Online Appendix

Beyond Risk: Firm Financing and Interest Rates

Contents

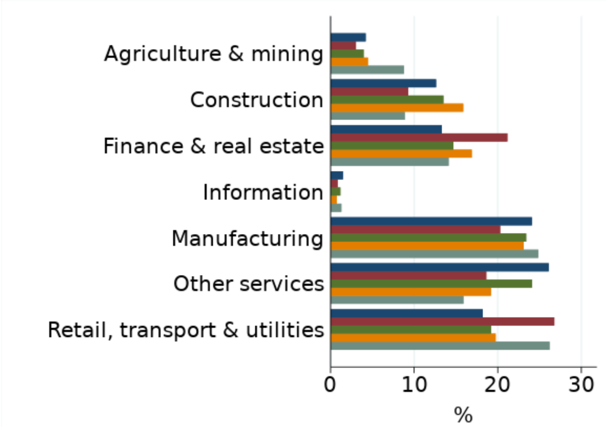
A Additional data details

A.1 Sample construction and composition

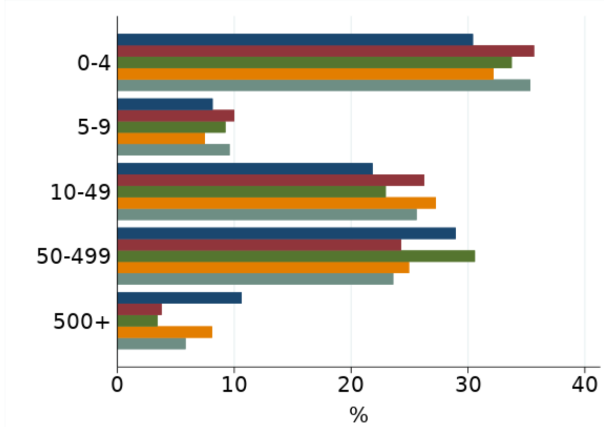
This section provides additional details about the distribution of loans issued by each of the five major banks, the comparison of our sample to the universe of Portuguese firms, how we compute the share of aggregate moments covered by our sample, the share of credit from banks covered by our sample,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

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

Table A.1 compares the size distribution of our sample of firms in 2019 with the size



(a) Industry distribution of credit, by bank



(b) Firm size distribution of credit, by bank

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

distribution of the universe of Portuguese firms in the same year.[1] Relative to Portugal, the sample substantially underrepresents zero-employee firms and overrepresents larger firms, indicating an upward size bias in employment relative to the population. To compute the shares of employer firms, total employment and aggregate output that are attributable to firms in our sample (presented in the main text), we use aggregate values from Eurostat.

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

Finally, we compare the firm size distribution of our sample with that of employer firms in the U.S. in 2019. The U.S. data comes from the Business Dynamics Statistics of the Census Bureau. Table A.2 reports the distribution of firms across employment size categories, as well as the corresponding employment shares. In both countries, the firm size distribution is skewed toward small firms, but more so in Portugal. Portugal

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

Size indicator	Number of firms		Employment Share	
	Sample	Portugal	Sample	Portugal
0 employees	12.3%	69.8%	-	-
1 - 9 employees	68.8%	26.2%	20.2%	44.3%
10 - 49 employees	16.0%	3.4%	27.6%	19.3%
50 - 249 employees	2.6%	0.5%	22.3%	15.2%
>250 employees	0.4%	0.1%	30.0%	21.2%

Table A.1: **Firm size distributions in our sample and in Portugal.** Distribution of firms and employment shares by firm size category in 2019. Our sample correspond to all firms with firm financing loans in 2019; Data for Portugal are from the Statistics Portugal, Integrated business accounts system.

Size indicator	Number of firms		Employment Share	
	Sample	USA	Sample	USA
1 - 4 employees	57.5%	58.6%	9.5%	5.2%
5 - 9 employees	20.9%	18.0%	10.7%	5.1%
10 - 19 employees	11.5%	11.4%	11.9%	6.5%
20 - 99 employees	8.6%	10.0%	25.9%	16.3%
100 - 499 employees	1.3%	1.7%	19.6%	14.0%
>500 employees	0.2%	0.4%	22.4%	52.9%

Table A.2: **Firm size distributions in our sample and in the U.S.** Distribution of firms and employment shares by firm size category. Data for Portugal correspond to 2019 data for firms in our sample with active firm financing loans in 2019; U.S. data are from Business Dynamics Statistics for 2019.

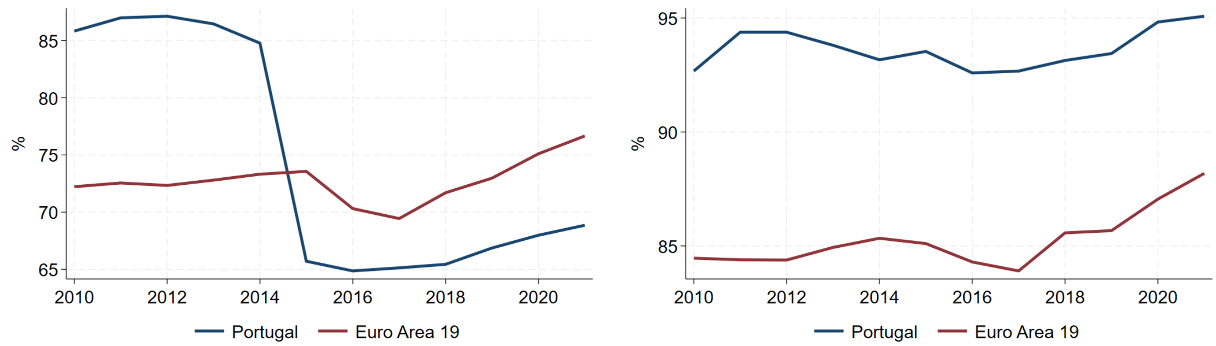
has more weight on all size categories below 500 employees, and less weight above this threshold. Overall, the average firm size is 11.3 employees in our sample, compared to 24.7 employees for U.S. employer firms.

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

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

For banking market shares, we compare the share of total banking sector assets held by the three and five largest banks between 2010 and 2021 in Portugal, with the average for the Euro Area 19 countries. The data are from the World Bank's Global Financial Development Indicators.[2] In both Portugal and the Euro Area, large banks have a similar

[2] The respective series are: assets of the three largest commercial banks as a share of total commercial banking assets ("bank concentration," code GFDD.OI.01); and assets of the five largest commercial banks



(a) Three largest banks

(b) Five largest banks

Figure A.2: **Share of banking sector assets held by largest banks.** Panels (a) and (b) show the share of total banking sector assets held by the three and five largest banks, respectively, between 2010 and 2021. The blue line corresponds to Portugal, and the red line is the average across the 19 Euro Area countries. Source: World Bank, Global Financial Development Indicators.

presence. They account for 84–88% of assets in Euro Area countries, and 92–95% in Portugal. The shares for the largest three banks are also similar. The sharp decline in this share in 2015 in Portugal was due the resolution of Banco Espírito Santo, one of the country's largest banks at the time.

The second moment to explain is the share of external firm financing accounted for by credit from banks in Portugal and the Euro Area 19 countries. Quarterly data covering 2010:Q1 to 2022:Q4 are drawn from the European Central Bank.[3] The measure of total external financing is the sum of credit from banks and debt securities. Figure A.3 shows that firms in both Portugal and the Euro Area rely heavily on bank credit. On average, it accounts for 86.2% of total external financing for Portuguese firms and 88.6% for firms across the Euro Area.

A.3 Interest rates vs. spreads

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

as a share of total commercial banking assets ("five-bank asset concentration," code GFDD.OI.06). Total assets include total earning assets, cash and due from banks, foreclosed real estate, fixed assets, goodwill, other intangibles, current and deferred tax assets, discontinued operations, and other assets.

[3]The series used are: Loans to non-financial corporations (codes QSA.Q.N.PT.W0.S11.S1.N.L.LE.F4.T._Z.XDC._T.S.V.N._T for Portugal and QSA.Q.N.I8.W0.S11.S1.N.L.LE.F4.T._Z.XDC._T.S.V.N._T for the Euro Area); and debt securities and loans of non-financial corporations (codes QSA.Q.N.PT.W0.S11.S1.N.L.LE.F3T4.T._Z.XDC._T.S.V.N._T for Portugal and QSA.Q.N.I8.W0.S11.S1.N.L.LE.F3T4.T._Z.XDC._T.S.V.N._T for the Euro Area).

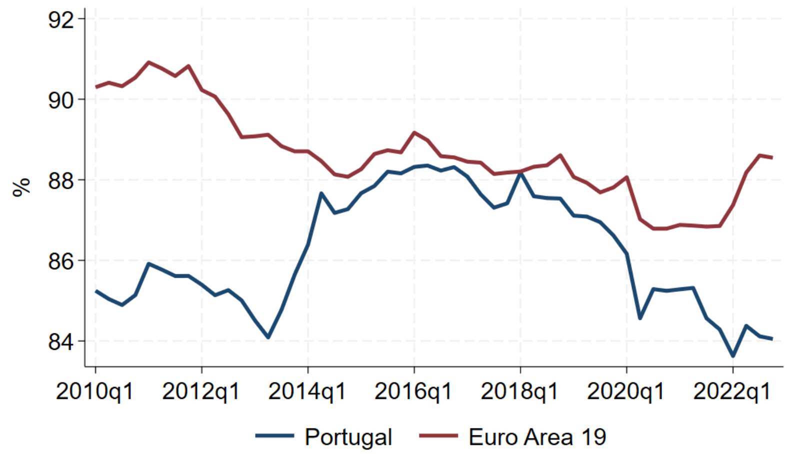


Figure A.3: **Share of bank loans in total firm credit.** Share of loans in total loans and debt securities issued by non-financial corporations, 2010–2022. The blue line represents Portugal, and the red line is the average across the 19 Euro Area countries. Source: European Central Bank.

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

$$\tilde{r}_{lbt} = r_{lbt} - s_{lbt},$$

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

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

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

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

Figure A.4 shows that the residual component of banks' financing costs is tightly concentrated around zero. Eighty percent of loans have a residual of approximately zero, and 90% of loans have a residual of less than or equal to 1 basis point. Thus, cross-bank heterogeneity in funding costs—after controlling for maturity and time fixed effects—is

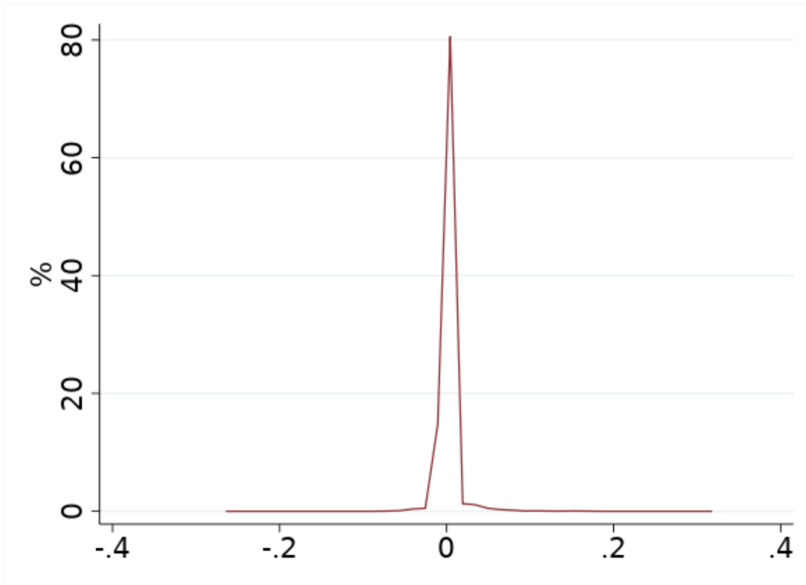


Figure A.4: **Variation in bank financing costs.** This figure presents the distribution of risk free interest rates after controlling for maturity-month-year fixed effects.

