



EIEF Working Paper 17/12
October 2017

**Asymmetric Information and Imperfect
Competition in Lending Markets**

by

Gregory S. Crawford

(University of Zürich, CEPR and CAGE)

Nicola Pavanini

(Tilburg University and CEPR)

Fabiano Schivardi

(LUISS University, EIEF and CEPR)

Asymmetric Information and Imperfect Competition in Lending Markets*

Gregory S. Crawford[†], Nicola Pavanini[‡], Fabiano Schivardi[§]

October 18, 2017

Abstract

We study the effects of asymmetric information and imperfect competition in the market for small business lines of credit. We estimate a structural model of credit demand, loan use, pricing, and firm default using matched firm-bank data from Italy. We find evidence of adverse selection in the form of a positive correlation between the unobserved determinants of demand for credit and default. Our counterfactual experiments show that while increases in adverse selection increase prices and defaults on average, reducing credit supply, banks' market power can mitigate these negative effects.

(*JEL* D82, G21, L13)

*We thank Daniel Akerberg, Jeff Campbell, Pierre-André Chiappori, Lorenzo Ciari, Valentino Dardanoni, Ramiro de Elejalde, Liran Einav, Ralph Koijen, Rocco Macchiavello, Gregor Matvos, Carlos Noton, Tommaso Oliviero, Steven Ongena, Ariel Pakes, Andrea Pozzi, Pasquale Schiraldi, Matt Shum, Michael Waterson, Chris Woodruff, Ali Yurukoglu, Christine Zulehner and various seminars' and conferences' participants for helpful comments. We thank for financial support the Research Centre Competitive Advantage in the Global Economy (CAGE), based in the Economics Department at University of Warwick. We thank Rafael Greminger for excellent research assistance.

[†]University of Zürich, CEPR and CAGE, gregory.crawford@econ.uzh.ch

[‡]Tilburg University and CEPR, N.Pavanini@tilburguniversity.edu

[§]LUISS University, EIEF and CEPR, fschivardi@luiss.it

1 Introduction

Following the seminal work of Akerlof (1970), Rothschild and Stiglitz (1976), and Stiglitz and Weiss (1981), a large theoretical literature has stressed the important role of asymmetric information in financial markets. This literature has shown that asymmetric information can generate market failures such as credit rationing, misspricing of risk and, in the limit, market breakdown. Indeed, the recent financial crisis can be seen as an extreme manifestation of these market failures, whose effects are likely to become more acute during recessions (Tirole 2006). Deepening our understanding of the extent and consequences of asymmetric information is critical for the design of regulatory frameworks that limit its negative effects.

Although the basic theoretical issues are well understood, empirical work analyzing asymmetric information is still uncommon. One reason is that, by definition, asymmetric information is hard to measure. If a borrower has better information than a lender, it is unlikely that a researcher can do better. While researchers cannot generally construct measures of ex-ante unobserved characteristics determining riskiness, they can often observe ex-post outcomes, such as loan defaults. For this reason, the empirical literature, both in credit and insurance markets, has analyzed how agents with different ex-post outcomes self-select ex-ante into contracts with different characteristics in terms of price, coverage, or deductibles (Chiappori and Salanié 2000, Einav, Jenkins and Levin 2012, Starc 2014).¹

The vast majority of this literature analyzes the consequences of asymmetric information using models of competitive markets. Assuming perfect competition in lending markets is not desirable, however, as market structure and asymmetric information can be intimately related. On the one side, informational frictions can constitute a barrier to entry and thus contribute to determining market structure (Dell’Ariccia, Friedman and Marquez 1999, Bofondi and Gobbi 2006); on the other, the effects of asymmetric information may depend on market structure itself (Petersen and Rajan 1994, Vives 2016). This is particularly important for the environment we analyze, the Italian market for small business loans: as shown by Guiso, Sapienza and Zingales (2004), legal entry barriers, whose effects persisted into the 1990s, have shaped its local nature and high degree of concentration.

A recent strand of theoretical research has focussed on the effects of adverse selection in the presence of market power (Lester, Shourideh, Venkateswaran and Zetlin-Jones 2017, Mahoney and Weyl 2017). With perfect competition, banks price at average cost (e.g., Einav and Finkelstein 2011). When adverse selection increases, prices rise, as a riskier pool of borrowers implies more defaults and higher average costs. When banks exert market power, however, greater adverse selection can *lower* prices, as it implies a riskier pool of borrowers at any given price, lowering infra-marginal benefits of a high price in the standard (e.g. monopoly) pricing equation. This implies both that adverse selection can moderate the welfare losses from market power and that imperfect competition can moderate the welfare consequences of adverse selection.

We measure the consequences of asymmetric information and imperfect competition in the market for small business lines of credit. We exploit detailed data on a representative sample of Italian firms, the population of medium and large Italian banks, individual lines of credit between them, and subsequent defaults. While our data include a measure of observable credit risk comparable to that available to a bank during the

¹ See Einav and Finkelstein (2011), Einav, Finkelstein and Levin (2010a), and Chiappori and Salanié (2013) for extensive surveys of this literature.

application process, we also allow firms to have private information about the underlying riskiness of the project they seek to finance. The market is characterized by adverse selection if riskier firms are more likely to either demand credit, use more of their loan, or both. Following Stiglitz and Weiss (1981), an increase in the interest rate exacerbates adverse selection, inducing a deterioration in the quality of the borrower pool. After providing reduced-form evidence of adverse selection and imperfect competition in this market, we formulate and estimate a model of credit demand, loan use, default, and bank pricing that allows us to estimate the extent of adverse selection and to run counterfactuals that approximate economic environments of likely concern to policymakers.

We begin by constructing a model in which banks offer loan contracts to firms. Banks are differentiated by their network of branches, the years in which they have been in a market, and the distance between a potential borrower and their closest branch. Banks compete Bertrand-Nash on interest rates, which also act as a screening device as in Stiglitz and Weiss (1981). Firms seek lines of credit to finance the ongoing activities associated with a particular business project, the riskiness of which is their private information. For their main line of credit, firms choose a bank from which to borrow, if any, according to a mixed logit demand system. They also choose how much of this credit line to use. Finally, they decide whether to repay the loan or default. There are two critical correlations in the model: that between the unobservable determinants of the choice to take up a loan and default (the extensive margin) and that between unobserved determinants of how much of that loan to use and default (the intensive margin). When these correlations are positive, we say that the market is characterized by adverse selection: riskier firms are more likely to demand and use credit.

The degree of competition can have significant consequences on the equilibrium effects of adverse selection in our model. We show that banks with higher market power have lower incentives to increase prices following an increase in adverse selection. This is confirmed by a Monte Carlo simulation: when markets are competitive, more adverse selection always leads to higher interest rates and less credit. As banks' market power increases, however, this relationship becomes weaker and eventually turns negative.²

We estimate the model on detailed microdata covering individual loans between firms and banks between 1988 and 1998. There are two key sources of data. The first, from the Italian Central Credit Register (*Centrale dei Rischi*), provides detailed information on all individual loans extended by the 94 largest Italian banks (which account for 80% of the loan market), including the identity of the borrower and interest rate charged. It also reports whether the firm subsequently defaulted. The second, from the *Centrale dei Bilanci* database, provides detailed information on borrowers' balance sheets and income statements. Critically, this second dataset includes an observable measure of each firm's default risk, which is called its "Score." Combining the two datasets yields a matched panel of borrowers and lenders. While the data span a 11-year period and most firms in the data take out multiple loans, in our empirical analysis we only use the first year of each firm's main line of credit. This avoids the need to model the dynamics of firm-bank relationships and the inferences available to subsequent lenders of existing lines of credit.³ We define local markets at

² Handel (2013), Lustig (2011), and Starc (2014) analyze adverse selection and imperfect competition in US health insurance markets. Each of these focuses on the price-reducing effect of asymmetric information in the presence of imperfect competition. None, however, articulates the non-monotonicity of these effects depending on the strength of competition, an empirically relevant result in our application.

³ A similar approach is followed by, among others, Chiappori and Salanié (2000). We model the dynamics of firm-bank

the level of Italian provinces, administrative units roughly comparable to a US county that, as discussed in detail by Guiso, Pistaferri and Schivardi (2013), constitute a natural geographical market for small business lending. We estimate individual firms' demand for lines of credit, banks' pricing of these lines, firms' loan use, and their subsequent default. We extend the econometric approach of Einav et al. (2012) to the case of multiple lenders by assuming unobserved tastes for credit independent of the specific bank chosen by the firm. We combine this framework with the literature on demand estimation for differentiated products (Berry 1994, Berry, Levinsohn and Pakes 1995). Data on default, loan use, demand, and pricing separately identify the distribution of firms' riskiness from heterogeneous firms' demand for credit.

We face two important challenges in identifying adverse selection in the structural model. First, we only observe prices for firm-bank pairs that actually established a loan relationship, while to estimate the model we also need prices charged by banks from whom firms chose not to borrow. Second, while we have extensive information about firms' characteristics, there may still be determinants of demand, loan use, and default that are observed by banks but not by us as econometricians, and such "soft information" (e.g. a bank's perception of a firm's creditworthiness) may determine loan pricing. We address these challenges using a unique feature of our data, multi-bank borrowing, to estimate a price prediction model with firm fixed effects. This allows us to predict prices accounting for any price-relevant firm characteristic that is common across banks and that they observe and we do not. This ensures that our estimates of adverse selection are not driven by informational differences between us as econometricians and banks. We also address the potential endogeneity of price in our three estimating equations using instrumental variable methods.

In our results, we find evidence of adverse selection in the form of a statistically significant correlation of 0.16 between the unobserved determinants of the choice to borrow and unobserved determinants of default, and of 0.14 between unobserved determinants of loan use and default. These results imply that firms with a higher unexplained propensity to borrow, on both the extensive and intensive margins, are also more likely to default. We also find a positive effect of interest rates on default, which we interpret as evidence of moral hazard.

We run three counterfactuals to quantify the effects of adverse selection and understand its interaction with imperfect competition. In the first experiment, we analyze how market outcomes vary with the degree of adverse selection and its interaction with market power. We do so by doubling the estimated correlation coefficients in the unobserved determinants of loan demand, loan use, and default, and looking at how equilibrium prices, demand, and default vary in response. We then relate these outcomes to banks' markups before the change, which we use as a measure of bank market power. This experiment illustrates the implications of adverse selection and market power in our estimated model and delivers two important findings. First, consistent with the majority of the theoretical literature analyzing adverse selection in competitive environments, we find that the average effect of an increase in adverse selection is to increase prices and reduce the supply of credit. Second, we find that market power significantly mitigates this effect: while increased adverse selection increases prices by an average of 12.9 percentage points, a one standard deviation increase in a bank's average markup lessens this increase by 6 percentage points.

In a second counterfactual, we simulate a potential effect of a financial crisis: an increase in banks' cost relationships in a companion paper (Pavanini and Schivardi 2017).

of capital. With this exercise, we seek to separately identify the effects of adverse selection and imperfect competition on the transmission mechanisms of higher capital costs to the economy. We find that, in the presence of adverse selection, banks with higher market power are less likely to raise prices following an increase in the cost of capital. In a final counterfactual, we investigate further the interaction between adverse selection and imperfect competition by simulating a merger between the two largest banks in each local market. We find that under high adverse selection, a larger fraction of prices declines as concentration rises.

All in all, our results show that asymmetric information and market power both play an important role in the Italian market for small business lines of credit. We find evidence of adverse selection and show that it negatively impacts market outcomes, leading to higher prices, less lending, and more default. At the same time, we also find that market power can mitigate the negative effects of adverse selection: when banks have higher markups, they moderate price increases to reduce the negative consequences of adverse selection on the quality of their borrower pools. These results speak to the debate on the cost and benefits of competition in financial markets. While competition is generally beneficial to borrowers, a more competitive market reduces banks' ability to absorb negative shocks, exacerbating the effects of adverse selection exactly when firms might most need credit. Banking regulators and competition policymakers should be aware of these effects when considering the impact of their decisions on small business lending.

Our paper is related to three main strands of research in economics. The first is a recent and growing theoretical literature analyzing markets with asymmetric information and imperfect competition. Lester et al. (2017) show that equilibrium contracts in insurance and credit markets are jointly determined by adverse selection and market power, and that increased competition and reduced informational asymmetries can be detrimental for welfare. Mahoney and Weyl (2017) show that when a monopolist insurer's market share is high, an increase in adverse selection drives prices down and quantities up, as the monopolist internalizes the increased default costs of its marginal customers. We add to this literature by taking a model with similar features to lending markets and measuring the relative importance of, and interaction between, the two frictions.

The second is the literature on empirical models of asymmetric information. While this literature has largely focused on insurance markets, we look at the less studied area of credit markets, where the most recent applications have followed both experimental (Karlan and Zinman 2009) and structural (Einav et al. 2012) approaches. Our empirical model is closest to that developed by Starc (2014). Her work looks at the welfare impact of imperfect competition in the US Medigap market accounting for customer self-selection into insurers' optimal pricing strategies. We share with Starc (2014) the identification of imperfect competition, through a structural model of demand for differentiated products. However, the market we analyze differs substantially along many important institutional dimensions. Insurers in the US Medigap market are heavily regulated for pricing, minimum loss ratios, and retaliatory taxes, whereas Italian banks face different kinds of regulations, but have almost no restriction on their pricing of loans.⁴ Moreover, firms in our data typically borrow from more than one bank, as there is no exclusivity in business lending. As a consequence, our empirical approach and identification strategy differs substantially from hers.

⁴ For example, insurers in her setting cannot price discriminate based on expected claims, whereas banks can and do price discriminate based on expected default.

Finally, we contribute to the empirical literature analyzing lending markets. One branch of this literature applies structural estimation techniques to analyze consumer and firm behavior in these markets (Ho and Ishii 2011, Koijen and Yogo 2017, Egan, Hortaçsu and Matvos 2017). No paper, however, studies market structure and its interaction with adverse selection. Another branch uses data similar to ours and provides reduced-form evidence consistent with various implications of adverse selection (Gobbi and Lotti 2004, Bofondi and Gobbi 2006, Panetta, Schivardi and Shum 2009, Albertazzi, Eramo, Gambacorta and Salleo 2015). To the best of our knowledge, ours is the first paper that uses structural methods to study how adverse selection and market structure interact in the market for banks' business lending.

The structure of the paper is the following. In Section 2 we describe the dataset and the market, and present reduced form tests of adverse selection and imperfect competition. Section 3 outlines the structural model and Section 4 describes our model of price prediction and the econometric specification of demand, loan use, default and supply. The estimation and the results are in Section 5, the counterfactuals are in Section 6, and Section 7 concludes.

2 Data and Institutional Details

We use a unique and comprehensive dataset of Italian small business lines of credit to study the effects of asymmetric information and imperfect competition. It is based on four main sources of data: data on individual loans from the Italian *Centrale dei Rischi*, or Central Credit Register; firm-level balance sheet data from the *Centrale dei Bilanci*, or Company Accounts Data Services; banks' balance-sheet and income-statement data from the Banking Supervision Register; and data on bank branches at the local level since 1959.⁵ By combining these data, we obtain a matched panel dataset of borrowers and lenders extending over an eleven-year period, between 1988 and 1998.

2.1 Loan Data

The Central Credit Register (hereafter "Credit Register") is a database that contains detailed information on individual loans extended by Italian banks. For each of a number of different types of loans, banks must report data for each individual borrower on both the amount granted and the amount used for all such loans if their total amount exceeds a given value threshold.⁶ In addition, a subgroup of around 90 banks (accounting for more than 80 percent of total bank lending) also provides detailed information on the interest rates they charge to individual borrowers on each loan. We restrict our attention to short-term lines of credit, which have ideal features for our analysis.⁷ First, the bank can change the interest rate at any time. This means that differences between interest rates on loans are not influenced by differences in loan maturity. Second,

⁵ The first three datasets were previously used in Panetta et al. (2009). Further information about each is available there. Detailed information about the last dataset is available in Ciari and Pavanini (2014).

