# Beyond Risk: Firm Financing and Interest Rates*

Miguel H. Ferreira † Nicholas Kozeniauskas ‡

February 2025

**Abstract**

Financial frictions are a prominent feature of theories for a broad range of macroeconomic phenomena. There are three dominant approaches to modeling these, but little empirical guidance for them. Using a novel loan-level dataset we show that these approaches are not consistent with the data. There is a lot of dispersion in rates paid by firms, even on observationally equivalent loans, and risk—the dominant theory for dispersion—can account for less than 15% of it. Alternative theories for dispersion, based on banks and bank-firm relationships, also have quantitatively little explanatory power. Instead the data dictates that theories should focus on factors that are idiosyncratic to firms, persistent, and not closely connected to economic size or performance. We provide a search-based theory of firm financing that can rationalize these facts.

**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 common 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.

There is also broad agreement on, or at least acceptance of, how to model these frictions. The vast majority of theories adopt one of three approaches. The most common approach is to impose a collateral constraint (e.g. Moll, 2014). Under this 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. All firms therefore pay the risk free rate. A general form of this constraint is $b < \lambda a$, where $a$ is the assets of the firm, $b$ is the amount that the firm can borrow and $\lambda \geq 0$ determines the tightness of the constraint. A second 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 for 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 (e.g. Bernanke et al., 1999). Under this approach

[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).

interests rate differ across firms depending on their risk.

This paper takes a step back and considers how we model financial frictions. We start by taking these dominant theories and evaluating them empirically. To date there has been little analysis of this sort. However, a detailed new credit registry, matched with comprehensive firm balance sheet data, enables us to undertake this analysis. We argue that none of the dominant theories is a good description of the data. For the first two theories, there is a lot of heterogeneity in interest rates paid by firms after controlling for a rich array of loan characteristics, bank effects and time-variation in interest rates. This is true whether we focus on collateralized loans (theory one) or uncollateralized loans (theory two). We evaluate the third theory using data on default probabilities and loss given default. These characteristics of a loan account for little of the variation in interest rates that we observe. Overall, once we control for the characteristics of a loan, the level of collateralization, default risk and expected losses given default, 80% of the variation in interest rates remains unexplained. These results imply that, at best, there are substantial frictions in financial markets that current macroeconomic models ignore.

The second stage of the analysis investigates potential explanations for the additional dispersion in interest rates. We start by considering 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 that the idiosyncratic features of firm-bank relationships can also matter (e.g. Bharath et al., 2011). We evaluate how quantitatively relevant these theories are for the total variation in interest rates and find that neither channel accounts for much of the variation in the data. Bank characteristics can account for only 10–15% of the variation and firm-bank relationships only an additional 2–5%. This leaves the majority of the variation, 65% 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 of this is due to idiosyncratic and highly per-

sistent 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 strong implications for models of firm financing. There needs to be a strong firm-specific component to interest rates that is not closely associated with the standard characteristics defining firms in models (e.g. productivity, capital and financial position) or other observable characteristics. This leaves two possibilities. Dispersion is due to an unobservable firm characteristic or is random. We develop a model that nests both possibilities. Firms have different abilities in searching for loans and also face idiosyncratic noise in the rates they are offered by banks. We show how to identify the strength of these mechanisms with the data, quantify them, and show that the model can rationalize the data.

**Contribution to the literature** The paper's main contribution is to provide a detailed decomposition of the sources of dispersion in interest rates 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 asso-

[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%).

ciated 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, 1995), 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., 2024), 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. (2024). 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. (2024) 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 measure, 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 against the data. Section 4 explores the importance of banks and bank-firm relationships for dispersion in interest rates and Section 5 investigates the source of the remaining dispersion. 6 develops and evaluates the model and Section 7 concludes.

# 2 Data

The analysis uses two primary datasets. A credit registry covering the universe of loans between banks and firms in Portugal, and balance sheet data for these firms.[3]

The credit registry is at the monthly frequency, starting in January 2019. For each month it provides detailed information about the characteristics off all outstanding loans that firms have from banks. The meaning of 'loans' in this context is broad. It includes 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 includes information about the current characteristics of each loan including the amount outstanding, the interest rate, the duration of the loan, whether the interest rate is fixed or variable and the time horizon for changes to it, as well as information about the collateral for a loan, if any. For credit lines there is also information about their total size and the amount that has been drawn. The information about loans extends to their past characteristics such as the time of origination, and characteristics at origination including the amount, interest rate, duration, the variability of the interest rate and collateral information. This backward-looking feature means that despite the dataset starting in 2019, it has considerable information about loans issued before this date. Over the life of a loan the registry records repayment information and data about loans that are in arrears or default.

The other dataset provides balance sheet information as well as some other operational variables for the universe of firms at an annual frequency. It is based on annual accounting data of individual firms. Firms have to fill out mandatory financial statements in order to comply with their statutory obligation. Consequently, this dataset covers the population of virtually all non-financial corporations in from 2010 onwards, so it effectively overlaps with all loans in the credit registry going back to their dates of origination.

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

Our analysis focuses on traditional bank loans that are analogous to firm financing in models. These loans account for 30% of total credit issued by banks to firms in the data. The other main classes of credit are factoring and confirming (40%), credit cards (21%), auto loans (4%) and mortgages (3%). We will also show that our main results are robust to using all credit instruments, with this analysis presented in the Appendix.

The market for firm financing loans consists of six main banks, which each account for 9% to 16% of the total value of credit (Figure 1(a)). There are 156 smaller banks that collectively provide the remaining 28%. The six main banks are all large national banks that are the main providers of both firm and consumer credit. Panels (b)–(c) of Figure 1 show that these banks have similar distributions of credit geographically, across industries and across firm size. So, it is not the case that these banks are specialized and have segmented the market in some way. They are large, general-purpose banks.

The sample consists of more than 376,438 firms across all years. To illustrate the type of firms in the data we present summary statistics from the loan portfolios and balance sheets of firms in 2019 in Table 1. The balance sheet variables reflect the characteristics of the population of firms. For the credit registry variables, the median firm has €125,620 thousand Euros in loans. The median number of loans is 12 and the median share of a firm's loans that are collateralized is 100%. The majority of firms have loans from only one bank, but many firms have more bank relationships. At the 75th, 90th and 99th percentiles firms have loans from 2, 3 and 4 banks.

Figure 2 presents summary statistics for all of the loans in the sample at their initiation date. There are 647,893 loans in total. We reiterate that the sample includes all outstanding loans from 2019 onward, even if initiated before this date. 88% of these were initiated from 2015 onward. The average interest rate is 3.01% with a standard deviation of 2.25%. The median loan is for €45,000. The right tail of the distribution is for much larger loans, with the 90th percentile having a




((a)) Distribution of credit across banks

((b)) Geographic distribution of credit by bank

((c)) Industry distribution of credit by bank

((d)) Firm size distribution of credit by bank

Figure 1: **Distribution of firm financing credit across banks.** This figure presents moments of the distribution of credit in December 2019. Panel (a) shows the share of firm financing credit accounted for by each bank. Banks other than the largest six are grouped together. Panels (b)–(d) show the distribution of credit across regions, industries and firm size categories for each of the six main banks. Firm size is measured with the number of employees.

value of €348, 455. Most loans are long term with 56% of credit in the form of loans with a maturity of more than 5 years. For collateral, 70% of credit is through fully collateralized loans and nearly all other loans are uncollateralized.

# 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

| | p25 | p50 | p75 | p90 | St. dev. | # obs. |
|---|---|---|---|---|---|---|
| Value of loans | 26.99 | 125.62 | 505.42 | 1907.75 | (25878.51) | 46620 |
| # of loans | 3.00 | 12.00 | 12.00 | 18.00 | (39.23) | 46620 |
| Av. loan value | 7.96 | 21.31 | 66.07 | 200.00 | (1237.45) | 46620 |
| % collateralized | 45.77 | 100.00 | 100.00 | 100.00 | (41.37) | 46620 |
| # of banks | 1.00 | 2.00 | 3.00 | 4.00 | (1.54) | 46620 |
| Employees | 4.00 | 10.00 | 27.00 | 75.00 | (425.57) | 46620 |
| Total assets | 3.14 | 9.24 | 31.67 | 115.49 | (2933.28) | 46620 |
| Revenue | 2.99 | 9.12 | 29.99 | 103.65 | (1065.06) | 46620 |
| Leverage (%) | 21.03 | 34.83 | 49.68 | 65.16 | (1563.34) | 46620 |
| Age | 8 | 15 | 15 | 15 | (4.66) | 46620 |

Table 1: **Descriptive statistics.** This table provides statistics about all loans outstanding at the end of 2019 and the firms that have these loans. The units for the value of loans and average loan value are thousands of Euros, while total assets and revenue are in thousands of 2016 Euros. % collateralized is the percentage of the firms loans that are collateralized and # of banks is the number of banks that a firm has loans from.

equilibrium. 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. This again implies that all firms pay the risk free rate. The third approach is to allow for 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 firms 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-

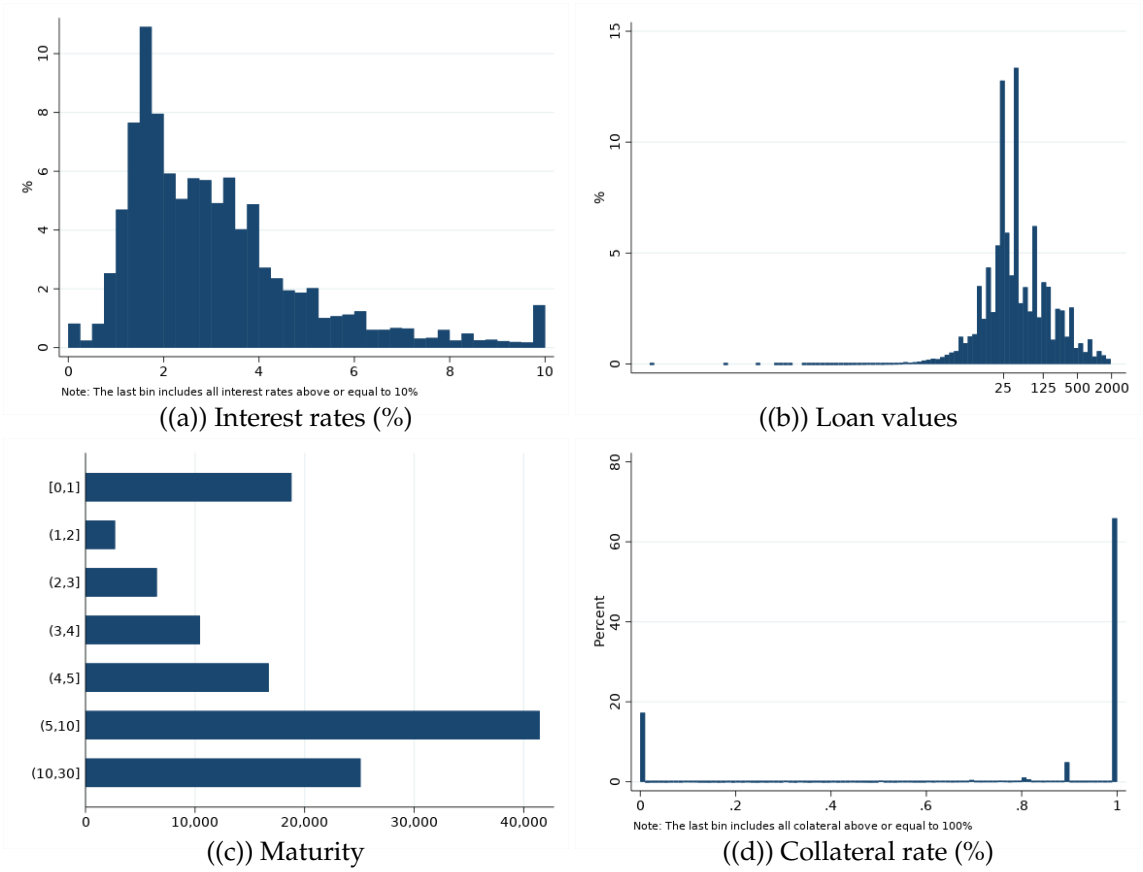



Figure 2: **Distributions of loan characteristics at initiation.** Panels (a), (b), (c) and (d) are the distributions of interest rates, loan values (in EUR thousands), maturity and collateral rates, respectively. The distributions in panels (a), (c) and (d) are weighted by loan value.