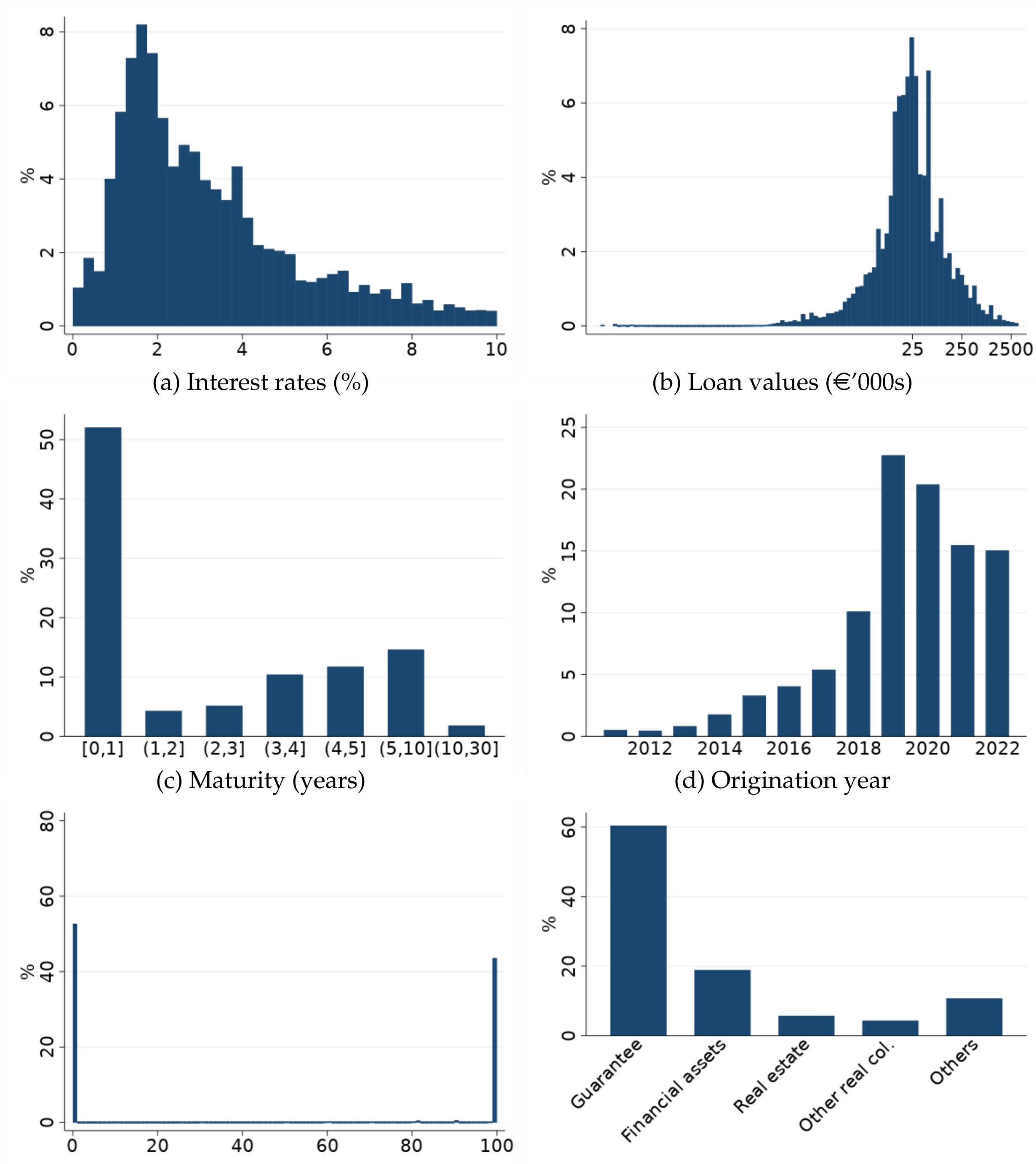
quantitatively small, and treating interest rates, rather than spreads, as the relevant object of analysis is unlikely to confound firm-level interest rate dispersion with variation in banks' own cost of funds.

A.4 Descriptive statistics for sample of all instruments

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

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

Figure A.5 replicates Figure 2 from the main text, but now includes all types of credit given by banks to firms. The interest rate and collateral rate distribution hardly change (panels a and d). These other types of loans are for lower values on average (panel b) and have a shorter maturity (panel c). This is not surprising as credit lines and credit cards are mostly short term loans with smaller amounts.



(a) Interest rates (%)

(b) Loan values (€'000s)

(c) Maturity (years)

(d) Origination year

(e) Collateral rate (%)

(f) Collateral type

Figure A.5: **Distributions of loan characteristics at origination: all credit instruments**
Panels (a), (b), (c), (d), (e) and (f) are the distributions of interest rates, loan values (in thousands of Euros), maturity (in years), origination years, collateral rates and collateral types, respectively. The scale for loan values in panel (b) is in log units. The distribution in panel (f) is conditional on loans having some collateral.

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

Table A.3: **Descriptive statistics for firms: all credit instruments.** This table provides statistics for firms in the sample. For each moment, four percentiles and the standard deviation are given. All moments are at the firm level and are averaged within firm over the years that each firm is in the sample. So, for example, the value of loans is the total value of a firm's loans, % collateralized is the share of a firm's loans that are collateralized, and # of banks is the number of banks that a firm has loans from. Value of loans, total assets and revenue are in thousands of 2016 Euros.

To illustrate the characteristics of the firms in the sample when considering all credit instruments we present summary statistics from their loan portfolios, balance sheets and income statements in Table A.3.[4] The statistics are within firm averages over the years that each firm is in the sample. The balance sheet and income statement variables reflect the characteristics of the population of firms. For the credit registry variables, the median firm has €63,400 in loans. The median number of loans is 6 and the median share of a firm's loan value that is collateralized is 50% The majority of firms have loans from only one bank, but many firms have more bank relationships. At the 90th percentile a firm averages 3 banking relationships over the years that it is in the sample. Overall, firms in this sample are slightly smaller and younger than firms used in the baseline sample, which considers only firm financing loans.

[4]Notice that there are firms that have no firm financing loans which have other credit instruments. This results in the sample of firms across the two the samples to differ.

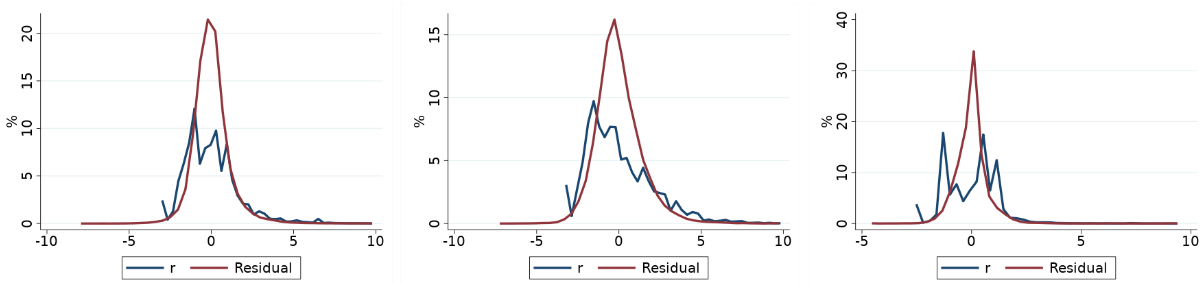
B Additional details for Section 3.1

B.1 Controls for loan characteristics

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

- *Loan value.* Log of the value of the loan in Euros.
- *Frequency of rate adjustment.* Loans can be fixed rate (52.4% of loans) or can be adjusted at a daily (1.6%), monthly (13.4%), quarterly (6.8%), biannual (7.6%) or annual (8.8%) frequencies, at other frequencies less than a year (0.03%), at frequencies greater than one year (0.02%), or at the lender's discretion (0.1%). There is also a category for all residual cases (9.2%).
- *Type of collateral.* Collateral are classified as guarantees (29.4%), financial assets (7.5%), real estate (2.7%), other collateral of real nature (2.4%) and a category for all residual cases (5.8%). Fifty-two per cent of loans have no collateral.
- *Number of collateral assets.* Among collateralized loans, 75.7% have one asset pledged, 14.6% have two, 4.8% have three, 2.2% have four and 2.8% have five or more.[5]
- *Amortization type.* The amortization method can be French (fixed installments with interest paid at the end of each period—23.9% of loans), German (fixed installments with interest paid in advance of each period—0.2%), fixed principal payments (6.6%), all principal paid at maturity (25.5%), increasing installments (0.01%), and a category for all residual cases (43.4%).
- *Purpose of the loan.* Loans can be for car purchases (14.5% of loans), liquidity management (13.8%), exports and imports (4.2%), equipment and other goods purchases (1.8%), real estate and land (2.6%) or no specific purpose (63.2%).
- *Frequency of repayment.* Loans can be repaid with monthly (42.5% of loans), quarterly (5.0%), biannual (0.4%) or annual (0.5%) installments, with a single principal payment (3.9%), with a single principal and interest payment (16.3%), and there is a category for all residual cases (31.3%).

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



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

Figure B.6: **Interest rate distributions before and after controlling for loan characteristics, by collateral type.** Panel (a), (b) and (c) present the distribution of interest rates for fully collateralized loans backed with real and financial assets, guarantees and other types of collateral, respectively. In both panels the blue line is the unconditional distribution with mean centered around zero, and the red line is the distribution after controlling for loan characteristics (i.e. the residuals from equation 1). All distributions are centered with their means at zero.

- *Right to immediate repayment.* For 16.2% of loans, the bank has the right to demand immediate repayment of the loan.
- *Securitized loan.* 97.7% of loans are securitized by banks.
- *Loan special characteristics.* The controls include dummy variables for whether loans were granted under Covid-related policies including guarantees (0.6%), are covered by consumer credit law (1.4%), or are a syndicated loans, related to real estate, or are involved in a default process (0.3%). The last three categories are grouped together as they include small numbers of loans. Loans can have more than one of these characteristics, but 98.2% of loans have none of them.

B.2 Analysis by collateral type

It is possible that some types of collateral could be more effective at insuring banks against risk, and that some of the residual interest rate variation is due to this. To assess this we estimate specification (1) separately for different classes of collateral: real and financial assets, guarantees, and other forms of collateral. Figure B.6 plots the distribution of interest rates and of the residuals from equation (1) for each class of collateral. The results show that distinguishing between different types of collateral is not materially important, with more than 41% of the variation in rates remaining unexplained in all cases.

C Additional details for Section 3.2

C.1 Additional details for Equation (3)

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

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

where $\mathcal{S}_{l,t+1}$ is the set of possible states in the next period, and $p(s_{l,t+1})$ is the probability of a state $s_{l,t+1} \in \mathcal{S}_{l,t+1}$.[6] For the case of a risk free lender, which is used as the baseline, this can be expressed as equation (3), with

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

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

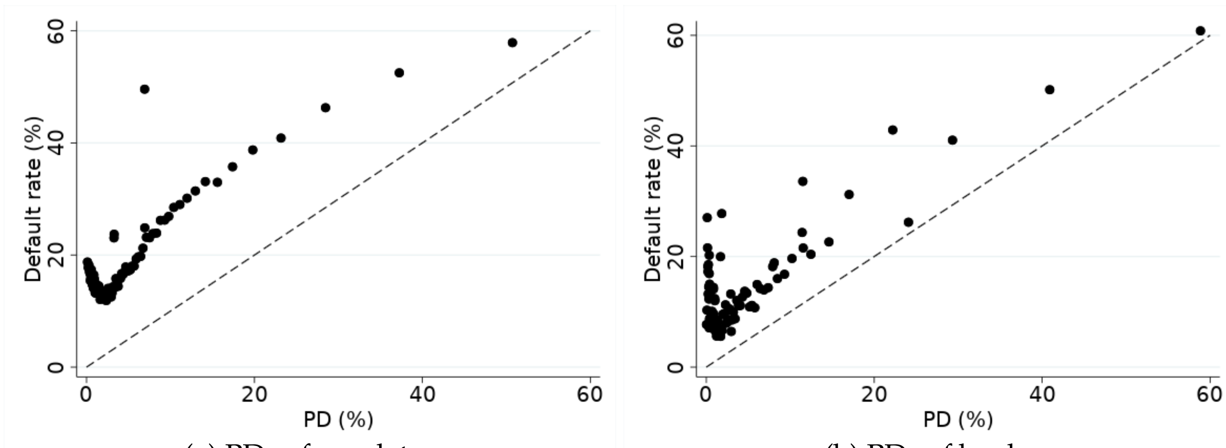
C.2 Validation of PDs

An important detail for the analysis is the quality of the reported PDs. We want to be confident that PDs are measured well in the data. In order to do this we use the fact that we observe loans through to completion, so that we can observe whether default occurs ex post. To evaluate the quality of the PD estimates we take their values from the data, rank them from lowest to highest, group them into 100 evenly sized bins based on this ranking, and then compare the values of the average default probabilities and ex post realized rates of default within each bin. The results are presented in Figure C.7 and show that the variables are high quality. Realized default rates are monotonically increasing in the ex ante estimates and quantitatively they closely track each other.

C.3 All credit instruments

To increase the number of observations we extend the analysis to include all credit instruments. This increases the number of observations in the baseline specification from

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



(a) PDs of regulator

(b) PDs of banks

Figure C.7: **Estimated vs. realized PDs.** In panel (a) the PDs from the regulator are ranked and sorted into 100 equally sized bins, with each bin assigned the average value of its elements. The vertical axis provides the share of loans in each bin that default. Panel (b) repeats the exercise for PDs reported by the banks. The dashed lines are 45° lines along which PDs equal ex post default rates.

25,460 to 108,847. Figure C.8 shows that results are unchanged by extending the sample to include all credit instruments. In this case, risk explains 0.7% of the total variation.

C.4 Bank PDs instead of regulator PDs

The regulator-provided PDs are estimated using balance sheet and income statement data and may not fully capture the information available to banks or their subjective assessments of a firm's probability of default. To ensure that our results are not driven by the use of regulator PDs, we reestimate equation (4) using the PDs reported by banks themselves. The sample size for this analysis is 31,458 loans.[7] Figure C.9 shows that relying on banks' own PDs does not increase the explanatory power of risk, confirming the robustness of our results.