⁶ The types of loans reported are lines of credit, financial and commercial paper, collateralized loans, medium and long-term loans, and personal guarantees. The loan value threshold was 41,000 euros until December 1995 and 75,000 euros thereafter.

⁷ A line of credit establishes a maximum loan balance that a lender permits a borrower to draw upon. The borrower can access funds up to this maximum at any time and pays interest only on the outstanding balance.

loan contracts in the Credit Register are homogeneous products, so that they can be compared across banks and firms. Third, lines of credit are not collateralized, a key feature for our analysis, as issues of adverse selection become less relevant for collateralized borrowing. Fourth, short-term bank loans are one of the main sources of borrowing by Italian firms. According to our data, short-term lines of credit represent over half of total bank lending to firms. We define the interest rate as the ratio of the payment made in each year by the firm to the bank to the average amount of the loan used.

We focus on firms' "main credit line" in the first year they open at least one line of credit. In Italy, firms have relationships with multiple banks to reduce liquidity risk (Detragiache, Garella and Guiso 2000). We define a firm's main credit line as the loan on which the firm borrows most. On average for the firms in our sample, it accounts for around 75% of the total share of credit (both credit extended and credit used). Since Chiappori and Salanié (2000), considering only the first year is common in empirical models of asymmetric information. We do so to avoid modeling challenging topics like heterogeneous experience ratings among borrowers, loan renegotiation, and learning by firms and/or banks, though these are very interesting avenues for future research. This means that we restrict our attention only to the first year in which we observe a firm in our data.⁸ This reduces the sample of firms from around 90,000 to just over 36,500. Panel A of Table 1 reports the loan-level information that we use in the empirical analysis. Out of these 36,520 firms, 69% take up a loan in our sample period and use on average 67% of the amount granted. The average amount granted is around €370,000, and the average interest rate is 14.2%.

Panel B of Table 1 shows summary statistics for the 94 banks that report detailed interest rate information. The average total asset level is almost 11 billion Euros and they employ on average 3,200 workers. The average bank is present in 34 provinces out of 95, but with significant variation across banks.

2.2 Firm Data

The *Centrale dei Bilanci* (hereafter CB) collects yearly data on the balance sheets and income statements of a sample of about 35,000 Italian non-financial and non-agricultural firms. This information is collected and standardized by the CB, who then sells these data back to banks' lending divisions. The unique feature of the CB dataset is that, unlike other widely used datasets on individual companies (such as the Compustat database of US companies), it has wide coverage of small and medium enterprises, almost all of which are unlisted. The coverage of these small firms makes the dataset particularly well-suited for our analysis, because informational asymmetries are potentially strongest for these firms. Initially, data were collected by banks themselves and transmitted to the CB. Over time, the CB has increasingly drawn from balance sheets deposited with local chambers of commerce, where limited liability companies are obliged to file. The firms in the CB sample represent about 30% of the total value added reported in the national accounting data for the Italian non-financial, non-agricultural sector.

In addition to collecting the data, the CB computes an indicator of the risk profile of each firm, which we refer to in the remainder of this paper as the "Score". The Score represents our measure of a firm's observable default risk. It takes values from 1 to 9 and is computed annually using discriminant analysis based on a series of balance sheet indicators (assets, rate of return, debts etc.) according to the methodology described

⁸ To avoid left censoring issues we drop the first year of our sample (1988) and just look at new relationships starting from 1989.

Table 1: Summary Statistics

	Variable	Nobs	Mean	Std. Dev.	Nobs	Mean	Std. Dev.
Panel A:	Demand	36,520	0.69	0.46			
Loan Level	Loan Use	25,351	246.1	444.8			
	Default	25,351	0.06	0.23			
	Amount Granted	25,351	367.3	476.7			
	Interest Rate	25,351	14.24	4.58			
Panel B:	Total Assets	900	10,727	16,966			
Bank Level	Employees	896	3,180	4,583			
	Number of Provinces	861	34.54	30.19			
Panel C:		Borrowing Firms			Non-Borrowing Firms		
Firm Level	Total Assets	25,351	11,336	19,825	11,169	3,622	8,601
	Net Assets	25,351	2,300	6,499	11,169	931	3,621
	Intangible Assets	25,351	360	2,066	11,169	131	1,079
	Intangible/Total Assets	25,351	0.16	0.23	11,169	0.22	0.29
	Profits	25,351	1,033	2,910	11,169	292	1,403
	Cash Flow	25,351	673	2,185	11,169	255	1,183
	Sales	25,351	14,478	25,014	11,169	5,029	12,084
	Trade Debit	25,351	1,710	3,546	11,169	814	3,617
	Short Term Debt	25,351	2,263	5,697	11,169	125	1,512
	Leverage	25,351	0.55	0.86	11,169	0.21	0.70
	Firm's Age	25,351	12.92	12.89	11,169	10.73	12.12
	Score	25,351	5.36	1.77	11,169	4.78	2.14
	Distance to Branch (km)	25,351	2.92	6.72			
	Number of Lenders	25,351	2.87	2.22			
	Share of Main Line	19,751	0.76	0.25			
Panel D:	Number of Banks	702	8.60	4.83			
Market level	Number of Branches	6,036	14.95	24.46			
	Share of Branches	6,036	0.06	0.08			
	Years in Market	6,036	21.06	14.23			
	Market Shares	6,036	0.08	0.08			
	Deposit Amount	2,566	21,113	18,883			
	Number Deposit Accounts	2,566	654	523			
	Deposit Interest Rate	2,566	6.51	1.72			

Note: This table reports sample statistics for the variables in our analysis. In Panel A, an observation is a firm for the first variable and a loan for the others. Demand is a dummy variable indicating whether a firm obtained a credit line, Loan Use is the amount of loan used in thousands of euros, Default is a dummy for a firm having any of its loans classified as bad within the next three years (see Section 2.3 below for further details), (Loan) Amount Granted is in thousands of euros, and Interest Rate is a percentage. In Panel B, an observation is a bank-year. Total Assets are in thousands of €, Employees is the number of employees at the end of the year, Number of Provinces is the number of provinces where a bank is actively lending. In Panel C, an observation is a firm. Total, Net, and Intangible Assets, Profits, Cash Flow, Sales, Trade Debit, and Short Term Debt are in thousands of €. Intangible/Total Assets is the ratio of intangible over total assets, Leverage is the ratio of debt over equity. Firm's Age measures the years since a firm's foundation. Score is an indicator of the risk of the firm computed each year by the CB (higher values indicate riskier companies, see Section 2.2 for more details). Distance to Branch is the distance in kilometers between the city council of each firm and the city council of the closest branch of the bank it borrows from, calculated using the geographic coordinates. Number of Lenders is the number of banks from which a firm opens a line of credit. Share of Main Line represents the ratio of credit used from a firm's main line of credit over total credit used, when credit used is positive. In Panel D, an observation is province-year for the Number of Banks, bank-province-year for the subsequent four variables, and bank-region-year for the last three variables. Number and Share of Branches are per bank-province-year, Years in Market are the number of years a bank has been in a province since 1959. Market Shares are in terms of number of borrowers. Deposit Amount is the total value of a bank's deposits in a region-year in thousands of €. Number of Deposit Accounts is in thousands. Deposit Interest Rate is a percentage.

in Altman (1968) and Altman, Marco and Varetto (1994). The inputs into a firm's Score approximate closely the information that a lending bank has available at the time a loan is granted, as reported in the survey by Albareto, Benvenuti, Mocetti, Pagnini and Rossi (2011) described in detail in Section 4.1.

We define a borrowing firm as one that is present in the Credit Register. Non-borrowing firms are defined according to two criteria: they are not in the Credit Register and report zero bank borrowing in their balance sheets. We use the second definition to exclude firms that are not in the Credit Register but are still borrowing from banks, either from one of the non-reporting banks or through loan types other than lines of credit. Panel C of Table 1 reports descriptive statistics for our sample of borrowing and non-borrowing firms. Borrowing firms have larger assets and sales and an average of 2.9 credit lines active every year. On average, the share of credit used from the main line is 76%.

There is ample evidence that firms, particularly small businesses like the ones in our sample, are tied to local credit markets. For instance, Petersen and Rajan (2002) and Degryse and Ongena (2005) show that lending to small businesses is a highly localized activity, as proximity between borrowers and lenders facilitates information acquisition and reduces borrowers' travel costs. Segmentation of local credit markets is thus very likely to occur. We use Italian provinces, administrative units roughly comparable to a US county, as our definition of banks' geographical markets.⁹ At the time of our data, there were 95 Italian provinces. We report summary statistics for these markets in Panel D of Table 1. There we show that there are 8.6 banks per province-year in our sample, each bank has on average just under 15 branches per province and a market share of 6% for branches and 8% for loans. The market share of the outside option, defined by those firms that choose not to borrow, is on average around 30%. On average a bank has been serving a province for at least 21 years.

Even though our dataset includes both borrowing and non-borrowing firms, we have no information on banks' loan approval decisions. For this reason we need to assume that all firms are offered an interest rate, or know the interest rate that each bank in their province would charge them, and then decide which bank is their best alternative, if any. In our model, a bank that classifies a firm as very risky and for which it does not wish to offer a loan cannot formally reject it, but instead offers it a sufficiently high interest rate to make the firm's demand probability very low (similarly its loan use if it ultimately chooses the bank). As such, it allows for an indirect form of loan rejection. Combined Credit Register datasets of loans and loan application have only recently become available to researchers, as in Jiménez, Ongena, Peydró and Saurina (2014) for the case of Spain. Analyzing the loan approval process is an important area for further research.

⁹ Provinces are a good measure of local markets in Italian banking for three reasons. First, this was the definition of a local market used by the Bank of Italy to decide whether to authorize the opening of new branches when entry was regulated. Second, according to the Italian Antitrust authority, the "relevant market" in banking for antitrust purposes is the province. Third, previous research has concluded that bankers' rule of thumb is to avoid lending to a client located at more than 1.4 (Degryse and Ongena 2005) or 4 (Petersen and Rajan 2002) miles from a branch. In our data firms are on average 2.9 km (1.8 miles) from the closest branch of their main bank.

2.3 Default

We define defaults in our data as follows. Banks must report to the Credit Register if they classify a loan as “bad debt”, meaning that they attach a low probability to the event that the firm will be able to repay the loan in full. This is done when firms are in liquidation or other bankruptcy proceedings, and for those loans that have not made payments for at least six months. This warning cannot be filed for a single overdue payment, but can only occur as the result of a negative evaluation from the bank about of the borrower’s overall financial situation, and usually occurs prior to a legally certified bankruptcy filing. There is institutional and anecdotal evidence that when one bank sends this kind of default warning to the Credit Register it has a “domino effect” on all other loans the defaulting firm has with other banks.¹⁰ In our data, 82.3% of firms receiving such a warning cease all bank borrowing in the same year, 15.1% in the following year, and all remaining firms within 4 years. According to the Italian Civil Code, information about firms’ defaults remains in the Credit Register for 10 years, compromising a defaulting firm’s access to credit from any bank for that period of time.

Following Panetta et al. (2009), we classify default as the event a firm’s main line of credit will be defined as bad debt within three years of being granted. We choose this window as we are interested in adverse selection and want the default event to be relatively close to when the loan was granted.¹¹ We choose this particular limit also because we can trace firm defaults until 2001, 3 years after the end of our sample, and this ensures that we have a uniform definition of failure for all firms, including those that start borrowing toward the end of our sample. According to this definition, 6% of new loans default during our sample period (see Table 1).

2.4 Preliminary Evidence of Imperfect Competition and Asymmetric Information

We provide some descriptive evidence of asymmetric information and imperfect competition before presenting the structural model. To save on space, we report the full analysis in Appendix A.4 and only summarize the main results here.

A positive correlation between bank concentration and interest rates in the Italian banking sector has previously been documented both for loans (Sapienza 2002) and deposits (Focarelli and Panetta 2003). This relationship also holds in our data. Using both the Herfindahl-Hirschman Index (HHI) and the 3-bank concentration ratio as our measure of market concentration, we find that higher concentration is generally associated with higher interest rates on loans (Appendix A.4.1).

As discussed in the Introduction, providing evidence of asymmetric information is more difficult as it is by definition unobserved. Following Chiappori and Salanié (2000), we conduct positive correlation tests between both the decision to take up a loan and default and between the decision of how much of a granted loan to use and default. A positive correlation between the unobservables is interpreted as evidence of asymmetric information, as it implies that firms that are more likely to demand credit are also more likely to default (in the first case) and that firms that use more of their loans are also more likely to default (in

¹⁰ Source: www.tuttocentraledeirisci.it, support web page for borrowers dealing with the Credit Register.

¹¹ This definition captures the majority of defaults: among the new borrowers on which we focus, we find that almost 70% of the firms that eventually default receive a default warning and no longer borrow from banks within 3 years of their first loan.

the second). We describe the details of the empirical framework underlying these positive correlation tests in Appendix A.4.2. Our results indicate a statistically significant positive correlation between both pairs of unobservables, suggesting that asymmetric information does play a role in this market. These results are later confirmed in our structural estimates presented in Section 5 below.

Based on these descriptive results, we formulate and estimate a structural model to measure the extent of asymmetric information and its consequences for market outcomes. The structural framework has four main advantages compared to the reduced form tests summarized in this section. First, it has a more flexible correlation structure for the residuals that allows us to estimate them jointly. Second, it delivers more accurate measures of market power than simple HHI and concentration indexes. Third, it allows us to distinguish between adverse selection and moral hazard. Finally, we can use the structural model to run counterfactual policy experiments to measure the consequences of both adverse selection and imperfect competition, and to understand how they interact with each other in Italian markets for small business lines of credit.

3 The Model

3.1 Overview and Key Assumptions

The model we construct aims at quantifying the effects of asymmetric information on the demand for and supply of small business lines of credit for Italian firms. We assume that each of $i = 1, \dots, I_{mt}$ firms in market m in year t is willing to invest in a project and is looking for credit to finance it. Firms select their main line of credit from among $j = 1, \dots, J_{mt}$ banks active in m in t , if any, that maximizes their benefits.¹² This determines the demand for credit. Conditional on taking a loan, firms also decide the amount of credit to use and whether or not to default. We assume that each bank j active in market m in year t sets interest rates, P_{ijmt} , for each firm i in that market-year based on a static model of Bertrand-Nash competition on interest rates.

The theoretical model we develop relies on three important assumptions. The first was described in detail in Section 2: we limit our analysis to the demand and pricing of firms' main line of credit in the first year they open at least one credit line. As motivated there, we do so to abstract from dynamic issues in firms' lending relationships and to simplify the scope of the empirical analysis. The second assumption relates to asymmetric information. Following Stiglitz and Weiss (1981), we assume that the informational asymmetry in this market concerns the riskiness of the firm. Specifically, conditional on observables, a firm's riskiness is known by that firm but not by any of the J_{mt} banks in its market; instead, banks are assumed to know the distribution of riskiness across firms. We also assume that both borrowers and lenders are risk neutral.