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 very common for loans to be collateralized, consistent with the first theory. Figure 2(d) in Section 2 shows that close to 70% of loans are fully collateralized and that only 20% have no collateral. In terms of the interest rates on loans, Figure 2(a) also shows that there is a lot of dispersion in interest rates. This is not necessarily inconsistent with these theories as the dispersion could come from factors other than risk that are outside the model. 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 all of these factors 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. This is another factor that is typically outside of models of of risk free lending. The data, however, is rich enough to control for this. Specifically we will use bank-time fixed effects as the control. We will assume that to the extent that two banks charge different interest rates for equivalent loans, the differences are constant across the loans. Note that these fixed effects will also control for time variation in the risk free rate at the frequency of the data (monthly).

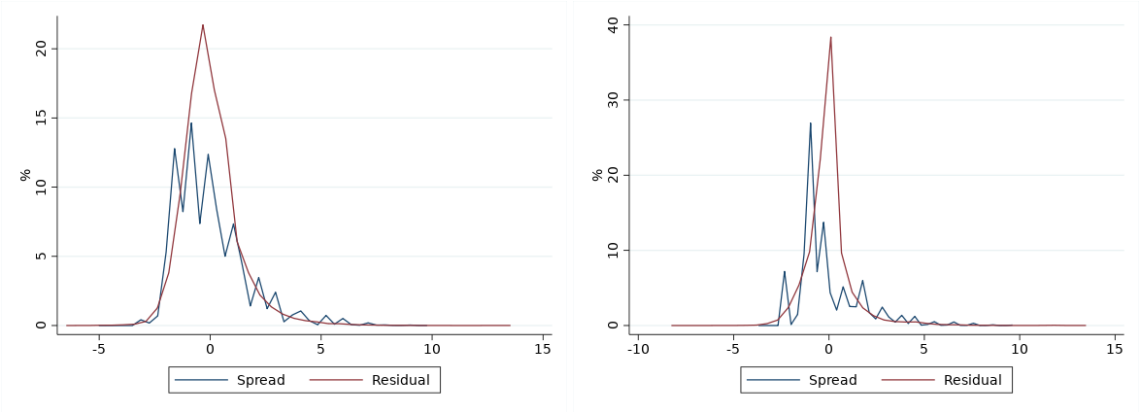
To assess how much variation in interest rates remains after controlling for loan characteristics and bank-time variation, we run the following regression

$$r_{lbt} = \beta_1 LC_{lt} + \beta_2 \theta_{bt} + \varepsilon_{lbt} \tag{1}$$

where $LC_{lt}$ is a vector of loan characteristics at initiation date $t$ and $\theta_{bt}$ is a vector of bank-time fixed effects. In the first vector we flexibly control for the adjustability of the interest rate and the maturity of the loan. 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.[4] To control for maturity we include fixed effects for whether the original maturity of the loan is $<$1, 1–2, 2–3, 3–4, 4–5, 5–10 or $>$10 years. To avoid history dependence of interest rates, we exclude the small share of renegotiated and renewed loans in the sample.

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 residual variation after we

[4]Three additional frequencies of adjustment that are included are another frequency less than a year, another frequency more than a year and other. These are uncommon, representing 0.05%, 0.03% and 7.75% of the sample, respectively.




((a)) Collateralized loans ((b)) Uncollateralized loans

Figure 3: **Spread distributions before and after controlling for loan characteristics, banks and time.** Panels (a) and (b) present the distribution of spreads 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 and bank-month fixed effects. All distributions are centered with their means at 0.

control for loan characteristics, bank effects and variation over time. The means of the two distributions are aligned at zero to make them visually comparable. It is clear that that the vast majority of interest rate variation remains after controlling for these factors. Overall, the standard deviation decreases from 1.59 to 1.24 as the distribution of residuals has some additional mass in the center, but otherwise the distributions are similar.

The results for the sample of uncollateralized loans are presented in Figure 3(b). The message is the same as for collateralized loans: nearly all dispersion in interest rates remains after controlling for loan characteristics, bank effects and time variation.

## 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

$$Q_{l,i,t} = \mathbb{E}_t \left[ \Lambda_{t+1} \left( 1 - \chi_{t+1} LGD_{t+1} \right) \right]. \tag{2}$$

$\Lambda_{t+1}$ is the stochastic discount factor, $\chi_{t+1}$ is a indicator for whether default occurs, $LGD_{t+1}$ (*loss given default*) is the share of the loan that is nor recovered in the case of default, and the expectation is taken over the set of possible states at time $t+1$. As examples of this type of setup, see Cooley & Quadrini (2001), Khan et al. (2021) or Ottonello & Winberry (2020).

To guarantee consistency with the data, in Appendix E we generalize the above equation for multi-period loans and re-write it in terms of interest rates, which is what we observe in the data. We get to the following equation

$$r = \frac{1 + \tilde{r}}{\left(1 - \chi LGD\right)^{\frac{1}{T}}} - 1, \tag{3}$$

where $\tilde{r}$ is the risk-free interest rate, $T$ the maturity of the loan, $\chi$ the expected probability of default over the lifespan of the loan and $LGD$ is the expected loss on the loan in case the firm defaults. [5]

There are different microfoundations for LGD. Ottonello & Winberry (2020) assumes that the lender does cannot recover the full value of the collateral. In Cooley & Quadrini (2001) there is asymmetric information and banks incur monitoring costs when default occurs and they need to enforce their right to collateral.

When taking equation (3) to the data, we have the LGD reported by the bank at the loan origination period, so we are agnostic on why the bank is unable to recover

[5] While we don't directly observe the risk-free rate, we observe the spread and the interest rate on the loan. The difference between the spread and the interest rate is the risk-free rate at the loan origination date.

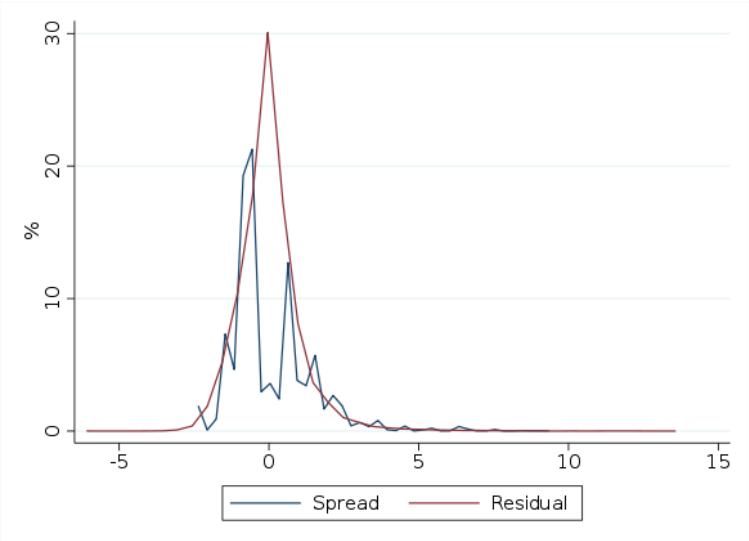



Figure 4: **Spread distributions before and after controlling for loan characteristics, banks and time for low risk firms.** Spread distribution for low-risk firms in blue and the residuals after controlling for loan characteristics and bank-month fixed effects in red.

the full value of the collateral and still incurs in losses, even for fully collateralized loans. We additionally use the probability of default (PD) reported by the bank at the loan origination date. The PDs are estimated at the firm level, as if a firm defaults on one loan, all loans are considered in default. Additionally, the PDs are estimated over one year period, in line with regulation implemented by the European Banking Authority. In Appendix D.1 we compare the reported PDs with different measures of observed defaults, and we find a strong positive correlation between the two, validating the reported PDs.

Before taking equation (3) to the data, to gauge if risk helps to explain the variation of spreads, we look at the their distribution for low-risk firms. We define these to be firms in the bottom 75% of the PD distribution, which means that they have a probably of default of less than 3.2%. If risk is a driving factor in explain the variance in distributions, after cleaning the spreads from loan characteristics and bank-time fixed effects, there shouldn't be much spread variance left to explain. Figure 4 shows this is not the case: after cleaning the spreads for low-risk firms, the spread variation is similar to the one in Figure 3, suggesting risk is not one of the main drives of the distribution of spreads.

Figure 5 compares the actual spread distribution in red, after controlling for

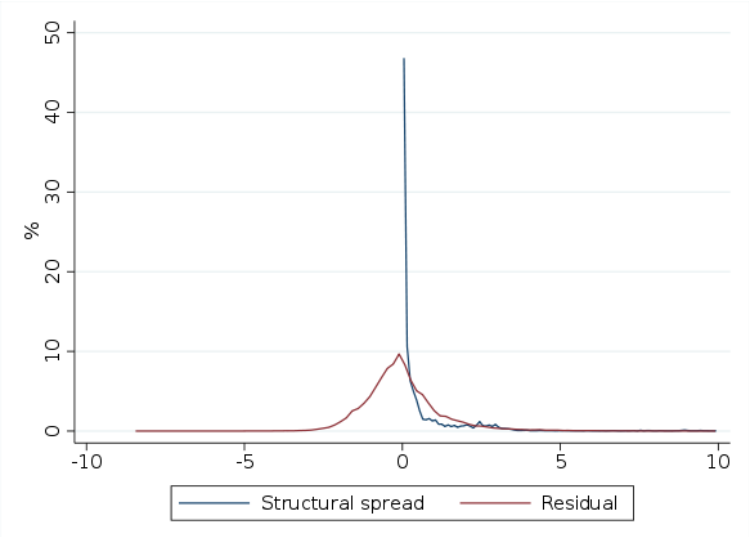



Figure 5: **Implied spread distribution from risk.** The blue line is the distribution of spreads implied by the loan pricing formula in equation (3). For comparison, the red line is distribution of spreads after controlling for loan characteristics and bank-month-year fixed effects.

loan characteristics and bank-month-year fixed effects, and the spread distribution implied by equation 3 in blue. It is possible to see that the spreads implied by the risk factors are strongly concentrated around zero.[6] This happens for two reasons: 1) large fraction of firms have a PD close to zero. 50% of the firms have a PD of 1% of lower and the PD at the 90th percentile is only 11.5%; 2) even for firms with high PD, the LGD is skewed to the left, with the median value being around 30%. For more details please see Table 5 in Appendix B, or Figure 11 in Appendix C for the full distribution of these variables.

As the banks do not report the PD or LGD for all the loans, to increase the number of observations and as a robustness test, we use the PD calculated by the Bank of Portugal (BoP). [7] The PD distribution, reported in Table 6 in Appendix B, is similar to the one reported by the commercial banks and the overall results for the implied spread are still concentrated around zero. Results are reported in Figure 9 in Appendix C.[8]

We go to the extreme of assuming that a firm that has previously default on a

[6]For details on some moments of the distribution, please see Table 7 in Appendix B.
[7]For more details on how the PD is estimated, please see Antunes et al. (2016).
[8]For details on some moments of the distribution, please see Table 8 in Appendix B.

loan has a PD of 100%. Even in this extreme scenario, the spreads implied by risk are not close to replicating the overall spread distribution. Results can be found in Figure 10 in Appendix C.

