Beyond Risk: Firm Financing and Interest Rates*

Miguel H. Ferreira † Nicholas Kozeniauskas ‡

June 2025

Abstract

Financial frictions are a prominent feature of theories for a broad range of macroeconomic phenomena. Using a novel loan-level dataset, we evaluate whether these theories are broadly consistent with observed interest rates. We find that 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 the major source of variation is 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 show that the data departs substantially from these theories. 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, the majority 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 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 of this 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 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., 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 against the data. Section 4 explores the importance of firms, banks and bank-firm relationships for dispersion in interest rates. Section 5 develops and evaluates the model and Section 6 concludes.

2 Data

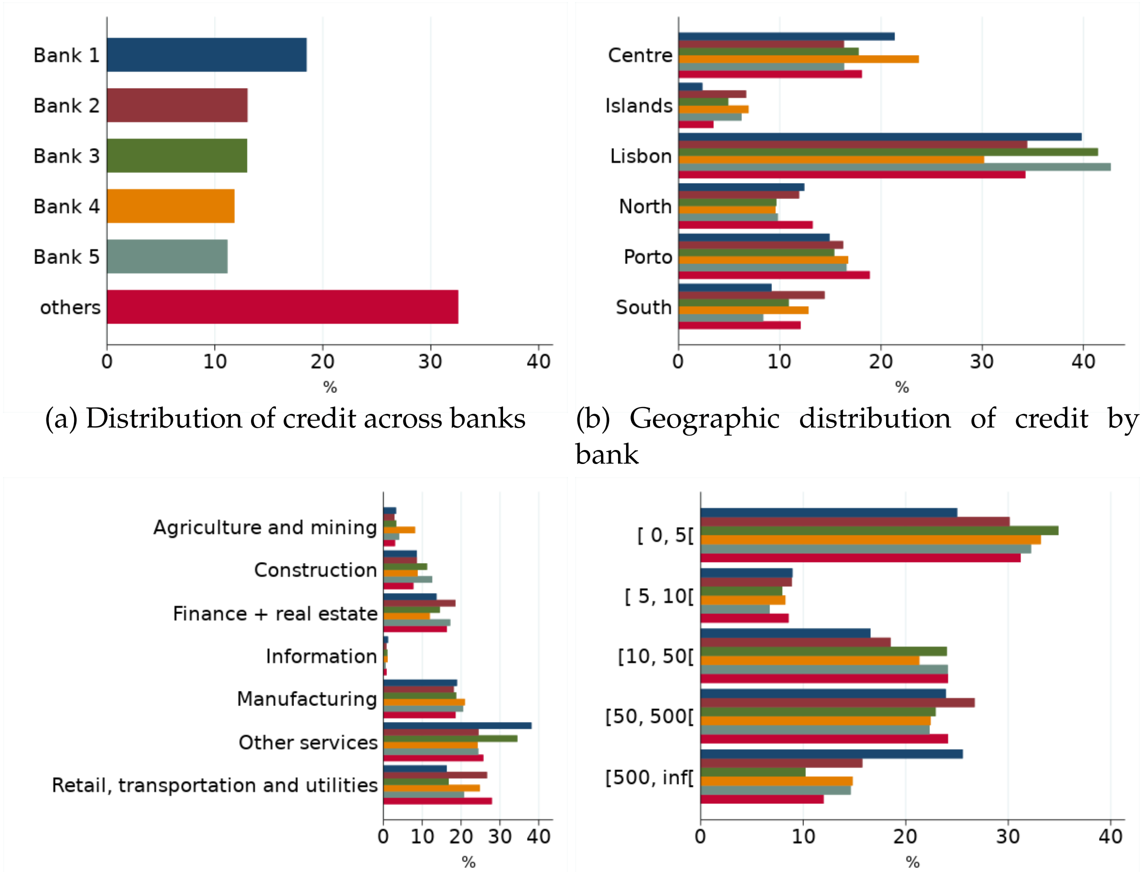
The analysis uses two primary datasets. A credit registry covering the universe of loans between banks and firms in Portugal, and balance sheet data for these firms.[3]

The credit registry has a monthly frequency, covering January 2019 to December 2022. 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, 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 operational variables for the universe of firms at an annual frequency. This data comes from mandatory reporting of financial statements to the Ministry of Finance each year. The data covers 2010 to 2022, so it effectively overlaps with all loans in the credit registry going back to their dates of origination.

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

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



(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.

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 also 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, which each accounting 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

	p25	p50	p75	p90	St. dev.	# firms.
Value of loans (€'000s)	14.68	41.16	123.39	437.50	(2991.35)	
# of loans	1.00	1.25	2.00	3.75	(3.23)	
Av. loan value (€'000s)	11.61	26.06	62.37	200.00	(1036.55)	
% collateralized	45.83	87.50	100.00	100.00	(35.54)	
# of banks	1.00	1.00	1.62	2.10	(0.77)	162,270
Employees	1.33	3.50	8.25	20.00	(122.24)	
Total assets (€'000s)	90.94	244.54	741.56	2395.45	(65036.51)	
Revenue (€'000s)	76.33	207.68	605.76	1815.54	(40084.66)	
Leverage (%)	0.50	0.72	0.93	1.37	(2.92)	
Age	5.00	10.50	20.50	31.50	(13.13)	

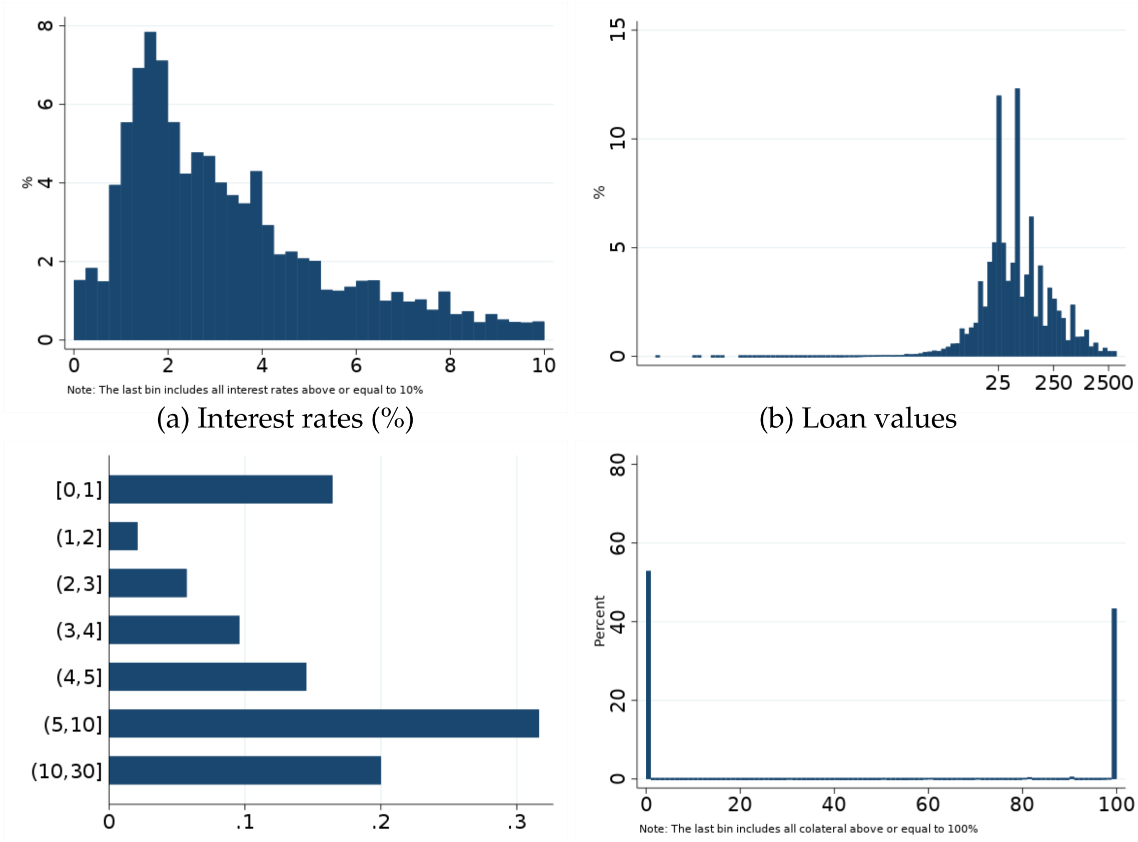
Table 1: **Descriptive statistics for firms.** This table provides statistics for firms in the sample. For each moment, four percentiles and the standard deviation are given. All moments are at the firm level and are averaged within firm over the years that each firm is in the sample. So, for example, 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. Total assets and revenue are in thousands of 2016 Euros.

case that these banks are specialized and have segmented the market in some way. They are large, general-purpose banks.

The sample consists of 162,270 firms. To illustrate their characteristics we present summary statistics from their loan portfolios and balance sheets in Table 1. Note that the statistics are within firm averages over the years that each firm is in the sample. The balance sheet variables reflect the characteristics of the population of firms. For the credit registry variables, the median firm has €123, 390 in loans. The median number of loans is 1.25 and the median share of a firm's loans that are collateralized is 87.5%. The majority of firms have loans from only one bank, but many firms have more bank relationships. At the 90th percentile a firm averages 2.1 banking relationships over the years that it is in the sample.

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. Eighty-eight percent of these were initiated from 2015 onward.[4] The average in-

[4]For a detailed breakdown of loans' origination year, see Figure 9 in Appendix C.



(a) Interest rates (%)

(b) Loan values

(c) Maturity (%)

(d) Collateral rate (%)

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 (in years) and collateral rates, respectively. The distributions in panels (a), (c) and (d) are weighted by loan value.

terest 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 value of €348, 455. Most loans are long term with 56% of credit coming from loans with a maturity of more than 5 years. For collateral, 50% of credit is through fully collateralized loans and nearly all other loans are uncollateralized. The most common type of collateral is a guarantee in terms of number of loans, followed by financial assets. While real estate is less used than the previous two categories it is used in larger loans. For more details on the type of assets used as collateral see Figure 10 in Appendix C.

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 2(d) 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 2(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. 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.