Our third assumption relates to how the amount of credit granted to a firm is determined. As in Stiglitz and Weiss (1981), we assume banks use interest rates as their only screening device. More specifically, we

¹² Firms make borrowing decisions based on the impact they have on their long-run profitability. We do not have enough information about borrowers to estimate these profits, however, and so represent them here as "utilities". This also helps distinguish them from banks' profits, which we are able to estimate.

assume that the amount of credit granted from bank j to firm i is exogenously given by the firm's project requirements, and that the bank offers an interest rate for that specific amount of credit to each firm i in each market m in year t . We justify this assumption based on the institutional features of the market we study. In a standard insurance or credit market with asymmetric information, insurers or banks can compete not only on prices, but also on other terms in the contract. Indeed, in environments with lending exclusivity, banks can offer menus of contracts that specify both the amount of credit granted and the associated interest rate, for example charging interest rates that increase with the amount of granted credit. This forces borrowers to self-select into contracts based on their unobserved riskiness, revealing some of their private information. Importantly, there is no contract exclusivity in the Italian market for small business lines of credit: borrowers can (and do) open multiple credit lines with different lenders. As explained in Chiappori and Salanié (2013), in the absence of contract exclusivity, no convex price schedule can be implemented.¹³ As such, we are comfortable that the exogeneity of granted credit is likely to hold in our application.

3.2 Demand, Loan Use and Default

3.2.1 Preliminaries

Given these assumptions, let there be $i = 1, \dots, I_{mt}$ firms and $j = 1, \dots, J_{mt}$ banks in $m = 1, \dots, M$ markets in years $t = 1, \dots, T$. Let firms have the following utility from their main line of credit, which determines their demand:

$$U_{ijmt}^D = \bar{\alpha}_0^D + X_{jmt}^D \beta^D + \xi_{jmt}^D + \alpha^D P_{ijmt} + Y_{ijmt}^D \eta^D + \varepsilon_i^D + \nu_{ijmt}, \quad (1)$$

where X_{jmt}^D is a vector of bank-market-year determinants of demand ("D"), P_{ijmt} is the interest rate offered by bank j to firm i in market m in year t , Y_{ijmt}^D is a vector of (non-price) firm-bank-market-year determinants of demand, ξ_{jmt}^D represents banks' unobservable (to the econometrician) attributes in market m in year t , and ν_{ijmt} represents unobserved shocks to i 's demand for bank j . Finally, ε_i^D represents firm i 's individual propensity to demand credit that is known to the firm but not the bank and is therefore the source of asymmetric information in the model. We model this as a random coefficient on the constant term, $\alpha_{0i}^D \equiv \bar{\alpha}_0^D + \varepsilon_i^D$, i.e. a shock to firm i 's demand for credit from any bank. We let $U_{i0mt}^D = \nu_{i0mt}$ be the utility from the outside option, which is not borrowing from any of the J_{mt} banks active in market m in year t . Firms choose their main credit line from the bank that maximizes their utility, or else they choose not to open a credit line at all ($j = 0$).

Conditional on borrowing, each firm chooses the amount of credit to use to maximize the following utility:

$$U_{ijmt}^L = \alpha_0^L + X_{jmt}^L \beta^L + \alpha^L P_{ijmt} + Y_{ijmt}^L \eta^L + \varepsilon_i^L, \quad (2)$$

¹³ If interest rates rise with the amount borrowed, borrowers can "linearize" the schedule by opening several credit lines with multiple banks. Indeed, in the pricing regressions we later use to predict prices for non-chosen banks, we find evidence of a negative relationship between interest rates and the amount of granted credit. We thank Pierre-André Chiappori for his suggestions on this point.

where X_{jmt}^L is a vector of bank-market-year determinants of loan use (“L”), Y_{ijmt}^L is a vector of firm-bank-market-year determinants of loan use, and ε_i^L represents the unobserved (to the bank) propensity of firm i to use credit.

Finally, conditional on borrowing, each firm chooses to default if its utility from doing so is greater than zero:

$$U_{ijmt}^F = \alpha_0^F + X_{jmt}^{LF} \beta^F + \alpha^F P_{ijmt} + Y_{ijmt}^{LF} \eta^F + \varepsilon_i^F, \quad (3)$$

where X_{jmt}^F is a vector of bank-market-year determinants of default (“F”), Y_{ijmt}^F is a vector of firm-bank-market-year determinants of default, and ε_i^F represents firm i ’s propensity to default, which is also observed by the firm but not by the bank. Note that while firm i has utility from each of the J_{mt} banks offering main lines of credit in market m in year t , it only has one such main credit line it can use and/or on which it can default.

We face two challenges when going from the economic to the econometric model. First, we only observe prices for bank-firm pairs that actually established a loan relationship, while to estimate the model we will also need prices charged by the banks that each firm did not choose. Second, while we have extensive information about firms’ characteristics, there might still be determinants of demand, loan use, and default that are observed by banks but not by us as econometricians. In Section 4 we explain how we use a specific feature of our data, multiple bank relationships, to address both issues.

3.2.2 Information Structure, Adverse Selection, and Moral Hazard

The main purpose of the model is to distinguish between information observed by both banks and firms, and information private to each firm. The former includes both hard information in the form of observable firm covariates, Y_{ijmt} (e.g., firm-specific income statement and balance sheet variables), as well as soft information known to the bank through its interaction with the firm. Private information known to firms but not to banks, by contrast, is captured by the unobservables ε_i^D , ε_i^L , and ε_i^F . We assume that ε_i^D , ε_i^L and ε_i^F are fixed firm attributes that don’t vary across banks, and are distributed according to the following multivariate normal distribution:

$$\begin{pmatrix} \varepsilon_i^D \\ \varepsilon_i^L \\ \varepsilon_i^F \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_D^2 & \rho_{DL}\sigma_D\sigma_L & \rho_{DF}\sigma_D \\ \rho_{DL}\sigma_D\sigma_L & \sigma_L^2 & \rho_{LF}\sigma_L \\ \rho_{DF}\sigma_D & \rho_{LF}\sigma_L & 1 \end{pmatrix} \right). \quad (4)$$

Following Stiglitz and Weiss (1981), we interpret a positive correlation between the firm-specific unobservables driving demand and default (ρ_{DF}) as evidence of adverse selection: a positive correlation between ε_i^D and ε_i^F implies that firms with a higher unobservable propensity to demand credit are also more likely to default. Following similar logic, we interpret a positive correlation between the unobservables driving loan use and default (ρ_{LF}) as further evidence of adverse selection.¹⁴ The correlation between unobservables

¹⁴ One concern with this interpretation is the possibility that two firms that are equally risky ex-ante take the same loan and one is hit by a negative shock after the contract has been signed that increases both the use of its loan and its probability of default. Given

driving demand and loan use (ρ_{DL}) does not have an interpretation in terms of adverse selection, but simply allows for the possibility that firms that are more likely to take up a loan are also more likely to draw more on it.

Our model is similar to Einav et al. (2012), but differs in the specification of both demand and supply. In our case, borrowers (firms) choose among multiple banks who compete for customers by setting prices (interest rates). This raises the issue of how best to correlate residuals from the demand model, which vary across both borrowers and alternatives (i.e. lenders), to the residuals from the loan use and default models, which instead vary only across borrowers. We resolve this issue by allowing the normally distributed random coefficient on the constant term to be correlated with the residuals from the loan use and default equations.¹⁵ This is a practical and intuitive solution, as it allows for a correlation between unobservables only at the level of the borrower. This implies that risky firms have high demand for credit from *all* lenders, and not differently across different lenders.¹⁶

While the focus of the paper is estimating adverse selection, we also allow for the presence of moral hazard. As shown by Holmstrom and Tirole (1997), high repayment requirements on loans can reduce the incentives to exert effort, thus increasing the default probability. Of course, firms that are observably riskier may also be offered higher prices. To account for this, in our econometric model we estimate α^F using the component of price variation that is orthogonal to firms' observable and unobservable characteristics. As a consequence, following Adams, Einav and Levin (2009), we interpret a positive effect of price on default ($\alpha^F > 0$ in equation (3)) as evidence of moral hazard. We do not, however, make this effect a focus of our counterfactual exercises.

3.3 Supply

On the supply side, we assume banks engage in Bertrand-Nash competition in prices (interest rates). In particular, bank j 's expected profits from charging a price P_{ijmt} offered to firm i in market m in year t are given by:

$$\Pi_{ijmt} = P_{ijmt}Q_{ijmt}(1 - F_{ijmt}) - MC_{ijmt}Q_{ijmt}, \quad (5)$$

that such a shock was not observed by either the bank or the firm ex-ante, such a positive correlation between loan use and default would not be related to adverse selection. There is also a further possibility that, following such a shock and an increase in the use of its loan, the firm's incentives to undertake risk could change. In this case, ρ_{LF} could also be interpreted as evidence of moral hazard. Although theoretically possible, we believe that these concerns are not likely to be important in our estimating framework as we measure the amount used in the first year in which the contract is signed, which limits the time in which such shocks to firm performance could occur.

¹⁵ Such random coefficients are common in the literature on demand estimation for differentiated products (Berry 1994, Berry et al. 1995). Following Nevo (2000b), we interpret $Y'_{ijmt}\eta^D$ as observed heterogeneity in the random coefficient. These firm-specific and firm-bank specific observable characteristics help us to control for the observable sources of the borrower's taste for credit (regardless of which bank it chooses), leaving ε_i^D as the unobserved taste for credit (and source of asymmetric information).

¹⁶ We also estimated the model with the random coefficient on the interest rate rather than on the intercept. We find similar results, in the sense that riskier firms have lower price elasticities and there is statistically significant evidence of adverse selection. We maintain the assumption of the random coefficient on the constant, as that is closer to the spirit of the Stiglitz and Weiss (1981) model in which firms have privately-observed differences in demand for credit, rather than privately-observed differences in price sensitivity. We present the results of this alternative specification in Appendix A.7.

where Q_{ijmt} and F_{ijmt} are banks' expectations of each firm's demand and default, P_{ijmt} is the interest rate on i 's loan, and MC_{ijmt} is bank j 's marginal cost of lending to firm i in market m in year t . Expected demand, Q_{ijmt} , is given by the product of the model's demand probability and expected loan use by i for a loan from j . This expected profit function corresponds to the standard one underlying Bertrand-Nash pricing (Berry et al. 1995), augmented to account for the probability of default, F_{ijmt} .

Note that expected default, F_{ijmt} , depends on the price charged by bank j through two channels. First, the default equation (3) allows for a direct impact of interest rates on firms' default probabilities (α^F). As described above, we interpret such a relationship as evidence of moral hazard. Second, a higher interest rate also changes the composition of borrowers: a higher price increases the expectation of ε_i^D , firm i 's unobserved demand for credit, conditional on a loan being taken as low-utility-from-borrowing firms are more likely to self-select out of the borrowing pool. If $\rho_{DF} > 0$, this implies in turn that an increase in price increases the average default probability of a bank's pool of borrowers.

The first-order condition of this profit function delivers the following pricing equation:

$$P_{ijmt} = \underbrace{\frac{MC_{ijmt}}{1 - F_{ijmt} + F'_{ijmt}\mathcal{M}_{ijmt}}}_{\text{Effective Marginal Cost}} + \underbrace{\frac{(1 - F_{ijmt})\mathcal{M}_{ijmt}}{1 - F_{ijmt} + F'_{ijmt}\mathcal{M}_{ijmt}}}_{\text{Effective Markup}}, \quad (6)$$

where F'_{ijmt} is the derivative of expected default with respect to price and $\mathcal{M}_{ijmt} = -Q_{ijmt}/Q'_{ijmt}$ is bank j 's markup on a loan to firm i (with Q'_{ijmt} the derivative of expected demand with respect to price). Much like a regular Bertrand-Nash pricing equation can be split into a marginal cost term and a markup, so too can ours. The possibility of default, however, changes the nature of each term and so we denote the first term in equation (6) bank j 's "Effective Marginal Cost" of serving firm i and the second term bank j 's "Effective Markup."¹⁷

The denominator in this pricing equation embodies the mechanisms by which adverse selection and imperfect competition interact to determine prices in lending markets. It has two terms: firm i 's repayment probability, given by one minus its default probability, $1 - F_{ijmt}$, and the derivative of this default probability with respect to price, F'_{ijmt} , multiplied by bank j 's (conventional) markup on its loan to i , \mathcal{M}_{ijmt} .

Consider first the impact of changes in adverse selection as measured by ρ_{DF} on the first term in this denominator, $1 - F_{ijmt}$.¹⁸ As discussed above, firms that borrow are more likely to have high unobservable demand for credit (ε_i^D). The essence of adverse selection, $\rho_{DF} > 0$, is that these firms are also more likely to default than an average firm.¹⁹ In such an environment, increases in adverse selection increase the selectivity of the default unobservables, increasing defaults, F_{ijmt} . This pushes down the denominator in equation (6) and tends to increase prices. We call this first effect of changes in adverse selection on prices

¹⁷ If default and its derivative are both zero, i.e. $F_{ijmt} = F'_{ijmt} = 0$, equation (6) simplifies to the standard Bertrand-Nash pricing equation, $P_{ijmt} = MC_{ijmt} - \frac{Q_{ijmt}}{Q'_{ijmt}} = MC_{ijmt} + \mathcal{M}_{ijmt}$, showing that price can be written as marginal cost plus a markup.

¹⁸ For convenience, we articulate the effects of changes in adverse selection via changes in the correlation in unobservable determinants of demand and default, ρ_{DF} . Similar effects obtain if we consider instead the correlation in unobservable determinants of loan use and default, ρ_{LF} .

¹⁹ See Appendix A.1 for proofs of the claims made in this paragraph.

the “*average borrower effect*”: an increase in adverse selection increases the riskiness of those that choose to borrow, increasing average default rates and thus prices.

Consider next the impact of changes in adverse selection on the second term in the denominator of the pricing equation, $F'_{ijmt} \times \mathcal{M}_{ijmt}$. While signing this term analytically is difficult, we show in a Monte Carlo exercise in Appendix A.2.2 that when there is adverse selection ($\rho_{DF} > 0$), not only does default increase with prices as banks lend to firms with higher unobservable demand for credit ($F'_{ijmt} > 0$), but also that increases in adverse selection exacerbate this effect (i.e., $\partial F'_{ijmt} / \partial \rho_{DF} > 0$). Because markups are also positive, this increases the denominator in equation (6) and tends to *reduce* prices. We call this second effect of changes in adverse selection on prices the markup-mediated “*marginal borrower effect*”: an increase in adverse selection increases the responsiveness of default to price changes (F'_{ijmt}), increasing the relative safety of marginal borrowers compared to average borrowers and increasing banks’ desire to keep them by lowering prices.

Which of the average borrower or marginal borrower effect dominates price changes when adverse selection increases depends on the level of competition, measured by the markup term \mathcal{M}_{ijmt} , as this influences the value of marginal borrowers. High levels of competition imply margins are low, lowering the value to the bank of marginal borrowers, and encouraging banks to respond to increased adverse selection by increasing prices. By contrast, low levels of competition imply margins are high, increasing the value to the bank of marginal borrowers, and encouraging them to respond to increased adverse selection with price reductions.

We further illustrate this non-monotonic response of prices to increases in adverse selection in a Monte Carlo exercise presented in Appendix A.2. We allow for both advantageous and adverse selection in the form of $\rho_{DF} \in [-1, 1]$ and analyze the pricing decisions of a monopolist facing a competitive fringe. We parameterize competitive intensity by varying the slope of the monopolist’s (residual) demand curve. Consistent with the economic effects described above, we show that when competition is strong (as measured by high absolute values of the slope of residual demand), increases in adverse selection increase prices and when competition is weak, they decrease prices.