C.5 Test for whether interest rate residuals contain information about default

To assess the possibility that banks have better information about default probabilities than they report in the data, we test whether the residuals from our regression in Sec-

[7]The baseline sample for this analysis is 89,896. This exercise has a smaller sample because banks do not always report a PD and LGD, and this analysis requires both.

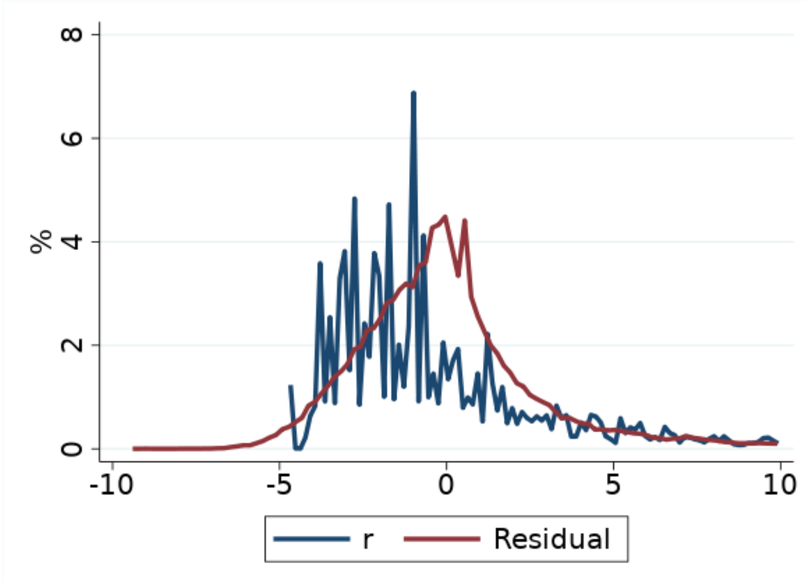


Figure C.8: **Interest rates and risk: all credit instruments.** The raw interest rate distribution (blue) and after controlling for loan characteristics and risk (red). Both distributions are centered with their means at 0.

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

Table C.4: **Explanatory power of interest rate residuals for default.** Standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

tion 3.2 have predictive power for defaults. Specifically we use the specification with a flexible semi-parametric relationship between *PD*, *LGD* and the interest rate (described in the main text), so that the regression extracts as much information as possible from these variables for interest rates. We take the residuals from the regression, and regress an indicator for whether a firm defaults on a loan on them. If banks have additional information about default that is not contained in their reported PDs, then these residuals should have explanatory power in this regression. The results are reported in Table C.4. While the coefficient of interest is positive and significant, indicating that banks are pricing some additional risk that is not captured by PDs and LGDs, quantitatively the effect is small. The regression has an R^2 of only 0.001.

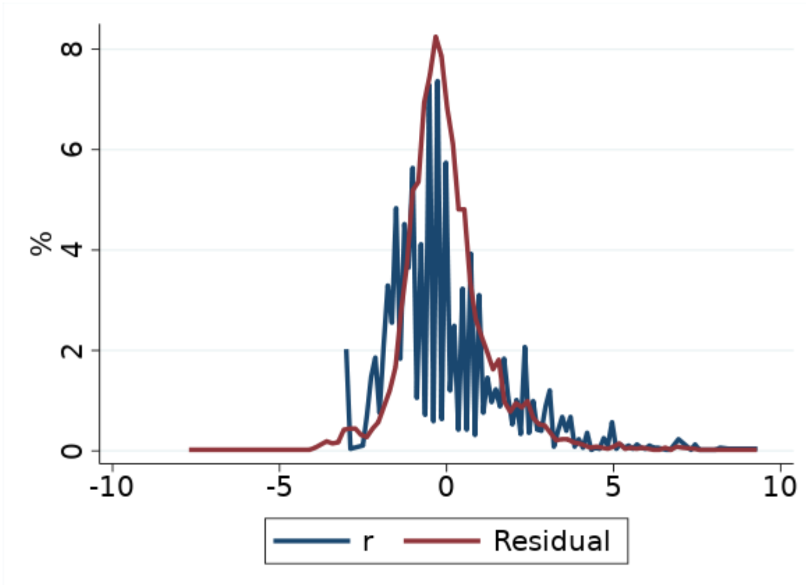


Figure C.9: **Interest rates and risk: bank PDs.** The raw interest rate distribution (blue) and after controlling for loan characteristics and risk (red). Both distributions are centered with their means at 0.

C.6 PD and LGD distributions that would be needed to explain rates

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

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

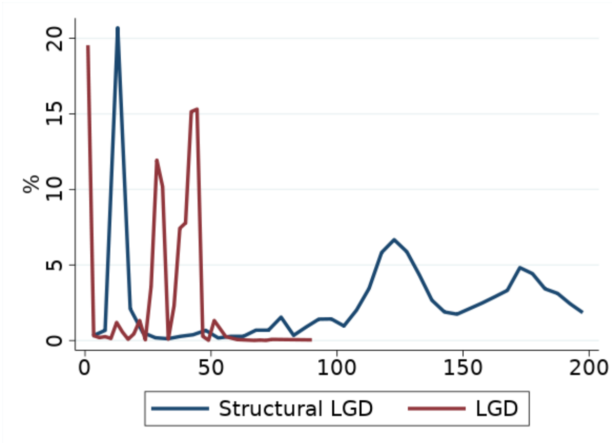
$$-\log(1 - PD_{lt}LGD_{lt}) = \epsilon_{lt}, \tag{1}$$

to determine what values of PD_{lt} and LGD_{lt} would be needed to fit the data.[8]

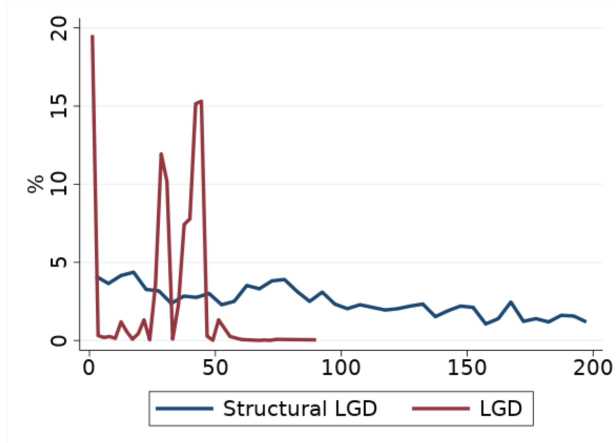
LGD We start by calculating the LGD needed to match the interest rate distribution, given the PD distribution. The LGD is given by

$$LGD_{lt} = \frac{1 - \exp(\epsilon_{lt})}{PD_{lt}}. \tag{2}$$

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



(a) PDs estimated by the Bank of Portugal.



(b) PDs reported by the commercial banks.

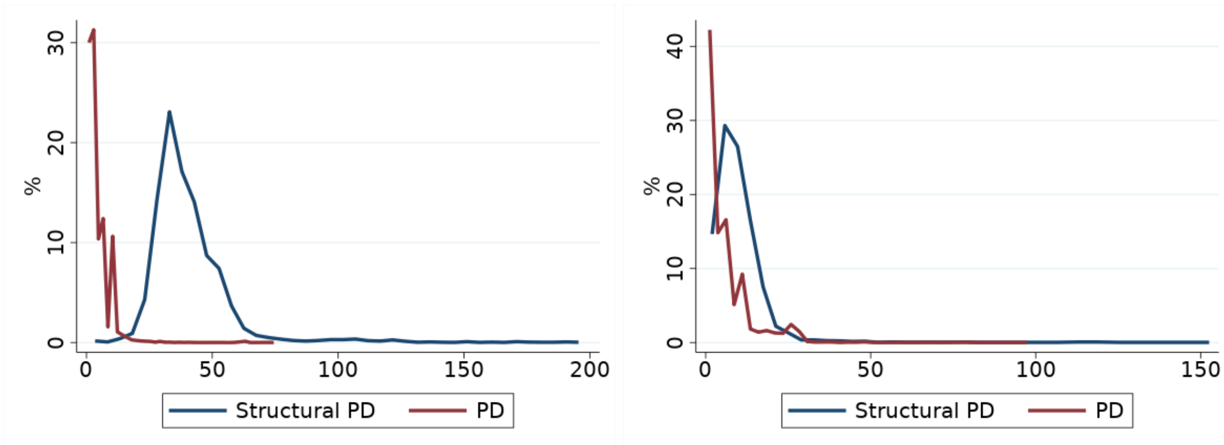
Figure C.10: **LGDs: actual vs. required to match interest rate distribution.** Probability density function of the required LGDs to match the interest rate distribution following equation (2) in blue, and of the actual LGD distribution in red.

Figure C.10 plots on the left hand panel the probability density function (PDF) of estimated LGD 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 losses 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 C.10 to 200, but the median is 545 and 129 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 an interest rate distribution in line with the observed one, given the LGD distribution. The PD is given by

$$PD_{lt} = \frac{1 - \exp(\epsilon_{lt})}{LGD_{lt}}. \tag{3}$$

Figure C.11 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



(a) PDs estimated by the Bank of Portugal.

(b) PDs reported by the commercial banks.

Figure C.11: **PDs: actual vs. required to match interest rate distribution.** Probability density function of the required PDs to match the interest rate distribution following equation (3) in blue, and of the actual PD distribution in red.

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 commercial banks. It is possible to observe in the Figure that the PDs required to generate an interest rate distribution in line with the empirically observed one, there would need to be much more default risk in the economy. While between 30% and 40% of the reported PDs – depending on the measure used – are close to 0%, the estimated median PD is 40%, which is higher than the 90th percentile in the distribution reported by the Bank of Portugal. Additionally, the estimated PDs have an average above 100%. So, similar to the conclusions when estimating the required LGDs, the required PDs to match the observed interest rate distribution are infeasibly high.

C.7 Alternative stochastic discount factors

The main specification assumes that lenders are risk neutral and price the loans according to the risk free rate. In principle, lenders could discount the future at a different rate, or have a stochastic discount factor that varies across states so that if two loans have payments with different correlations with the stochastic discount factor, then they will be priced differently. To account for this we want to control for the correlation between the likelihood that a firm defaults in a given state and the valuation of the lender in the state when default occurs. For a variable with information about potential firm

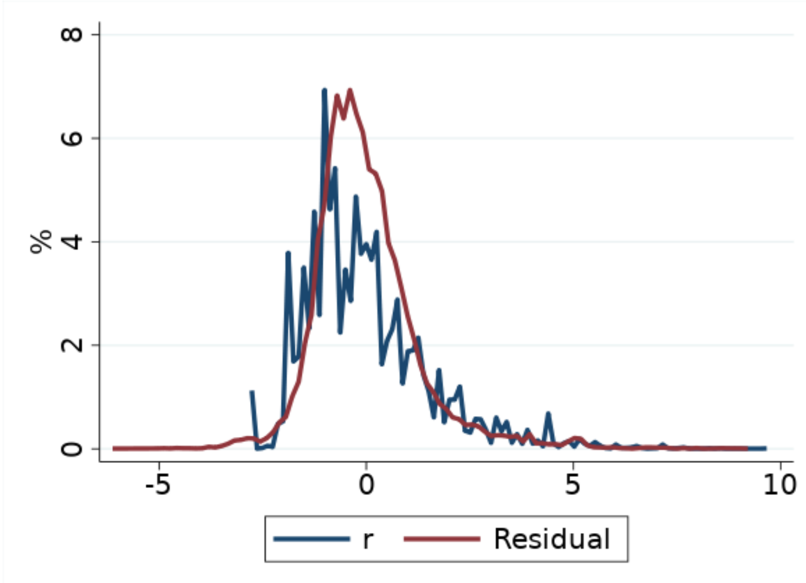


Figure C.12: **Interest rates and risk: alternative SDF.** The raw interest rate distribution (blue) and after controlling for loan characteristics, risk and the correlation coefficient between state of the firm and aggregate state (red). Both distributions are centered with their means at 0.

default we choose firm revenues. For the value of states by lender, we take guidance from consumption-based asset pricing theories and use the value of aggregate consumption. We then take the correlation between these two variables and add it to equation (4) interacted with the risk term. Therefore, we estimate the following specification

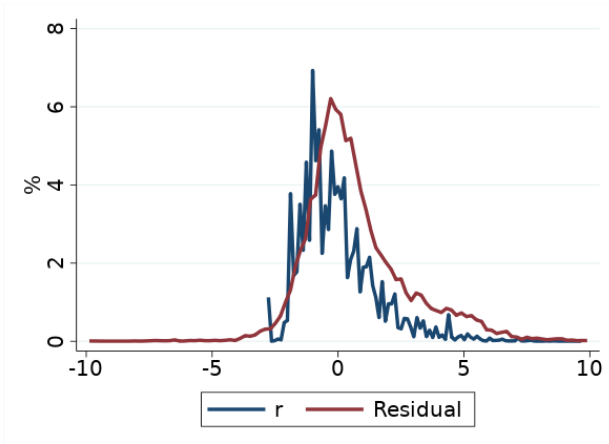
$$r_{lt} = -\log(1 - PD_{lt}LGD_{lt}) + \beta_1 \Lambda_f \log(1 - PD_{lt}LGD_{lt}) + \beta_2 \mathbf{L}_l + \mathcal{M}_{lt} + \varepsilon_{lt}, \tag{4}$$

where Λ_f is the correlation between firm f's revenues and aggregate consumption during the period the firm is in the sample. Results in Figure C.12 show that the addition of the correlation coefficient Λ_f does not increase significantly the explanatory power of the risk component.