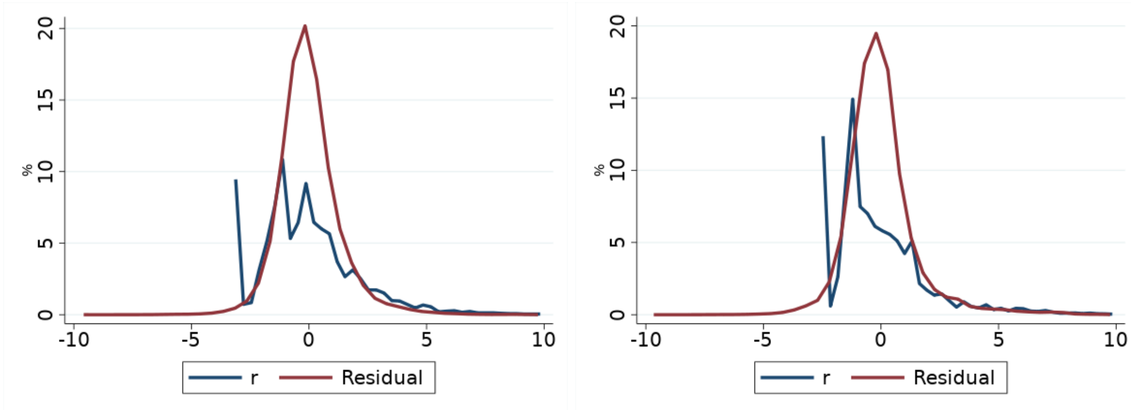
An additional possibility is that interest rates could vary because firms are lending from different banks. There are several reasons why different banks could charge different rates, which we will be agnostic about for now. This is another factor that is typically outside of models of risk free lending. The data, however, is rich enough to control for this. Specifically we will use bank fixed effects. We assume that to the extent that two banks charge different interest rates for equivalent loans, the differences are constant across the loans. Note that we will also control for maturity-time fixed effects to capture 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 and maturity-time variation, we run the following regression

$$r_{lbt} = \beta_1 L_{lt} + \beta_2 F_{lt} + \theta_b + \lambda_{mt} + \varepsilon_{lbt}, \tag{1}$$

where r_{lbt} is the interest rate on loan l from bank b at its initiation date t, L_{lt} is the loan amount, F_{lt} is a set of fixed effects for the adjustability of the rate, and θ_b and λ_{mt} are bank and maturity-time fixed effects. The time period for the analysis is monthly. 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.[5] The maturity-time fixed effects,

[5]There are three additional frequencies of adjustment in the data, which we also include fixed effects for. They are another frequency less than a year, another frequency more than a year, and a



(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 with mean centered around zero, 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 0.

we include 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 distribution of residuals from equation (1), once it is estimated on this sample. Comparing these distributions tells us how much variation in interest rates remains after controlling for all of the loan, bank, maturity and time characteristics that we have discussed. The means of the two distributions are aligned at zero to make them visually comparable. It is clear that the vast majority of interest rate variation remains. Overall, the standard deviation decreases from 1.99 to 1.31. The distribution of residuals has some additional mass in the center, but otherwise the distributions are similar.

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

category for all residual cases. These are categories are uncommon, representing 0.05%, 0.03% and 7.75% of the sample, respectively.

is due to this. To assess this, we re-do the previous exercise separately for different classes of collateral: guarantees, real and financial assets, and other forms of collateral. The results, provided in Figure 11 in Appendix C, show that distinguishing between different types of collateral is not materially important for the conclusion, with more than 60% of the distribution remaining unexplained in each case.

The second main approach to modeling financial frictions departs from the assumption that loans are collateralized. Instead 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 them. Since there is no risk, firms all pay the same risk free rate.

Finally, we perform the analysis on the set of loans that are uncollateralized. These loans fit the second theory of interest rates, under which there are endogenous borrow limits that ensure that lending is risk free even in the absence of collateral. The results fort his sample are in Figure 3(b). The message is the same as for collateralized loans: more than 70% 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. Loan, bank, maturity and time characteristics can only explain around 35% 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

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

Λ_{t+1} is the stochastic discount factor, χ_{t+1} is a indicator for whether default occurs, LGD_{t+1} (*loss given default*) is the share of the loan that is not 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

$$r = \frac{1 + \tilde{r}}{(1 - PD \times LGD)^{\frac{1}{T}}} - 1, \tag{3}$$

where $\tilde{r}$ is the risk-free interest rate, T the maturity of the loan, PD is 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. [6]

There are different microfoundations for LGD. Ottonello & Winberry (2020) assumes that the lender cannot recover the full value of the collateral. In Cooley & Quadrini (2001) there is asymmetric information and banks incur monitoring costs 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 time of loan origination, so we are agnostic about what the source of losses is in the case of default, even for fully collateralized loans. Additionally, we use two different measures of the probability of default (PD): 1) the value reported by the bank at the loan origination date; 2) the one estimated by Bank of Portugal,

[6] 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.

which follows the method described in Antunes et al. (2016). Our baseline results use the PD estimated by the Bank of Portugal, as the commercial bank ones is only reported by a subset of the sample.[7] 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 a one year period, in line with regulation implemented by the European Banking Authority. For this reason, we restrict the baseline exercise to loans with maturity up to one year.

To validate the reported PDs, we compare both PD measures with three different measures of observed default, within one year of the reported PD: 1) if any firm loan is in arrears - firm falls behind on payments by one month or more; 2) if any firm loans is overdue - firm falls behind on payments by three months or more; and 3) if the firm's share of overdue to total credit is above 2.5%, similar to the definition used by Antunes et al. (2016). Results in Figure 20 in Appendix D.1 show that any of the observed default measure is strongly and positively correlated with the two PD measures.

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 to have a PD below 1%. If risk is a driving factor in explaining the variance in distributions, after cleaning the interest rates from the loan characteristics, bank and maturity-time fixed effects, as in equation (1), there shouldn't be much interest rate variance left to explain. Figure 4 shows this is not the case: after cleaning the interest rates for low-risk firms, the variation is similar to the raw interest rate distribution, suggesting risk is not one of the main drives of the interest rate variation.

Panel (a) of Figure 12 in Appendix C reports a similar distribution but considering firms in the bottom 75% of the PD distribution, which means that they have a probably of default of less than 3.2%. Panels (b) and (c) of the same figure redo the

[7] Results using the commercial banks reported PDs are qualitatively and quantitatively in line across all the different specifications.

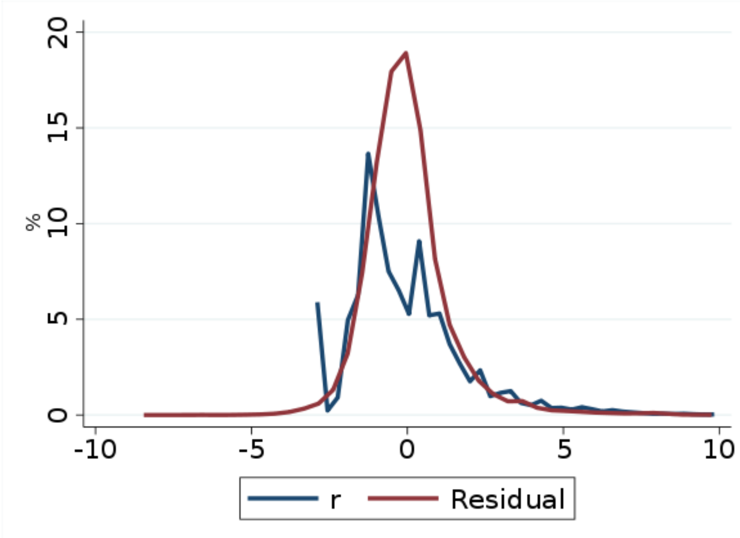


Figure 4: **Interest rate distributions before and after controlling for loan characteristics, banks and time for firms with $PD < 1\%$.** Interest rate distribution for firms with $PD < 1\%$ in blue and the residuals after controlling for loan characteristics, bank and maturity-time fixed effects in red. All distributions are centered with their means at 0.

same exercise but using the PD reported by commercial banks. Results are robust across the different measures and thresholds for low risk firms.

To more formally test the role of risk in explaining the interest rate variation, we start by taking logs of equation (3) and approximating the interest rate as

$$r \approx \tilde{r} - \log(1 - \chi LGD). \tag{4}$$

In specification (1) the maturity-time fixed effects control for the risk-free rate $\tilde{r}$ in the above equation. We then augment equation (1) with the extra risk variable to get:

$$r_{lbt} = -\log(1 - \chi LGD) + \beta_1 L_{lt} + \beta_2 F_{lt} + \beta_3 Col_{lt} + \theta_b + \lambda_{mt} + \varepsilon_{lbt}. \tag{5}$$

Additionally we include a vector Col_{lt} which includes the collateral rate and the type of collateral provided for loan l.

Figure 5 compares the actual interest rate distribution in blue, and the distribution after controlling for loan characteristics, the risk component, bank and maturity-times fixed effects in red. As it is possible to see from the figure, the two distributions overlay each other, suggesting that risk is not a major factor in ex-

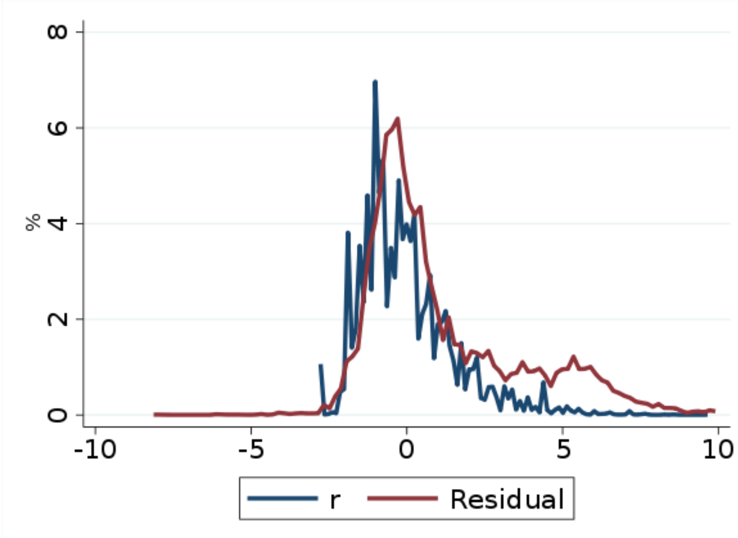


Figure 5: **Interest rate distribution before and after controlling for loan characteristics, risk, banks and time.** The blue line is the unconditional distribution and the red line is distribution after controlling for loan characteristics, risk, bank and maturity-time fixed effects. All distributions are centered with their means at 0.

plaining the interest rate distribution. This happens for two reasons: 1) a large fraction of firms have a PD close to zero. 50% of firms have a PD of 2% or lower and the PD at the 90th percentile is only 14%; 2) the LGD distribution is equally concentrated to the left, with a mean of 16% and the 90th percentile being 43%. On top of this, the overall joint PD and LGD distribution is concentrated to the left, with jointly low values of PDs and LGDs. For more details please see Table 6 in Appendix B, and panel (a) of Figure 19 in Appendix C for the joint distribution of these variables.