3.4 Theoretical Work on Imperfect Competition in Selection Markets

Theoretical work on the relationship between asymmetric information and imperfect competition is still very limited. The paper that analyzes a setting most comparable to ours is Mahoney and Weyl (2017).²⁰ They study the interaction of adverse selection and imperfect competition in insurance markets using graphical price-theoretic reasoning in the spirit of Einav, Finkelstein and Cullen (2010b), but extended to allow for imperfect competition. They model adverse selection as a correlation between consumers’ willingness to pay and insurers’ costs, causing marginal costs to be downward-sloping, and parameterize changes in adverse selection as a rotation of the industry marginal cost curve holding population average costs constant. They show that, when insurers have market power, increases in adverse selection can either raise or lower prices

²⁰ Lester et al. (2017) also analyze the interaction of asymmetric information and imperfect competition, but do so in markets with contract exclusivity, meaning firms can (and do) offer menus of contracts in order to encourage buyers to self-select according to their private information. As argued above, there is no exclusivity in the Italian business lending market.

depending on how the expected cost of the marginal consumer changes. Using linear demand curves, they show that a monopolist facing an increase in adverse selection will raise prices if, before the change, it served less than half the population and will reduce them otherwise.

The mechanisms underlying the price response of changes in adverse selection in our setting are different and stem from fundamental differences between insurance and lending markets. In the insurance markets that are the main focus of Mahoney and Weyl (2017), adverse selection manifests itself in its impact on insurers' marginal costs. By contrast, in lending markets like ours, adverse selection manifests itself in default rates that impact banks' *marginal revenue*. This is evident in the banks' profit equation (5): banks must pay the cost of providing every loan, but only receive revenue (and thus marginal revenue) on those loans that are repaid. Whereas adverse selection can therefore be parameterized as a rotation in an insurer's marginal cost curve in insurance settings, it can be parameterized as a rotation in a bank's marginal revenue curve in lending environments. Further theoretical work analyzing how imperfect competition and adverse selection interact in the variety of markets characterized by informational asymmetry is an important area for further research.

4 Econometric Specifications

A key challenge in estimating the model of the previous section is to account for the differences in the information set of firms, banks, and us as econometricians. Specifically, we assume that there are factors observed by all of firms, banks, and us as econometricians (which we call "hard information"), factors observed by firms and banks, but not us as econometricians (which we call "soft information"), as well as factors observed by firms, but not banks or us as econometricians (which we call "private information," the correlations between which are our measures of adverse selection). We need to ensure that we can distinguish between soft and private information to properly identify the latter. We explain how we do so in what follows.

4.1 Price Prediction

4.1.1 Overview

A crucial challenge that we face when implementing our empirical model is that we only observe prices (interest rates) on loans from banks from which a firm chose to borrow. As such, for our demand model in equation (1), we must predict the prices each firm faces at all other banks offering loans in its market.²¹ By

²¹ While predicting prices would seem to necessarily introduce measurement error into our econometric framework, we believe the consequences of any such errors are likely to be small. In fact, firms might need to predict prices as well: loan applications require time and effort, so firms might form expectations for some of the prices rather than asking for a quote from each bank in their market. The model of price prediction we present below can therefore be interpreted not only as a way to recover the price that bank j charges firm i in market m in year t , but also our best estimate of *firms'* price predictions. Any measurement error then reflects differences between firms' price predictions and ours, and may therefore not be a source of econometric bias. That being said, to be conservative we approach our environment as we would with conventional measurement error problems.

contrast, for our loan use and default models, equations (2) and (3), we observe and use in the estimation the actual prices paid by firms.

One of the main determinants of loan prices is borrowers' riskiness as perceived by banks, which is predicted by lenders using a combination of "hard" and "soft" information (e.g. financial data versus a loan officer's perceptions of a borrower's creditworthiness). Whether there is an information gap between us as econometricians and banks as lenders, and whether any such gap is a problem for our analysis, depends on how well we can capture banks' actual pricing decisions.

We adopt several strategies to limit the potential extent of this problem. First, we discuss the evidence regarding how banks price loan contracts, showing that, particularly for large banks such as those in our sample, the hard information we observe in our data is the key determinant of prices. Second, as described in Section 2, we only consider the first year in which a firm borrows. In addition to allowing us to abstract from dynamic considerations in lending relationships, this focus also lessens the information gap between us and the lender, as we only consider loans from borrowers that approach a bank for the first time, and thus for whom banks are less likely to have soft information. Third, we select the best model for price prediction among a variety of alternatives based on both institutional and econometric evidence. Importantly, in our preferred specification, we exploit the fact that Italian firms frequently have multiple banking relationships, which allows us to include firm fixed effects in our price prediction model. The firm fixed effects capture any feature unobservable to the econometrician but observable to and common across banks, including soft information that we do not directly observe. Fourth, we test the statistical and economic significance of the residuals from this pricing regression in predicting default. If the residuals were correlated with defaults, our prediction model would be systematically missing a component of firms' riskiness taken into account by banks when pricing loans. In our preferred specification, we find no such correlation. Finally, while we are confident that our price predictions do not adversely impact our measurement of adverse selection in this market, we also discuss the implications for our results of inaccurate price predictions.

4.1.2 Institutional Features of Banks' Pricing Decisions

Before describing the modeling strategy we use to predict prices, we give an institutional overview of how banks determine interest rates for new borrowers in this market. The datasets we use are the main sources of hard information used by the banks in our sample. The Credit Register provides banks with information about firms' current set of loans, whereas the Centrale dei Bilanci (CB) provides banks with a detailed archive of firms' balance sheet information. As described in Cerqueiro, Degryse and Ongena (2011) using US data, banks use both hard and soft information to determine their lending policies. The importance of each factor depends on loan and borrower characteristics, as well as the nature of local lending markets and borrower-lender relationships.

To describe the institutional features of the Italian lending market, we rely on the results of a survey conducted by the Bank of Italy, summarized in Albareto et al. (2011), of over 300 Italian banks in 2007 about the organization of their lending activities. Several features of this survey are relevant for our analysis. First, it shows that larger banks, which are the ones we have in our data, tend to rely more on hard information and standardized scoring techniques that we are likely to capture with our econometric model. Second,

large banks have on average twice the number of layers of hierarchy between the top management and the branch managers compared to small banks. Therefore, as predicted by Stein (2002), large banks are likely to give less independence to branch managers in lending policies due to the difficulties both in monitoring managers' actions and in managers' ability to credibly transmit soft information about borrowers to top management. Multiple layers of hierarchy also result in large banks having shorter terms in office for branch managers, in part to avoid branch managers developing relationships with local borrowers and deriving private benefits from these. Both of these aspects limit the extent to which soft information can be used by large banks in their lending policies. Last, large banks are asked to list in order of importance the factors they consider in assessing creditworthiness of a new loan applicant. The most important factors are: (i) Financial statement data (i.e. hard information from the Centrale dei Bilanci), (ii) Credit relations with the entire system (i.e. hard information from the Credit Register), (iii) Statistical-quantitative methods, (iv) Qualitative information (i.e. bank-specific soft information codifiable as data), (v) Availability of guarantees, and (vi) First-hand information (i.e. branch-specific soft information). This ranking portrays the key role played by hard information for large banks when dealing with new borrowers. The survey also shows that small banks do rely more on soft information, although even for these it is still less important than the first two forms of hard information.

4.1.3 Price Prediction Model

Our model of price prediction is based on OLS regressions of prices on an increasing set of control variables.²² In our preferred specification, we use firm fixed effects to account for information observed by banks but not by us as econometricians. As we are interested in the degree of asymmetric information at the time a loan is granted, we continue to use only the first year in which a firm obtains credit from its main bank, just as in the estimation sample. However, unlike the estimation sample, where we only rely on the main credit line, to predict prices we use the observations for *all* firm-bank relationships in place in this first year.

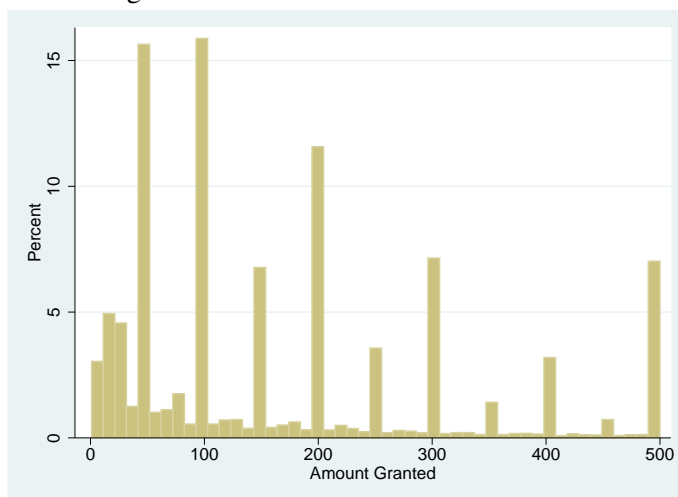
Firm-level controls in our model of price prediction include information from firms' balance sheets (measures of assets and debts) and income statements (measures of profitability and sales), the distance between a firm and a bank's nearest branch, and year, sector, area, and bank fixed effects (as well as their interactions, depending on the specification).²³ Given that we use uncollateralized credit lines, which exhibit no heterogeneity in maturity, collateral, covenants and/or other contract features important in other types of loans, the only loan-level control variable is the amount of granted credit, entered linearly or as amount dummies, depending on the specification. The decision to discretize the distribution of granted amounts of credit comes from the shape of the empirical distribution of the loans in our data, presented in Figure 1 below, which exhibits a significant number of observations around a few mass points. For example, over 40% of the loans

²² We also experimented with LASSO regressions, but it didn't improve our results as in our preferred specification we predominantly rely on fixed effects for which LASSO methods do not offer an improved fit. For alternative ways of predicting prices, see Gerakos and Syverson (2015).

²³ In our price regressions, we adopt a parsimonious definition of geographic regions in terms of four macro areas rather than 95 provinces. In specifications that interact area, bank, and year effects, these interactions increase exponentially with the number of geographic areas: from 1,313 with bank-area-year interactions to 10,802 with bank-province-year interactions. As the R-squared improves only marginally (from 0.72 to 0.76), and the adjusted R-squared does not change, we choose the smaller number of areas.

we consider are exactly €50,000, €100,000, or €200,000 and 71% are exact multiples of €50,000.

Figure 1: Distribution of Amount Granted



Note: Amount Granted is in thousands of €. The observations above €500,000 (15% of sample) have been excluded to simplify the interpretation of the graph.

The results of these regressions are shown in Table 2, with increasing numbers of controls included as one moves across the columns in the table. As anticipated earlier, the coefficient on the amount of granted credit in columns (1) and (2) is negative, showing that banks in the Italian market for small business lines of credit do not use convex price/credit-limit schedules as a screening device. Indeed, controlling non-parametrically for the amount granted using fixed effects (Column 3) confirms that interest rates monotonically decrease with loan size. The fit of the regression increases marginally when going from separate bank, year, and macro area fixed effects (Column 1) to dummies for the interaction of the three variables (Column 2), to the same with granted credit fixed effects (Column 3). The largest increase in the R-squared occurs when we introduce firm fixed effects (Column 4), indicating that fixed firm attributes observed by banks but not by us are an important element in the determination of prices. In this specification, we are able to explain over 71% of the variation in observed prices, higher than that typically obtained in the empirical banking literature.²⁴ We interpret such effects as evidence of “soft information” in this market, a feature we are careful to account for in our econometric model presented in the next subsection.

As we are concerned that banks may set prices based on unobserved firm characteristics that may be correlated with risk and that may be missing from our model of price prediction, we investigated whether unexplained variation in prices is a predictor of firms’ subsequent default. To do so, we used the residuals from each of the regressions presented in Table 2 as an explanatory variable in a regression of default on these plus the same controls used in each pricing equation.²⁵ The results of this exercise are shown in Table

²⁴ Petersen and Rajan (1994), Degryse and Ongena (2005), and Cerqueiro et al. (2011) all measure the dispersion in prices charged by banks to small and medium enterprises. Each estimates a loan-pricing model using lender, borrower, and loan-level information, finding R-squared of 14.5% (Petersen and Rajan 1994), 25% (Cerqueiro et al. 2011), and 22% (Degryse and Ongena 2005) (67% for loans over €50,000).

²⁵ We use a linear probability model for ease of interpretation, but estimates from a discrete choice regression yield similar results. The controls were the same in each column as in Table 2, apart from those in the final column: as firms default on all their

Table 2: Price Regressions

Variables	(1)	(2)	(3)	(4)
Amount Granted	-2.37	-2.39	-	-
	(0.07)	(0.07)		
50,001-100,000	-	-	-1.47	-0.92
			(0.07)	(0.09)
100,001-150,000	-	-	-2.44	-1.55
			(0.08)	(0.09)
150,001-200,000	-	-	-2.77	-1.98
			(0.10)	(0.10)
200,001-300,000	-	-	-3.18	-2.19
			(0.10)	(0.10)
300,001-400,000	-	-	-3.72	-2.63
			(0.11)	(0.10)
400,001-500,000	-	-	-3.99	-2.88
			(0.12)	(0.11)
500,001-1,000,000	-	-	-4.37	-3.06
			(0.12)	(0.10)
1,000,001-3,000,000	-	-	-5.02	-3.44
			(0.13)	(0.12)
Distance to Branch	-0.94	-0.70	-1.33	-0.40
	(0.21)	(0.20)	(0.20)	(0.30)
Constant	16.80	15.53	17.52	15.49
	(0.18)	(0.13)	(0.15)	(0.81)
Firm Controls	Yes	Yes	Yes	No
Bank FE	Yes	No	No	No
Area FE	Yes	No	No	No
Year FE	Yes	No	No	No
Bank-Area-Year FE	No	Yes	Yes	Yes
Firm FE	No	No	No	Yes
R ²	0.294	0.319	0.365	0.717
N obs.	92,596	92,596	92,596	92,602

Note: The table shows the results of OLS regressions of the interest rate (in percentage points) on a series of controls and dummies. An observation is a firm-bank. The sample only includes the first year in which a firm borrows, excluding 1988. See Table 1 for variables' definition. Firm controls include Sector and Score fixed effects, Sales, Total Assets, Net Assets, Profits, Cashflow, Leverage, and Short Term Debt. Firm controls and distance are rescaled to interpret the coefficients more easily: the linear term for Amount Granted is in €10,000, and Distance to Branch is in 100 km. Sector fixed effects group sectors into 3 categories: *Primary Sectors* (primary goods, minerals' extraction, chemicals, metals, and energy), *Manufacturing and construction* and *Commerce and Services*. Area fixed effects are based on four geographic areas of similar size in terms of population: North-West, North-East, Center, and South. Standard errors are clustered at the bank-area-year level in columns (1)-(3), and at the firm level in column (4).

Table 3: The Ability of Pricing Residuals to Predict Default

Variables	(1)	(2)	(3)	(4)
Residual	0.26 (0.03)	0.27 (0.03)	0.26 (0.03)	0.04 (0.05)
Residual SD	4.22	4.15	4.05	2.11
Default SD	0.24	0.24	0.24	0.24
1 Residual SD vs % of 1 Default SD	4.7%	4.7%	4.4%	0.3%
Firm Controls	Yes	Yes	Yes	Yes
Bank FE	Yes	No	No	No
Area FE	Yes	No	No	No
Year FE	Yes	No	No	No
Bank-Area-Year FE	No	Yes	Yes	Yes
Amount Granted FE	No	No	Yes	Yes
R ²	0.088	0.123	0.123	0.121
N obs.	35,316	35,316	35,316	35,316

Note: The table shows the results of OLS regressions of a dummy equal to one if the firm defaults on a loan, and zero otherwise, on the residuals of the pricing regressions in Table 2 and other controls. An observation is a firm. All the controls used are the same as in each respective column in Table 2, with the exception column (4) where we cannot include firm fixed effects. Residual is divided by 100 for ease of interpretation. See Table 1 for variables' definition. Standard errors are clustered at the bank-year-province level.