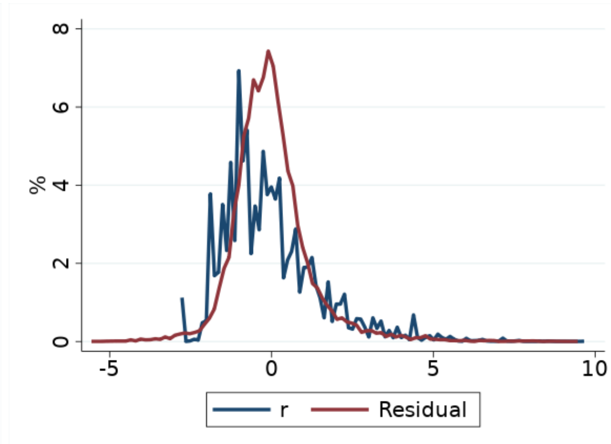
C.8 More flexible specifications for risk

C.8.1 Imposing risk coefficient to be minus one

In regression (4) we allow the coefficient on the risk term to be freely estimated, while the theory suggests the coefficient should be -1. While this is the dominant approach in macroeconomic models, it is possible that this is misspecified and the true relationship



(a) Imposing coefficient to -1



(b) Semi-parametric

Figure C.13: **Interest rates and risk: alternative specifications.** The raw interest rate distribution (blue) and after controlling for loan characteristics and risk (red). Both distributions are centered with their means at 0. Panel (a) when estimating the coefficient on the risk term. Panel (b) for the semi-parametric estimation.

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

$$r_{lt} = -\log(1 - PD_{lt}LGD_{lt}) + \beta_2 \mathbf{L}_l + \mathcal{M}_{lt} + \varepsilon_{lt}. \tag{5}$$

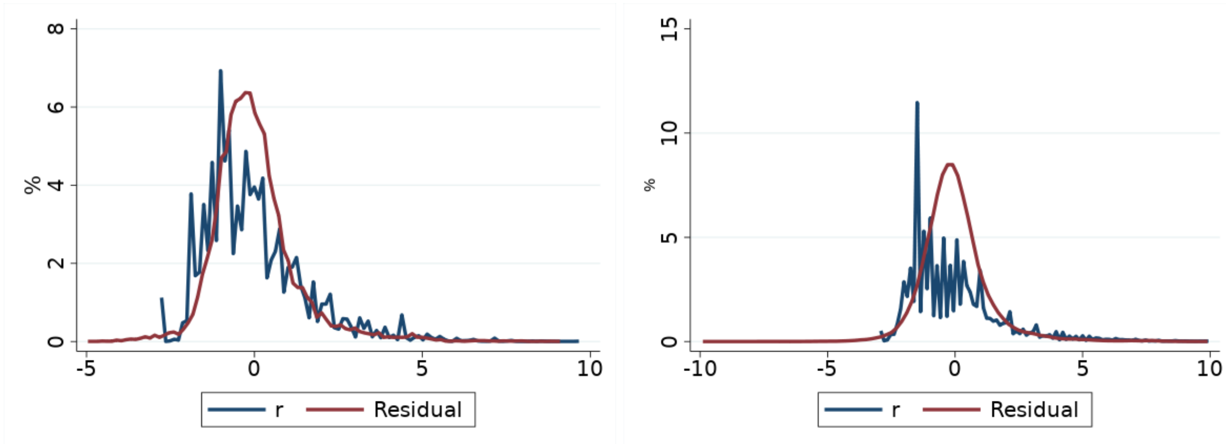
Results in Figure C.13(a) illustrate that imposing the coefficient on the risk term to be -1 does not improve the explanatory power of risk.

C.8.2 Semi-parametric estimation

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

$$\begin{aligned} r_{lt} = & \sum_{p \in \mathcal{PD}} \gamma_p PD_{lt} \mathbf{1}_{i \in \mathcal{S}^{(P)}} + \sum_{d \in \mathcal{LGD}} \theta_d LGD_{lt} \mathbf{1}_{i \in \mathcal{S}^{(D)}} + \sum_{d \in \mathcal{LGD}} \sum_{p \in \mathcal{PD}} \lambda_{dp} PD_{lt} LGD_{lt} \mathbf{1}_{i \in \mathcal{S}^{(P)}} \mathbf{1}_{i \in \mathcal{S}^{(D)}} \\ & + \sum_{p \in \mathcal{PD}} \sigma_p PD_{lt}^2 \mathbf{1}_{i \in \mathcal{S}^{(P)}} + \sum_{d \in \mathcal{LGD}} \alpha_d LGD_{lt}^2 \mathbf{1}_{i \in \mathcal{S}^{(D)}} + \sum_{d \in \mathcal{LGD}} \sum_{p \in \mathcal{PD}} \mu_{dp} PD_{lt}^2 LGD_{lt}^2 \mathbf{1}_{i \in \mathcal{S}^{(P)}} \mathbf{1}_{i \in \mathcal{S}^{(D)}} \\ & + \beta_1 \mathbf{L}_l + \mathcal{M}_{lt} + \varepsilon_{lt}. \end{aligned} \tag{6}$$

The terms $\mathcal{S}^{(P)}$ and $\mathcal{S}^{(D)}$ represent the p^{th} PD and d^{th} LGD group, respectively. We partition the sample into ten deciles for each of the PD and LGD distribution to capture po-



(a) LGD sample (b) Full sample

Figure C.14: **Interest rates and risk: PDs only.** The raw interest rate distribution (blue) and after controlling for loan characteristics and risk (red). Both distributions are centered with their means at 0. Panel (a) when estimating the semi-parametric on the sample for which we have LGD values. Panel (b) for the semi-parametric estimation on the full sample.

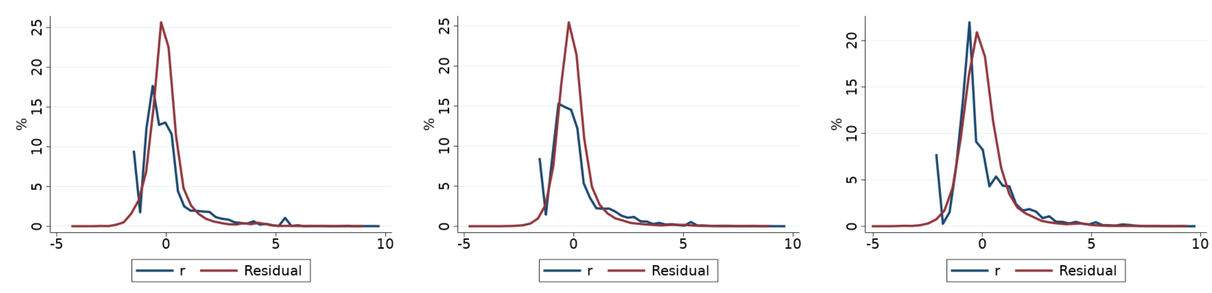
tential nonlinear relationships between the risk terms and interest rates. We additionally include the squared terms of both PD and LGD variables to further account for potential nonlinear relations. Results in Figure C.13(b) illustrate that even with this more flexible specification, risk still doesn't explain much of the interest rate variation.

C.9 Validation that selective LGD reporting does not drive results

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

$$r_{lt} = \sum_{p \in \mathcal{PD}} \gamma_p PD_{lt} \mathbf{1}_{i \in \mathcal{S}^{(P)}} + \sum_{p \in \mathcal{PD}} \sigma_p PD_{lt}^2 \mathbf{1}_{i \in \mathcal{S}^{(P)}} + \beta_1 \mathbf{L}_l + \mathcal{M}_{lt} + \varepsilon_{lt}. \tag{7}$$

First, to guarantee our results in the full sample are not driven by the omission of the LGD variable, we estimate the above specification in the subsample for which we have LGD values. Figure C.14(a) illustrates that results are similar to when also including LGD. Then, we extend the analysis to the entire sample. Figure C.14(b) illustrates that results are robust to including the full sample in the risk analysis.



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

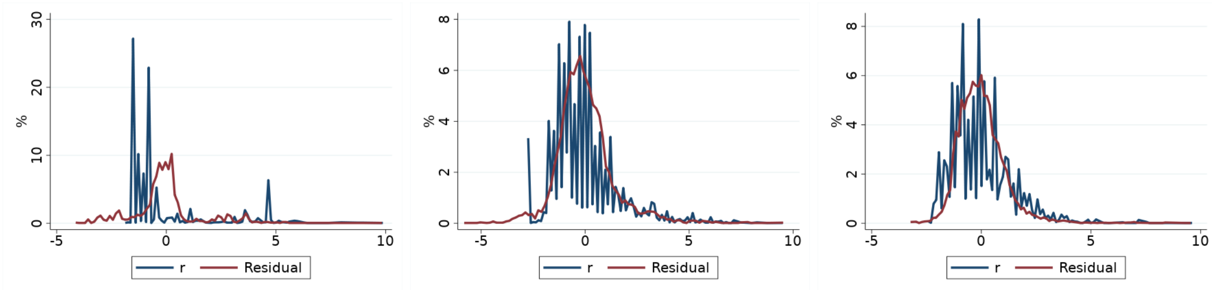
Figure C.15: **Interest rates and risk: low PD loans only.** The raw interest rate distribution (blue) and after controlling for loan characteristics (red). Both distributions are centered with their means at 0. Panel (a) PDs in the bottom 5^{th} percentile. Panel (b) PDs in the bottom 10^{th} percentile. Panel (c) PDs in the bottom 25^{th} percentile

C.10 Low risk loans only

To circumvent concerns about how risk is specified in the regression, as well reducing concerns about measurement that we discussed earlier, we restrict the sample to loans with a low PD. To implement this we restrict the sample to loans with a PD below 0.29%, 0.55% and 1.14%, respectively the 5th, 10th and 25th percentile of the PD distribution. Then we estimate equation (1) for the three different groups. Results for the three groups in Figures C.15(a), C.15(b) and C.15(c) show than even for low-risk loans, after controlling for loan characteristics, the majority of the variance in interest rates remains.

C.11 Amortization type

The theory assumes that both the interest and the principal are only repaid when the loan matures. In the data, for 28.6% of loans, the principal is repaid in a unique repayment. To guarantee as much consistency with the theory as possible, we reestimate (4) for loans with a unique repayment of principal. Additionally, to guarantee the results are not driven by a specific amortization type, we reestimate (4) restricting to the other two largest amortization types, French (fixed install- ments, with interest and principal changing over time - 20.9% of loans) and category for all residual cases (45%). Figure C.16 illustrates that results are robust across the different amortization types.



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

Figure C.16: **Interest rates and risk: by amortization type.** The raw interest rate distribution (blue) and after controlling for loan characteristics and risk (red). Both distributions are centered with their means at 0. Panel (a) for loans with a unique payment of capital. Panel (b) for loans with a French type amortization. Panel (c) for loans with other amortization types.

C.12 Value of exposure

We directly assess whether payments on loans prior to maturity affect the results by accounting for a bank's expected exposure at the time of default.[9] To do so, we scale the LGD by expected exposure at default, denoted by VE_{lt}, which is measured as a share of the initial loan value:

$$r_{lt} = \beta_1 \log(1 - PD_{lt}LGD_{lt}VE_{lt}) + \beta_2 \mathbf{L}_l + \mathcal{M}_{lt} + \varepsilon_{lt}. \tag{8}$$

This adjustment accounts for the fact that repayments before default reduce a bank's potential losses. In principle, heterogeneity in exposure across loans could generate additional heterogeneity in rates. Figure C.17 illustrates that results remain robust to including VE_{lt}. If anything, the contribution of risk is smaller than in the baseline specification, as the estimated coefficient on the risk term is close to zero and statistically insignificant.

C.13 Exclusion of Covid period

To guarantee results are not driven by the Covid period and potential guarantees offered by the government offered during this period that would eliminate risk for banks, we reestimate equation (4) for loans initiated prior to 2020. Figure C.18 shows that results are robust to the exclusion of the Covid period.

[9]Exposure is the total outstanding amount on a loan (principal and interest) at a given point in time. In this exercise, we measure exposure relative to the initial loan value.