We conduct a battery of robustness tests to guarantee the validity of results. First, we test if results hold when using the PD reported by the commercial banks. Figure 13 in Appendix C illustrates that results are very similar to the baseline ones, with the raw distribution and the residual one overlaying each other. The reason is similar to before, as the distribution of PDs reported by commercial banks is also concentrated at low values. For more details please see Table 5 in Appendix B, and panel (b) of Figure 19 in Appendix C for the joint distribution of these variables.

We next extend the exercise to consider all loans independently of their maturity. For this, we assume the probability of default to be i.i.d. each period, which allows us to express the probability that a firm defaults within T years as $1-(1-PD_{1-year})^T$. Figure 14 in Appendix C illustrates that the results are not altered by extending the analysis to multi-period loans.

In regression (5) we impose the coefficient on the risk term to be -1, in line with what the theory dictates. In Figure 15 in Appendix C we relax that assumption and allow the coefficient on the risk term to be freely estimated.

Equation (3) for interest rate pricing is derived assuming linear utility and risk neutrality of the financial intermediary. While estimating the specification (5) we use maturity-time fixed effects to absorb the SDF, or risk-free interest rate, it can be the case that the SDF or risk-free interest rate does not enter linearly in the equation if one deviates from linear utility and risk-neutrality assumptions. To control for this we extend the baseline specification in two different directions. First we estimate the SDF directly, which under household's log utility takes the form of $\frac{c_t}{\mathbb{E}c_{t+1}}$, by taking today's and tomorrow's consumption directly from the data. Second, we estimate for every firm in the dataset a correlation coefficient between firm's revenues and aggregate consumption, to assess the risk's cyclicality component of each firm. We then introduce these terms, one at the time, linearly in the regression and equally interacted with the baseline risk term. Results in Figure 16 in Appendix C show results are robust to including the extra terms.

In addition, to guarantee the results are not driven by any structure we are imposing on the data, we estimating the specification semi-parametrically. For that we enter the PD and LGD variables separately and interacted with ten decile dummies for each variable. Additionally, we include a interaction term between PD, LGD and the dummies for the deciles of the two variables. Results in Figure 17 in Appendix C remain robust to this semi-parametric estimation.

To guarantee results are not driven by the Covid-period and potential guarantees offered by the government that would eliminate loan's risk for banks we redo the baseline exercise for the pre 2020-period. Results in Figure 18 in Appendix C illustrate results are robust to dropping the Covid-period.

Lastly, in Appendix D.2 we invert the problem and estimate the required PD and LGD distribution to match the observed interest rate distribution. We conclude that the required PD and LGD distribution would need to be infeasibly high, with a large fraction of loans needing to have PDs and LGDs above 100% to match the interest rate on the loan.

4 Interest rate determinants

The previous section established that it is not risk or loan characteristics that explain the interest rate distribution. In addition to the heterogeneous firms literature, which focuses either on collateral or risk as drivers of interest rate variation, there are branches of literature that offer different explanations for interest rate 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. (2024b), for example, focus on the bank-firm relationship as a potential explanatory factor. In this section we start by exploring the contribution of both banks and bank-firm relationships to explaining the interest rate distribution. We instead find that a firm fixed component is the major driver of the dispersion in interest rates. We conclude the section by adopting a statistical model to quantify the role of the firm fixed component.

4.1 Banks and interest rate variation

The earlier results have already suggested that bank heterogeneity cannot explain much of the variation in interest rates that we observe in the data. The results in Figure 3 in Section 3.1 showed that bank fixed effects along with controls for loan characteristics and maturity-time fixed effects leave most variation in interest rates unexplained. In this section we make this point more explicitly, and quantify how much of the variation in the data can be attributed to bank heterogeneity.

To do this, we use an extension of the regression from equation (1) in which we add firm fixed effects. Specifically, we use the following regression:

$$r_{lfbt} = \beta LC_{lfbt} + \alpha_f + \theta_b + \lambda_{mt} + \epsilon_{lfbt}, \tag{6}$$

where α_f is the firm fixed effect. LC is a vector of loan characteristics, with the same controls as it contained in equation (1). We emphasize that these fixed effects capture the effects of *all* relevant firm and bank fixed characteristics that are common over time. We are including the firm fixed effect to avoid attributing variation in interest rates due to this source to banks, which would occur if there is sorting of firms across banks. The inclusion of these fixed effects requires firms and banks to have multiple new loans to provide variation for identification of other coefficients. This specification also allows us to quantify the relevance of firm characteristics for interest rate variation, which we will build on in subsequent sections.

In the baseline regression we avoid the inclusion of bank-time and firm-time fixed effects for three main reasons: 1) to avoid estimating too many fixed effects, which can lead to noisy estimates of these effects as pointed by Kwon (2023) and Best et al. (2023); 2) the sample size reduces particularly with the time dimension of the firm fixed effect, as it requires firms to have multiple loans originating in the same time period; and 3) the source of variation is narrower as it must be within time period. Nonetheless, we later show that results are quantitatively robust to the inclusion of bank-time and firm-time fixed effects, that capture any bank and firm characteristics which may vary over time and affect interest rate dispersion.

Once we have estimated the regression in equation (6), we quantify the contribution of each component on the right-hand side in accounting for interest rate variation. Specifically, the five components are the loan characteristic controls, the bank fixed effects, the firm fixed effects, the maturity-time fixed effects, and the residuals. For each loan-firm-bank-time observation, let C^p_{lfbt} denote the value of component p from the right-hand side of the estimate regression. The overall contribution of this component in accounting for the variation in interest rates in the

	Contribution
(1) Loan Characteristics	0.041
(2) Firm FE	0.392
(3) Bank FE	0.079
(4) Maturity-Time FE	0.212
(5) Residual	0.277
# Obs.	552,041

Table 2: **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).

data is

$$\frac{cov(i_{lfbt}, C^p_{lfbt})}{var(i_{lfbt})}. \tag{7}$$

By definition, the sum of the contributions of the four components is equal to one.

Table 2 reports the results. In line with the results from Section 3.1, the loan characteristics and the bank fixed effects are not major drivers of interest rate dispersions, accounting for 4.1% and 7.9% respectively. The maturity-time fixed effects has more explanatory power, by accounting for 21.1% of the dispersion. But the major driver of interest rate dispersion is the firm fixed component, accounting for close to 40% of the overall variation. We have a large residual, with 27.7% of the variation not being accounted for by any of the components in the regression.

The large residual could be driven by several factors, which can affect our overall conclusions. First, it could be related to time varying firm characteristics. To account for this, we include a rich array of firm characteristics in three categories: First, balance sheet variables: age; log of the number of employees; log total assets log of revenue; growth rate of revenue; log of intangible assets; TFP, estimated following Ackerberg et al. (2015); leverage ratio; liquidity ratio; interest rate coverage ratio; an indicator for whether the firm was late on any loan repayments in the year the loan was granted. The rationale for including this variables is to capture the state of the firm at the time the loan is granted and indicators of past performance that can be use to gauge the firm future performance, in line with the theory proposed by Petersen & Rajan (1995), that banks price loans factoring in the future value of the relationship with the firm.

Second, despite having already established risk is not a major driver of interest rate dispersion we include the probability of default estimated by the Bank of Portugal.[8] Third, to capture potential bank-firm relationship effects that may get mixed up with a firm fixed effect, we include variables that may capture the strength of the firm relationship with bank b: log value of all loans received from bank b over the last 5 years; number of loans received from bank b over the last 5 years; percentage of loans coming from bank b over the last 5 years; total value of loans coming from bank b over the last 5 years; duration of the relationship. Lastly, we include variables capturing how open a firm is to working with different banks: number of bank relationships; concentration of bank relationships (measured with a HHI index).

Table 9 in Appendix B reports the results when including the firm control variables. Notice, that due to the inclusion of the additional controls, the sample size decreases considerably. Despite the addition of the controls and the different sample size, results are qualitatively in line with the baseline ones. The firm fixed component is still the most important component, accounting for close to 37% of the variation, while banks account for slightly more than 5% of the variation.

Next, we test the inclusion of bank-time and firm-time fixed effects. To still try to avoid estimating too many fixed effects, we now include only a maturity fixed effect, instead of maturity-time. On the bank side, we use bank-month-year fixed effects.[9] This does not restrict the sample since all banks issue multiple loans each month. For firms, the fixed effect is more restrictive since most firms do not take out multiple new loans within a year, much less a month. Therefore for the firm-time fixed effects, we consider time periods of one year, two years, three years,

[8] As banks do not report PDs and LGDs for all the loans, in our benchmark scenario we include only the PDs estimated by the Bank of Portugal.

[9] The results are robust to using bank-year fixed effects as well

or and also construct time periods that corresponds to phases of the business cycle.[10] This makes our estimates of the firm-time fixed effect conservative as the firm characteristics may change within the time period that we consider. As results illustrate, the firm-year fixed effect will account for considerably more variation of interest rates than the firm-cycle fixed effect, in support of the idea that firm-time fixed effects estimates are conservative estimates of the overall importance of firm fixed plus time varying characteristics.[11]

Table 10 in Appendix B reports the results, with separate columns for each of the time periods used for the firm-time fixed effects. Focusing on the results for banks first, we see that the component becomes more important than when using only bank fixed effects. Despite the increased importance, when considering firm-year fixed effects, the bank-month-year fixed effect accounts for 15% of overall variation, still considerably less than the estimates for the firm fixed component and for the firm-time fixed effect. Results for the importance of loan characteristics are quantitatively similarl to previous estimates, accounting for only 4.9–5.9% of the variation. By far the most important factor for interest rate variation is firm characteristics. They explain 53.4–63.6% of the total variation. It is also important to note that the unexplained variation is smaller than in the baseline results. When considering firm-year fixed effects the importance of the residual is reduced to 16.5% overall.

Lastly, we go back to the baseline regression (6) and test three additional robustness tests: 1) including all types of loans; 2) using spreads on the left hand side of the regression instead of the interest rate;[12] and 3) excluding loans initiated after 2019, to exclude the Covid-period and any persistent effects of policies adopted during that period. Results in Tables 11-13 show results are robust across the dif-

[10]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.

[11]As we control for firm-time fixed effects we do not include the time varying firm observables.

