - \chi(s^T) LGD(s^T)) \\
&= 1 - \sum_{s^T \in \mathcal{S}_T} p(s^T)\chi(s^T) LGD(s^T) \\
\implies \left(\frac{1+\tilde{r}}{1+r}\right)^T &= 1 - \chi LGD \\
\implies r &= \frac{1+\tilde{r}}{(1-\chi LGD)^{\frac{1}{T}}} - 1,
\end{aligned}$$

where

$$\begin{aligned}
LGD &\equiv \frac{\sum_{s^T \in \mathcal{S}_T} p(s^T)\chi(s^T) LGD(s^T)}{\sum_{s^T \in \mathcal{S}_T} p(s^T)\chi(s^T)} \\
\chi &\equiv \sum_{s^T \in \mathcal{S}_T} p(s^T)\chi(s^T).
\end{aligned}$$

One can recover the implied LGD or PD, conditional on the observed spread, from this

equation as

$$\chi LDG = 1 - \left(\frac{1+\tilde{r}}{1+r}\right)^{T}$$

An approximation of the formula (which looks more like a spread) is

$$r - \tilde{r} \approx -\frac{1}{T}\log(1 - \chi LGD).$$

One can recover the implied LGD or PD, conditional on the observed spread, from this equation as

$$\chi LDG = 1 - exp^{(\tilde{r}-r)T}.$$