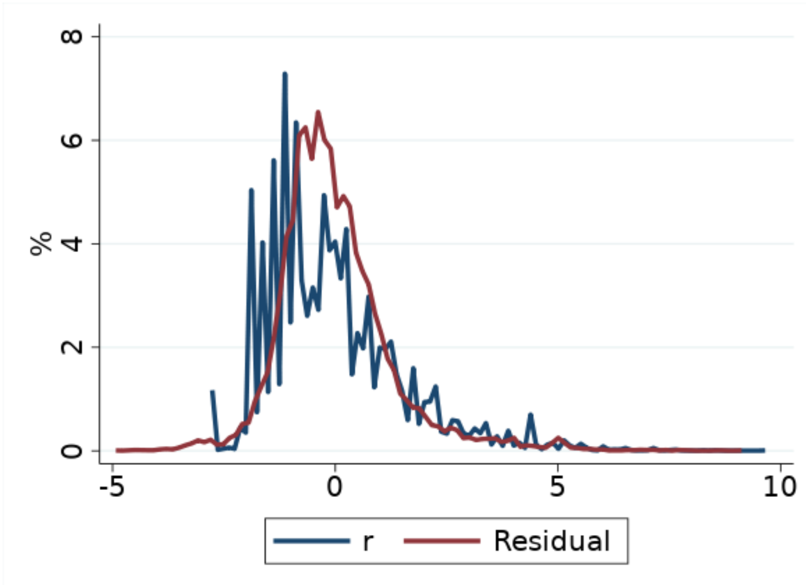


Figure C.17: **Interest rates and risk: controlling for value of exposure.** The raw interest rate distribution (blue) and after controlling for loan characteristics and risk (red). Both distributions are centered with their means at 0.

C.14 Firm groups

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

C.14.1 Age groups

We split firms into three different age groups, following Clymo & Rozsypal (2023). Young firms, below 5 years of age, mature firms, between 5 and 15 years of age, and old firms, older than 15 years. We then reestimate equation (4) for the different groups of firms. Figure C.19 illustrates that results are robust across firms of different ages.

C.14.2 Size groups

We split firms according to number of employees, following Clymo & Rozsypal (2023). Small firms are below 5 employees, small-medium firms between 5 and 12 employees, medium firms between 13 and 40 employees, medium-large firms between 40 and 120 employees and large firms above 120 employees. We then reestimate equation (4) for the different groups of firms. Figure C.20 shows that results are robust across firms of

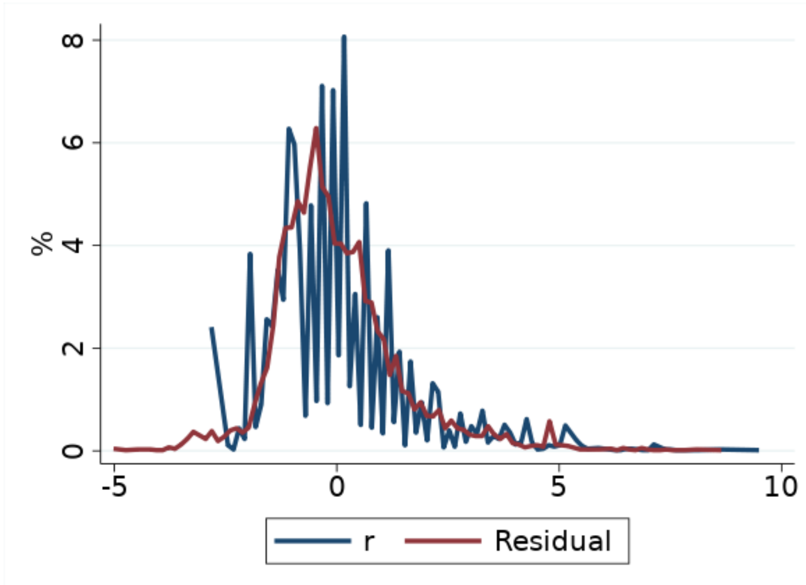
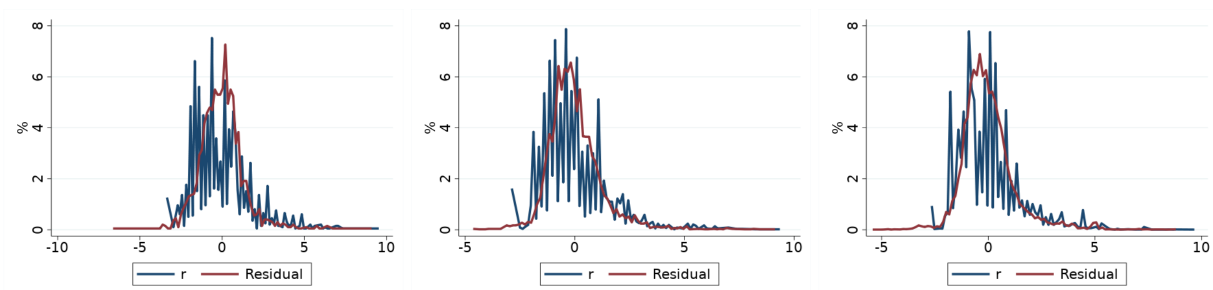


Figure C.18: **Interest rates and risk: loans initiated pre-2020.** The raw interest rate distribution (blue) and after controlling for loan characteristics and risk (red). Both distributions are centered with their means at 0.



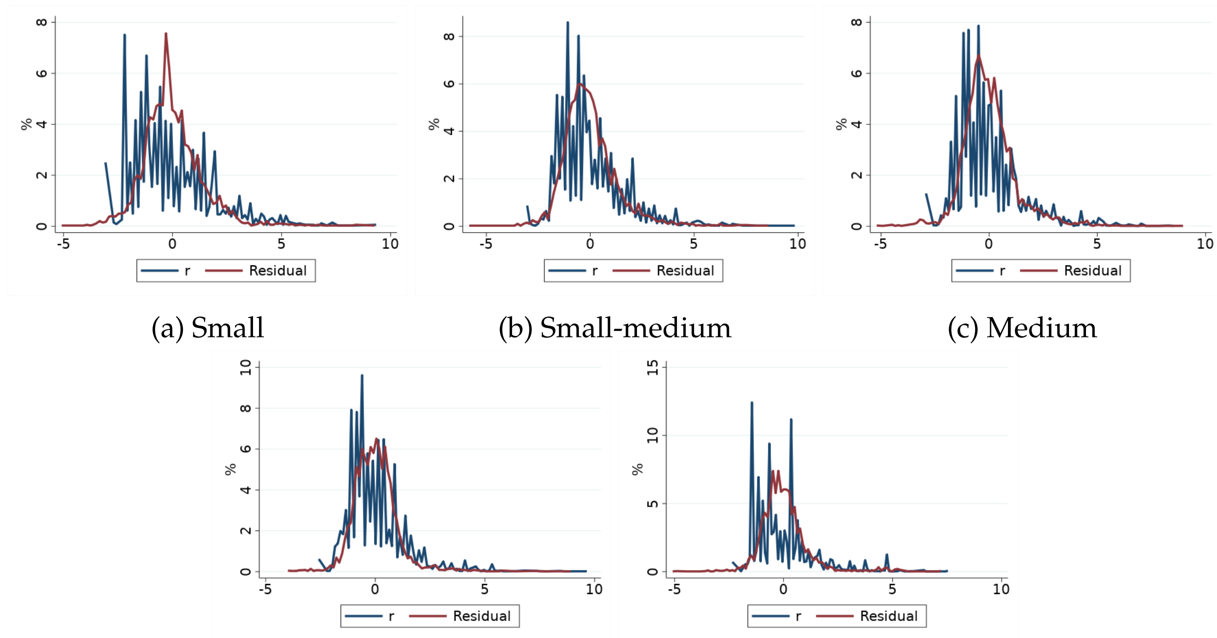
(a) Young (b) Mature (c) Old

Figure C.19: **Interest rates and risk: firm age groups.** The raw interest rate distribution (blue) and after controlling for loan characteristics and risk (red). Both distributions are centered with their means at 0. Panel (a) for young firms. Panel (b) for mature firms. Panel (c) for old firms.

different sizes.

C.15 Extension to longer maturity loans

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



(a) Small (b) Small-medium (c) Medium

(d) Medium-large (e) Large

Figure C.20: **Interest rates and risk: firm size groups.** The raw interest rate distribution (blue) and after controlling for loan characteristics and risk (red). Both distributions are centered with their means at 0. Panel (a) for small firms. Panel (b) for small-medium firms. Panel (c) for medium firms. Panel (d) for medium-large firms. Panel (e) for large firms.

interest rate on the loan satisfies

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

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

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

$$PD_{ltT} \equiv \sum_{s_{lt}^T \in \mathcal{S}_{lt}^T} p(s_{lt}^T) \chi(s_{lt}^T).$$

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

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

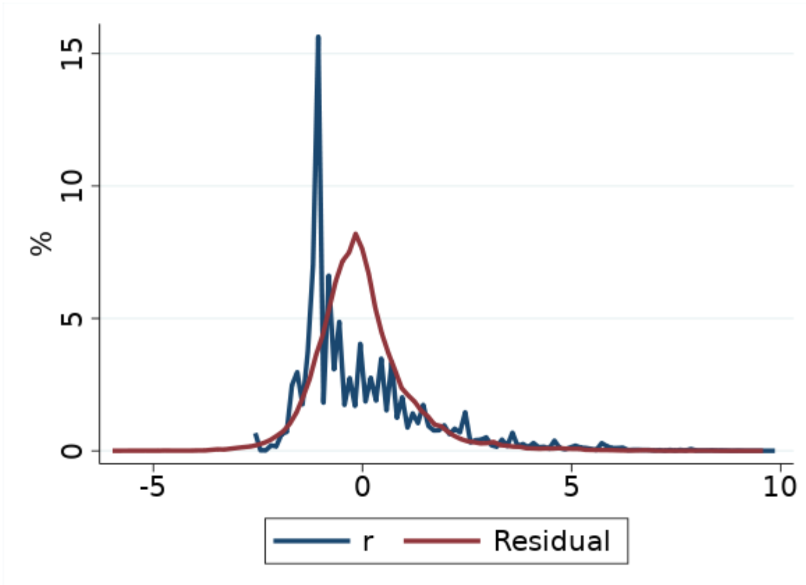


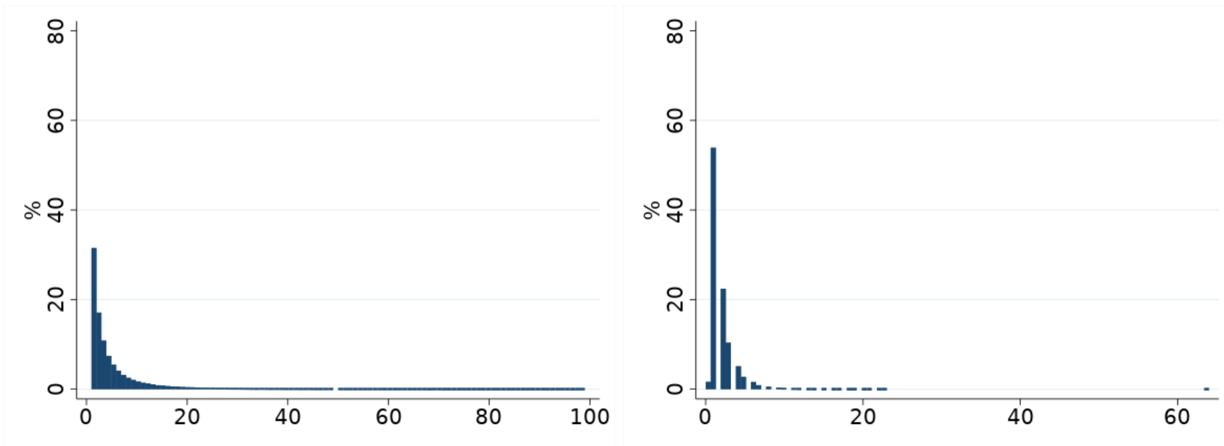
Figure C.21: **Interest rates and risk: all maturities.** The raw interest rate distribution (blue) and after controlling for loan characteristics and risk (red). Both distributions are centered with their means at 0.

Figure C.21 shows the raw distribution of rates and the residual variation, confirming that after controlling for default risk and loan characteristics, substantial variation remains.