3. In specifications (1)-(3), we find that the residuals are estimated to have a positive and significant effect on default: a 1 standard deviation increase in the residuals is estimated to increase default probabilities by between 4.7% and 4.2% of its standard deviation. It is only in the last specification which includes firm fixed effects that we cannot reject the hypothesis that the residuals have no effect on default.²⁶

Based on this last result, we adopt the pricing model with firm fixed effects as our preferred specification. Formally, this specification assumes that the price charged to firm i borrowing from bank j in market m in year t , P_{ijmt} , takes the following form:

$$P_{ijmt} = \gamma_0 + \gamma_1 \mathcal{D}_{ijmt} + \gamma_2 \mathcal{L}_{ijmt} + \lambda_{jmt} + \omega_i^P + \tau_{ijmt}, \quad (7)$$

where ω_i^P and λ_{jmt} are firm and bank-area-year fixed effects, \mathcal{D}_{ijmt} is the distance between firm i and the nearest branch of bank j , \mathcal{L}_{ijmt} are dummies for the size of the granted loan amount, and τ_{ijmt} are prediction errors.²⁷ Using combinations of the estimated coefficients $\tilde{\gamma}_0$, $\tilde{\gamma}_1$, $\tilde{\gamma}_2$, $\tilde{\lambda}_{jmt}$, and $\tilde{\omega}_i^P$ we are able to predict prices \tilde{P}_{ijmt} offered to borrowing firms from banks they could have chosen but did not.

lines almost simultaneously, we have only one observation per firm in the default regressions and cannot therefore include firm fixed effects.

²⁶ The 95% confidence interval on the residual in this last specification is (-0.06,0.14). Furthermore, its economic magnitude is also estimated to be substantially smaller: a 1 standard deviation increase in the residuals would increase default by 0.3% of its standard deviation, or 1.4% of its mean.

²⁷ With a slight abuse of notation we use the market (province) subscript m also for the four geographic areas defined above, despite their being aggregations of provinces.

To predict prices offered to *non-borrowing* firms, we use propensity score matching: we match several borrowing firms to non-borrowing firms that are similar in observable characteristics, and then randomly assign a borrowing firm’s fixed effect, $\tilde{\omega}_i^P$, to a matched non-borrowing firm. We assign the granted loan amount to non-borrowing firms using the same approach. A detailed description of the matching model is presented in Appendix A.5. A similar method was used in Adams et al. (2009).²⁸

4.2 Econometric Model

As in the theoretical model presented in Section 3 above, let $m = 1, \dots, M$ index a market (province), $t = 1, \dots, T$ a year, $i = 1, \dots, I_{mt}$ a firm, and $j = 1, \dots, J_{mt}$ a bank in market m in year t . Y_{ijmt} is a vector of non-price firm and firm-bank specific characteristics, some observed by both banks and us as econometricians (e.g. “hard information” like a firm’s balance sheet and income statement data), and some observed by banks but not by us as econometricians (e.g. “soft information” like the perceived riskiness of the borrower). X_{jmt} is a vector of bank-market-year-specific attributes (number and share of branches in the market, years of presence in the market).

We estimate a system of three equations: demand for credit lines, amount of loan used, and default. To do so, we use a 2-step method based on maximum simulated likelihood and instrumental variables estimation (Train 2009). In the first step, using data on firms’ choices of bank, loan use, and default, we estimate the firm-level parameters across all three equations, $\eta = \{\alpha^L, \alpha^F, \eta^D, \eta^L, \eta^F\}$, the bank-level parameters for the loan use and default equations, $\beta^{LF} = \{\alpha_0^L, \alpha_0^F, \beta^L, \beta^F\}$, and the variance-covariance matrix of the errors in the system, including our key measures of adverse selection, $\Sigma = \{\sigma_D, \sigma_L, \rho_{DF}, \rho_{DL}, \rho_{LF}\}$. We also recover the bank-market-year specific constants (“mean utilities”) in the demand model ($\delta_{jmt}^D = \bar{\alpha}_0^D + X_{jmt}^{LD} \beta^D + \xi_{jmt}^D$) using the contraction method introduced by Berry et al. (1995). These serve as the dependent variables in the second step of the estimation, where we recover the price coefficient in the demand equation, α^D , as described further below.²⁹

Our estimation exercise faces two main challenges. First, we need to account for the endogeneity of prices in the three estimating equations. Second, as in the price prediction model, we need to allow for the possibility that the banks’ information set is richer than the one that we use, i.e., that we accurately account for soft information known to banks but not to us as econometricians. This problem goes beyond that of an accurate price prediction. To see this, consider a firm characteristic that is observed by the bank and it is therefore accounted for in its pricing, but that also has an independent effect on the probability to demand for credit and to default (e.g., the “reputation” of the borrower). If we do not control for such a characteristic in the demand and default equations, we could generate an omitted variable problem that would end up in the residuals of both models, possibly generating a correlation between them not due to the adverse selection we seek to measure. We explain how we address these issues in what follows.

²⁸ One potential criticism of our price prediction for non-borrowing firms is that we can match borrowers to non-borrowers that are similar in terms of observables, but the firm fixed effect we assign to non-borrowers is a combination of the borrower’s observables and unobservables, and of course we cannot observe the latter. This can induce measurement error in the price prediction. We discuss possible consequences of measurement error on our estimates in Subsection 4.2.1.

²⁹ We are unable to use the contraction method for loan use and default as we have a smaller number of observations and many zeros in the loan use and default market shares. We explain this in detail in Section 4.3.2.

4.2.1 First-stage Estimation

The goal of our estimation is to identify and estimate adverse selection, measured by correlation in the unobservable determinants of firms' demand for credit, loan use, and default. To do so, we need to consistently estimate the three equations describing firms' behavior. We discuss them each in turn.

Consider first the demand equation. Our price prediction model allows us to decompose the price charged to firm i by bank j in market m in year t as follows:

$$\begin{aligned} P_{ijmt} &= \tilde{P}_{ijmt} + \tilde{\tau}_{ijmt} \\ &= \tilde{P}_{jmt} + \tilde{\gamma}_1 \mathcal{D}_{ijmt} + \tilde{\gamma}_2 \mathcal{L}_{ijmt} + \tilde{\omega}_i^P + \tilde{\tau}_{ijmt}, \end{aligned} \quad (8)$$

where $\tilde{P}_{jmt} = \tilde{\gamma}_0 + \tilde{\lambda}_{jmt}$ is a bank-market-year specific component of the predicted price for all firms i in market m in year t , \mathcal{D}_{ijmt} is the distance between firm i and the nearest branch of bank j , \mathcal{L}_{ijmt} are dummies for the size of the granted loan amount, $\tilde{\omega}_i^P$ is the estimated firm fixed effect from our pricing regression, and $\tilde{\tau}_{ijmt}$ are fitted prediction errors.

We assume that any determinants of demand observed by banks but unobserved by us as econometricians, including soft information relevant for demand, will be taken into account by banks when setting interest rates. We also assume that such information can be summarized by an (unobserved to us) variable ω_i^D . Because a firm chooses only one main line of credit, we cannot estimate ω_i^D directly. However, we can use the firm fixed effect estimated in the price prediction by $\tilde{\omega}_i^P$ as a proxy for this demand unobservable. We can do so because any soft information influencing firm i 's demand for credit should also impact banks' pricing to firm i and will be measured by $\tilde{\omega}_i^P$.³⁰ Therefore, in the spirit of a control function approach, we relate soft information influencing demand to soft information influencing pricing through the following equation:

$$\omega_i^D = \eta_4^D \tilde{\omega}_i^P. \quad (9)$$

Using this relationship, we can define all firm-level covariates influencing demand as:

$$\begin{aligned} Y_{ijmt}^D &= \eta_1^D \mathcal{D}_{ijmt} + \eta_2^D \mathcal{L}_{ijmt} + \eta_3^D Y_i + \omega_i^D \\ &= \eta_1^D \mathcal{D}_{ijmt} + \eta_2^D \mathcal{L}_{ijmt} + \eta_3^D Y_i + \eta_4^D \tilde{\omega}_i^P, \end{aligned} \quad (10)$$

where Y_i are observable firm covariates and the second line imposes our assumption on the relationship between demand and pricing unobservables. Substituting (8) and (10) into the demand equation (1), we obtain the following demand utility:

³⁰ For example, suppose that older entrepreneurs have a lower propensity to demand credit due to private savings and this is reflected in lower prices offered by banks. Given that we do not observe entrepreneurs' ages, this effect will be captured in the price prediction model by $\tilde{\omega}_i^P$ which, in turn, can proxy for this unobserved (to us) attribute in the demand equation.

$$\begin{aligned}
U_{ijmt}^D &= \delta_{jmt}^D + \alpha^D (\tilde{P}_{jmt} + \tilde{\gamma}_1 \mathcal{D}_{ijmt} + \tilde{\gamma}_2 \mathcal{L}_{ijmt} + \tilde{\omega}_i^P + \tilde{\tau}_{ijmt}) + \eta_1^D \mathcal{D}_{ijmt} + \eta_2^D \mathcal{L}_{ijmt} + \eta_3^D Y_i + \eta_4^D \tilde{\omega}_i^P + \varepsilon_i^D + \nu_{ijmt} \\
&= \underbrace{(\delta_{jmt}^D + \alpha^D \tilde{P}_{jmt})}_{\tilde{\delta}_{jmt}^D} + \underbrace{(\eta_1^D + \alpha^D \tilde{\gamma}_1)}_{\tilde{\eta}_1^D} \mathcal{D}_{ijmt} + \underbrace{(\eta_2^D + \alpha^D \tilde{\gamma}_2)}_{\tilde{\eta}_2^D} \mathcal{L}_{ijmt} + \eta_3^D Y_i + \underbrace{(\eta_4^D + \alpha^D)}_{\tilde{\eta}_4^D} \tilde{\omega}_i^P + \varepsilon_i^D + \underbrace{\alpha^D \tilde{\tau}_{ijmt} + \nu_{ijmt}}_{\zeta_{ijmt}} \\
&= \tilde{\delta}_{jmt}^D + \underbrace{Y_{ijmt}^D \tilde{\eta}^D}_{V_{ijmt}^D} + \varepsilon_i^D + \zeta_{ijmt},
\end{aligned} \tag{11}$$

where $Y_{ijmt}^D = \{\mathcal{D}_{ijmt}, \mathcal{L}_{ijmt}, Y_i, \tilde{\omega}_i^P\}$ and $\tilde{\eta}^D = \{\tilde{\eta}_1^D, \tilde{\eta}_2^D, \eta_3^D, \tilde{\eta}_4^D\}$.

Three aspects of equation (11) merit discussion. First, we cannot exclude that any of the determinants of prices also affects demand and, as such, we have included in the demand equation all the variables that we have used to predict prices. For example, the distance between a firm and a bank can influence the price that the bank offers but also the likelihood that a firm borrows from the bank, conditional on the price. The parameters that we estimate on these firm- and firm-bank covariates are therefore a mixture of the direct effect of that covariate on demand and an indirect effect through prices. We denote these composite effects $\tilde{\eta}_1^D$, $\tilde{\eta}_2^D$, and $\tilde{\eta}_4^D$. Given a consistent estimate of α^D from the second-stage estimation (described below), we can then back out the (demand-only) parameters, η_1^D , η_2^D , and η_4^D . Second, the error term is a composite of the structural demand and predicted price errors, $\zeta_{ijmt} = \alpha^D \tilde{\tau}_{ijmt} + \nu_{ijmt}$, which we assume is distributed as a Type I Extreme Value.³¹ Third, because the demand price parameter, α^D , does not enter equation (11) independently except as part of the composite parameters of the structural error term, we cannot identify it in the first-stage estimation. Instead, we estimate it in the second-stage IV procedure that uses variation in average prices at the bank-market-year level \tilde{P}_{jmt} .³²

Based on these assumptions, the probability that borrower i in market m in year t chooses bank j is given by:

$$\Pr_{ijmt}^D = \int \left[\frac{\exp \left(\tilde{\delta}_{jmt}^D \left(X_{jmt}^D, \tilde{P}_{jmt}, \xi_{jmt}^D, \bar{\alpha}_0^D, \alpha^D, \beta^D \right) + V_{ijmt}^D \left(Y_{ijmt}^D, \sigma_D, \tilde{\eta}^D \right) \right)}{1 + \sum_{\ell} \exp \left(\tilde{\delta}_{\ell mt}^D \left(X_{\ell mt}^D, \tilde{P}_{\ell mt}, \xi_{\ell mt}^D, \bar{\alpha}_0^D, \alpha^D, \beta^D \right) + V_{i\ell mt}^D \left(Y_{i\ell mt}^D, \sigma_D, \tilde{\eta}^D \right) \right)} \right] f(\varepsilon_i^D) d\varepsilon_i^D, \tag{12}$$

where $f(\varepsilon_i^D)$ is the density of ε_i^D , and $\tilde{\delta}_{jmt}^D$ are the bank-market-year specific constants that we recover using the contraction method of Berry et al. (1995).

Estimation of the equations (2) and (3) describing loan use and default respectively is more standard, as in this case we can directly use actual prices. For these equations, all parameters are estimated directly in

³¹ This assumption rules out any effect of measurement error on our estimates of adverse selection, but our procedure is robust to the possibility that it does not hold. First, as discussed above, firms themselves need to predict prices, so measurement error, if any, is likely to be small. Second, even in the case that measurement error is not absorbed in the logit error, we show in Appendix A.6 that, as long as the residuals in the pricing regression of equation (7) are uncorrelated with default unobservables, it can at most result in a conservative estimate of the degree of adverse selection.