[12]Spreads are defined as the interest rate minus the risk free interest rate. The variable is directly reported by the commercial banks.

ferent robustness tests, with the firm fixed-component always being the most important source of interest rate variation and the bank component maxing at 10.8%.

4.2 Bank-firm relationships and interest rate variation

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

$$r_{lfbt} = \beta LC_{lfbt} + \gamma_{bf} + \lambda_{mt} + \epsilon_{lfbt}, \tag{8}$$

where γ_{fb} is a vector of bank-firm fixed effects. Noticed the inclusion of bank-firm fixed effects reduces the sample to firms that switch banks or have multiple bank relationships. With these fixed effects we can measure the total variation coming from firms, banks, and the relationships between them. We want to isolate the component of this that is due to *relationships*, above and beyond the variation to firm and bank characteristics independently. Our method for this is as follows. After estimating the above regression, we take exactly the same sample and re-estimate the baseline equation (6) using firm and bank fixed effects instead of bank-firm fixed effects. This provides us with estimates of the independent firm and bank effects. We can then compute the share of variation in interest rates accounted for by the fixed effects in the two specifications, and the *additional* variation accounted for bank-firm effects is the contribution of bank-firm relationships.

The results are presented in Table 3. Bank-firm relationships accounts for less than 10% of the overall variation in spreads, indicating that it is an important component for overall variation, but still considerably smaller than the firm fixed component. If we repeat the analysis using all credit instruments instead of just firm financing loans, or spreads on the left hand side of equation (8) the conclusions are similar. Bank-firm relationships account for 9.5–11.1% of the variation, as presented in Tables 14 and 15 in Appendix B.

Lastly, as the firm-bank relationship can change over time, we are potentially missing an important time component of the relationship by not including time

	Contribution
(1) Firm FE	0.392
(2) Bank FE	0.052
(3) Firm-Bank FE	0.094
# Obs.	150,294

Table 3: **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 Firm-Bank fixed effect 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.

fixed effects. As such, we extend equation (8) to include firm-bank-time fixed effects. The time dimension of these uses the same set of periods as the firm-time fixed effects in section 4.1. Results are presented in Table 16 in Appendix B and are supportive of the above conclusions, with the bank-firm-time component explaining less than 5% of the overall variation in interest rates.

4.3 Permanent component of spreads

So far, we have established that firm fixed characteristics are the major driver of interest rate variation. Besides the loan, bank and bank-firm 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 6 plots these two statistics in our data, with panel (a) depicting the standard deviation and panel (b) depicting the autocorrelation profile of interest rates within firm over different loans.[13] First, the standard deviation is decreasing only mildly over loans,

[13]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

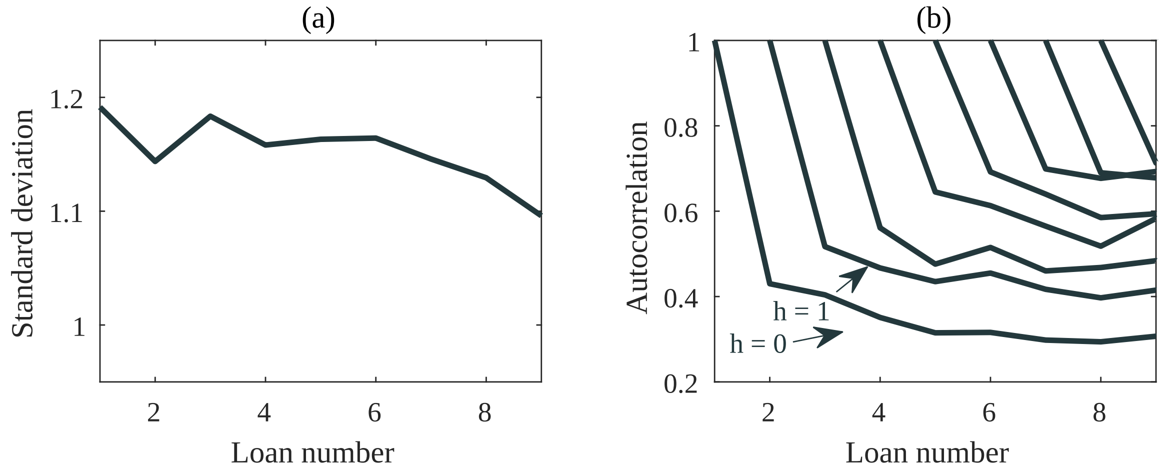


Figure 6: **Standard deviation and autocorrelation of interest rates across loans, within firm.** The left panel presents the standard deviation of interest rates by loan number, after controlling for loan characteristics, bank, maturity-time and sector fixed effects. The right panel presents the autocorrelation of spreads by loan number l and loan number $h \leq l$. Across lines h changes, while l changes along the lines.

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.[14] 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:

$$\underbrace{r_{i,l}}_{\text{Interest rate}} = \underbrace{u_{i,l} + v_{i,l}}_{\text{Ex-ante component}} + \underbrace{w_{i,l} + z_{i,l}}_{\text{Ex-post component}} , \tag{9}$$

characteristics, bank, maturity-time and sector fixed effects and then use the interest rate residuals.

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

where

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

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 for by minimizing the sum of squared differences between the model and empirical autocovariance. Table 4 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. The model fit to the data can be found in panel (a) of

	ρ_u	ρ_v	ρ_w	σ_θ	σ_u	σ_v	σ_ϵ	σ_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 4: Equally-weighted minimum distance estimates of the statistical model. Standard deviation in parenthesis.

Figure 7, which plots the interest rate autocovariance generated by the model in blue and the empirical counterpart in orange.

Panel (a) of Figure 7 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 1.1 percentage points. This again demonstrates that there seems to be differences across 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 Figure 7 illustrates the importance of the ex-ante component for the variance as a function of loan number. The ex-ante component

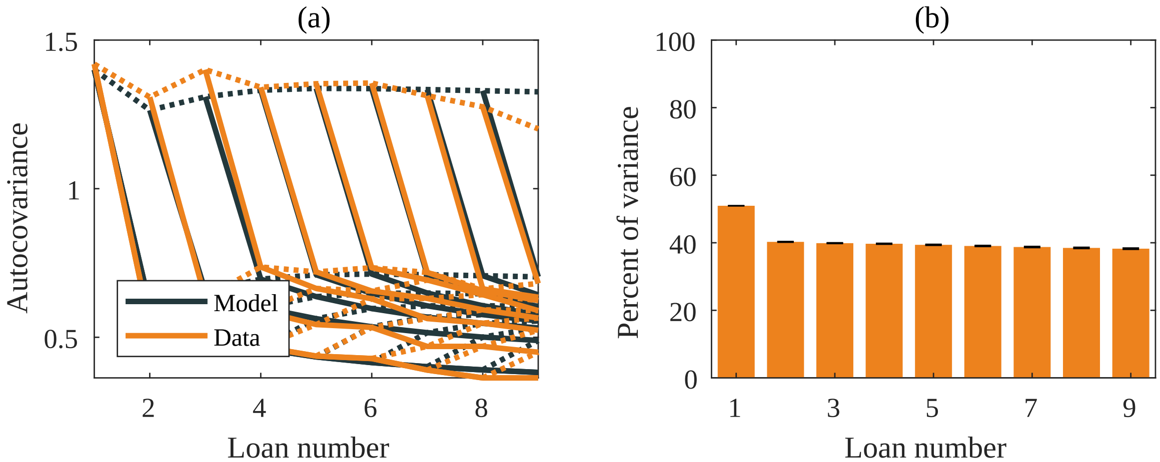


Figure 7: **Model fit to the data and ex-ante contribution.** Panel (a) plots both the autocovariance structure of the spread in the calibrated statistical model (blue) and in the data (orange). Panel (b) presents the percent of variance explained by ex-ante factors as a function of the loan number, according to the statistical model.

contribution is close to 50% at the first loan, and around 40% of the variance of interest rates after the ninth loan, which is quantitatively in line with what we find the firm fixed component to account for in Section 4.1.

Lastly, we take another alternative approach to guarantee the quantitative robustness of our results. We take the the firm-time fixed effects α_{ft} from the robustness exercise we ran in Section 4.1, in which we include firm-time and bank-year fixed effects, and ran the following regression:

$$\alpha_{ft} = \beta X_{ft} + \theta_{ft} + \sigma_f + \epsilon_{ft}. \tag{10}$$

where θ_{ft} is a sector time fixed-effect, to capture differences in sector cyclicality, σ_f is a firm fixed effect and X_{ft} is a vector of firm controls. The controls include are the same as the ones discussed at the end of Section 4.1.

Results, in Table 17 in Appendix B, show the firm fixed component to represent between 30% and 35% of overall interest rate distribution, in line with the results presented in Section 4.1 and from the decomposition of the autocovariance of interest rates into ex-ante and transitory components.

5 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.

5.1 Environment

Time is discrete and infinite. There is a continuum of firms and a continuum of identical banks.[15] 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 $f(0) = 0$, $f'(0) > 0$ 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

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

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 $U[0, \hat{r}/e]$,[16] where $\hat{r} > 0$ is a parameter that determines the upper bound of the distribution for a firm exerting maximum effort. The cost of effort is θe.

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.

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

[16] U denotes the uniform distribution.

that dissolves its bank relationship, which occurs with probability $\delta_b \in (0,1)$, then it enters the next period without a bank.

5.2 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

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

$$\begin{aligned}
\text{s.t. } k' &= (1-\delta_k)k + i, \\
z' &\sim G_z(z'|z),
\end{aligned}$$

where $\mathbb{1}_i$ is an indicator for whether a firm invests ($i > 0$). The solution to this problem is an investment function $i(z,k,\theta,r')$. Investment is increasing in z, decreasing in k and decreasing in r. These relationships are standard.[17] The effect of θ on the investment decision has several components. A higher θ means that search is more costly for a firm, so it expects a higher interest rate on average and to invest less. A high θ also decreases the probability of obtaining a lower interest rate than r' in the future, leading to more investment in the current period.

The search problem for a firm with a bank is

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

[17] 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.

and the problem for a firm without a bank is

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

These problems yield policy functions for effort for firms with a bank, $e(z,k,\theta,r)$, 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.

5.3 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.

6 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.[18] 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 in the data, including measures of economic size and performance.

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 ratio-

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

nalizing 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).

Best, M. C., Hjort, J., & Szakonyi, D. (2023). Individuals and organizations as sources of state effectiveness. *American Economic Review*, 113(8), 2121–2167.