### 3.3 LGD and PD distributions

To have a better idea of what the LGD and PD distributions would need to be in order for equation (3) to match the spread distribution, we invert the problem and given the spread and LGD (PD) distribution, we calculate the implied PD (LGD) distribution.

**LGD** We start by calculating the implied LGD distribution. For this we follow OW20 by assuming that LGD is given by:

$$LGD_{t+1} = 1 - \min\left(\frac{\theta k_{t+1}}{b_{t+1}}, 1\right), \tag{4}$$

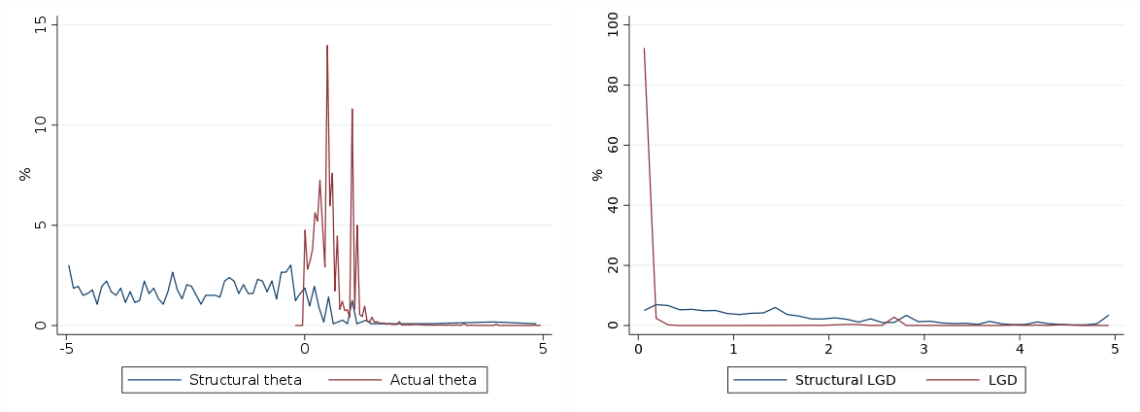
where $\theta$ is the recovery rate. Given that we have data on the market value of the collateral, we can calculate the implied recovery rate to match the observed distribution of spreads. To calculate the implied recovery rate, given the observed spread, PD, and collateral rate, we replace in equation (3) the LGD formula from equation (4) and then we isolate $\theta$ to get:

$$\hat{\theta} = \left(1 - \frac{1 - \frac{1+\tilde{r}}{1+r}}{\chi}\right)\frac{b'}{k'}, \tag{5}$$

where $\frac{b'}{k'}$ is the inverse of the collateral rate. Then using the estimated recovery rate $\hat{\theta}$, we use it to estimate the implied loss given default, using equation (4) and replacing $\theta$ with $\hat{\theta}$.

To calculate the actual $\theta$, we use the fact that we observe LGD and collateral rate, and using equation (4), we get:

$$\theta = (1 - LGD)\frac{b'}{k'}, \tag{6}$$




((a)) Recovery rates ((b)) Residual

Figure 6: **Recovery rates and LGDs: actual vs. required to match loan spreads.** In panel (a) the recovery rate implied by equation (2) is presented in blue and the actual recovery rate is in red. The implied recrovery rates are using the collateral rates and PDs in the data to match observed spreads. The observed recovery rates are calculated using observed LGDs and collateral rates. In panel (b) the observed LGD is presented in red and the LGD needed to generate loan spreads is in blue.

Figure 6 plots on the left hand panel both the structurally estimated $\hat{\theta}$ and the actual $\theta$. It is possible to notice that the implied recovery rate needs to be negative for a large fraction of loans, in order for equation 3 to generate a spread in line with the observed one. This would mean that the banks not only are not able to recover anything from the collateral, would even lose money when recovering the collateral. This translates into LGDs that would be above 1, as it is possible to see on the right hand panel of Figure 6, which compares the LGD implied by $\hat{\theta}$ and the observed LGD. An LGD value above 1 means the banks would lose more money than the amount of the original loan. Figure 12 in Appendix plots the same objects but using the PD calculated by the BoP, for which the conclusions stand.

**PD** We next calculate, given the observed LGD, what would the PD distribution need to be for equation (3) to generate a spread distribution in line with the observed one. To do this, we take equation (3) and isolate the default probability on the left hand side to get

$$\chi = \frac{1 - \frac{1+\tilde{r}}{1+r}}{LGD}. \tag{7}$$

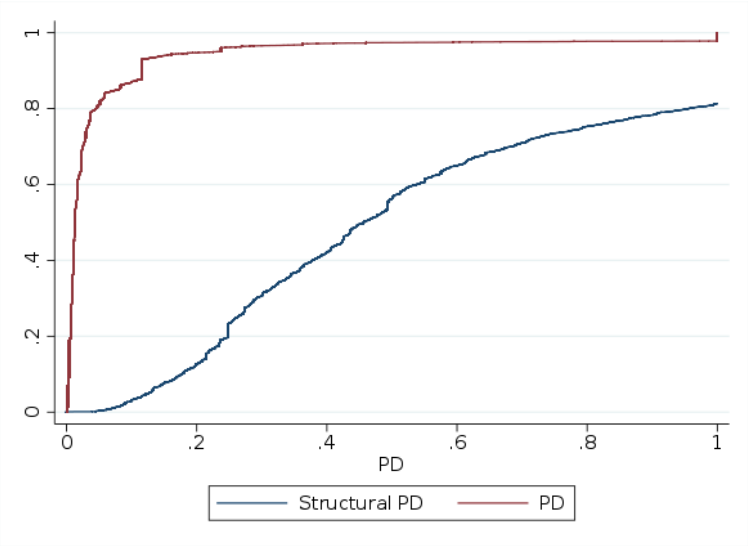



Figure 7: **PD distributions: actual vs. required to match loan spreads.** The red line is the CDF for PDs from the data. The blue line is the CDF of the PDs that would be necessary to generate the interest rates observed in the data, taking the LGD as given.

Figure 7 plots the CDFs for both the structural PD and the observed PD. It is possible to observe in the Figure that in order for equation (2) to generate a spread distribution in line with the empirically observed one, there would need to be much more default risk in the economy. While more than 50% of the reported PDs are close to 0%, the estimated median PD is 46%, which corresponds to the 95th percentile in the reported distribution. Given that the average loan default rate over the entire sample is 8%, the reported probability of default, with an average of 6%, is much more in line with actual default, than the estimated PD, which has an average above 100%. Figure 13 in the Appendix plots the structural PD against the PD reported by the BoP, for which the average PD is 11% and even closer to the actual default rate. Even when comparing to the PD from the BoP, the conclusions are robust, with much more risk being need in the economy to justify the observed spreads.

# 4 Banks and bank-firm relationships

In addition to the heterogeneous firms literature, which focuses either on collateral or risk as drivers of spread variation, there are branches of literature that offer

different explanations for spread distribution. Papers such as Bellifemine et al. (2022), Dempsey & Faria-e-Castro (2024) focus on bank characteristics, including market power, net worth, and other relevant factors. Bolton et al. (2016) and Faria-e-Castro et al. (2024), for example, focus on the bank-firm relationship as a potential explanatory factor. In this section we explore the contribution of both banks and bank-firm relationship to explaining the spread distribution.

## 4.1 Banks and interest rate variation

To measure the contribution of loan, firm and bank characteristics to spread variation we estimate the following regression:

$$i_{lfbt} = \beta LC_{lfbt} + \alpha_{ft} + \lambda_{bt} + \epsilon_{lfbt} = \sum_{p} C^{p}_{lfbt}, \tag{8}$$

where $i_{lfbt}$ is the interest rate $i$ charged in loan $l$ provided by bank $b$ to firm $f$ at the time of the loan origination date $t$. LC is a vector of loan characteristics, similar to the one included in equation (1). Then $\alpha_{ft}$, $\lambda_{bt}$, are the firm-time and bank-time fixed effects. These fixed effects should capture the firm and bank characteristics at the time of the loan origination.

The inclusion of these fixed effects require firms and banks to have multiple new loans created in each time period. From the bank side, as banks grant multiple loans every year, or even month, this fixed effect is not to restrictive, and as such we assume a time period to be a year. For firms, the fixed effect is more restrictive since few firms take out multiple new loans annually. As such, we are more loose with the firm-time fixed effects, considering time to be year, two-years or three-year periods, or corresponding to the economic cycle.[9]

The contribution of each component in explaining the interest rate variance can be measured as:

[9]The cycle periods are broken into the following periods: 1) before 2008 and the financial crisis; 2) the financial and sovereign debt crisis from 2008 to 2012; 3) from 2013 to 2019; 4) the Covid period 2020-21; and lastly 5) post-Covid from 2022 onwards.

$$\frac{cov(i_{lfbt}, C^p_{lfbt})}{var(i_{lfbt})}, \tag{9}$$

where $C^p$ represents each component in regression (8), and by definition these contributions sum to one.

Table 2 reports the contribution of each component to the interest rate variation for the firm-financing loans, with each column referring to the time dimension on the firm-time fixed effect. Firm characteristics are the most important component, explaining between 60% to 70% of the variation, depending on how the time component of the fixed effect is defined. The bank-time explains between 10% and 15%, with the loan characteristics explaining only around 3% and the remaining being explained by the residual. Tables 9 and 10 in Appendix show the results are robust to first, including all the loan types in the sample and also controlling for bank-month-year fixed effects. Lastly, results remain robust when considering only fully collateralized loans, as seen in Table 11 in the Appendix. Overall the bank component is never a major driver of the spread distribution, explaining between 10% and 15%, far from the firm's role, which account for always more than 50% of the overall distribution.

| | Cycles | Triannual | Biannual | Annual |
|---|---|---|---|---|
| (1) Loan characteristics | 0.036 | 0.037 | 0.036 | 0.027 |
| (2) Firm-Time FE | 0.614 | 0.657 | 0.657 | 0.717 |
| (3) Bank-Time FE | 0.141 | 0.099 | 0.111 | 0.095 |
| (4) Residual | 0.209 | 0.207 | 0.196 | 0.161 |
| # Obs. | 319312 | 309840 | 274567 | 211803 |

Table 2: **Contribution of loan characteristics, firm-time and bank-time fixed effects to interest rate variation.** Each column represents the periodicity of the time fixed effect and sums to 1. The different lines represent the different components included in equation (8). Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (9).

## 4.2 Bank-firm relationships and interest rate variation

To assess the importance of the bank-firm relationship we first run the following regression:

$$i_{lfbt} = \beta LC_{lfbt} + \gamma_{fbt} + \epsilon_{lfbt} = \sum_{p} C^{p}_{lfbt}, \tag{10}$$

where $\gamma_{fbt}$ is the bank-firm-time fixed effect, where the time dimension takes on the same periodicity as in the previous exercise. We then, on exactly the same sample of loans, regress equation 8, with the difference that the time component of the bank-time fixed effect takes on the same periodicity as that of the bank-firm-time fixed effect. The contribution of the the bank-firm relationship component will then be given by the difference between $\gamma_{fbt}$ and $\alpha_{ft} + \lambda_{bt}$. Results are presented in Table 3 and show that the firm-bank relationship accounts for only between 2% and 5% of the overall variation in spreads, and is not a major driver of the spread variation. When including all the instruments the conclusions are similar, with the firm-bank relationship accounting for between 4% and 6%, as presented in Table 12 in the Appendix.

| | Cycles | Triannual | Biannual | Annual |
|---|---|---|---|---|
| (1) Firm-Bank-Time FE | 0.793 | 0.819 | 0.830 | 0.893 |
| (2) Firm-Time FE | 0.659 | 0.689 | 0.689 | 0.762 |
| (3) Bank-Time FE | 0.089 | 0.092 | 0.107 | 0.110 |
| (4) Firm-Bank relationship | 0.046 | 0.037 | 0.034 | 0.022 |
| # Obs. | 246477 | 235314 | 200228 | 146730 |

Table 3: **Contribution of bank-firm relationships to interest rate variation.** Each column represents the periodicity of the time fixed effect. The first line represents the contribution of $\gamma_{fbt}$ from equation (10). The firm-time FE and bank-time FE represent the contribution of the terms $\alpha_{ft}$ and $\lambda_{bt}$ from equation (8), estimated on the same sample as equation (10). The firm-bank relationship contribution is calculated as the difference between the first line and the sum of lines two and three.