³² Note that, given that the correlation coefficients that measure adverse selection are estimated in the first stage, they are not influenced by this second-stage IV procedure.

the first stage. As we explain in detail in Section 4.3.2, we address endogeneity of prices and unobserved determinants (to us but not to banks) of loan use and default using various fixed effects and a control function approach. The probability of observing a utilization of L conditional on borrowing ($D = 1$), is given by:

$$\begin{aligned}\Pr_{ijmt,L|D=1}^L &= E [\Pr (L_{ijmt} = \alpha_0^L + X_{jmt}^L \beta^L + \alpha^L P_{ijmt} + Y_{ijmt}^L \eta^L + \varepsilon_i^L | \varepsilon_i^D) | D = 1] \\ &= \int \phi_{\varepsilon_i^L | \varepsilon_i^D} \left(\frac{L_{ijmt} - \alpha_0^L - X_{jmt}^L \beta^L - \alpha^L P_{ijmt} - Y_{ijmt}^L \eta^L - \tilde{\mu}_{\varepsilon_i^L | \varepsilon_i^D}}{\tilde{\sigma}_{\varepsilon_i^L | \varepsilon_i^D}} \right) f(\varepsilon_i^D | D = 1) d\varepsilon_i^D,\end{aligned}\quad (13)$$

where:

$$\varepsilon_i^L | \varepsilon_i^D \sim N \left(\underbrace{\frac{\sigma_L}{\sigma_D} \rho_{DL} \varepsilon_i^D}_{\tilde{\mu}_{\varepsilon_i^L | \varepsilon_i^D}}, \underbrace{\sigma_L^2 (1 - \rho_{DL}^2)}_{\tilde{\sigma}_{\varepsilon_i^L | \varepsilon_i^D}^2} \right),$$

ϕ is the PDF of a standard normal distribution, and $f(\varepsilon_i^D | D = 1)$ is the density of ε_i^D conditional on borrowing. Finally, the probability of default ($F = 1$) conditional on borrowing and loan utilization is:

$$\Pr_{ijmt,F=1|D=1,L}^F = \int \Phi_{\varepsilon_i^F | \varepsilon_i^D, \varepsilon_i^L} \left(\frac{\alpha_0^F + X_{jmt}^F \beta^F + \alpha^F P_{ijmt} + Y_{ijmt}^F \eta^F - \tilde{\mu}_{\varepsilon_i^F | \varepsilon_i^D, \varepsilon_i^L}}{\tilde{\sigma}_{\varepsilon_i^F | \varepsilon_i^D, \varepsilon_i^L}} \right) f(\varepsilon_i^D | D = 1) d\varepsilon_i^D,\quad (14)$$

where:

$$\begin{aligned}\varepsilon_i^F | \varepsilon_i^D, \varepsilon_i^L &\sim N \left(\underbrace{A \varepsilon_i^D + B \varepsilon_i^L}_{\tilde{\mu}_{\varepsilon_i^F | \varepsilon_i^D, \varepsilon_i^L}}, \underbrace{\sigma_F^2 - (A \rho_{DF} + B \rho_{LF})}_{\tilde{\sigma}_{\varepsilon_i^F | \varepsilon_i^D, \varepsilon_i^L}^2} \right), \\ A &= \frac{\rho_{DF} \sigma_L^2 - \rho_{LF} \rho_{DL}}{\sigma_D^2 \sigma_L^2 - \rho_{DL}^2}, \\ B &= \frac{-\rho_{DF} \rho_{DL} + \rho_{LF} \sigma_D^2}{\sigma_D^2 \sigma_L^2 - \rho_{DL}^2}.\end{aligned}$$

The joint estimation of these three choice equations through maximum simulated likelihood delivers estimates of the parameters in η , β^{LF} and Σ , based on the following log-likelihood function:

$$\log L = \sum_i d_{ijmt} \{ \log(\Pr_{ijmt}^D) + \log(\Pr_{ijmt}^L) + f_{ijmt} \log(\Pr_{ijmt}^F) + (1 - f_{ijmt}) \log(1 - \Pr_{ijmt}^F) \}, \quad (15)$$

where d_{ijmt} is the dummy for the choice by firm i of bank j in market m in year t , and f_{ijmt} is the dummy identifying its default.³³

³³ We use 100 Halton draws for simulation. According to Train and Winston (2007), 100 Halton draws achieve greater accuracy

4.2.2 Second-stage Estimation

We use instrumental variables estimation to recover the structural parameters in the demand equation, including the demand price coefficient, α^D . The contraction method in the first-stage of the demand estimation finds the bank-market-year-specific values, $\tilde{\delta}_{jmt}^D$, that equate predicted market shares \widehat{S}_{jmt}^D to actual market shares S_{jmt}^D . This iterative process is defined by:

$$\tilde{\delta}_{jmt}^{D,r+1} = \tilde{\delta}_{jmt}^{D,r} + \ln \left(\frac{S_{jmt}^D}{\widehat{S}_{jmt}^D(\tilde{\delta}_{jmt}^{D,r})} \right), \quad (16)$$

where r is the iteration number, the predicted market shares are $\widehat{S}_{jmt}^D = \sum_i \text{Pr}_{ijmt}^D / I_{mt}$, where I_{mt} are the number of firms in market m in year t . Once recovered, we define these bank-market-year specific values as $\widetilde{\delta}_{jmt}^D$. These constants contain the bank-market-year covariates, X_{jmt}^D , as well as the bank-market-year specific component of predicted prices, \widetilde{P}_{jmt} . We use these estimated constants as dependent variables in an IV regression on X_{jmt}^D and \widetilde{P}_{jmt} , using cost shifters as instruments:

$$\widetilde{\delta}_{jmt}^D = \bar{\alpha}_0^D + \alpha^D \widetilde{P}_{jmt} + X_{jmt}^D \beta^D + \xi_{jmt}^D, \quad (17)$$

where ξ_{jmt}^D is the structural econometric error term. A detailed explanation of our identification strategy for all the equations follows.

4.3 Identification

We address potential endogeneity bias in the price coefficients in our three estimating equations using instrumental variables in the second stage of the demand equation estimation and a control function approach in the loan use and default equation estimation.

4.3.1 Demand Instruments

The unobserved attributes ξ_{jmt}^D in our second-stage IV demand estimation (equation (17)) can be interpreted as borrowers' (firms') unobserved valuation of a bank's brand, quality, or stability which are known to the bank and therefore affect the bank's interest rates in market m in year t . For example, ξ_{jmt}^D could capture a bank's reputation for offering valuable and helpful assistance to its borrowers in their business projects, unobserved to us as econometricians but known to firms. Moreover, because we predict the prices that enter demand, ξ_{jmt}^D could also include market-specific measurement error in prices. As a first step to address these endogeneity problems, we include bank, market, and year fixed effects in our second-stage demand estimation. Bank fixed effects capture bank characteristics that do not vary by market and year that might otherwise influence pricing, market fixed effects control for time-invariant market characteristics that affect all banks within a market equally, and year fixed effects control for any macroeconomic changes affecting

in mixed logit estimations than 1,000 pseudo-random draws.

lending in Italy over time. We can therefore rewrite equation (17) as:

$$\widetilde{\delta}_{jmt}^D = \bar{\alpha}_0^D + \alpha^D \widetilde{P}_{jmt} + X_{jmt}^D \beta^D + \xi_j^D + \xi_m^D + \xi_t^D + \Delta \xi_{jmt}^D, \quad (18)$$

where $\xi_j^D, \xi_m^D, \xi_t^D$ are bank, market, and year fixed effects, and $\Delta \xi_{jmt}^D$ represents market-year specific deviations from the national mean valuation of the demand unobservable for each bank.

To account for the potential correlation between interest rates and these bank-market-year-specific errors, we use information about household deposits at the local bank level as instruments. Households' bank deposits represent an important source of capital that banks invest in loan products. The high degree of autonomy that local branch managers have in their lending decisions, as described in Albareto et al. (2011), implies that the ability of local branches to collect deposits at more favorable conditions affects the loan conditions they can offer to borrowers. The validity of the exclusion restriction rests on the fact that deposit conditions are determined in a market with different buyers (households, not firms) for products with different demand characteristics than the loan market (bank accounts and payment services, not lines of credit). These market differences, and the inclusion of bank fixed effects to account for any effects of a bank's "brand" in both markets, as well as market and year fixed effects, lessen concerns that variation in deposits at the bank-region-year level are correlated with the error term in demand for banks' loans, $\Delta \xi_{jmt}^D$.

Formally, we use information on interest rates on household deposits, the euro value of collected deposits, and the number of deposit accounts at the bank-region-year level as demand instruments.³⁴ First, higher interest rates on deposit are likely associated with higher loan rates, as higher deposit rates mean it is more costly for a bank to raise funds. A higher value of deposits is likely associated with lower loan rates, as the bank has higher availability of funds. Finally, more deposit accounts are likely associated with higher loan rates, because managing deposit accounts involves fixed costs, so that, for given value of deposits, a higher number of accounts implies a higher total cost. These predictions are verified in our first-stage estimates, reported in Appendix A.7. We also find that the instruments are jointly significant in explaining bank-market-year level interest rates. Finally, as expected, instrumenting for price increases the estimate of the elasticity of demand.

4.3.2 Loan Use and Default Instruments

We use a different set of instruments for prices in the loan use and default equations. In this case, endogeneity comes from potential heterogeneity (observed by banks) in firms' need for external funds and likelihood of repayment which may be correlated with interest rates P_{ijmt} . In particular, banks may price based on soft information, unobserved to us as econometricians, which determines loan use and default. To correct for any potential bias, we need price variation that is orthogonal to a firms' unobserved riskiness and need for external funds.

We adopt a two-part strategy to address this potential bias. First, we include bank and market-year fixed effects, which capture, respectively, market- and time-constant bank characteristics and any market-year-

³⁴ We only have information about deposits at the bank-region-year level, where regions are geographic areas that include on average 5 markets (provinces).

specific unobservables. Second, following Nevo (2001) and Hausman and Taylor (1981), we use prices in other markets as an instrument in the loan use and default equations. The logic of this instrument is that banks face cost shocks that are common across markets and are reflected in their interest rates. For example, idiosyncratic costs of obtaining funds for lending from the interbank market are common across markets within a bank-year. Furthermore, conditional on the large set of controls we include in our estimating equations, including the bank and market-year fixed effects described above, we think it unlikely that prices charged by banks in other markets are correlated with a firm’s decision regarding the amount of its loan to use or to default.³⁵ Thus we instrument the prices charged by a bank j in a market m in year t with the average of the prices that the same bank charges in all the other markets in the same year. The first-stage regression reported in Appendix A.7 shows that prices in other markets are indeed relevant: an increase in interest rates in other markets implies statistically significantly higher interest rates in a bank’s home market.

Given the nonlinearity of the first-stage estimating equations, we address potential endogeneity concerns in the loan use and default equations using a control function approach (Train 2009). We regress observed interest rates P_{ijmt} on the same observables that we use for loan use and default, as well as the instrument (prices in other markets). We then use the residuals from this pricing regression \hat{u}_{ijmt} as controls in the utility from choosing how much of a loan to use and the utility from defaulting. In practice, this amounts to including \hat{u}_{ijmt} in the estimating equations (13) and (14).

This control function approach has two benefits. First, conditional on the residuals, \hat{u}_{ijmt} , and other controls, the remaining price variability is attributable to variation in the instrument (prices in other markets) that is orthogonal to the unobserved heterogeneity in the propensity to draw and to default on the credit line, resolving any endogeneity concerns. Second, it also resolves any concerns about soft information that is observed to banks but not to us as econometricians. As long as any soft information used by bank j to price the loan offered to firm i is orthogonal to prices in other markets (as is likely), the control function residual, \hat{u}_{ijmt} , will include this soft information, preventing it from biasing our estimates of adverse selection.

5 Estimation and Results

The observables that we include in firms’ demand, loan use, and default equations are the firm and bank characteristics summarized in Table 1. These are selected based on statistical testing and insights from the literature. Among firm characteristics, we control for different measures of firm size, in the form of assets and sales, but also for measures of firms’ current performance, in terms of profits and cash flow (Albaretto et al. 2011). We also control for other specific forms of finance that firms have access to, such as credit from suppliers (trade debit), as well as for the ratio of intangible assets to total assets, a measure of asset

³⁵ The common concern using prices in other markets as an instrument in demand estimation is that there can be unobserved demand shocks that are correlated across markets, for example national advertising strategies, that invalidate the exclusion restriction. It is for this reason that we rely on household deposits instead as instruments in the demand equation. Such issues are unlikely to be important for firms’ loan use and default decisions, however. These choices are made after a loan is secured and are likely to be determined by firms’ financial situation, not whether and how much their particular bank advertises.

pledgeability (Petersen and Rajan 1995). We include the firm’s age and the distance between the city council where the firm is located and the city council where the closest branch of each bank in the firm’s choice set is located (Degryse and Ongena 2005). We also include fixed effects for the Score, the firm’s industrial sector (primary, secondary, or tertiary), the granted loan amount, and various combinations of bank, market, and year dummies.³⁶ Finally, we include the predicted interest rate in the demand equation, and the actual interest rate in the loan use and default equations. Among bank characteristics, we include the number and the share of branches that a bank has in a market-year, as well as the number of years that it has been in the market. These variables capture the experience a bank has in a market as well as the density of its network of branches with respect to its competitors, both of which can be factors influencing firms’ decisions.

We estimate our structural model on a subset of the full data for computational and institutional reasons. Following Cohen and Mazzeo (2007), we define a local banking market as the geographic area outside which borrowers don’t choose lenders, meaning the markets should not be too small, and also within which there are no overlapping markets, meaning that markets should not be too big. Based on this and on our assumption that the choice set of a borrowing firm is given by the banks actively lending in its market, out of a sample of 977 market-year combinations in our original data, we drop the first and last decile of the market size distribution.³⁷ This leaves us with 702 market-year combinations in our estimation sample.

5.1 Results

The estimates of the structural model are presented in Table 4. The three columns of results refer to the demand, loan use, and default equations, respectively. The top panel in the table shows the effect of firm characteristics, the middle panel the effect of bank characteristics, and the bottom panel the covariance matrix of unobservable determinants of demand for credit, loan use, and default. Our measures of adverse selection – the estimated correlation between the demand and default unobservables (ρ_{DF}) and the loan use and default unobservables (ρ_{LF}) – are reported in a box at the bottom of the table.

We find that, as expected, higher interest rates have a negative impact on demand for loans from a given bank. Furthermore, as described in Section 4.3 and reported in Appendix A.7, instrumenting for price makes demand more elastic. Using these estimates, we calculate the mean own- and cross-price elasticities for the five largest banks in the sample and find that a 10% increase in interest rates reduces a bank’s own market share by slightly more than 10% and increases competitor banks’ shares by slightly less than 1%. Firms with more cash flow and trade debit are less likely to borrow, but firms with more assets, profits, and sales are more likely to borrow. Older firms are also more likely to borrow. Firms tend to favor banks that are closer, and with a higher number and share of branches in the market. They are also more likely to

³⁶ As explained in Section 4.2, we estimate the demand model in two steps, and the loan use and default models in one step. In the first step we estimate jointly some parameters of demand and all parameters of loan use and default. In this stage we can rely on variation across firms as well as variation across markets and time and therefore can include a richer array of bank, market, and year controls (bank-market-year fixed effects for demand and bank and market-year fixed effects for loan use and default). In the second step, we estimate the remaining parameters of the demand model, including separate bank, market, and year fixed effects along with other bank-market-year level controls.

³⁷ To improve the convergence of the contraction mapping, we also eliminated a few markets with zero or near-zero market shares of the outside option.

choose loans from older banks.

The estimated effects of these same covariates on loan use follow the same pattern as for demand for most of the relevant variables. As far as the default probability is concerned, it increases with interest rates. Firms with more trade debit face a higher rate, while the opposite is true for cash flow and sales. Note that, given our control function approach, the price variability used to estimate the price coefficient is orthogonal to both observed or unobserved firm characteristics. As such, it measures how the default probability increases when a firm faces an exogenously higher interest rate. Following Adams et al. (2009), we interpret its positive coefficient as evidence of moral hazard. The marginal effect of price on default implies that a one standard deviation increase in interest rate (a 4.6 percentage point increase) translates into a 0.12 percentage points increase in default probability, or 2% of the average default rate.

The box in the third panel of Table 4 presents our primary parameters of interest: estimates of adverse selection between both demand and default and between loan use and default. The correlation coefficient between unobservables in the demand and in the default equations is 0.16, that between unobservables in the loan use and in the default equations is 0.14, and both are highly significant. In line with the insight of Stiglitz and Weiss (1981), we find that firms with an unexpectedly high propensity to borrow are also unexpectedly more likely to default. We also find that firms that use an unexpectedly high amount of their granted credit are also unexpectedly more likely to default. We present counterfactual exercises that help interpret the economic significance of these estimated magnitudes after discussing the fit of the model and alternative explanations for our interpretation of these correlations as adverse selection.

5.2 Fit of the Model

In Table 5 we provide some descriptive statistics on the fit of the model. We choose to focus on the main objects of interest of the model: the predicted demand probabilities, amounts of loan used, default probabilities, effective marginal costs, and effective markups.³⁸ We recover each bank's borrower-specific marginal cost using the pricing equation (6) as follows:

$$\widehat{MC}_{ijmt} = \tilde{P}_{ijmt} \left[1 - \widehat{F}_{ijmt} + \widehat{F}'_{ijmt} \widehat{\mathcal{M}}_{ijmt} \right] - \frac{(1 - \widehat{F}_{ijmt}) \widehat{\mathcal{M}}_{ijmt}}{1 - \widehat{F}_{ijmt} + \widehat{F}'_{ijmt} \widehat{\mathcal{M}}_{ijmt}}, \quad (19)$$

where $\widehat{\mathcal{M}}_{ijmt} = -\widehat{Q}_{ijmt}/\widehat{Q}'_{ijmt}$ is the predicted markup for bank j on a loan to firm i , \widehat{F}_{ijmt} is the predicted expected default probability of firm i on that loan, and \widehat{F}'_{ijmt} is its derivative with respect to the price. Given \widehat{MC}_{ijmt} , we calculate effective marginal costs and effective markups as shown in equation (6).