D Additional details for Section 4.1

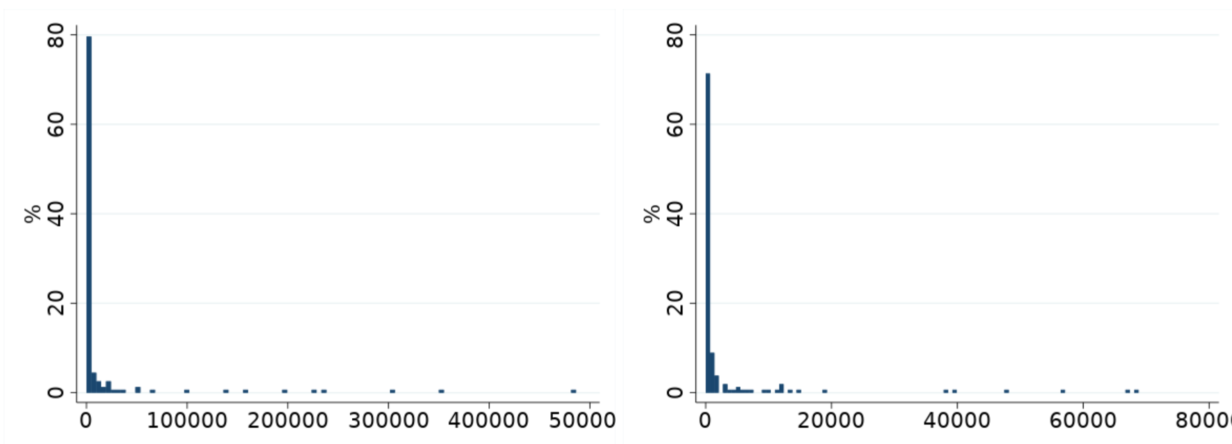
D.1 Distribution of loans across banks and firms

Figure D.22 presents the distribution of number of loans and number of bank relationships by firm over time. The majority of firms have a small number of loans issued by a small number of banks. There is then a right tail, indicating that some firms have many loans originating from multiple banks. Figure D.23 plots the distribution of number of loans and firm relationships by bank over time. In contrast to the corresponding firm-level distributions, the majority of banks issue thousands of loans and have relationships with thousands of firms.



(a) Loans (b) Banks

Figure D.22: **Loan and bank relationship distribution across firms.** On the left hand side the distribution of number of loans by firm over time. On the right hand side the distribution of number of bank relationships by firm over time.



(a) Loans (b) Firms

Figure D.23: **Loan and firm relationship distributions across banks.** On the left hand side the distribution of number of loans by bank over time. On the right hand side the distribution of number of firm relationships by bank over time.

D.2 Bank-time and firm-time fixed effects specification

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

$$r_{lfbt} = \beta_1 \mathbf{R}_{lt} + \beta_2 \mathbf{L}_l + \mathcal{M}_{lt} + \mathcal{B}_{bt} + \mathcal{F}_{ft} + \varepsilon_{lfbt}. \tag{9}$$

To estimate the above specification, we need firms that originate multiple loans in the same time period. We use a one year time period, enough to capture changes in firm and

	Contribution
(1) Default risk	0.0001
(2) Loan characteristics	0.1288
(3) Bank-time fixed effects	0.0947
(4) Firm-time fixed effects	0.6099
(5) Residual	0.1705
(6) Covariances	-0.0039
# Obs.	336,098

Table D.5: **Contributions of risk, loan characteristics, firm-time and bank-time fixed effects to interest rate variation.** The different lines represent the different components included in equation (9). Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (5) in the main text.

time characteristics while preserving close to 50% of the original sample.[10]

Table D.5 presents the results. Overall results are qualitatively in line with the results presented in Section 4.1, with the firm component still having the largest explanatory power. Quantitatively, the main difference comes from the increased explanatory power of the firm-time fixed effect. While the firm component accounted for close to 33% of the overall variation in the baseline regression, the firm-time component accounts for 61% of the variation, absorbing part of the explanatory power previously attributed to the loan-characteristics component and diminishing the variance of the residual.

D.3 Firms with at least 10 loans

To address concerns that the analysis in Section 4.1 might have upwardly biased estimates of the role of firm fixed effects if there are not enough loans per firm in the sample, we re-estimate specification (7), restricting the sample to firms with at least 10 loans. In total, 6,955 satisfy this restriction. This increases the number of loans per firm from 6 to 36. The results are presented in Table D.6 and are very similar to the baseline. The share of variation due to bank fixed effects falls from 9.6% to 8.4%, while the share due to firm fixed effects rises from 32.7% to 41.5%.

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

	Contribution
(1) Default risk	0.0033
(2) Loan characteristics	0.1707
(3) Bank fixed effects	0.0843
(4) Firm fixed effects	0.4153
(5) Residual	0.2833
(6) Covariances	0.0465
# Obs.	251,847

Table D.6: **Contributions of risk, loan characteristics, firm and bank fixed effects to interest rate variation: firms with 10 loans or more.** The different lines represent the different components included in equation (7) in the main text. Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (5) in the main text.

D.4 All instruments

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

$$r_{lfbt} = \beta_1 \mathbf{R}_{lt} + \beta_2 \mathbf{L}_l + \mathcal{M}_{lt} + \mathcal{B}_b + \mathcal{F}_f + \varepsilon_{lfbt}. \tag{10}$$

The instrument type is now additionally included as a fixed effect in the vector $\mathbf{L}_{lt}$. In this case the number of firms in the sample increases from 84,535 to 135,350 and the number of observations goes from 521,856 to 1,598,466. This strengthens the identification of the fixed effects even further, as we now have almost twice as many observations per firm and more within firm variation. Nonetheless, results in Table D.7 are qualitatively and quantitatively very stable. The firm component is still the most important, with its importance almost identical to the baseline results—changing from 32.7% to 37.3%. The largest change came from the increase in importance of the loan characteristics, which can be explained by the inclusion of an instrument-type fixed effect.

D.5 Amortization type

We here repeat the robustness exercise done for risk and reestimate equation (7) for three different amortization types: unique repayment of principal, French amortization and all residual cases. Results in Table D.8 are similar across the different amortization types, with the firm fixed component always being the most important and the bank component

	Contribution
(1) Default risk	0.0021
(2) Loan characteristics	0.3287
(3) Bank fixed effects	0.0967
(4) Firm fixed effects	0.3728
(5) Residual	0.2525
(6) Covariances	-0.0507
# Obs.	1,598,466

Table D.7: **Contributions of risk, loan characteristics, firm and bank fixed effects to interest rate variation: all instruments.** The different lines represent the different components included in equation (10). Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (5) in the main text.

	Contribution		
	Unique repayment	French	Other
(1) Default risk	0.0017	0.0005	0.0022
(2) Loan characteristics	0.1802	0.2447	0.2161
(3) Bank fixed effects	0.0632	0.0932	0.0920
(4) Firm fixed effects	0.5421	0.4001	0.4121
(5) Residual	0.2139	0.1662	0.1418
(6) Covariances	0.0006	0.0958	0.1380
# Obs.	89,488	56,823	163,153

Table D.8: **Contribution of risk, loan characteristics, firm and bank fixed effects to interest rate variation: by amortization type.** The different lines represent the different components included in equation (7) in the main text. The different columns represent the estimates for the three different amortization types: unique repayment of capital, French and residual cases. Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (5) in the main text.

not a major driver of interest rate variation.

D.6 Exclusion of Covid period

To guarantee results are not driven by the Covid period and any potential compression of rates or changes in the banking market that may have been caused by government policies, we reestimate equation (7) for loans initiated prior to 2020. Results in Table D.9 indicate that the overall qualitative and quantitative conclusions do not change by excluding loans originated during the Covid period.

	Contribution
(1) Default risk	0.0055
(2) Loan characteristics	0.1372
(3) Bank fixed effects	0.1006
(4) Firm fixed effects	0.3908
(5) Residual	0.2576
(6) Covariances	0.1138
# Obs.	248,490

Table D.9: **Contributions of risk, loan characteristics, firm and bank fixed effects to interest rate variation: loans initiated pre-2020.** The different lines represent the different components included in equation (7) in the main text. Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (5) in the main text.

D.7 Firm groups

To guarantee results are not driven by specific groups of firms, we reestimate equation (7) for firms of different sizes and ages.

D.7.1 Age groups

We split firms into the same three groups as in Appendix C.14.1 and estimate equation (7) separately for each of the three different groups. Results in Table D.10 indicate that the conclusions are robust across the three age groups. For young, mature and old firms, the firm fixed component is always the most important component. Results indicate that this firm fixed component is increasing with age, while the loan characteristics and bank component are more important for young firms.

D.7.2 Size groups

We split firms into five different size groups, the same as in Appendix C.14.2. We then estimate equation (7) for each of the five groups separately. Results in Table D.11 illustrate that the results are qualitatively similar across groups, with the firm fixed component always being the most important. Similar to the results by age groups, the firm component increases with firm size, accounting for 59% of the overall variation in rates for the largest firms. Additionally, while the covariance terms are relatively small and unimportant across the four groups of smaller firms, they become important for the largest

	Contribution		
	Young	Mature	Old
(1) Default risk	0.0024	0.0065	0.0049
(2) Loan characteristics	0.3733	0.2909	0.2254
(3) Bank fixed effects	0.1514	0.1198	0.0788
(4) Firm fixed effects	0.2880	0.3099	0.3840
(5) Residual	0.2038	0.2562	0.2826
(6) Covariances	-0.0165	0.0233	0.0293
# Obs.	72,390	151,635	273,319

Table D.10: **Contributions of risk, loan characteristics, firm and bank fixed effects to interest rate variation: by firm age groups.** The different lines represent the different components included in equation (7) in the main text. The different columns represent the estimates for the three different age groups: young, mature and old. Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (5) in the main text.

	Contribution				
	Small	Medium-small	Medium	Medium-Large	Large
(1) Default risk	0.0039	0.0056	0.0051	0.0043	0.0172
(2) Loan characteristics	0.3461	0.3146	0.2835	0.2496	0.1762
(3) Bank fixed effects	0.1469	0.1350	0.1116	0.0801	0.0579
(4) Firm fixed effects	0.2640	0.2772	0.3247	0.3933	0.5879
(5) Residual	0.2284	0.2708	0.2955	0.2303	0.2660
(6) Covariances	0.0146	0.0024	-0.0154	0.0467	-0.0879
# Obs.	137,468	115,493	119,218	73,346	54,023

Table D.11: **Contributions of risk, loan characteristics, firm and bank fixed effects to interest rate variation: by firm size groups.** The different lines represent the different components included in equation (7) in the main text. The different columns represent the estimates for the five different size groups: small, medium-small, medium, medium-large and large. Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (5) in the main text.

firms. The increased importance of the covariance terms comes from a strong sorting of the largest firms into specific banks (-8.8%).

E Robustness exercises and extensions for Section 4.2

E.1 Extension to firm-bank-time fixed effects

As the firm-bank relationship may evolve over time, our baseline specification may underestimate the contribution of the relationship component. To address this, we further

	Contribution	
	Firm financing	All instruments
(1) Bank-time effects	0.096	0.087
(2) Firm-time effects	0.705	0.535
(3) Relationship-time effects	0.006	0.164
# Obs.	262,876	1,159,644

Table E.12: **Contributions of firm-time, bank-time and firm-bank-time fixed effects to interest rate variation.** The different lines represent the different components. The different columns represent the results when estimating equation (11) on the sample of firm financing loans only and on the sample with all credit instruments. Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (5) in the main text. The Relationship-Time fixed effect is presented as the difference between the estimated firm-bank-time fixed effect and the separate firm-time and bank-time fixed effects estimated on the same sample.

test the results when including a firm-bank-time fixed effect $\mathcal{FB}_{fbt}$:

$$r_{lfbt} = \beta_1 \mathbf{R}_{lt} + \beta_2 \mathbf{L}_l + \mathcal{M}_{lt} + \mathcal{FB}_{fbt} + \varepsilon_{lfbt}. \tag{11}$$

This places additional restrictions on the sample, as this requires firms to initiate multiple loans from different banks in the same time period. The frequency for the time is one year. To guarantee the validity of the results with the reduced sample, we estimate specification (11) for firm financing loans only and for all credit instruments. Results in Table E.12 illustrate that the firm component is still the most important one, with the bank and relationship components still accounting for less of the overall variation than the firm one.

E.2 All credit instruments

We extend the baseline analysis in Section 4.2 to include all instruments in the sample. We estimate the following specification

$$r_{lfbt} = \beta_1 \mathbf{R}_{lt} + \beta_2 \mathbf{L}_l + \mathcal{M}_{lt} + \mathcal{FB}_{fb} + \varepsilon_{lfbt}. \tag{12}$$

The instrument type is now also included as a fixed effect in the vector $\mathbf{L}_{lt}$. The number of firms in the sample increases from 73,858 to 115,673 and the number of observations goes from 463,624 to 1,469,225. This strengthens the identification of fixed effects, as we now have almost twice as many observations per firm and more within firm variation. The

	Contribution
(1) Bank effects	0.079
(2) Firm effects	0.381
(3) Firm-bank relationship effects	0.176
# Obs.	1,469,225

Table E.13: **Contributions of firm, bank and firm-bank fixed effects to interest rate variation: all credit instruments.** The different lines represent the different components. Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (5) in the main text. The Relationship Component is presented as the difference between the estimated firm-bank fixed effect and the separate firm and bank fixed effects estimated on the same sample.

results are presented in Table E.13. Firm-bank relationships now account for a larger share of rate variation. The overall importance of the relationship component increases slightly compared to the baseline scenario, but the bank and relationship component accounts for 25.5% of the overall variation, still considerably less than the firm component.

F Additional details for Section 5

F.1 Definitions of firm and relationship banking control variables

This section provides definitions of firm-level variables and the relationship-banking controls used in the empirical analysis. Variables measured in monetary units are deflated using the GDP deflator. The following variables are specified in log units in the regression: capital, number of employees, revenue, wage bill, TFP, financial assets, total assets, total liabilities, and the number of R&D employees. For inventories, there is an indicator variable for values of zero, and otherwise it is in log units.