# 5 Firm characteristics

So far, we have established that the firm component is the major driver of the overall spread variation. However, firm default risk does not explain the importance of this component. To further understand what are the firm characteristics that help explain the interest variation we further decompose the firm-time fixed effect from equation 8 into time-varying observables, time-varying unobservables, time-sector fixed effect (657 sectors) and the firm fixed effect, which should capture firm characteristics that are not time-varying. More formally, we estimate the following regression:

$$\alpha_{fst} = \beta X_{fst} + \theta_{st} + \sigma_f + \epsilon_{fst}, \tag{11}$$

where $X_{fst}$ is a vector of firm observable characteristics, $\theta_{st}$ the sector s time t fixed effect and $\sigma_f$ is the firm fixed effect.[10] We break the vector of firm observable characteristics into four different categories: 1) balance sheet variables - employees; leverage ratio; revenues in log and growth rate; log of intangible assets; liquidity ratio; interest rate coverage ratio; age; and TFP, estimated following Ackerberg et al. (2015) ; 2) bank-firm relationship variables - log value of all loans and number of loans received from bank b over the last 5 years; share of the number of total loans and the total value of loans coming from bank b; duration of the relationship; and 3) variables capturing how open a firm is to working with different banks - number of bank relationships; concentration of bank relationships.[11]

Table 4 reports the contribution of each of these components. The firm fixed effect is the most significant factor in explaining the firm-time fixed effect, explaining

[10] Notice that we have firm balance sheet data at the annual level. When using the firm-time fixed effect at a different time frequency than annual, we average the firm characteristics over the same time period.

[11] As banks do not report PDs and LGDs for all the loans, in our benchmark scenario we exclude the risk variables from the analysis. We then show in the Appendix that adding the risk variables do not change the overall conclusions.

between 30% and 40% of the overall interest rate variance. Time-varying unobservable characteristics further explain around 10% of the total interest rate variation. Despite including a variety of time-varying observables, these explain little of the firm-time FE, less 10% of the total interest rate variation.

Lastly, we further decompose the firm fixed effect from equation (11) into unobservable and observable characteristics that are time invariant, which include the location of the firm, the sector, the within firm standard deviation of TFP and log revenues. The unobservable component explains more than 90% of the firm fixed effect from equation (11) and between 30% and 40% of the overall interest rate variation.

| | Cycles | Triannual | Biannual | Annual |
|---|---|---|---|---|
| (1) Firm-Time FE | 0.614 | 0.657 | 0.657 | 0.717 |
| (1.1) Time-varying firm obs. | 0.034 | 0.124 | -0.028 | 0.068 |
| (1.2) Time-varying firm unobs. | 0.087 | 0.097 | 0.092 | 0.100 |
| (1.3) Time-Sector FE | 0.122 | 0.145 | 0.217 | 0.163 |
| (1.4) Firm FE | 0.371 | 0.291 | 0.376 | 0.385 |
| (1.4.1) Firm FE: obs component | 0.017 | 0.002 | 0.024 | 0.013 |
| (1.4.2) Firm FE:unobs. component | 0.353 | 0.289 | 0.352 | 0.373 |
| # Obs. | 3110 | 3876 | 2851 | 2098 |
| # Obs. | 6239 | 8882 | 6681 | 5688 |

Table 4: **Decomposition of the contribution of firm characeristics to interest rate variation.** Each column represents the periodicity of the time fixed effect. The firm-time FE is the same as in Table 2. (1.1) to (1.4) represent different components in equation (11), with component (1.2) representing the contribution of the residual. (1.4.1) and (1.4.2) break down the firm FE component into firm fixed observable characteristics and a residual.

Table 13 in Appendix B reports the results when including all different loan types. Table 14 in Appendix B includes the PD calculated by the Bank of Portugal, for which we have data for the universe of firms, when explaining the Firm-Time FE. The additional explanatory power of this variable is included in the time-varying firm observables. We additionally include the average PD when trying to explain the firm fixed-effect, to see if it is permanent risk differences across firms that explain the importance of the firm fixed component. Lastly, Table 15

in Appendix B includes the probability of default reported by commercial banks, a dummy indicating if the firm has defaulted on any loan in the past, and if the firm ever defaulted, how long it took to start repaying the loan on time, and the average expected loss given default. We also include the following risk variables when trying to explain the firm fixed component: if the firm defaults at any point in time in our sample, the average probability of default and the average difference between the probability of default in normal times vs recessions, to capture aggregate exposure. Overall, results consistently show that firm fixed effects account for over 30% of the variation in spreads across all specifications.

**Robustness tests:** In the appendix we test some additional theories for spread variation. On such theory is that banks price loans factoring in the future value of the relationship, similar to Petersen & Rajan (1995). To test this we do a decomposition of the spread variation across the loans that initiate a firm-bank relationship. We include the bank-year fixed effects and variables to capture the future value of the relationship, namely the total number of loans and the total amount that the bank will grant to the firm in the future. To capture the firm's growth potential we additionally include the 5 year growth rate of assets, employees and revenues. Table 16 presents the contributions of each component, with the future value component accounts for only 2.5% of the spread variation in loans that initiate a firm-bank relationship.

Additionally, we implement a statistical model similar to the one by Sterk et al. (2021). The model uses information provided by the autocovariance structure to decompose the spreads into transitory (or ex-post) and permanent (or ex-ante) components. We assume the spread follows a process with two ex-ante components, that are determined before the firm obtains its first loan, and two ex-post components, a persistent shock and a i.i.d. innovation. We calibrate the model to match the autocovariance structure of the spreads. We then decompose the importance of heterogeneity in the ex-ante component for the variance of spreads as a function of the loan number. Even after nine loans, heterogeneity in the ex-ante

component explain around 40% of the spread variance, quantitatively in line with the importance of the firm fixed component found in Table 4. For more details, please see Appendix F.

# 6 Model

The data indicates that the most important driver of differences in interest rates across firms is a very persistent firm-level factor that is only weakly correlated with firm observables. In this section we develop a theory to explain this type of variation and evaluate the theory with additional empirical tests.

Since the source of variation cannot be related to any observable economic fundamentals, we focus on two possible alternative sources. The first is random variation in interest rates that is persistent due to features of the financing environment and the second is unobservable differences in the ability of firms to obtain low rates. We will nest both theories in our model, explain their different empirical predictions and use the data to distinguish between them.

## 6.1 Environment

Time is discrete and infinite. There is a continuum of firms and a continuum of identical banks.[12] Firms are profit-maximizing. When a firm is born it receives an initial capital stock $k_0$, a productivity level $z$, and a cost parameter $\theta > 0$ for searching for loans. Productivity is drawn from a distribution $G_z^0(z)$ and the search cost from $G_\theta(\theta)$. After birth productivity follows a Markov process, $G_z(z'|z)$, with stationary distribution $G_z(z)$. We assume that $\mathbb{E}_{G_z^0}[z] < \mathbb{E}_{G_z}[z]$ so that on average firm productivity increases after birth. Each period firms face an exogenous death probability, $\delta_f \in (0,1)$.

Firms produce with a production technology $zf(k)$ with the properties $f(0) =$

[12] We abstract from difference between banks since the data indicates that this is not a primary source of interest rate variation.

0, $f'(0) > 0$ and $f''(0) < 0$. The capital stock of a firm follows the law of motion

$$k_{t+1} = (1 - \delta_k)k_t + i_t,$$

where $\delta_k \in (0, 1)$ is the depreciation rate. In order to invest a firm needs to pay a fixed cost $\gamma > 0$ to borrow. Firms start their lives without a bank and each period they decide how much effort to exert in finding a bank, or replacing their existing bank. The idea for the search process is that if a firm exerts more effort then it will receive a lower rate, on average, and there will be less variance in the rate. Additionally, some firms are better at searching than others. To capture these ideas we assume that each period a firm chooses an effort level $e \in [0, 1]$ and then gets an interest rate draw from the distribution $U[0, \bar{r}/e]$,[13] where $\bar{r} > 0$ is a parameter that determines the upper bound of the distribution for a firm exerting maximum effort. The cost of effort is $\theta e$.

In practical terms, think of this process as capturing the fact that some entrepreneurs find it easier to obtain a better bank loan than others. This could be because they have more financial knowledge and are better able to navigate the banking system, they have better information about available rates, or they are simply more efficient at contacting banks and eliciting offers. Then, once an entrepreneur is in contact with a bank, the bank makes subjective assessments that result in randomness in the interest rates offered to entrepreneurs that look the same on paper. This could be because loan officers make different judgments based on the same information, or personal characteristics that are not economically or financially relevant affect a bank's decision.

When a firm meets its first bank it can either accept the interest rate offer and choose how much to borrow, or continue without a bank until the next period. If a firm accepts the offer then we denote its interest rate by $r'$. Everything that is borrowed is invested, so we denote the size of the loan in period $t$ by $i_t$. A firm needs to repay its loan at the end of period $t$.

[13] $U$ denotes the uniform distribution.

If a firm enters a period with a bank relationship, denote its interest rate from the previous period with $r$. The firm observes its new productivity and chooses its effort level for searching for a new bank. It receives a new interest rate draw $\tilde{r}$ and if $\tilde{r} < r$ then it switches banks. From here, borrowing and investing proceed as described above. If a firm chooses not to invest in a period or it receives a shock that dissolves its bank relationship, which occurs with probability $\delta_b \in (0,1)$, then it enters the next period without a bank.

## 6.2 Solution

A firm has two decisions to make each period: how much effort to exert in searching for a new bank and how much to invest. Denote the best interest rate that a firm has at the time of investment by $r'$. If a firm entered the period without a bank then $r' = \tilde{r}$, and if it did have a bank then $r' = \min\{r, \tilde{r}\}$. A firm's investment problem is

$$V_I(z,k,r',\theta) = \max_{i \geq 0} f(k) - (1+r')i - \mathbb{1}_i\gamma + (1-\delta_f)\Big(((1-\mathbb{1}_i) + \mathbb{1}_i\delta_b)\mathbb{E}[V(z',k',\theta)] + \mathbb{1}_i(1-\delta_b)\mathbb{E}[V(z',k',r',\theta)]\Big),$$

$$\text{s.t. } k' = (1-\delta_k)k + i,$$

$$z' \sim G_z(z'|z),$$

where $\mathbb{1}_i$ is an indicator for whether a firm invests ($i > 0$). The solution to this problem is an investment function $i(z,k,r',\theta)$. Investment is increasing in $z$, decreasing in $k$ and decreasing in $r$. These relationships are standard.[14] The effect of $\theta$ on the investment decision has several components. A higher $\theta$ means that search is more costly for a firm, so it expects a higher interest rate on average and to invest less. A high $\theta$ also decreases the probability of obtaining a lower interest rate than $r'$ in the future, leading to more investment in the current period.

[14] A new element of the relationship between $r$ and $i$ is that a firm needs to invest in order to maintain its banking relationship. This leads some firms to invest when otherwise they would not.