Table 5 shows that the model fits the mean of the data well, but predicts less variation than that in the data. Despite the relative inelastic own-firm demand, we estimate effective markups of 71 basis points (.71 percentage points, or 5% of the average interest rate of 14.48 percent).³⁹ To help validate our model estimates, we investigated whether our predicted marginal costs correlate with information about banks'

³⁸ When predicting demand probabilities we set the econometric error ($\Delta \xi_{jmt}^D$) to zero as otherwise we would perfectly predict market shares regardless of our parameter estimates.

³⁹ We find that for 957 firms out of 36,520 (2.6%), there is at least one firm-bank observation with a negative predicted marginal cost. We omit those firms from this table and from the counterfactual analysis.

Table 4: Structural Estimates

			Demand	Loan Use	Default	
Firm Level	Price	Interest Rate	-1.45 (0.62)	-0.01 (0.00)	1.06 (0.02)	
		Assets	Total Assets	5.84 (0.08)	0.09 (0.00)	-0.04 (0.03)
	Intangible/Total Assets		-0.82 (0.05)	-0.01 (0.01)	0.08 (0.05)	
	Profitability	Profits	1.12 (0.05)	0.01 (0.00)	0.03 (0.02)	
		Cash Flow	-0.93 (0.05)	-0.05 (0.00)	-0.12 (0.02)	
		Sales	7.16 (0.07)	-0.01 (0.00)	-0.34 (0.04)	
	Debt	Trade Debit	-3.44 (0.05)	-0.04 (0.00)	0.12 (0.02)	
		Others	Firm's Age	0.23 (0.03)	0.00 (0.00)	0.02 (0.03)
	Distance to Branch		-1.22 (0.03)	-0.01 (0.01)	-0.04 (0.06)	
	Bank Level		Number of Branches	4.38 (0.15)	0.01 (0.00)	0.03 (0.02)
			Share of Branches	0.53 (0.04)	-0.07 (0.01)	-0.27 (0.10)
			Years in Market	0.06 (0.03)	0.01 (0.00)	-0.13 (0.02)
First-stage Fixed Effects		Score	Yes	Yes	Yes	
		Sector	Yes	Yes	Yes	
		Loan Amount	Yes	Yes	Yes	
		Bank-Market-Year	Yes	No	No	
		Bank	No	Yes	Yes	
		Market-Year	No	Yes	Yes	
		Observations	506,230	25,351	25,351	
Covariance Matrix (Σ)	Adverse Selection	$\sigma_D = 0.34$ (0.00)		$\sigma_L = 0.30$ (0.00)	$\sigma_F = 1$	
		$\rho_{DL} = 0.10$ (0.00)				
		$\rho_{DF} = 0.16$ (0.00)		$\rho_{LF} = 0.14$ (0.00)		

Note: All coefficients are estimated in the first stage, with the exception of the Interest Rate, the Number of Branches, the Share of Branches and the Years in Market for the demand equation, that are estimated in the second stage. Second stage fixed effects, only for the demand equation, are at the bank, market and year level. See Table 1 for variables' definition. Standard errors are in brackets. First stage standard errors are calculated by the inverse of the Information matrix, obtained providing the solver with analytical gradient and hessian. Second stage standard errors are computed with 200 bootstrap replications.

Table 5: Descriptives on Model Fit

Variables	Nobs	Mean	Std. Dev.
Actual Demand	506,230	7.21	25.87
Predicted Demand	506,230	7.21	15.16
Actual Loan Use	25,351	246.06	444.75
Predicted Loan Use	25,351	245.42	297.34
Actual Default	25,351	5.83	23.44
Predicted Default	25,351	5.82	9.74
Predicted Price	452,594	14.48	4.54
Predicted Effective Marginal Cost	452,594	13.77	4.57
Predicted Effective Markup	452,594	0.71	0.48

Note: All variables are predicted at the firm level. Actual Demand is a dummy equal to one if the firm takes the loan from a bank and zero otherwise, multiplied by 100. Predicted Demand is the respective demand probability from our model, again multiplied by 100. Actual and Predicted Loan Use refer to loan use in the data and in our model, in thousands of €. Actual Default is a dummy equal to one if the firm defaults and zero otherwise, multiplied by 100 for ease of interpretation. Predicted Default is the respective default probability from our model, again multiplied by 100. Predicted Price is the interest rate predicted by our model as of equation (6). Predicted Effective Marginal Cost and Predictive Effective Markup are computed based on the two right hand side terms in equation (6), with Effective Markup being the negative of the second term. Predicted Price, Effective Marginal Cost, and Effective Markup have fewer observations than Actual and Predicted Demand as there is no predicted price for the outside good.

costs to which we have access. We find that conditional on bank, region, and year fixed effects, our estimated marginal costs are negatively correlated with the value of deposits, positively correlated with the number of deposit accounts, and positively correlated (but with no statistical significance) with deposit interest rates. These support our use of these variables as instruments in our second-stage demand estimation.

5.3 Alternative Explanations for Positive ρ_{DF} and ρ_{LF}

We interpret a positive correlation between unobserved determinants of the demand for credit and firms' decision to default, ρ_{DF} and ρ_{LF} , as evidence of adverse selection, but alternative explanations for such a correlation are possible. One alternative is that it arises because of agency issues between equity and debt holders. Jensen and Meckling (1976) show that debt financing gives rise to agency costs, causing firms to make decisions that don't serve shareholders' interests. In our context, this would imply risk-shifting by firms with more debt, which in turn would decrease the quality of firms' projects and increase their default probabilities. More generally, firms and banks could adjust their financing structure as asymmetric information varies. Another possible explanation follows from Myers (1977), who argues that firms with more debt are more likely to run into debt overhang, declining to fund good projects and increasing their default probabilities.

While theoretically sound, we believe that these alternative explanations are unlikely to hold in our data. With respect to agency issues and risk-shifting, we rely on a sample of small and medium enterprises (SMEs), most of which are owned and controlled by an individual or a family.⁴⁰ In these firms, ownership

⁴⁰ Bugamelli, Cannari, Lotti and Magri (2012) show that 85.6% of Italian SMEs are family businesses, and in 66.3% of these

is concentrated and bank debt is, with trade debit, the main source of finance.⁴¹ Typically, owners of family firms hold a large portion of their overall wealth in the firm itself (Moskowitz and Vissing-Jorgensen 2002) and, as a consequence, undertake less risky, more conservative projects than a well-diversified owner or an external CEO (Michelacci and Schivardi 2013). Moreover, in all our regressions we control for various indicators of the incentives to engage in risk-shifting, including net worth, cash flows, profits, and trade debit. We feel that excessive risk-taking is therefore unlikely to arise due to firms' financial structure.

With respect to theories of debt overhang, we look only at firms' first entry into credit markets, when firms have relatively low levels of debt.⁴² Furthermore, the solution proposed by Myers (1977) to suboptimal investments caused by agency costs is to shorten debt maturity, concluding that permanent debt capital is best implemented by rolling over short-maturity debt claims. In our setting, credit lines can be closed at any time by the bank, thus the specific loans we consider are less likely to give rise to these agency issues.

6 Counterfactuals

We run three counterfactual policy experiments to quantify the effects of adverse selection, as well as to understand the relationship between adverse selection and imperfect competition. In the first counterfactual, we analyze the impact of an increase in adverse selection on equilibrium prices, demand, loan use, and default. Its goal is to quantify the economic implications of our estimates of adverse selection in terms of relevant outcome variables. We examine how prices, demand, loan use, and default vary as we increase the correlation between unobserved determinants of demand, loan use, and default, as well as how these effects change with banks' market power.

In our second counterfactual exercise we increase banks' marginal costs under our estimated baseline level of adverse selection and analyze its consequences for equilibrium prices, demand, loan use, and default. Increases in the cost of funding can be related to situations of distress in financial markets, to changes in monetary policy, and to macro-prudential policies imposing tougher capital requirements. With this exercise we seek to identify how adverse selection and imperfect competition, and their interactions, affect the transmission mechanisms of banks' higher capital costs to lending activity.

With the last counterfactual, we investigate further the interaction between adverse selection and imperfect competition by simulating a merger between the two largest banks in each local market, both under our estimated baseline and a higher level of adverse selection.

cases the family also manages the firm, compared to 25.8% of such cases in France, 28.0% in Germany, 35.5% in Spain, and 10.4% in the UK.

⁴¹ Equity markets were very underdeveloped in Italy during the years of our data (Demekas, Potter and Pradhan 1995). There were less than 400 firms on the stock market and SMEs were very unlikely to list. Furthermore, Barba Navaretti, Bugamelli, Schivardi, Altomonte, Horgos and Maggioni (2011) show that the share of firms financed by venture capital is between 0.35% and 0.52%, and the bond market for small firms is non-existent due to legal restrictions on bond issuances for SMEs.

⁴² The average leverage of new borrowers is below 50%, while that of borrowers beyond the first year is above 55%.

6.1 The Effects of Increased Adverse Selection

In our first counterfactual exercise, we analyze the consequences for market outcomes of an increase in adverse selection, implemented by doubling the estimated correlation coefficients between unobserved determinants of demand, loan use, and default, ρ_{DF} and ρ_{LF} .⁴³ We first predict market outcomes with greater adverse selection and then investigate whether and how the changes that we observe from our (estimated) baseline levels are correlated with measures of banks’ market power. As is typical in counterfactual exercises, we assume that marginal costs remain the same in the counterfactual scenario, although “effective marginal costs” change with changes in counterfactual default rates. We re-calculate firms’ demand probabilities, loan use, and default probabilities with the counterfactual level of adverse selection, and derive new equilibrium prices as:

$$\bar{P}_{ijmt} = \frac{\widehat{MC}_{ijmt}}{1 - \bar{F}_{ijmt} + \bar{F}'_{ijmt}\bar{\mathcal{M}}_{ijmt}} + \frac{(1 - \bar{F}_{ijmt})\bar{\mathcal{M}}_{ijmt}}{1 - \bar{F}_{ijmt} + \bar{F}'_{ijmt}\bar{\mathcal{M}}_{ijmt}}, \quad (20)$$

where \widehat{MC}_{ijmt} is each bank’s borrower-specific marginal cost from equation (19), and \bar{F}_{ijmt} and $\bar{\mathcal{M}}_{ijmt} = -\bar{Q}_{ijmt}/\bar{Q}'_{ijmt}$ are the new equilibrium default probabilities and markups in the counterfactual setting, with equilibrium quantities \bar{Q}_{ijmt} given by the counterfactual demand probabilities (\bar{Q}_{ijmt}^D) multiplied by the counterfactual amount of loan used (\bar{Q}_{ijmt}^L).⁴⁴ We then compute the changes in these variables with respect to the same variables predicted by the model at our estimated parameters, what we call the “baseline” case. We report these changes at the firm-market-year level (i.e., aggregated across banks for each firm), both in percentage points (P.P.) and as a change relative to predicted baseline levels (%).⁴⁵

The first group of columns in Table 6 shows that a doubling in the correlation between unobserved determinants of demand, loan use, and default (ρ_{DF} and ρ_{LF}) causes a 1.87 percentage points (or 12.9%) increase in the average interest rates offered by banks. There is substantial heterogeneity in these price responses, however, with interest rates to some firms more than doubling and many not changing at all.⁴⁶ Demand probabilities decline by 1.28 percentage points (4.1%), as does (slightly) loan use (by €700 or 0.3%). On average, higher adverse selection significantly worsens banks’ pool of borrowers, more than doubling average default probabilities (from 5.5 to 11.4 percent).⁴⁷

⁴³ We have experimented with larger and smaller changes in these correlation coefficients and obtain results scaled in proportion to the change in the ρ ’s. We increase both correlation coefficients as they are both measures of adverse selection, and from a policy perspective we are more interested in their combined effect than in their relative importance.

⁴⁴ In some cases, our model predicts negative loan use. Given that we compute growth rates, to avoid loosing such observations in those cases we set loan use to 1€.

⁴⁵ The percentage variation in prices is measured as $\Delta P_{ijmt} = 100 \times \frac{\bar{P}_{ijmt} - \tilde{P}_{ijmt}}{\tilde{P}_{ijmt}}$, where \bar{P}_{ijmt} is the new counterfactual price and \tilde{P}_{ijmt} is the price predicted by our model in the baseline case. We define similarly percentage variations in demand probabilities and loan use. The percentage point change in prices is measured as $\Delta P_{ijmt} = \bar{P}_{ijmt} - \tilde{P}_{ijmt}$, whereas for demand probabilities it is measured as $\Delta Q_{ijmt}^D = 100 \times (\bar{Q}_{ijmt}^D - \tilde{Q}_{ijmt}^D)$, and similarly for default. We only analyze and report percentage point variation for default, given that the very small baseline probabilities can give rise to very large percentage changes that are harder to interpret.

⁴⁶ The large increase in some rates represents an implicit way for the model to predict that a bank “rejects” a risky borrower by ensuring the firm has both a low demand probability and a low loan use if a loan is taken. A richer dataset including loan approval information would allow us to address this aspect of lending decisions.

⁴⁷ In Appendix Figure A.5, we show that increased adverse selection predicts more mass in both tails of the default distribution.

Table 6: Summary Statistics for Counterfactual Changes in Outcomes
for Higher Adverse Selection and Higher Marginal Costs

Variables	Nobs	Higher Adverse Selection		Higher Marginal Costs	
		Mean	Std. Dev.	Mean	Std. Dev.
P.P. Price Variation	35,563	1.87	5.76	1.78	3.50
% Price Variation	35,563	12.9	40.9	11.0	18.6
P.P. Demand Variation	35,563	-1.28	6.17	-1.47	4.96
% Demand Variation	35,563	-4.1	14.5	-5.0	10.0
Loan Use Variation	35,563	-0.7	2.4	-0.7	1.4
% Loan Use Variation	35,563	-0.3	2.0	-0.6	2.2
P.P. Default Variation	35,563	5.84	22.37	1.84	4.58

Note: These variables are average changes at the firm-market-year level. See footnote 45 for variable definitions. Loan Use changes are in thousands of €. Higher Adverse Selection corresponds to a doubling of the estimated values for ρ_{DF} and ρ_{LF} from Table 4; Higher Marginal Costs corresponds to a 5% (approximately 70 basis points) increase in banks' marginal costs.

We are particularly interested in how these predicted changes in prices, quantities (both demand and loan use), and default vary with banks' market power. To do so, we use as a measure of market power the fitted "effective markups" facing each firm at our estimated baseline levels of adverse selection. We then run OLS regressions of our predicted changes in firms' interest rates, demand probabilities, loan use, and default probabilities on these measures of banks' market power. We aggregate this information to the firm level (i.e. across banks for each firm).⁴⁸ The top panel in Table 7 shows that higher effective markups are negatively correlated with changes in interest rates and default probabilities, and positively correlated with the changes in demand probabilities and loan use. In particular, we find that a one standard deviation increase in banks' average effective markup is associated with a 6 percentage points lower change in price (ΔP_{ijmt}), a 2.2 percentage points higher change in the probability of taking a loan (ΔQ_{ijmt}^D), a 0.2 percentage points higher change in loan use (ΔQ_{ijmt}^L), and a 2.1 percentage points lower change in default probability (ΔF_{ijmt}).