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. (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.

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.

Kwon, S. (2023). Optimal shrinkage estimation of fixed effects in linear panel data models. *arXiv preprint arXiv:2308.12485*.

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 (%)	32	0	8	13	26	36	40	43	45	51	(12)	32561
PD (%)	4	0	0	0	1	1	3	12	14	80	(11)	32561

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

	(1)											
	mean	p1	p5	p10	p25	p50	p75	p90	p95	p99	sd	count
LGD (%)	16	0	0	0	0	0	36	43	45	47	(19)	90688
PD (%)	11	0	0	1	1	2	7	14	100	1.00	(26)	90688

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.

	Contribution
(1) Loan+Firm Characteristics	0.069
(2) Firm FE	0.369
(3) Bank FE	0.053
(4) Maturity-Time FE	0.197
(5) Residual	0.311
# Obs.	165,222

Table 9: **Contribution of loan and firm characteristics, firm, bank and maturity-time fixed effects to interest rate variation.** The different lines represent the different components included in equation (6) plus the firm characteristics. Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (7).

	Cycles	Triannual	Biannual	Annual
(1) Loan characteristics	0.049	0.056	0.059	0.049
(2) Firm-Time FE	0.534	0.554	0.563	0.636
(3) Bank-Month-Year FE	0.207	0.190	0.186	0.150
(4) Residual	0.210	0.200	0.193	0.165
# Obs.	465528	454327	416571	344495

Table 10: **Contribution of loan characteristics, firm-time and bank-time fixed effects to interest rate variation.** Each column represents the periodicity of the time component of the firm-time fixed effect and sums to 1. The different lines represent the different components included in the regression. Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (7).

	Contribution
(1) Loan Characteristics	0.222
(2) Firm FE	0.332
(3) Bank FE	0.108
(4) Maturity-Time FE	0.096
(5) Residual	0.242
# Obs.	1,735,054

Table 11: **Contribution of loan characteristics, firm, bank and maturity-time fixed effects to interest rate variation for all loan types.** 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).

	Contribution
(1) Loan Characteristics	0.042
(2) Firm FE	0.430
(3) Bank FE	0.080
(4) Maturity-Time FE	0.174
(5) Residual	0.274
# Obs.	552,027

Table 12: **Contribution of loan characteristics, firm, bank and maturity-time fixed effects to spread variation.** The different lines represent the different components included in equation (6). Each value represents the contribution of the respective component to the overall spread variance, given by equation (7).

	Contribution
(1) Loan Characteristics	0.043
(2) Firm FE	0.481
(3) Bank FE	0.107
(4) Maturity-Time FE	0.117
(5) Residual	0.253
# Obs.	269,155

Table 13: **Contribution of loan characteristics, firm, bank and maturity-time fixed effects to interest rate variation excluding the Covid-period.** 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).

	Contribution
(1) Firm FE	0.363
(2) Bank FE	0.086
(3) Firm-Bank FE	0.095
# Obs.	496,306

Table 14: **Contribution of firm, bank and firm-bank fixed effects to interest rate variation for all loan types.** 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-Bank fixed effect 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.

	Contribution
(1) Firm FE	0.332
(2) Bank FE	0.075
(3) Firm-Bank FE	0.111
# Obs.	150,294

Table 15: **Contribution of firm, bank and firm-bank fixed effects to spread variation.** The different lines represent the different components. Each value represents the contribution of the respective component to the overall spread variance, given by equation (7). The Firm-Bank fixed effect 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.

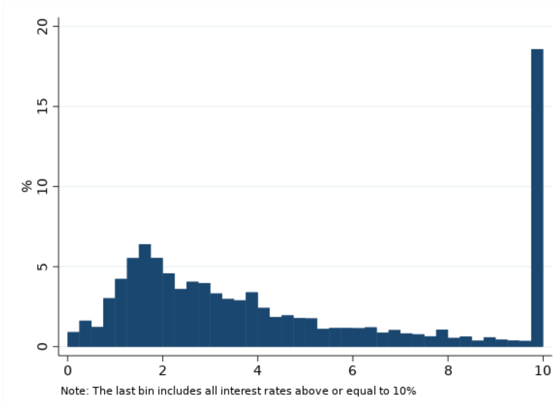
	Cycles	Triannual	Biannual	Annual
(1) Firm-Time	0.628	0.649	0.655	0.730
(2) Bank-Time FE	0.069	0.055	0.056	0.049
(3) Firm-Bank-Time FE	0.040	0.033	0.028	0.023
# Obs.	397964	383578	344999	280528

Table 16: **Contribution of firm-time, bank-time and firm-bank-time fixed effects to interest rate variation.** Each column represents the periodicity of the time component of the fixed effects. 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-Bank fixed effect 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.

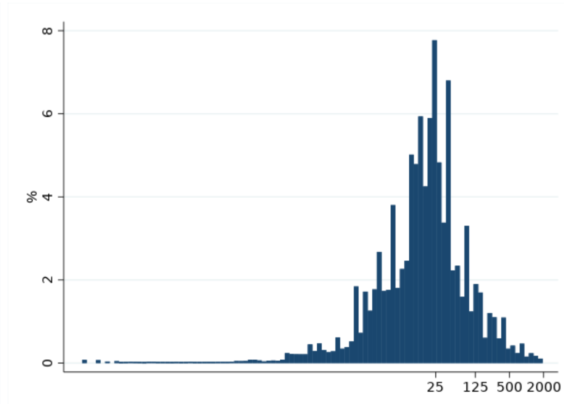
	Cycles	Triannual	Biannual	Annual
(1) Firm-Time FE	0.549	0.572	0.583	0.655
(1.1) Time-varying firm obs.	-0.022	0.082	-0.028	0.058
(1.2) Time-varying firm unobs.	0.106	0.105	0.111	0.125
(1.3) Time-Sector FE	0.117	0.149	0.182	0.132
(1.4) Firm FE	0.348	0.236	0.318	0.339
# Firm-time Obs.	9933	11550	12459	11895

Table 17: **Decomposition of the contribution of firm characteristics to interest rate variation.** Each column represents the periodicity of the time fixed effect. (1.1) to (1.4) represent different components in equation (10), with component (1.2) representing the contribution of the residual, and sum up to the value of the Firm-Time FE in line (1).

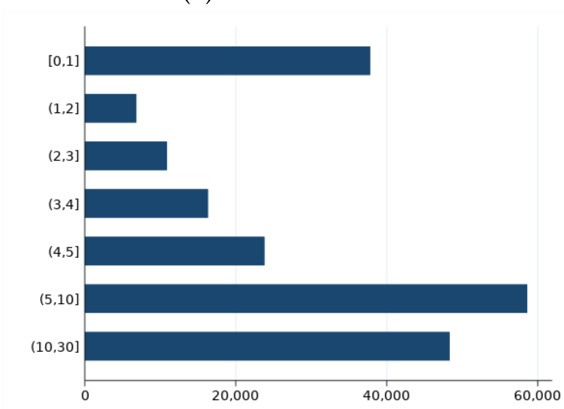
C Additional Figures



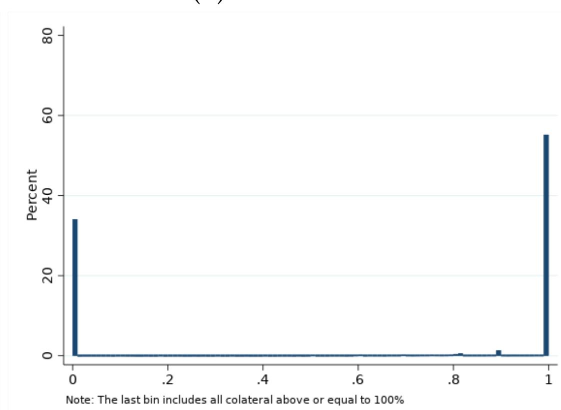
(a) Interest rates



(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.

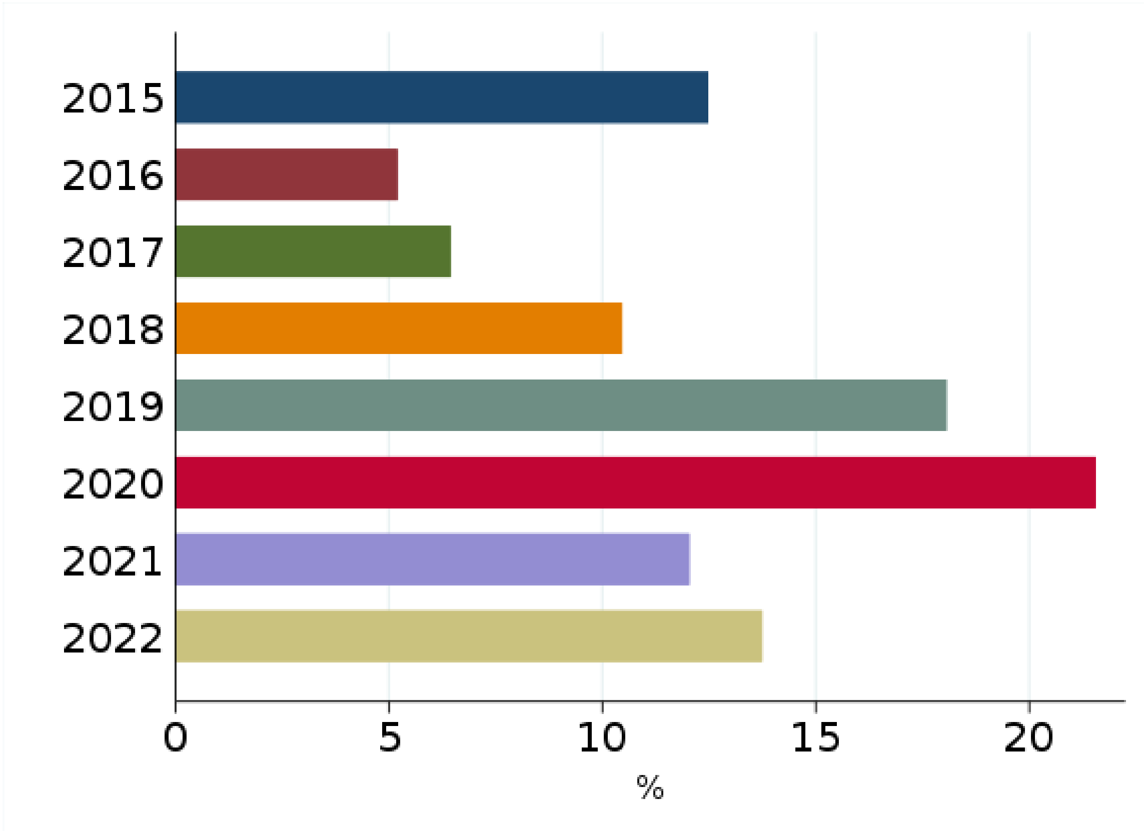
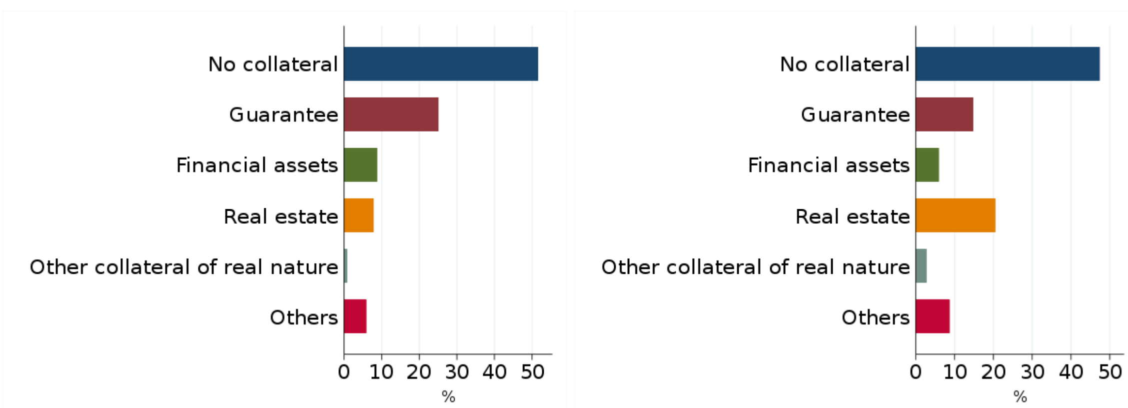