The search problem for a firm with a bank is

$$\begin{aligned} V(z,k,r,\theta) &= \max_{e\in[0,1]} \mathbb{E}[V_I(z,k,r',\theta)] - \theta e, \\ \text{s.t. } r' &= \min\{r,\tilde{r}\}, \\ \tilde{r} &= U[0,\bar{r}/e], \end{aligned}$$

and the problem for a firm without a bank is

$$\begin{aligned} V(z,k,\theta) &= \max_{e\in[0,1]} \mathbb{E}[V_I(z,k,r',\theta)] - \theta e, \\ \text{s.t. } r' &= U[0,\bar{r}/e]. \end{aligned}$$

These problems yield policy functions for effort for firms with a bank, $e(z,k,r,\theta)$, and those without, $e(z,k,\theta)$. A firm with a higher $z$ will invest more over time and therefore has a higher value from a lower interest rate. It will put more effort into search. Similarly, a firm with a smaller capital stock has more investment to do, and has more to gain from search effort. Effort also increases in the current interest rate, because the chance of obtaining a better interest rate is higher, and decreases in the cost of search.

## 6.3 Interest rate heterogeneity and dynamics

The model has persistence in firm-level interest rates for several reasons. There is initial dispersion in rates because firms have different search costs leading to different effort choices. Consequently, the distributions from which rates are drawn differ across firms. Conditional on the distribution there is also randomness. Importantly, this dispersion in rates is not necessarily correlated with the observable characteristics of firms: capital and productivity. Search costs are independent of these variables and there is randomness in the interest rate draws. Correlation between rates and observable characteristics will only arise because firms with a higher $z$ have an incentive to put more effort into searching for a lower rate, and a lower rate will increase a firm's capital stock. The data dictates that these effects

need to be relatively small.

In terms of the dynamics of interest rates, the model has several predictions. As long as a firm maintains a bank relationship, its interest rate will gradually decrease. When a relationship dissolves there will be a reset of the interest rate, and the variance of rates will increase. We can compute moments from the data related to these dynamics and quantitatively test the theory with them.

The data also allows us to quantify the roles of the two mechanisms in the model. The key to this is that the two sources of interest rate heterogeneity have some differences in their empirical predictions. To the extent that differences in interest rates are random, interest rates across firms will gradually converge over time, and there will be no persistence in a firm's rates when its relationships with a bank dissolves. In contrast, when dispersion in interest rates is due to differences in search effort there is weaker convergence in rates across firms over time, and greater persistence in rates when bank-firm relationships dissolve.

# 7 Conclusion

Financial frictions are a core feature of modern macroeconomic theory. They are used to explain a wide variety of phenomena, ranging from business cycles and monetary policy transmission, to misallocation and economic development, to inequality, and aspects of international trade.[15] While there are dominant methods for modeling these frictions, so far empirical guidance has been limited. This paper addresses this by using a detailed new credit registry and firm balance sheet data to show that there is substantial variation in interest rates beyond what the dominant theories can explain. Theories based on banks and bank-firm relationships also can't explain much of the variation. Instead we show that the largest component of this variation is due to highly persistent firm characteristics that are not closely associated with the observables into the data, including measures of economic size and performance.

[15] See footnote 1 for references to these branches of the literature.

A narrow implication of the results is that they provide empirical constraints for how much variation in interest rates various theories can generate. For example, bank heterogeneity can only generate 10–15% of total variation.

More importantly, the results imply that, at best, there are substantial frictions in financial markets that current macroeconomic models ignore. We provide a search-based theory for these frictions and show that it does a good job of rationalizing the data. The theory provides a basis for exploring the macroeconomic implications of these frictions.

# References


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.

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.

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

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).

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

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.

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.

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

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.

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

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

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

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

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

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.

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

behavior 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.

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.

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

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

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.

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. (1995). The effect of credit market competition on lending relationships. *Quarterly Journal of Economics*, 110(2), 407–443.

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

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.

# Appendix

## A Other credit instruments

The analysis in the main text focuses on traditional bank loans to firms. In this section we show that our main results are robust to the inclusion of other types of financial instruments. Figure 8 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), although most are still for over 5 years.

# B Additional Tables

| | (1) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | p1 | p5 | p10 | p25 | p50 | p75 | p90 | p95 | p99 | sd | count |
| LGD | 27.81 | 0.00 | 0.00 | 2.66 | 13.59 | 31.56 | 39.88 | 43.17 | 45.00 | 50.53 | (15.42) | 52229 |
| PD | 5.64 | 0.10 | 0.20 | 0.40 | 0.70 | 1.30 | 3.18 | 11.50 | 22.76 | 100.00 | (15.64) | 52229 |

Table 5: **PD and LGD distribution.**

| | (1) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | p1 | p5 | p10 | p25 | p50 | p75 | p90 | p95 | p99 | sd | count |
| LGD | 15.81 | 0.00 | 0.00 | 0.00 | 0.00 | 0.17 | 36.25 | 42.16 | 45.00 | 60.43 | (19.08) | 129400 |
| PD | 10.97 | 0.11 | 0.25 | 0.57 | 1.22 | 2.45 | 6.92 | 16.12 | 100.00 | 100.00 | (25.36) | 129400 |

Table 6: **PD reported by BoP and LGD distribution.**

| | (1) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | mean | p10 | p25 | p50 | p75 | p90 | p99 | sd | count |
| Residual | 0.07 | -1.26 | -0.63 | -0.11 | 0.50 | 1.49 | 4.81 | (1.29) | 48268 |
| Implied spread | 0.02 | 0.00 | 0.00 | 0.00 | 0.01 | 0.03 | 0.51 | (0.12) | 48268 |

Table 7: **Structural equation statistics.** This table provides statistics about the spread implied by the structural equation when using the PD reported by the banks.

| | (1) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | mean | p10 | p25 | p50 | p75 | p90 | p99 | sd | count |
| res | 0.00 | -1.33 | -0.74 | -0.16 | 0.50 | 1.46 | 4.64 | (1.29) | 106608 |
| implied spread | 0.02 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.71 | (0.18) | 106608 |

Table 8: **Structural equation statistics.** This table provides statistics about the spread implied by the structural equation when using the PD reported by the Central Bank.

Table 9: Contributions to interest rate variation: First stage, all instruments

| | Cycles | Triannual | Biannual | Annual |
|---|---|---|---|---|
| Loan characteristics | 0.137 | 0.135 | 0.132 | 0.130 |
| Firm-Time FE | 0.476 | 0.505 | 0.523 | 0.558 |
| Bank-Time FE | 0.159 | 0.138 | 0.134 | 0.130 |
| Residual | 0.229 | 0.222 | 0.210 | 0.183 |
| # Obs. | 806610 | 794502 | 748453 | 655475 |

Table 10: Contributions to interest rate variation: First stage, with bank-month-year fixed effects

| | Cycles | Triannual | Biannual | Annual |
|---|---|---|---|---|
| Loan characteristics | 0.035 | 0.035 | 0.034 | 0.026 |
| Firm-Time FE | 0.613 | 0.648 | 0.652 | 0.707 |
| Bank-Time FE | 0.155 | 0.121 | 0.129 | 0.115 |
| Residual | 0.198 | 0.196 | 0.184 | 0.151 |
| # Obs. | 319312 | 309840 | 274567 | 211803 |

Table 11: Contributions to interest rate variation: First stage for fully collateralized loans

| | Cycles | Triannual | Biannual | Annual |
|---|---|---|---|---|
| Loan characteristics | 0.074 | 0.079 | 0.078 | 0.079 |
| Firm-Time FE | 0.606 | 0.638 | 0.627 | 0.689 |
| Bank-Time FE | 0.128 | 0.095 | 0.119 | 0.094 |
| Residual | 0.192 | 0.189 | 0.175 | 0.138 |
| # Obs. | 226362 | 220984 | 190520 | 138778 |

Table 12: Contributions to interest rate variation: Firm-bank relationship all instruments

| | Cycles | Triannual | Biannual | Annual |
|---|---|---|---|---|
| Firm-Bank-Time FE | 0.670 | 0.696 | 0.716 | 0.748 |
| Firm-Time FE | 0.504 | 0.534 | 0.548 | 0.583 |
| Bank-Time FE | 0.104 | 0.102 | 0.117 | 0.124 |
| Firm-Bank relationship | 0.062 | 0.060 | 0.052 | 0.041 |
| # Obs. | 682344 | 666584 | 615174 | 523174 |

Table 13: Contributions to interest rate variation: Second-stage all instruments

| | Cycles | Triannual | Biannual | Annual |
|---|---|---|---|---|
| Firm-Time FE | 0.476 | 0.505 | 0.523 | 0.558 |
| Time-varying firm obs. | 0.049 | 0.069 | 0.008 | 0.040 |
| Time-varying firm unobs. | 0.081 | 0.100 | 0.109 | 0.122 |
| Time-Sector FE | 0.056 | 0.092 | 0.111 | 0.104 |
| Firm FE | 0.290 | 0.243 | 0.297 | 0.292 |
| Firm FE: obs component | 0.006 | 0.004 | 0.021 | 0.007 |
| Firm FE:unobs. component | 0.284 | 0.239 | 0.275 | 0.284 |
| # Obs. | 5055 | 6784 | 5758 | 4866 |
| # Obs. | 10344 | 16284 | 14475 | 15218 |

Table 14: Contributions to interest rate variation: Second-stage firm-financing with PDs

| | Cycles | Triannual | Biannual | Annual |
|---|---|---|---|---|
| Firm-Time FE | 0.614 | 0.657 | 0.657 | 0.717 |
| Time-varying firm obs. | 0.036 | 0.125 | -0.048 | 0.066 |
| Time-varying firm unobs. | 0.087 | 0.096 | 0.091 | 0.099 |
| Time-Sector FE | 0.122 | 0.147 | 0.226 | 0.165 |
| Firm FE | 0.369 | 0.290 | 0.388 | 0.386 |
| Firm FE: obs component | 0.054 | 0.020 | 0.030 | 0.057 |
| Firm FE:unobs. component | 0.316 | 0.269 | 0.358 | 0.329 |
| # Obs. | 3110 | 3875 | 2851 | 2097 |
| # Obs. | 6239 | 8880 | 6681 | 5686 |

Table 15: Contributions to interest rate variation Second-stage firm-financing with risk variables

| | Cycles | Triannual | Biannual | Annual |
|---|---|---|---|---|
| Firm-Time FE | 0.614 | 0.657 | 0.657 | 0.717 |
| Time-varying firm obs. | 0.119 | 0.035 | -0.406 | 0.048 |
| Time-varying firm unobs. | 0.065 | 0.076 | 0.078 | 0.061 |
| Time-Sector FE | 0.088 | 0.162 | 0.272 | 0.102 |
| Firm FE | 0.342 | 0.385 | 0.712 | 0.505 |
| Firm FE: obs component | 0.022 | 0.035 | 0.036 | 0.074 |
| Firm FE:unobs. component | 0.320 | 0.350 | 0.676 | 0.431 |
| # Obs. | 732 | 887 | 444 | 292 |
| # Obs. | 1558 | 2026 | 997 | 694 |

Table 16: Contributions to interest rate variation across the loans that initiate a firm-bank relation, including variables to capture the future value of the relationship.

| Component | |
|---|---|
| Future value | 0.025 |
| Bank-Time FE | 0.263 |
| Residual | 0.712 |
| # Obs. | 216659 |

Table 17: Firm Time FE decomposition

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Observations | 762 | 762 | 762 | 762 | 762 |
| R-squared | 0.182 | 0.532 | 0.836 | 0.886 | 0.892 |
| Firm FE | NO | NO | YES | YES | YES |
| Sector $\times$ Time FE | NO | YES | NO | YES | YES |
| Firm controls | YES | YES | YES | NO | YES |
| Risk controls | YES | YES | YES | NO | YES |
| Relation controls | YES | YES | YES | NO | YES |
| Competition controls | YES | YES | YES | NO | YES |