Time-varying characteristics

- *Capital stock.* Book value of fixed tangible assets.
- *Number of employees.* Total number of employees.
- *Wage bill.* Total labor costs.
- *Revenue.* Total operating revenue (turnover).
- *Revenue growth rate.* Log difference of revenue from previous year.
- *Profit share.* Operating profits as a share of total revenue.
- *Total factor productivity (TFP).* Firm-level productivity estimated using the methodology of Ackerberg et al. (2015), based on capital, intermediate inputs, and labor costs.

- *Investment rate.* Annual change in fixed tangible assets divided by the previous year's level of fixed tangible assets.
- *Inventories.* Value of inventories held.
- *Equity, financial assets, total assets and total liabilities.* Values reported in balance sheet.
- *Past default indicator.* Indicator equal to one if the firm has had overdue credit exceeding 2.5% of total credit for three consecutive months between September 2018 and the date at which the loan was issued.[11]
- *Past late repayment indicator.* Indicator equal to one if the firm has been at least 30 days late on a loan repayment between September 2018 and the date at which the loan was issued.
- *Trade payables over total assets.* Trade credit received from suppliers relative to total assets.
- *Trade receivables over total assets.* Trade credit extended to customers relative to total assets.

Time-invariant characteristics

- *Industry.* Five-digit industry classification according to the Portuguese Classification of Economic Activities. There are 833 such industries.
- *Location.* Firm headquarter location measured with four-digit postcodes.
- *Number of local banks (top 5).* Number of the five largest banks that have branches located in a firm's five and seven-digit postcode areas (two variables).
- *Number of local banks (other banks).* Number of other banks with branches located in a firm's five and seven-digit postcode areas (two variables).
- *Number of establishments.* Maximum number of establishments operated by the firm during the sample period.
- *Home country.* Country of origin of the firm.
- *Legal form.* Legal structure of firm. The options are: private limited liability company (87% of firms), public limited company (6%) and other legal structures.[12]
- *Number of R&D employees.* Maximum number of employees engaged in research and development activities during the sample period.
- *Exporter indicator.* Indicator equal to one if the firm reports having exports in any year in the sample period.
- *Number of foreign establishments.* Maximum number of foreign establishments operated by the firm during the sample period.
- *Birth year indicators.* Indicator variable for the founding years of firms. Firms born before 1943 are grouped together because for these years there are less than 50 firms founded per year in our sample.

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

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

- *Revenue growth standard deviation.* Standard deviation of the annual log difference of revenue.
- *Revenue growth skewness.* Skewness of annual log difference of revenue.
- *TFP growth standard deviation.* Standard deviation of the annual log difference of TFP.
- *TFP growth skewness.* Skewness of the annual log difference of TFP.

Relationship banking controls

- *Number of banks.* Total number of banks from which the firm has obtained firm financing loans from prior to the month of loan initiation.[13]
- *Herfindahl–Hirschman index for bank credit.* The Herfindahl–Hirschman index for all of a firm's outstanding firm financing loans in the month prior to loan initiation.[14]
- *Duration of relationship.* Months since the firm's first firm financing loan from a given bank.
- *Number of loans.* Number of firm financing loans that a firm has received from the bank issuing the new loan over the past five years.
- *Value of loans.* Log total value of firm financing loans the firm received from the bank issuing the new loan over the past five years.
- *Share of loans.* Share of a firm's total number of loans that have been issued by the bank issuing the new loan over the past five years.
- *Share of loans by value.* Share of a firm's total value of loans that have been issued by the bank issuing the new loan over the past five years.

F.2 All credit instruments

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

Table F.14 reports the results. When including all loans, firm fixed effects account for 49% of the total variation in interest rates. Of this, 3.6 percentage points are explained by observable firm characteristics, while 45.2 percentage points remain attributable to the unobservable component.

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

[13]As is the case throughout the paper, we cannot observe loans that were not active in September 2018 or later, so these are implicitly omitted.

[14]For measuring the value of each loan we use the current outstanding amount that a firm owes the relevant bank.

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

Table F.14: **Contributions of firm fixed effects and time-invariant observable firm characteristics: all credit instruments** The different lines represent the different components. Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (5) in the main text.

the firm fixed component, which cannot be explained by time-invariant observable firm characteristics.

F.3 Ex ante vs. ex post heterogeneity

The results in the main text up to Table 6 show that there are highly persistent differences in interest rates across firms, which are not accounted for by loan characteristics, default risk, bank heterogeneity and relationship banking, or observable firm characteristics. In the exercise in this section we evaluate to what extent this heterogeneity is present from the early years of firms, or whether it arises as a result of shocks that they experience later on. To do this we apply the statistical methodology of Sterk et al. (2021) to decompose the paths of interest rates for firm-level loan sequences into ex ante heterogeneity and ex post persistent shocks.[15]

For this exercise we use residualized interest rates. Specifically, we estimate a simplified version of equation (9) that omits the firm fixed effects and the controls for firm observables in $\mathbf{F}_{ft}$, to give us interest rates purged of variation due to loan characteristics, default risk, and bank and firm-bank relationship effects. Our focus is on understanding the nature of the remaining variation in interest rates.

Consider a firm f for which we observe a sequence of loans numbered $n = 1, 2, 3, \ldots$ in the data. As for the earlier analysis, we use the interest rates on those loans at the time of issuance, with $\check{r}_{f,n}$ denoting the residualized interest rate on the n^{th} loan of firm f. The

[15]The methodology is designed to decompose the path of firm-level variables into these components. The Sterk et al. (2021) application is to employment, while here it is interest rates.

statistical model is a stochastic process that describes the evolution of these rates:

$$\check{r}_{f,n} = \underbrace{u_{f,n} + v_{f,n}}_{\text{ex ante component}} + \underbrace{w_{f,n} + z_{f,n}}_{\text{ex post component}}, \tag{13}$$

where

$$\begin{aligned}
u_{f,n} &= \rho_u u_{f,n-1} + \theta_f, & u_{f,0} &\sim i.i.d.\left(\mu_u, \sigma_u^2\right), & \theta_f &\sim i.i.d.\left(\mu_\theta, \sigma_\theta^2\right), & |\rho_u| &\leq 1, \\
v_{f,n} &= \rho_v v_{f,n-1}, & v_{f,0} &\sim i.i.d.\left(\mu_v, \sigma_v^2\right), & & & |\rho_v| &\leq 1, \\
w_{f,n} &= \rho_w w_{f,n-1} + \varepsilon_{f,n}, & w_{f,0} &= 0, & \varepsilon_{f,n} &\sim i.i.d.N\left(0, \sigma_\varepsilon^2\right), & |\rho_w| &\leq 1, \\
z_{f,n} &\sim i.i.d.\left(0, \sigma_z^2\right).
\end{aligned}$$

$u_{f,n}$ and $v_{f,n}$ capture the ex ante component of the process because their shocks are fully determined before the firm obtains its first rate, while $w_{f,n}$ and $z_{f,n}$ capture the ex post component because they are subject to shocks for every rate.[16] For the ex ante component, there are three shocks—$u_{f,0}$, $v_{f,0}$ and θ_f—representing the initial conditions of firms.[17] The effects of $u_{f,0}$ and $v_{f,0}$ decay to zero over time, while θ_f has a permanent effect that accumulates over loans at speed ρ_u. In particular, with $\rho_u < 1$, the long-run effect of a firm's initial conditions on its interest rates is $\frac{\theta_f}{1-\rho_u}$. This specification allows for rich heterogeneity in the initial rates of firms, long run rates, the gaps between these, and the speeds of convergence. Even firms with the same initial and long run rates can follow different paths due to heterogeneity in $u_{f,0}$ and $v_{f,0}$, and the different decay rates for these shocks.

The ex post component is composed of two shocks: an i.i.d. shock with expected value of zero, $z_{f,n}$, and a persistent shock, $w_{f,n}$, that follows an AR(1) process with persistence ρ_w.[18] To ensure that all variation coming from $w_{f,n}$ is ex post, we impose $w_{f,0} = 0$.

Estimation To parameterize the interest rate process we follow the approach of Sterk et al. (2021) by estimating the eight parameters so that simulated interest rates match the autocovariance matrix for interest rates paid by firms from the data. To construct this

[16]The notation $i.i.d(\mu, \sigma^2)$ denotes a random variable with an i.i.d. distribution that has mean μ and variance σ^2.

[17]These shocks are i.i.d. across firms.

[18]All shocks for these two processes are i.i.d. across firms and over time.

autocovariance matrix, we need to observe a sequence of loans for a large enough sample of firms from their first loan. In our data, we can be sure to observe all loans issued to firms born from September 2018 onwards, however this sample is too small to generate reliable autocovariance estimates. We therefore extend the sample to include all firms born from 2010 onwards and use the interest rates on the first seven loans issued by each firm.[19] While this gives us a larger sample, it comes with the limitation that we do not observe loans that matured before September 2018. So, for firms born before this date, we will observe loans that have a long enough maturity, but will miss shorter maturity loans. To verify that this does not substantially affect the results, we replicate the analysis restricting the sample to firms born in 2015 or later and find very similar results.[20] We then estimate the autocovariance matrix for loans one to seven for the firms in our sample. We use an unbalanced panel, allowing firms that have issued fewer than seven loans to be included, in order to maximize the sample size. We then estimate the parameters of the model by simulated method of moments, minimizing the sum of squared deviations of the 28 moments of the autocovariance matrix implied by the following equation from their empirical counterparts:[21]

$$\operatorname{cov}[\check{r}_{f,n}, \check{r}_{f,n-j}] = \underbrace{\left(\sum_{k=0}^{n-1} \rho_u^k\right)\left(\sum_{k=0}^{n-1-j} \rho_u^k\right)\sigma_\theta^2 + \rho_u^{2n-j}\sigma_u^2 + \rho_v^{2n-j}\sigma_v^2}_{\text{ex ante component}}$$

$$+ \underbrace{\sigma_\varepsilon^2 \rho_w^j \sum_{k=0}^{n-1-j} \rho_w^{2k} + \sigma_z^2 \mathbf{1}_{j=0}}_{\text{ex post component}}. \tag{14}$$

Table F.15 presents the parameter values. We repeat the estimation process 1000 times and report the standard deviation of the parameters. The fit of the model to the data is presented in Figure F.24.

Two key parameters of the process are ρ_u and σ_θ. Together, they determine how much

[19]To maximize the sample size, we use an unbalanced panel, including all loans issued by firms up to their 7$^{\text{th}}$ loan. We choose seven loans to maintain a large enough sample for the autocovariance estimates.

[20]For firms born from 2015 to August 2018, there are still some loans that we do not observe, but this issue is reduced relative to the baseline analysis. We do not impose the tighter restrictions of firms being born in 2018 or later because the sample is too small for reliable estimates of the autocovariances.

[21]This equation corresponds to equation (2) in Sterk et al. (2021)

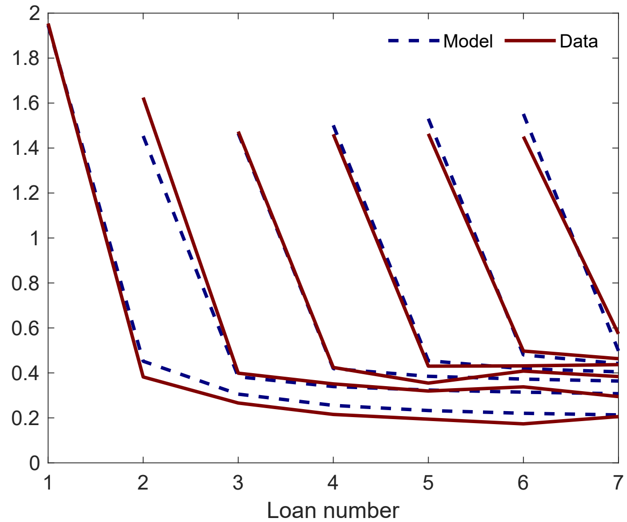


Figure F.24: **Autocovariance of interest rates: model and data.** This figure presents the autocovariances of interest rates from the model and the data. The line that starts at loan one presents the variance of loan one interest rates (the value in the loan one column) and the covariance of these rates with the rates for each future loan number. The line starting at loan two does the same for loan two, etc.

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

Table F.15: **Parameter values.** Parameter values for the interest rate process in equation (13). Standard deviations are in parenthesis.

long run heterogeneity in interest rates results from ex ante factors. The point estimates imply a long-run standard deviation of spreads of 0.55 percentage points. For comparison, our empirical estimate of the standard deviation in interest rates due to firm fixed effects is 1.01.