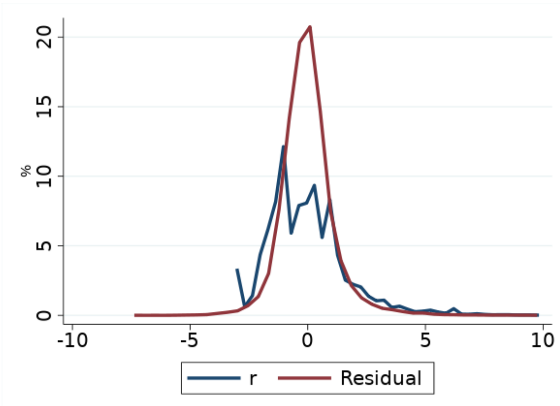
Figure 9: Distribution of loans by origination year. 2015 stands for all the loans initiated in or before 2015.



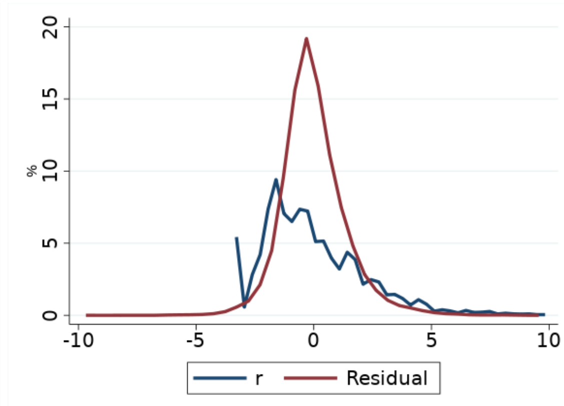
(a) Share of loans

(b) Loan values

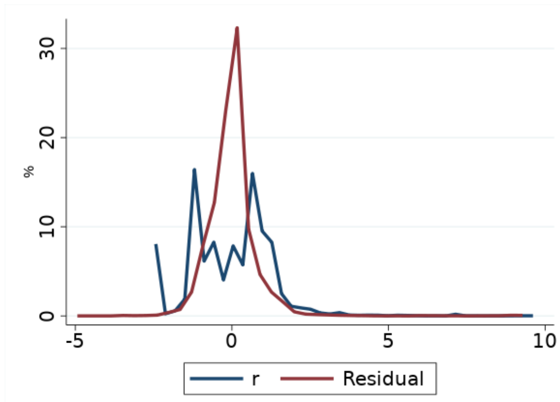
Figure 10: **Distributions of collateral types.** Panels (a) and (b) are the distributions of collateral types by number of loans and by loan values.



(a) Real and financial asset.

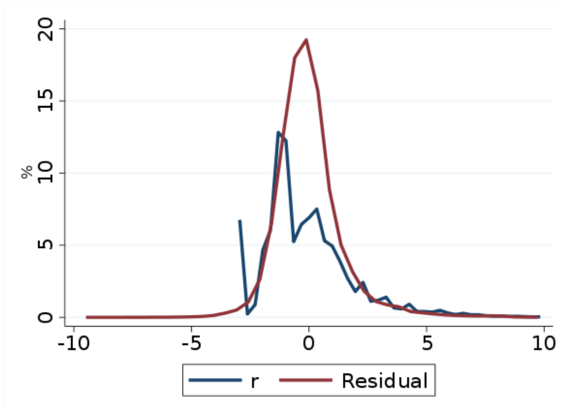


(b) Guarantees.

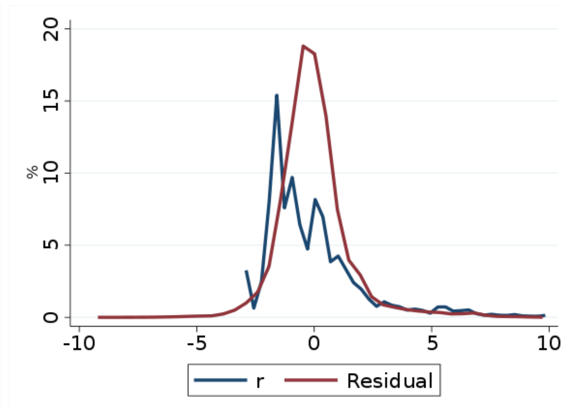


(c) Other types.

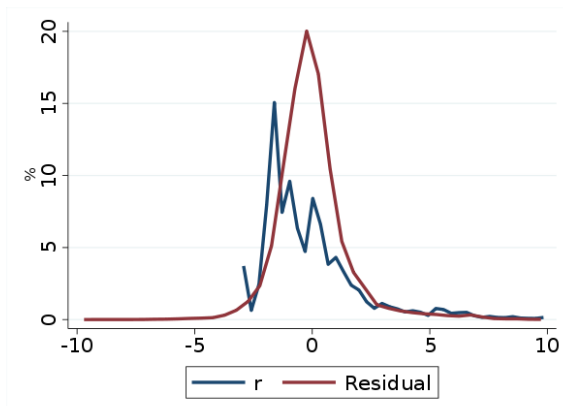
Figure 11: **Spread distributions before and after controlling for loan characteristics, banks and maturity-time.** Panel (a), (b) and (c) present the distribution of spreads for fully collateralized loans back with real and financial assets, guarantees and other types of collateral respectively. In both panels the blue line is the unconditional distribution with mean centered around zero, 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 0.



(a) BoP $PD < 3\%$.



(b) $PD < 1\%$.



(c) $PD < 3\%$.

Figure 12: **Spread distributions before and after controlling for loan characteristics, bank and maturity-time fixed effects.** Spread distribution for low risk firms in blue and the residuals after controlling for loan characteristics, bank and maturity-time fixed effects in red. Top left figure considers firms with a $PD < 3\%$, using the PD reported by the Bank of Portugal. Top right figure and bottom figure consider firms with $PD < 1\%$ and $PD < 3\%$ respectively, using the PD reported by the commercial banks.

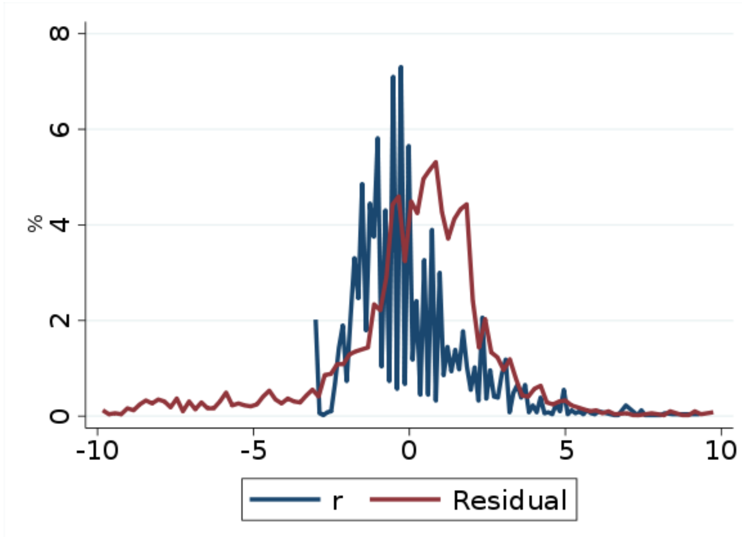
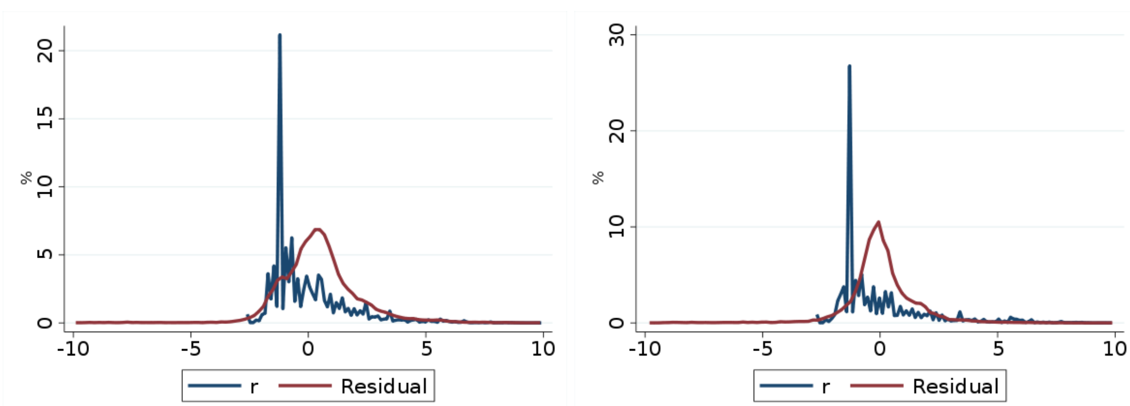
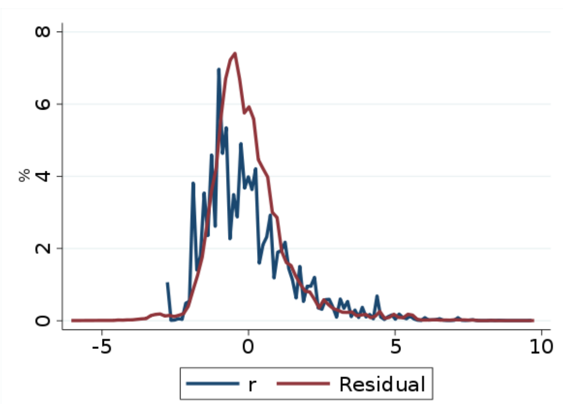


Figure 13: **Interest rate distribution before and after controlling for loan characteristics, risk, banks and time.** Distribution of interest rates implied by specification (5) after controlling for loan characteristics, risk, bank and maturity-time fixed effects in red and using the probability of default reported by commercial banks, and the raw distribution of interest rates in blue. All distributions are centered with their means at 0.