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

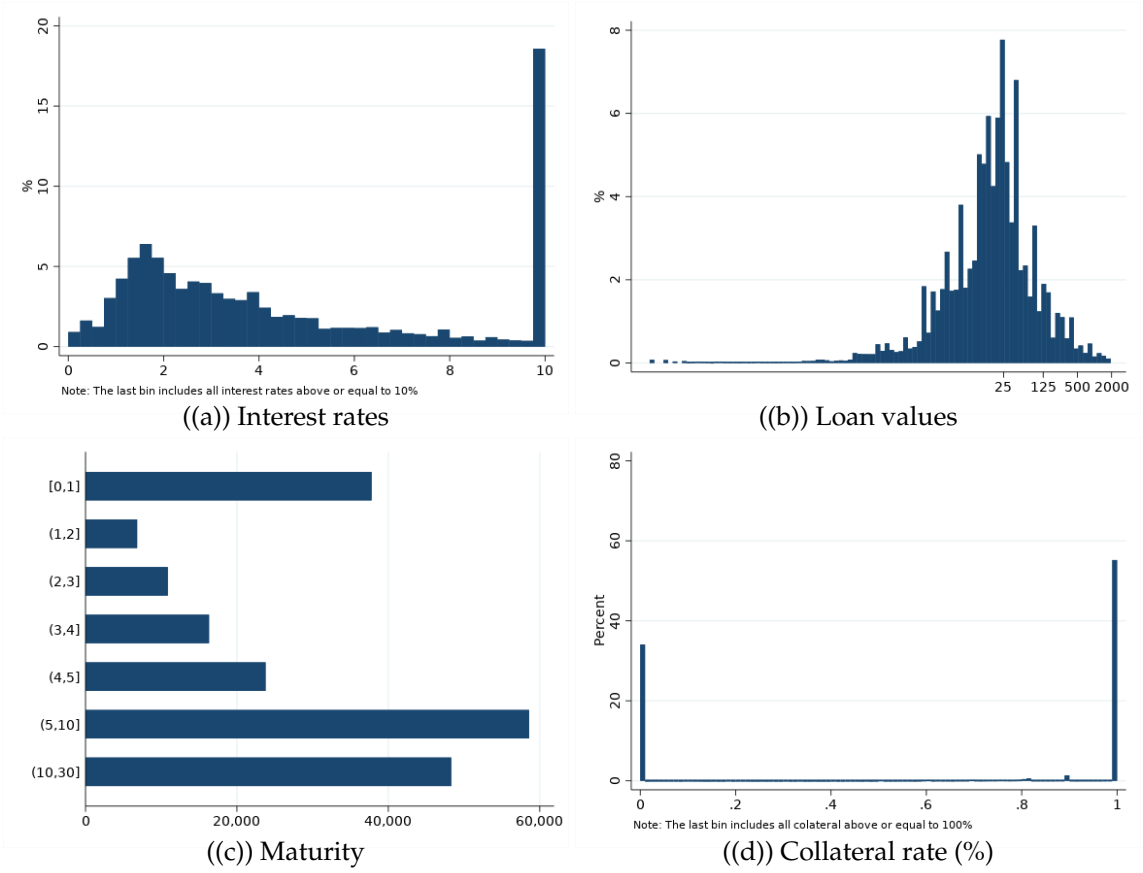
Table 18: Firm FE decomposition

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| VARIABLES | | | | | |
| Sector | 0.000*** | 0.000*** | 0.000*** | 0.000*** | 0.000*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Location | 0.001*** | 0.001*** | 0.001*** | 0.001*** | 0.001*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Frequency bank rel | 0.416 | 0.391 | 0.401 | 0.410 | 0.420 |
| | (0.339) | (0.292) | (0.316) | (0.292) | (0.313) |
| Stdev(log revenues) | | 0.130 | 0.128 | 0.116 | 0.114 |
| | | (0.194) | (0.196) | (0.190) | (0.192) |
| Stdev (TFP) | | 0.206 | 0.207 | 0.170 | 0.171 |
| | | (0.499) | (0.501) | (0.501) | (0.502) |
| PD | | 0.026*** | 0.026*** | 0.024*** | 0.024*** |
| | | (0.004) | (0.004) | (0.004) | (0.004) |
| Max(default) | | | -0.011 | | -0.011 |
| | | | (0.109) | | (0.107) |
| Aggregate exposure | | | | 0.026** | 0.026** |
| | | | | (0.012) | (0.012) |
| Observations | 339 | 339 | 339 | 339 | 339 |
| R-squared | 0.352 | 0.488 | 0.489 | 0.497 | 0.497 |

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

# C Additional Figures




((a)) Interest rates

((b)) Loan values

((c)) Maturity

((d)) Collateral rate (%)

Figure 8: **Distributions of loan characteristics at initiation.** Panels (a), (b), (c) and (d) are the distributions of interest rates, loan values, maturity and collateral rates, respectively. The distributions in panels (a), (c) and (d) are weighted by loan value.

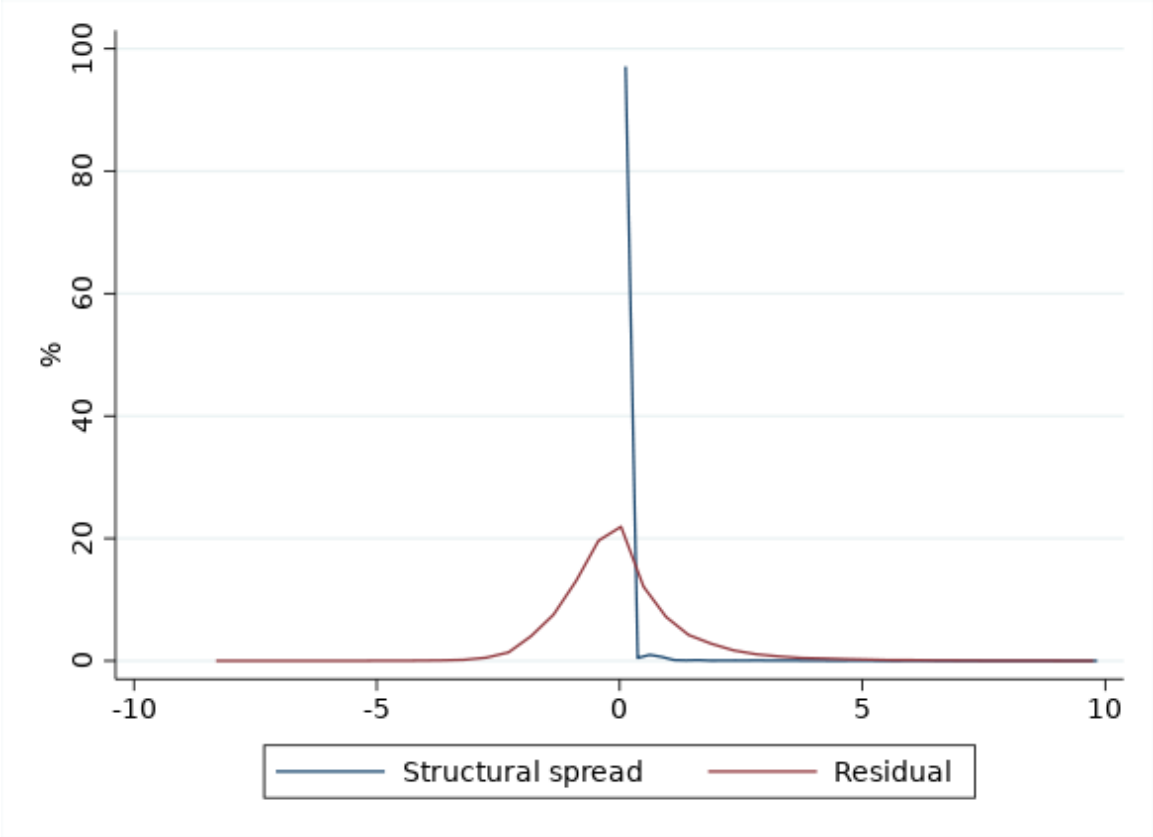



Figure 9: Distribution of spreads implied by equation (2) after controlling for loan characteristics and bank-month fixed effects and using the probability of default calculated by the Bank of Portugal.

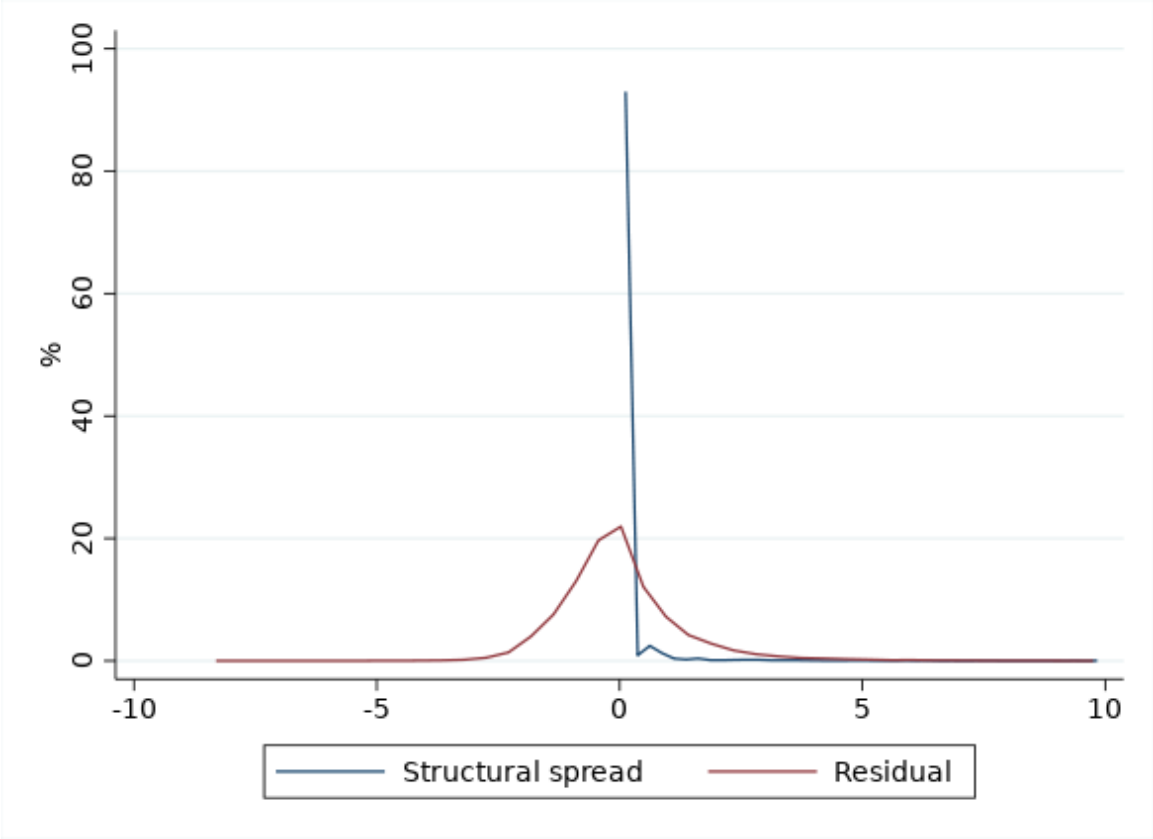
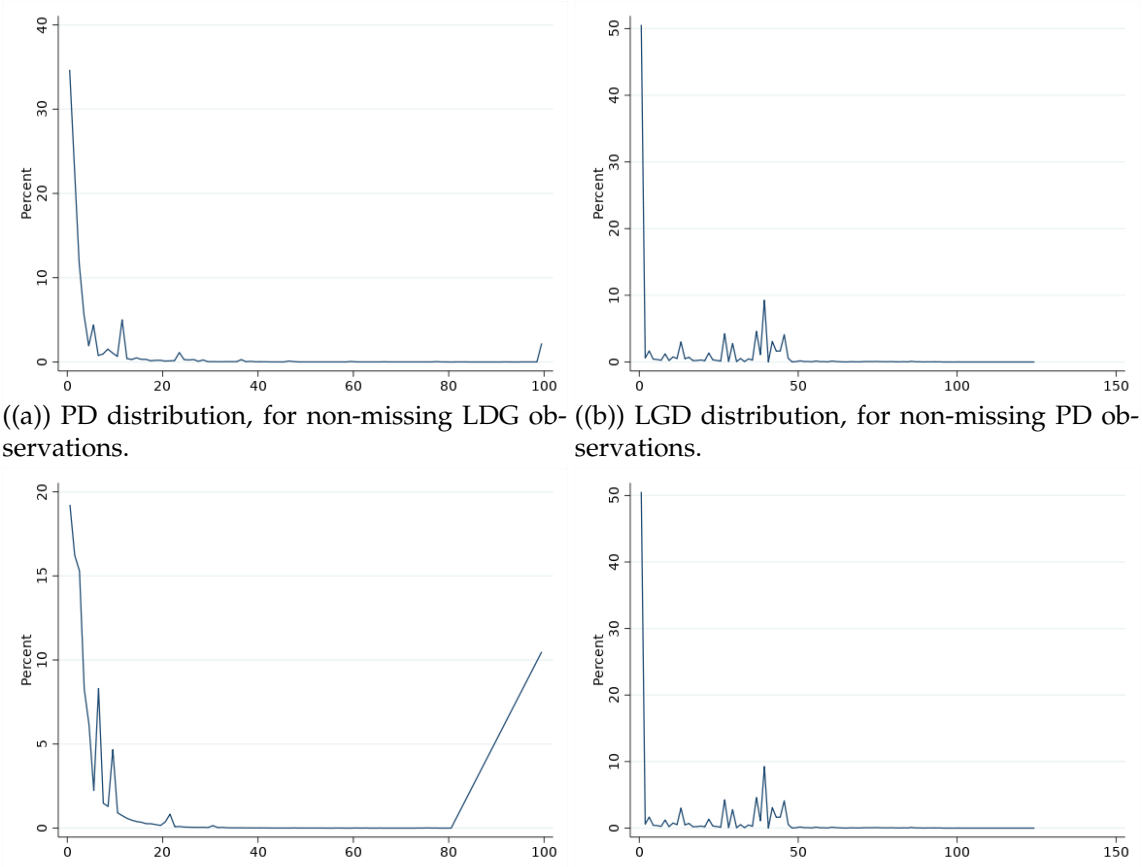