Results Figure F.25 presents the share of the variance of interest rates across firms that is due to the ex ante component, for loan numbers one to seven. The ex ante component is particularly important for determining first loan interest rates, accounting for approximately 43% of the variance. This contribution halves for the second loan, but then persists at around 21% for all loans thereafter. This indicates that firm heterogeneity prior to first loan issuance is an important source of persistent rate variation. To tie this to our earlier estimate of the contribution of firm fixed effects, this translates to ex ante factors accounting for approximately 15 of the 31 percentage points of total interest rate variation due to

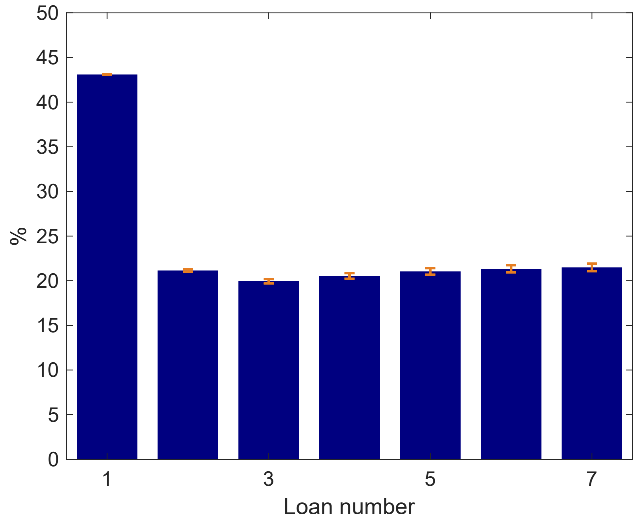


Figure F.25: **Ex ante contribution to interest rate variance.** This figure presents the percentage of the variance of interest rates across firms that is accounted for by ex ante factors as a function of the loan number.

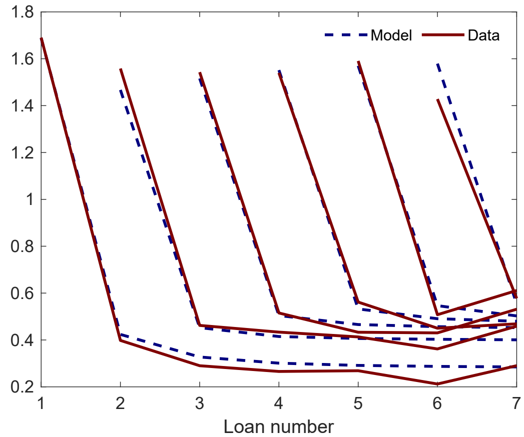
ρ_u	ρ_v	ρ_w	σ_θ	σ_u	σ_v	σ_ϵ	σ_z
0.413	0.140	0.317	0.405	1.321	3.927	0.581	0.753
(0.099)	(0.086)	(0.256)	(0.033)	(0.667)	(2.397)	(0.237)	(0.313)

Table F.16: **Parameter values for firms born from 2015 onwards.** Parameter values for the interest rate process in equation (13), estimated using data for firms born from 2015 onwards. Standard deviations are in parenthesis.

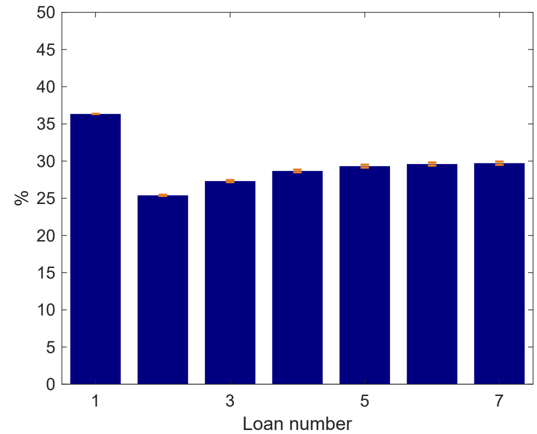
the fixed firm component in Table 6.[22]

Analysis for firms born from 2015 onwards To assess the extent to which the results are affected by us not observing all loans for some firms born before September 2018, we repeat the analysis for firms born from 2015 onwards. There are still some omitted loans for this sample, but it is less of an issue. The estimated parameter values are presented in Table F.16, the fit of the model to the data is in Figure F.26(a), and the share of the variance for each loan number that is due to ex ante factors is in Figure F.26(b). With a smaller sample, the autocovariance values are less stable. However, quantitatively, there is a similar share of the variance of interest rates due to ex ante factors.

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



(a) Autocovariance of interest rates



(b) Ex ante contribution to rate variance

Figure F.26: **Autocovariances and ex ante contribution to rate variation for firms born from 2015 onwards.** Panel (a) replicates Figure F.24 for firms born from 2015 onwards, and panel (b) replicates Figure F.25 for the same sample.

G Additional details for Section 6

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

In Section 5 we take the firm fixed effects and evaluate how much can be explained by firm time invariant characteristics, including location at the level of 4-digit postcodes. Here, we repeat the analysis but using location controls at the 5-digit postcode level. Results in Table G.17 show that increasing the precision of the location controls does not contribute to explaining the firm fixed component in interest rates.

G.2 Bank-location fixed effects

To test the possibility that bank market power shows up at the firm-location level, we estimate the following specification:

$$r_{lfbt} = \beta_1 \mathbf{R}_{lt} + \beta_2 \mathbf{L}_l + \mathcal{M}_{lt} + \mathcal{B}_{bp} + \beta_3 \mathbf{F}_{ft} + \mathcal{F}_f + \beta_4 \mathbf{BF}_{bft} + \varepsilon_{lfbt}, \tag{15}$$

where $\mathcal{B}_{bp}$ contains fixed effects for each bank b and postcode p pair. As the location component of the bank-postcode fixed effect is the location of the firm receiving the loan, we present the results when including and when excluding the firm fixed effect. We do this for two reasons: (1) including bank-location and firm fixed effects causes a substantial

	Contribution
(1) Default risk	0.0002
(2) Loan characteristics	0.2319
(3) Bank fixed effects	0.0933
(4) Time-varying firm characteristics	0.0105
(5) Relationship banking controls	0.0008
(6) Firm fixed effects	0.3006
(6a) Fixed firm characteristics	0.057
(6b) Latent firm fixed effect	0.243
(7) Residual	0.2872
(8) Covariances	0.0870
# Loans	412,199

Table G.17: **Contributions of risk, loan, firm and firm-bank relationship characteristics, firm and bank fixed effects to interest rate variation: location controls at 5-digit postcode level.** The different lines represent the different components included in equation (9) in the main text. Each value represents the contribution of the respective component to the overall interest rate variance, given by equation (5) in the main text.

number of fixed effects to be estimated, which may cause poor identification and biased results; and (2) the bank-location and firm fixed effects may be correlated, as we use the location of the firm. We enter location fixed effects at the 4–digit postcode level. Table G.18 illustrates that even with the location component, the bank fixed effect accounts for less than the firm fixed effect in the baseline scenario. In addition, including the firm fixed effect does not quantitatively change the size of the bank-location component, which is almost 50% smaller than the firm fixed effect.

G.3 Pricing loan portfolios

To test whether banks price loan portfolios jointly, and in a way that matters for rate variation, we compute residualized interest rates after controlling for loan characteristics, borrower risk, and bank fixed effects. We then compare residuals across banks with which a firm has more than 10 loans and those with which it has fewer than 5 loans. If banks price loans jointly, residuals should vary within a firm across banks as a function of the intensity of the lending relationship. The left panel of Figure G.27 shows that residuals are slightly more dispersed for banks with which the firm has 10 or more loans—consistent with potential omitted variables related to joint pricing—but the magnitude is small and

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

Table G.18: **Contributions of risk, loan, firm and firm-bank relationship characteristics, firm and bank-location fixed effects to interest rate variation.** Each row is the share of the total variance in interest rates due to each component of regression (15), calculated with equation (5) in the main text. The first column indicates the results without the firm fixed component, while the second column indicates the results when including the firm fixed component.

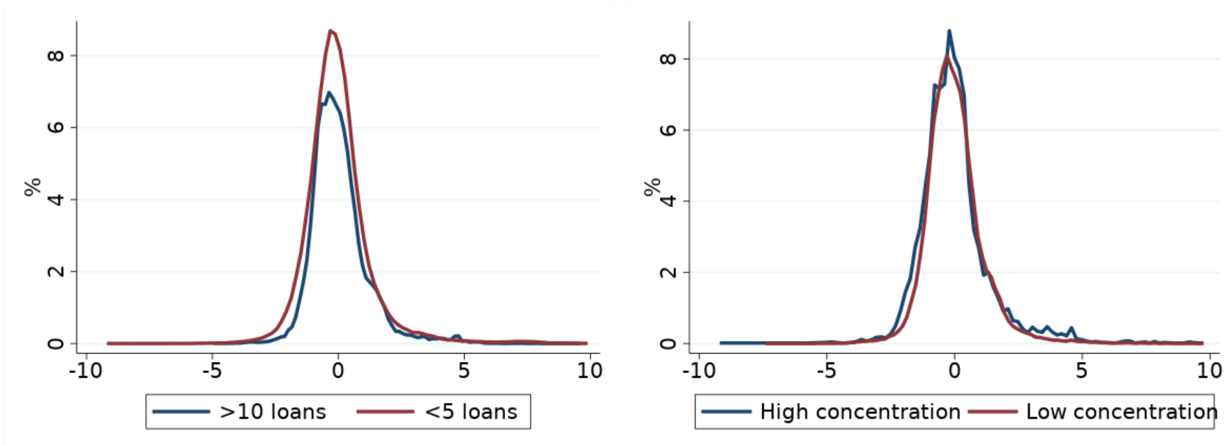
the two distributions largely overlap.

In addition, we split the sample into firms with an HHI above the 70^{th} percentile and below the 30^{th} percentile of the HHI distribution, capturing highly concentrated and more dispersed borrowing structures, respectively. The right panel of Figure G.27 shows that residuals are not quantitatively different across these groups, providing further evidence against economically meaningful differences in interest rates due to loans in a portfolio being priced jointly.

G.4 Variance of firm fixed effects by firm size

To assess the relevance of personal financial relationships between the owner or manager of a firm and a bank, we compare firm fixed effects across firm size groups. The underlying hypothesis is that such relationships should be less relevant for larger firms. In Appendix D.7.2, we have already shown that the share of interest rate variance due to firm fixed effects increases with firm size, which already challenges the importance of personal financial connections.

A potential concern, however, is that this pattern could be driven by lower overall dispersion in interest rates among larger firms. To address this, we estimate the firm fixed effects from equation (9) and examine their standard deviation directly, rather than



(a) Loans (b) HHI

Figure G.27: **Interest rate distribution: by number of loans from a bank and concentration of loans.** The interest rate distribution after controlling for loan characteristics, risk and bank fixed effects. Left panel splits the sample into firms with more than 10 loans from a bank (blue) and less than 5 loans from a bank (red). Right panel splits the sample into firms in the top 30% of loan concentration (blue) and bottom 30% of loan concentration (red).

	Standard deviation				
	Small	Medium-small	Medium	Medium-Large	Large
(1) Firm fixed effects	1.05	0.96	0.83	0.83	0.96
# Obs.	79,404	103,238	110,671	67,772	51,114

Table G.19: **Standard deviation of firm fixed effects by firm size group.** The different columns represent the estimates for the five different size groups: small, medium-small, medium, medium-large and large. Each value represents the standard deviation, in percentage points, of the firm fixed effect.

their contribution to total variance.

Table G.19 shows that, although the standard deviation declines slightly with firm size, the differences are quantitatively small. Moreover, it increases for the largest firms, reaching levels comparable to those of medium-small firms. Overall, this provides additional evidence that the firm fixed effect is unlikely to be driven by entrepreneur-specific relationships.

G.5 Public vs. private firms

Alternatively, we assess the importance of owner or manager relationships by estimating equation (9) in the main text separately for private and public firms. If entrepreneur-specific relationships drive the firm fixed effects, these should be smaller for public firms.

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

Table G.20: **Contributions of risk, loan, firm and firm-bank relationship characteristics, firm and bank fixed effects to interest rate variation: public vs. private firms.** Each row is the share of the total variance in interest rates due to each component of regression (15), calculated with equation (5) in the main text. The first column indicates the results for the sample of private firms, while the second column indicates the results for the sample of public firms.

However, Table G.20 shows the opposite pattern: the rate variation due to firm fixed effects is nearly twice as large for public firms as for private firms, providing further evidence against an entrepreneur-driven explanation.

G.6 Demand shocks

To test if firms have persistent differences in credit demand, which could explain the importance of latent firm fixed effects for the variance of interest rates, we estimate credit demand and supply shocks following Amiti & Weinstein (2018). We then average the demand shocks within firms over time and add these as a control in the second stage regression of the Section 5 analysis, to see if the firm average demand shock has explanatory power for the firm fixed effects from the first stage. The results in Table G.21 show that the addition of the average demand shocks provides no additional explanatory power for the firm fixed effects.

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

Table G.21: **Contribution of average demand shocks to explaining the firm fixed effects.** The different lines represent the different components, calculated with equation (5) in the main text. The first column indicates the results when not including the average demand shocks, while the second includes the average demand shocks.

References

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

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

Clymo, A. & Rozsypal, F. (2023). Firm cyclicality and financial frictions. Working paper.

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