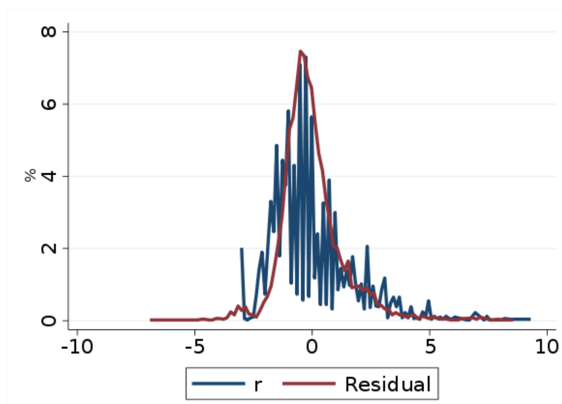


(a) PDs estimated by the Bank of Portugal. (b) PDs reported by the commercial banks.

Figure 14: **Interest rate distribution before and after controlling for loan characteristics, risk, banks and time for multi-period loans.** Distribution of interest rates implied by specification (5) after controlling for loan characteristics, risk, bank and maturity-time fixed effects in red, and the raw distribution of interest rates in blue. Both distributions consider multi-period loans. All distributions are centered with their means at 0.

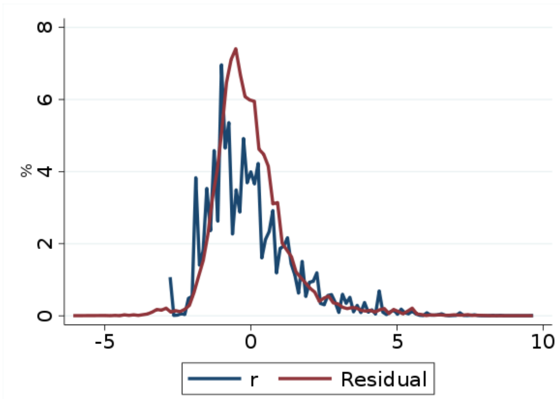


(a) PDs estimated by the Bank of Portugal.

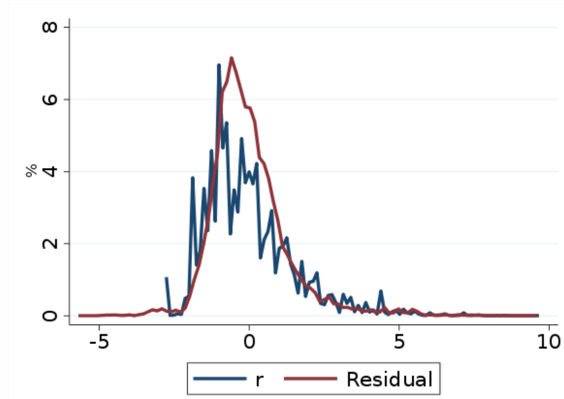


(b) PDs reported by the commercial banks.

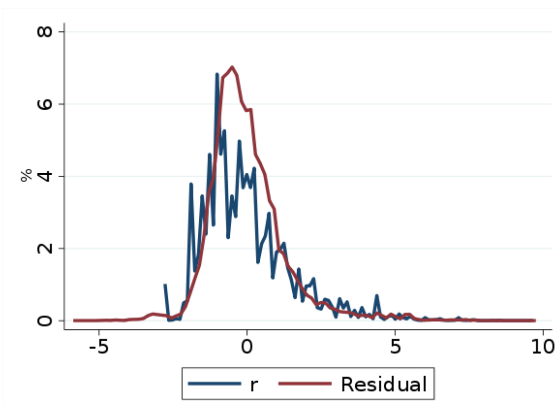
Figure 15: **Interest rate distribution before and after controlling for loan characteristics, risk, banks and time.** Distribution of interest rates implied by specification (5) after controlling for loan characteristics, risk, bank and maturity-time fixed effects in red while allowing the coefficient on the risk component to be freely estimated, and the raw distribution of interest rates in blue. All distributions are centered with their means at 0.



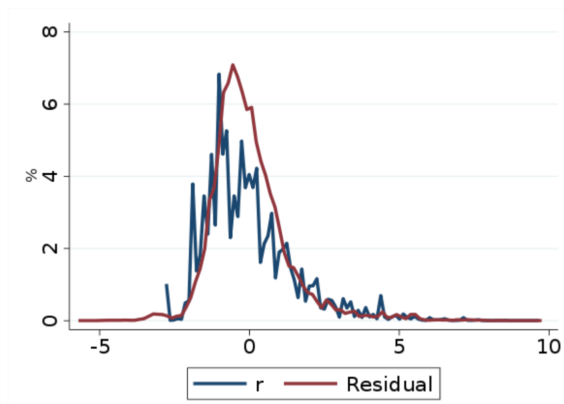
(a) Correlation coefficient.



(b) Correlation coefficient interacted with risk term.



(c) SDF.



(d) SDF interacted with risk term.

Figure 16: **Interest rate distribution before and after controlling for loan characteristics, risk, banks and time including SDF or correlation coefficient.** Distribution of interest rates implied by specification (5) after controlling for loan characteristics, risk, bank and maturity-time fixed effects in red while allowing the coefficient on the risk component to be freely estimated and including the SDF or the correlation coefficient, and the raw distribution of interest rates in blue. All distributions are centered with their means at 0. PD estimated by the Bank of Portugal.

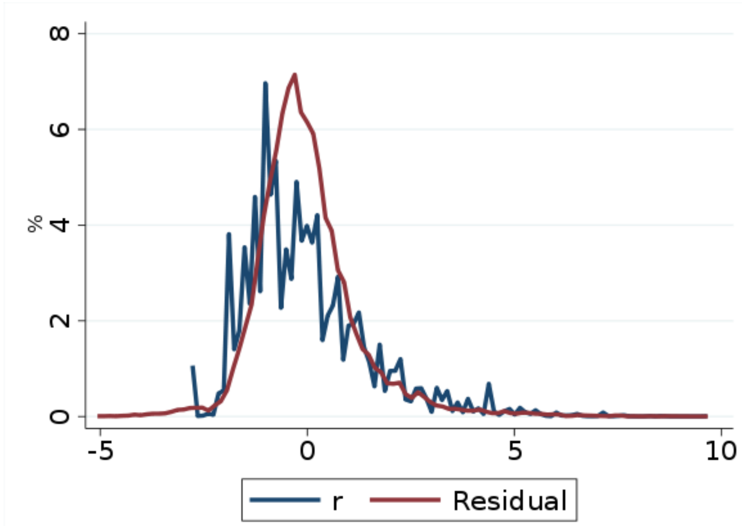


Figure 17: **Interest rate distribution before and after controlling for loan characteristics, risk, banks and time.** The blue line is the unconditional distribution and the red line is distribution after controlling for loan characteristics, risk component in a semi-parametric way, bank and maturity-time fixed effects. All distributions are centered with their means at 0.

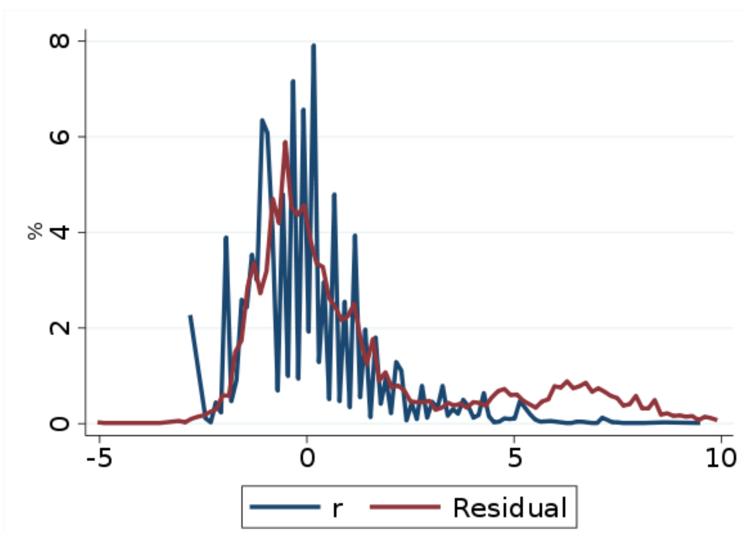
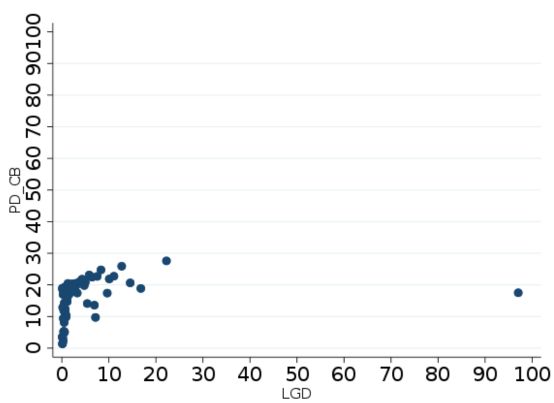
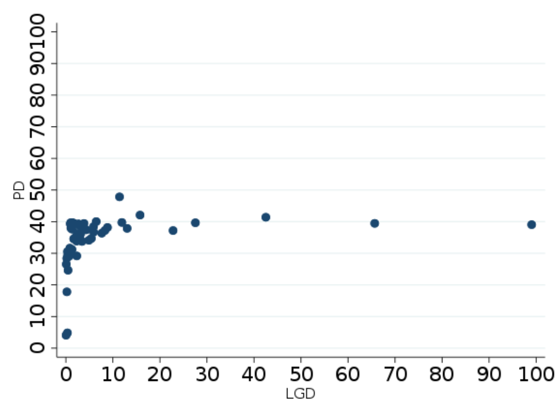


Figure 18: **Interest rate distribution before and after controlling for loan characteristics, risk, banks and time.** The blue line is the unconditional distribution and the red line is distribution after controlling for loan characteristics, risk, bank and maturity-time fixed effects. All distributions are centered with their means at 0. Only post 2021 period considered.