Figure 10: Distribution of spreads after controlling for loan characteristics and bank-month fixed effects and the distribution of spreads implied by equation (2) assuming that firms which previously defaulted on a loan have PD set to 100%.



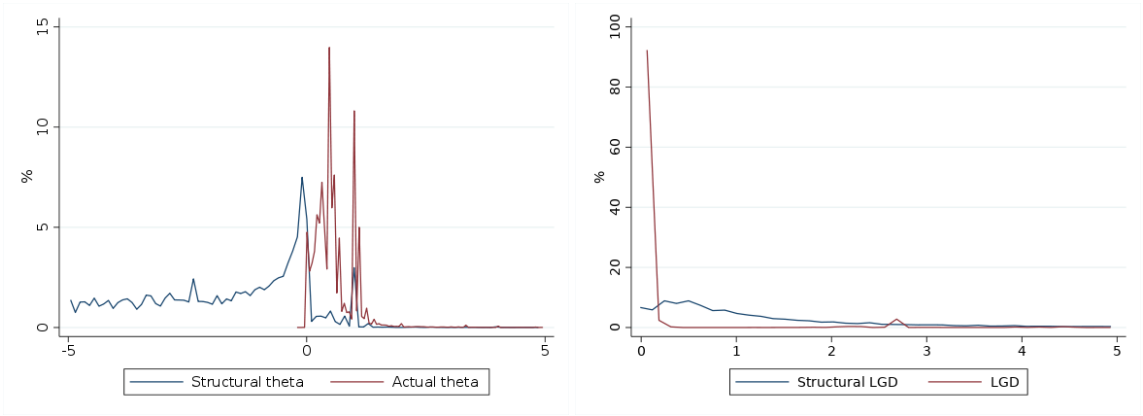

((a)) PD distribution, for non-missing LDG observations.

((b)) LGD distribution, for non-missing PD observations.

((c)) PD calculated by BoP distribution, for non-missing LDG observations.

((d)) LGD distribution, for non-missing PD observations reported by the BoP.

Figure 11: **PDs and LGDs.**




((a)) Interest rate ((b)) Residual

Figure 12: On the left panel the recovery rate implied by equation 2 in blue and the actual recovery rate in red. The structural theta is calculated using the collateral rates and the PDs in the data to match the observed spread. The actual theta is calculated using the reported LGDs and the collateral rates. On the right hand side, the reported LGD in red and the LGD implied by the structural theta in blue. PDs reported by the Bank of Portugal are used in this case.

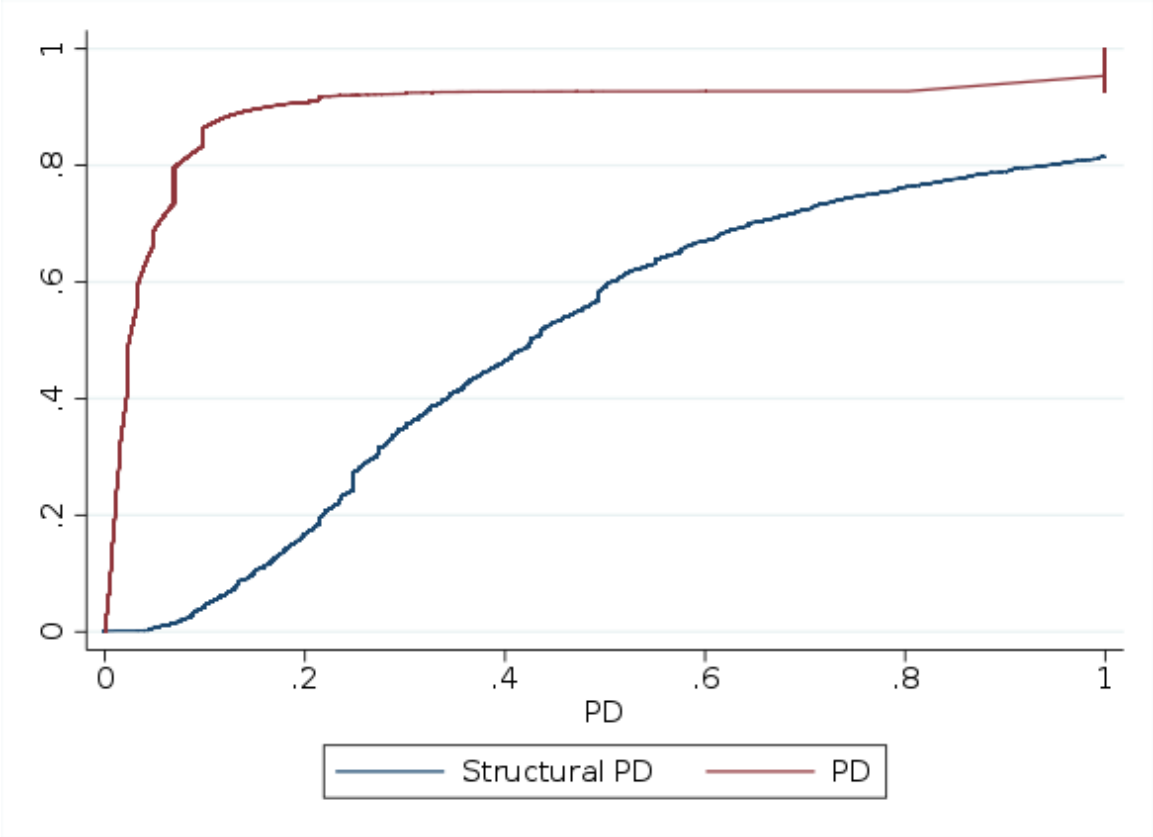



Figure 13: CDF of structural PD

Figure 14: CDF of structural estimated PD (blue line) and the actual PD (red line).

# D Validating reported PDs

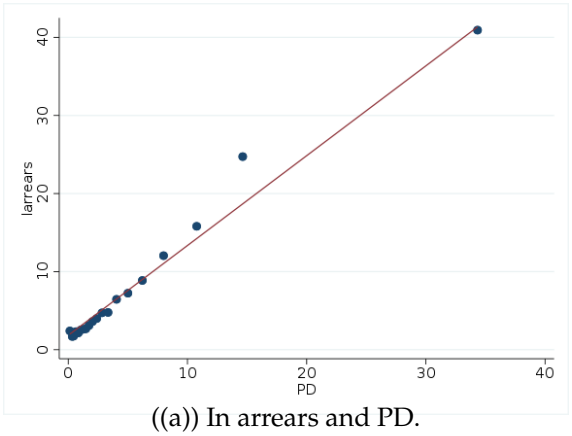
## D.1 PDs

In this section we compare the reported one year probabilities of default by both the commercial banks and the Bank of Portugal to the actual default rates, one year after the reported PD. We define different measures of default:

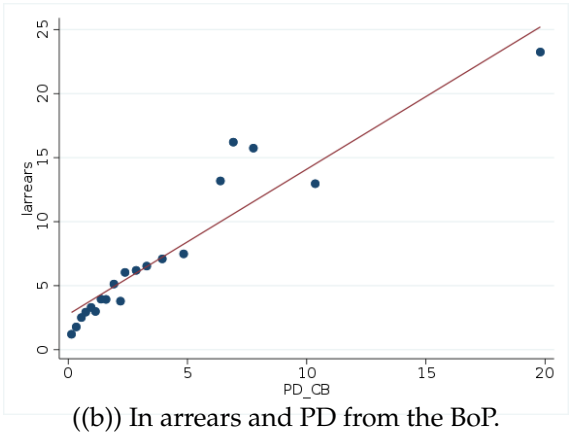
- In arrears: If at any point in time the firm falls behind on payments for at least one month;
- In default: If at any point in time the firm falls behind on payments for at least three months;
- Default_amount: If the share of overdue to total credit is above 2.5%, similar to Antunes et al. (2016); [16]
- Default_ex_post: Loans which are not fully repaid by original maturity.

Figure 15 plots the correlation between PDs and PDs reported by the BoP and the different actual default measures. Overall, the reported PDs predict well actual default.

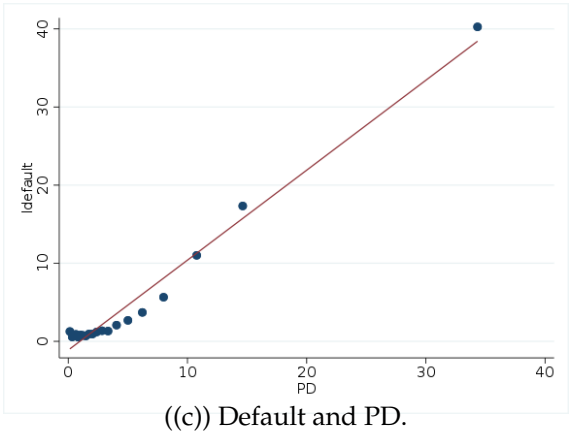
[16] Overdue credit is define as credit for which the borrower as not made payments by at least 90 days.



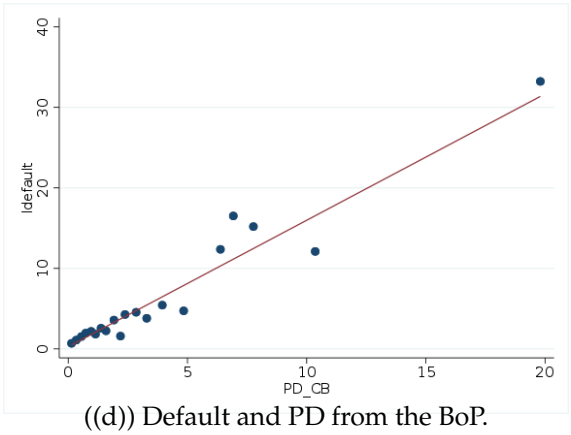

((a)) In arrears and PD.