(a) Joint PD reported by the Bank of Portugal and LGD.



(b) Joint PD reported by the commercial banks and LGD.

Figure 19: **Each point represents a LGD percentile and the respective PD average.**

D PD and LGD distribution

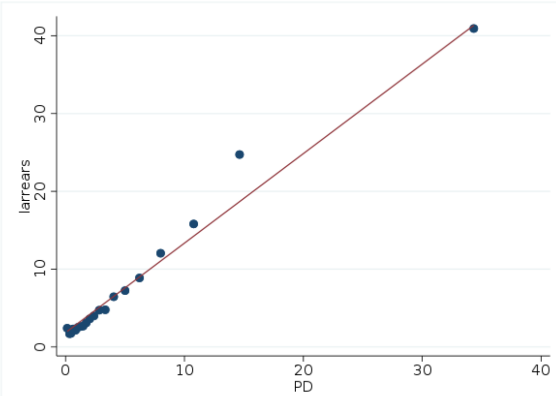
D.1 Validating the PD distribution

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:

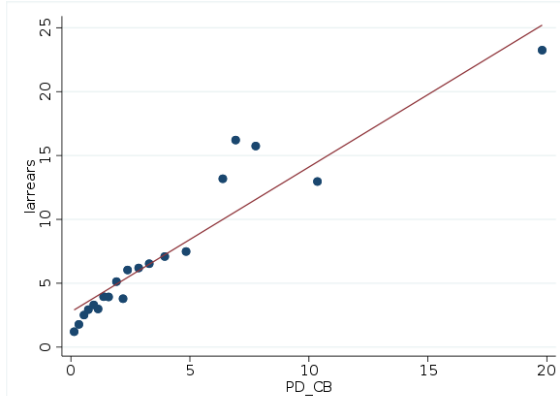
- In arrears: If at any point in time the firm falls behind on payments for at least one month;
- In default: If at any point in time the firm falls behind on payments for at least three months;
- Default_amount: If the share of overdue to total credit is above 2.5%, similar to Antunes et al. (2016); [19]

Figure 20 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.

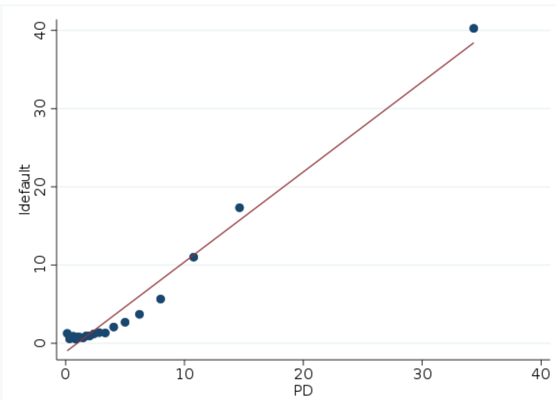
[19] Overdue credit is define as credit for which the borrower as not made payments by at least 90 days.



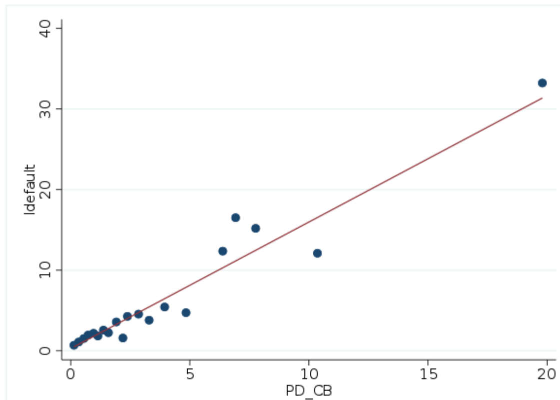
(a) In arrears and PD.



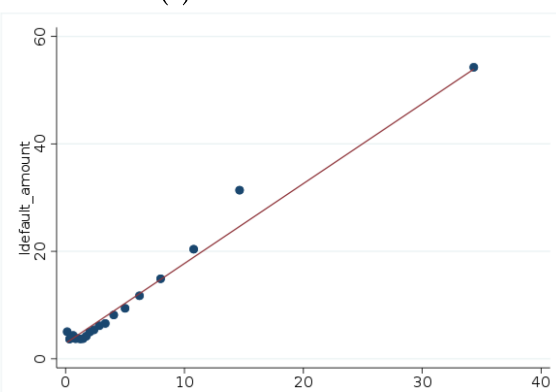
(b) In arrears and PD from the BoP.



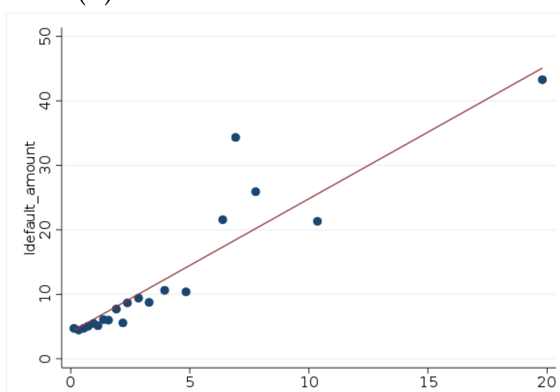
(c) Default and PD.



(d) Default and PD from the BoP.



(e) Default amount and PD.



(f) Default amount and PD from the BoP.

Figure 20: **PDs and actual default rates.**

D.2 Required LGD and PD distributions

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 (5) 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

$$-\log(1 - PD_{lt}LGD_{lt}) = \epsilon_{lbt}, \tag{11}$$

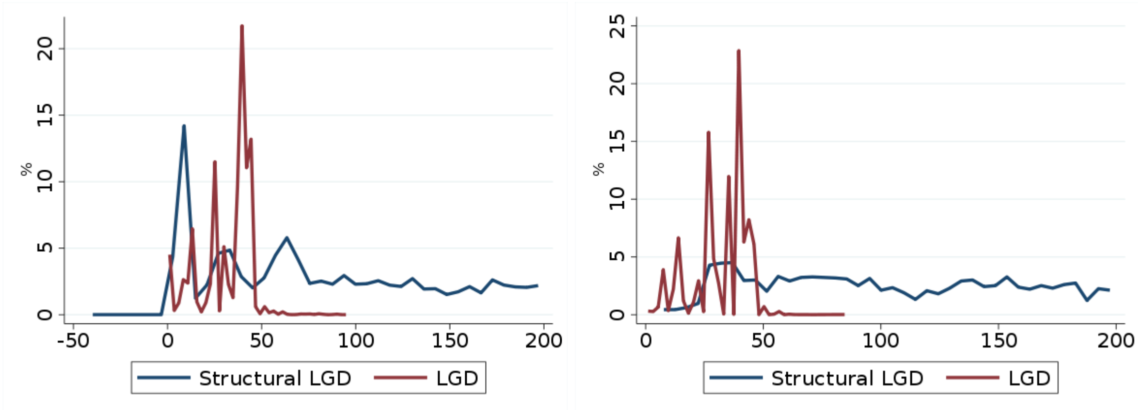
to determine what values of PD_{lt} and LGD_{lt} would be needed to fit the data.[20]

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

$$LGD_{lt} = \frac{1 - \exp(\epsilon_{lbt})}{PD_{lt}}. \tag{12}$$

Figure 21 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 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 21 to 200, but the median is 177 and

[20]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 estimated from equation (5).



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

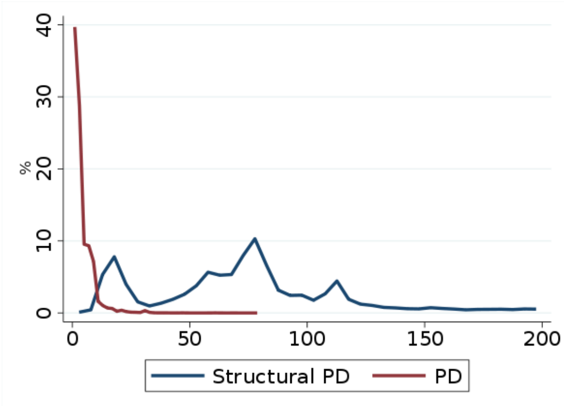
Figure 21: **LGDs: actual vs. required to match interest rate distribution.** Probability density function of the required LGDs to match the interest rate distribution following equation (13) in blue, and of the actual LGD distribution in red.

156 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.

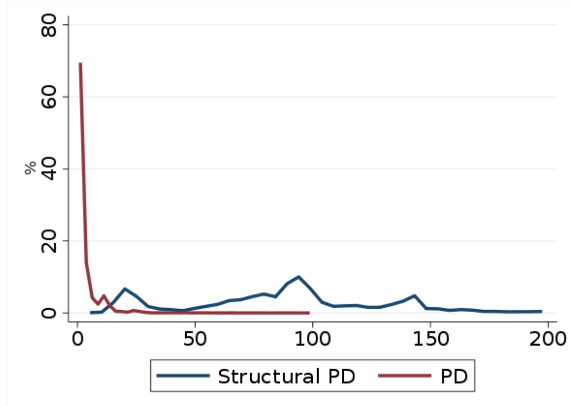
PD We next calculate, given the observed LGD, what would the PD distribution need to be to generate a interest rate distribution in line with the observed one, given the LGD distribution. The PD is given by

$$PD_{lt} = \frac{1 - \exp(\epsilon_{lbt})}{LGD_{lt}}. \tag{13}$$

Figure 22 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 40% and 60% of the reported PDs – depending on the measure used – are close to 0%, the estimated median PD is 18%, which corresponds to the 90th percentile in the distribution



(a) PDs estimated by the Bank of Portugal.



(b) PDs reported by the commercial banks.

Figure 22: **PDs: actual vs. required to match interest rate distribution.** Probability density function of the required PDs to match the interest rate distribution following equation (13) in blue, and of the actual LGD distribution in red.

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.

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:

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

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

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

E.2 Spreads

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

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

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

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

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