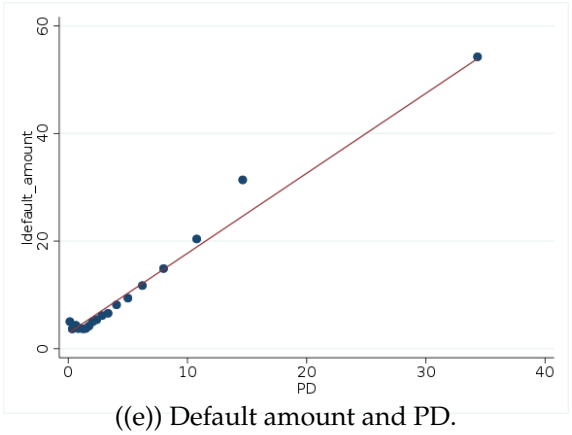

((b)) In arrears and PD from the BoP.



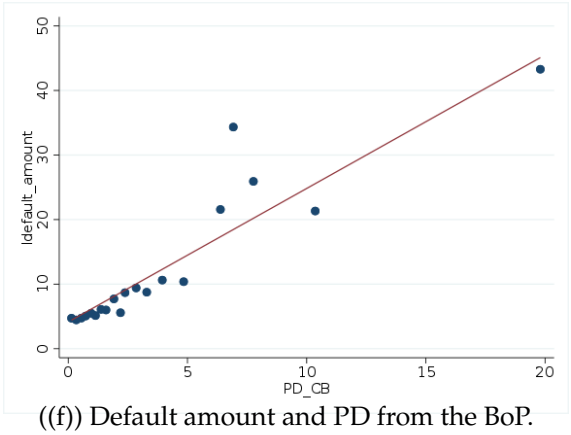

((c)) Default and PD.



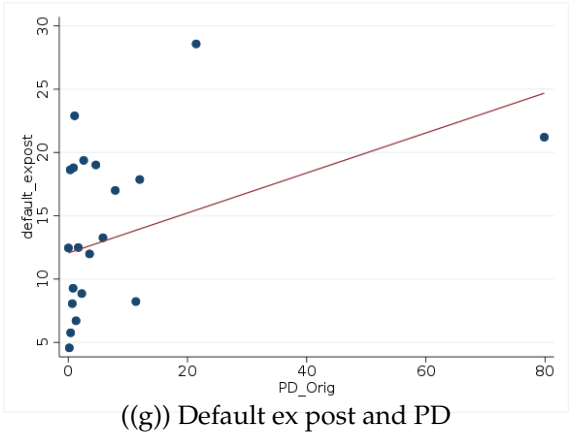

((d)) Default and PD from the BoP.



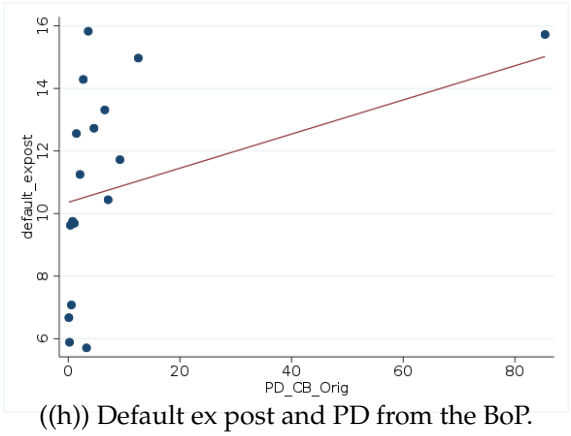

((e)) Default amount and PD.




((f)) Default amount and PD from the BoP.




((g)) Default ex post and PD




((h)) Default ex post and PD from the BoP.



Figure 15: **PDs and actual default rates.**

# E Multi-period loan price with risk

## E.1 Relative prices

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 contrast 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 perspective 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:

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

$\chi_t$ 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

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

## E.2 Spreads

An issue with the formula derived in the previous section is that we don't observe loans 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

$$LGD \equiv \frac{\sum_{s^T \in \mathcal{S}_T} p(s^T)\chi(s^T) LGD(s^T)}{\sum_{s^T \in \mathcal{S}_T} p(s^T)\chi(s^T)}$$

$$\chi \equiv \sum_{s^T \in \mathcal{S}_T} p(s^T)\chi(s^T).$$

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

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

# F Permanent component of spreads

## F.1 Autocorrelation and standard deviation

If, as suggested by the evidence in Section 5, there is heterogeneous firm fixed effects, we should observe a relatively constant profile of standard deviation of spreads and relatively high autocorrelation between spreads over different loans. 16 plots these two statistics in our data, with panel (a) depicting the standard deviation and panel (b) depicting the autocorrelation profile.[17] First, the standard deviation is decreasing only mildly over loans and stabilizes at high values. Second, the autocorrelation is relatively high between loan numbers, suggesting that initial conditions are also very important for spreads.

[17] 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-month-year of origination and sector fixed effects and then use the spread residuals.

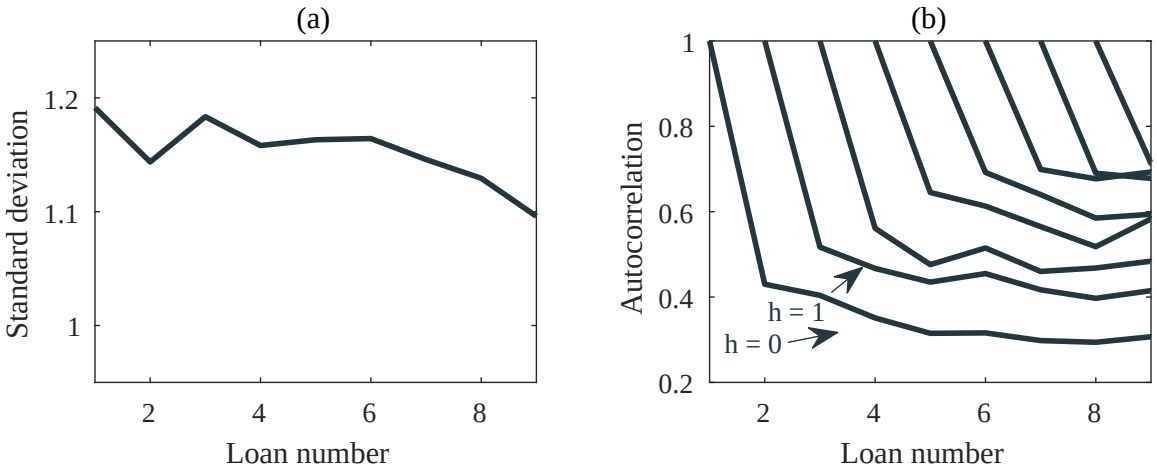



Figure 16: **Standard deviation and autocorrelation of spreads across loans, within firm.** The left panel presents the standard deviation of spreads by loan number, after controlling for bank-month-year and sector fixed effects. The right panel presents the autocorrelation of spreads by loan number $a$ and loan number $h \leq a$. Across lines $h$ changes, while $a$ changes along the lines.

## F.2 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.[18] This model uses the information provided by the autocovariance structure of spreads to capture the importance of both types of heterogeneity.

Consider the following decomposition for spread $r$ by firm $i$ at loan number $a$:

$$\underbrace{r_{i,a}}_{\text{Spread}} = \underbrace{u_{i,a} + v_{i,a}}_{\text{Ex-ante component}} + \underbrace{w_{i,a} + z_{i,a}}_{\text{Ex-post component}}, \tag{12}$$

where

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

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

In this spread process, the terms $u_{i,a}$ and $v_{i,a}$ capture the ex-ante profile while $w_{i,a}$ and $z_{i,a}$ 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 $a = -1$. 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. $\theta_i$ is the permanent component, which will accumulate over the different loans at speed $\rho_u$. In particular, with $\rho_u < 1$, the long-run steady state level of employment will be given by $\frac{\theta_i}{1-\rho_u}$. Further, this specification allows for rich heterogeneity not only in terms of long term spreads the firms will face, depending on the distribution of $\theta_i$, 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 $\rho_w$. 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 for by minimizing the sum of squared differences between the model and empirical autocovariance. Table 19 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 report the standard deviation of the parameters.

Panel (a) of Figure 17 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 $\rho_u$ and $\sigma_\theta$, as, together, they imply that permanent heterogeneity exists. The point estimates imply that ex-ante conditions matter, as both $\rho_u$ and $\sigma_\theta$ 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,

| $\rho_u$ | $\rho_v$ | $\rho_w$ | $\sigma_\theta$ | $\sigma_u$ | $\sigma_v$ | $\sigma_\epsilon$ | $\sigma_z$ |
|---|---|---|---|---|---|---|---|
| 0.344 | 0.494 | 0.660 | 0.461 | 1.871 | 3.248 | 0.409 | 0.709 |
| (0.122) | (0.412) | (0.132) | (0.095) | (0.604) | (3.348) | (0.091) | (0.096) |

Table 19: Equally-weighted minimum distance estimates of the statistical model. Standard deviation in parenthesis.

the standard deviation is 1.1 percentage points. This again demonstrates that there seem to be differences between firms that originate from ex-ante conditions.

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 17 illustrates the importance of the ex-ante component for the variance as a function of loan number. The ex-ante component contribution is above 60% at the first loan, and 40% of the variance of spreads after the first loan, which is quantitatively in line with what we find the firm fixed component to account for in Section 5.

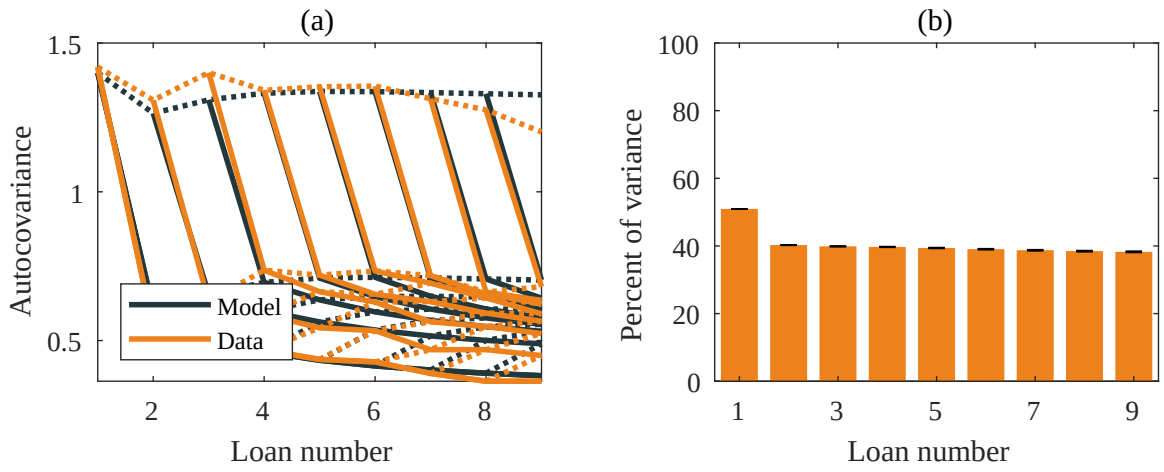



Figure 17: **Model fit to the data and ex-ante contribution.** Panel (a) plots both the autocovariance structure of the spread in the calibrated statistical model (blue) and in the data (orange). Panel (b) presents the percent of variance explained by ex-ante factors as a function of the loan number, according to the statistical model.