These results confirm the intuition of the effects described in Section 3.3: banks with higher market power respond to an increase in adverse selection by either lowering their prices or increasing their prices by less than banks with lower market power. As in Stiglitz and Weiss (1981), this price reduction (or smaller price increase) in response to increases in adverse selection attracts more (or loses fewer) of the banks' increasingly safe marginal borrowers, lowering the average default probability of banks' borrowers.

6.2 The Effects of Increased Costs of Capital

In our second counterfactual, we explore the consequences of increased capital costs. To do so, we analyze the counterfactual effects of a 5% increase (roughly 70 basis points) in banks' marginal costs and compare these to the effects of increased adverse selection that we simulated in the last subsection. We simulate

⁴⁸ Appendix Table A.9 shows that we obtain very similar results when we conduct our analysis at the more disaggregate firm-bank-market-year level.

Table 7: Regressions of Counterfactual Outcomes' Changes on Markups for Higher Adverse Selection and Higher Marginal Costs

Variables	ΔP_{ijmt}	ΔQ_{ijmt}^D	ΔQ_{ijmt}^L	ΔF_{ijmt}
Higher Adverse Selection				
Effective Markup	-0.38 (0.06)	0.15 (0.02)	0.01 (0.00)	-0.13 (0.02)
Bank-Market-Year FE	Yes	Yes	Yes	Yes
R ²	0.134	0.155	0.092	0.097
N obs.	35,563	35,563	35,563	35,563
Higher Marginal Costs				
Effective Markup	-0.42 (0.05)	0.22 (0.03)	0.01 (0.00)	-0.11 (0.01)
Bank-Market-Year FE	Yes	Yes	Yes	Yes
R ²	0.433	0.451	0.112	0.447
N obs.	35,563	35,563	35,563	35,563

Note: An observation is a firm-market-year. Price, demand probabilities, and loan use changes are measured in percentages. Default changes are measured in percentage points. See footnote 45 for dependent variables' definition. Effective Markup is constructed as the negative of the second term on the right hand side of equation (6). Standard errors are clustered at the bank-market-year level.

this higher cost of capital under the baseline level of adverse selection and examine the price, quantity, and default responses across firm-bank relationships. As for the previous policy experiment, we then regress the changes in these outcomes on the effective markups from the baseline model.

The second group of columns in Table 6 shows the change in price, demand, loan use, and default associated with this 5% increase in banks' cost of capital. We find that prices always increase when marginal costs rise. The average increase is 1.78 percentage points (11.0%), comparable to that arising under a doubling of adverse selection, but with half the standard deviation. As expected, as average prices rise, average demand and loan use decline. Unlike the case of higher adverse selection, however, default rates only increase by 1.84 percentage points, less than one-third the increase found there. The explanation for this difference is that while increases in marginal costs increase prices, they do not modify the sensitivity of defaults to price changes. By contrast, an increase in adverse selection not only increases prices, but also makes borrowers with high willingness to pay riskier, further decreasing the quality of banks' borrower pools. For a given change in price levels, this implies a larger increase in the percentage of defaults.

The bottom panel of Table 7 presents regression results relating these changes in outcomes to our measure of banks' market power (their effective markups in our baseline results), where we show that higher effective markups are associated with reductions in price and default changes and increases in quantity changes. In particular, we find that a one standard deviation increase in the effective markup reduces the price variation (ΔP_{ijmt}) by around 6.6 percentage points, increases the variation in demand probabilities (ΔQ_{ijmt}^D) by 3.5 percentage points and in loan use (ΔQ_{ijmt}^L) by around 0.2 percentage points, and reduces variation in

defaults (ΔF_{ijmt}) by 1.7 percentage points.⁴⁹ As was the case for an environment in which adverse selection increases, banks with higher market power can better absorb an increase in the cost of capital in the presence of adverse selection.

6.3 Merger Simulation

In a final counterfactual, we explore the effects of an increase in concentration by simulating a merger between the two banks with the highest market shares in each of our 702 market-year combinations. On average, the largest bank in a market has a market share of 22%, while the second-largest has a market share of 13%. In the counterfactual we change only the ownership structure of the newly merged banks, allowing them to offer two different prices to each borrower as a result of joint profit maximization.

We simulate this merger under two different scenarios to illustrate the impact of increased adverse selection on the effects of mergers. First, we allow the two main banks to merge under the baseline level of adverse selection. We then compare this to outcomes when they merge at the doubled level of adverse selection we analyzed in our first counterfactual. Table 8 shows the changes in prices, demand, loan use, and default under each scenario. In the first group of columns, changes in outcomes are with respect to the case of no merger and our baseline (estimated) levels of adverse selection, whereas in the second case, they are with respect to the case of no merger and twice our estimated levels.

Under baseline levels of adverse selection, interest rates rise by an estimated 0.15 percentage points (1.0%) in response to higher concentration. Similar to the findings of Nevo (2000a) in markets for differentiated cereal products, only a few prices increase in response to the mergers, depending on the degree of substitutability between the differentiated banks. Most of the market-wide average price increase is driven by price increases of the merged banks, which rise by 0.73 percentage points (or 5.7%) on average. There is important heterogeneity across markets in this effect, however, with prices decreasing in 16.4% of cases (and only ever for the merged banks). As expected, given the predicted average price increase, average loan use declines and average default probabilities rise.

Under higher levels of adverse selection, we find even smaller average price effects of merger. Average prices decline on a percentage-point basis and over 33% of prices decrease. As a consequence, average demand probabilities increase slightly and average default probabilities decline. Again, there are differences across banks: average prices increase slightly in merged banks (0.15 percentage points, or 4.5%), and decline slightly in non-merged banks. This exercise further confirms that higher adverse selection mitigates the consequences of increased market power and can even reverse them.

⁴⁹ We find very similar results when we run these regressions at the firm-bank-market-year level, as reported in Appendix Table A.9.

Table 8: Summary Statistics for Counterfactual Changes in Outcomes
for Merger under Baseline and High Adverse Selection

Variables	Nobs	Baseline Adverse Selection			High Adverse Selection		
		Mean	Std. Dev.	< 0	Mean	Std. Dev.	< 0
All Banks							
P.P. Price Variation	35,563	0.15	1.52		-0.02	1.57	
% Price Variation	35,563	1.0	9.9	16.4%	0.4	9.3	33.2%
P.P. Demand Variation	35,563	-0.19	2.67		0.08	2.98	
% Loan Use Variation	35,563	-0.1	1.1		-0.1	1.7	
P.P. Default Variation	35,563	0.14	1.18		-0.03	0.96	
Merged Banks							
P.P. Price Variation	35,563	0.73	6.45		0.15	7.64	
% Price Variation	35,563	5.7	47.1	17.1%	4.5	47.3	33.1%
Non-Merged Banks							
P.P. Price Variation	35,196	0.06	0.95		-0.04	0.13	
% Price Variation	35,196	0.3	4.4	0%	-0.0	0.4	11.4%

Note: These variables are average changes at the firm-market-year level. See footnote 45 for variables' definition. Loan Use changes are in thousands of €. There are less firms in this table compared to Table 6 because in three markets there is only one bank in a market so no merger takes place. Note that in the High Adverse Selection case the mean of the % Price Variation is positive, but the mean of the P.P. Price Variation is negative. This is driven by the different skewness of the two distributions.

7 Conclusion

In this paper we analyze the interaction between imperfect competition and asymmetric information in the Italian market for small business lines of credit. We use a rich dataset with detailed information about credit contracts between firms and banks, including all the main Italian credit institutions and a highly representative sample of firms. We estimate a structural model of firms' demand for credit, loan use, and default, and join with it a model of bank pricing to individual firms. We find evidence of adverse selection, in the form of a positive correlation between unobservables determining both demand and default as well as loan use and default. We also find a causal effect of interest rates on borrowers' default, which we interpret as evidence of moral hazard.

We conduct three counterfactual policy experiments to quantify the importance of adverse selection and investigate its interaction with market power for prices and credit supply. We show that increases in adverse selection and in banks' cost of capital cause prices to increase, demand and loan use to fall, and default to rise. Higher market power, however, *moderates* these effects. Similarly, higher adverse selection moderates and can even reverse the effects of banks' consolidation on prices, demand probabilities, loan use, and default.

These findings have several important policy implications. They confirm that, as theory predicts and taken in isolation, both market power and adverse selection worsen lending conditions. That being said, we

also document that imperfect competition moderates the effects of adverse selection and vice versa. This suggests that competition and banking policymakers should jointly consider the two factors, particularly in those contexts where either is likely to be strong. In practice, the idea that financial markets are characterized by a trade-off between competition and stability due to informational frictions is diffuse in the policy circles. Our structural estimates offer a quantitative assessment of the costs and benefits of market power in banking in the presence of adverse selection.

Of course, these conclusions are predicated on the scope of our study. In particular, our modeling strategy focuses on a subset of the choices firms and banks make in credit markets and our conclusions should be interpreted with these choices in mind. First, we have neglected the fact that many Italian firms borrow from multiple banks, concentrating on each's largest line of credit. Second, we only focus on the first year in which a firm borrows from a bank, and have not considered how lending relationships between firms and banks evolve over time. Third, our estimates do not allow the degree of asymmetric information to vary over the business cycle. This is important if adverse selection is stronger during recessions, in which case a very competitive financial sector could amplify the effects of negative aggregate shocks. Finally, given that we do not estimate borrowing firms' profits, we cannot directly measure social welfare; as such, our framework cannot address the issue of the optimal supply of credit. Each of these topics is interesting and worthy of further research.

References

- Adams, W., L. Einav, and J. Levin**, “Liquidity Constraints and Imperfect Information in Subprime Lending,” *American Economic Review*, 2009, 99 (1), 49–84.
- Akerlof, G.**, “The Market for ‘Lemons’: Quality Uncertainty and the Market Mechanism,” *Quarterly Journal of Economics*, 1970, 84 (3), 488–500.
- Albareto, G., M. Benvenuti, S. Mocetti, M. Pagnini, and P. Rossi**, “The Organization of Lending and the Use of Credit Scoring Techniques in Italian Banks,” *The Capco Institute Journal of Financial Transformation*, 2011, 32, 143–157.
- Albertazzi, U., G. Eramo, L. Gambacorta, and C. Salleo**, “Asymmetric Information in Securitization: An Empirical Assessment,” *Journal of Monetary Economics*, 2015, 71, 33–49.
- Altman, E. I.**, “Financial ratios, discriminant analysis and the prediction of corporate bankruptcy,” *Journal of Finance*, 1968, XXI, 589–609.
- , **G. Marco, and F. Varetto**, “Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience),” *Journal of Banking and Finance*, 1994, 18, 505–529.
- Barba Navaretti, G., M. Bugamelli, F. Schivardi, C. Altomonte, D. Horgos, and D. Maggioni**, “The Global Operations of European Firms - The Second EIFGE Policy Report,” *Brugel Blueprint Series*, 2011, XII.
- Berry, S.**, “Estimating Discrete-Choice Models of Product Differentiation,” *RAND Journal of Economics*, 1994, 25 (2), 242–262.
- , **J. Levinsohn, and A. Pakes**, “Automobile Prices in Market Equilibrium,” *Econometrica*, 1995, 63 (4), 841–890.
- Bofondi, M. and G. Gobbi**, “Informational Barriers to Entry into Credit Markets,” *Review of Finance*, 2006, 10, 39–67.
- Bugamelli, M., L. Cannari, F. Lotti, and S. Magri**, “Il Gap Innovativo del Sistema Produttivo Italiano: Radici e Possibili Rimedi,” *Bank of Italy Occasional Papers: Questioni di Economia e Finanza*, 2012, 1 (121).
- Caliendo, M. and S. Kopeinig**, “Some Practical Guidance for the Implementation of Propensity Score Matching,” *Journal of Economic Surveys*, 2008, 22 (1), 31–72.
- Cerqueiro, G., H. Degryse, and S. Ongena**, “Rules versus Discretion in Loan Rate Setting,” *Journal of Financial Intermediation*, 2011, 20, 503–529.
- Chiappori, P. A. and B. Salanié**, “Testing for Asymmetric Information in Insurance Markets,” *Journal of Political Economy*, 2000, 108 (1), 56–78.
- and —, “Asymmetric Information in Insurance Markets: Empirical Assessments,” *Handbook of Insurance - Second Edition*, 2013, pp. 397–422.
- Ciari, L. and N. Pavanini**, “Market Structure and Multi Market Contact,” 2014. Working Paper.

- Cohen, A. M. and M. Mazzeo**, “Market Structure and Competition among Retail Depository Institutions,” *Review of Economics and Statistics*, 2007, 89 (1), 60–74.
- Degryse, H. and S. Ongena**, “Distance, Lending Relationships, and Competition,” *Journal of Finance*, 2005, LX (1), 231–266.
- Dell’Ariccia, G., E. Friedman, and R. Marquez**, “Adverse Selection as a Barrier to Entry in the Banking Industry,” *RAND Journal of Economics*, 1999, 30 (3), 515–534.
- Demekas, D. G., B. Potter, and M. Pradhan**, *Italy: Background Economic Developments and Issues (IMF Staff Country Report 95/37)*, International Monetary Fund (IMF), 1995.
- Detragiache, E., P. Garella, and L. Guiso**, “Multiple versus Single Banking Relationships: Theory and Evidence,” *Journal of Finance*, 2000, LV (3), 1133–1161.
- Egan, M., A. Hortaçsu, and G. Matvos**, “Deposit Competition and Financial Fragility: Evidence from the US Banking Sector,” *American Economic Review*, 2017, 107 (1), 169–216.
- Einav, L., A. Finkelstein, and J. Levin**, “Beyond Testing: Empirical Models of Insurance Markets,” *Annual Review of Economics*, 2010, 2, 311–336.
- , —, and **M. R. Cullen**, “Estimating Welfare in Insurance Markets using Variation in Prices,” *Quarterly Journal of Economics*, 2010, 125 (3), 877–921.
- and —, “Selection in Insurance Markets: Theory and Empirics in Pictures,” *Journal of Economic Perspectives*, 2011, 25 (1), 115–138.
- , **M. Jenkins, and J. Levin**, “Contract Pricing in Consumer Credit Markets,” *Econometrica*, 2012, 80 (4), 1387–1432.
- Focarelli, D. and F. Panetta**, “Are Mergers Beneficial to Consumers? Evidence from the Market for Bank Deposits,” *American Economic Review*, 2003, 93 (4), 1152–1172.
- Gerakos, J. and C. Syverson**, “Competition in the Audit Market: Policy Implications,” *Journal of Accounting Research*, 2015, 53 (4), 725–775.
- Gobbi, G. and F. Lotti**, “Entry Decisions and Adverse Selection: An Empirical Analysis of Local Credit Markets,” *Journal of Financial Services Research*, 2004, 26 (3), 225–244.
- Guiso, L., L. Pistaferri, and F. Schivardi**, “Credit within the Firm,” *Review of Economic Studies*, 2013, 80 (1), 211–247.
- , **P. Sapienza, and L. Zingales**, “Does Local Financial Development Matter?,” *Quarterly Journal of Economics*, 2004, 119 (3), 929–969.
- Handel, B.**, “Adverse Selection and Switching Costs in Health Insurance Markets: When Nudging Hurts,” *American Economic Review*, 2013, 103 (7), 2643–2682.
- Hausman, J. A. and W. E. Taylor**, “Panel Data and Unobservable Individual Effects,” *Econometrica*, 1981, 49 (6), 1377–1398.
- Heckman, J. J.**, “Sample Selection Bias as a Specification Error,” *Econometrica*, 1979, 47 (1), 153–161.
- , **H. Ichimura, and P. E. Todd**, “Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme,” *Review of Economic Studies*, 1997, 64 (4), 605–654.

- Ho, K. and J. Ishii**, “Location and Competition in Retail Banking,” *International Journal of Industrial Organization*, 2011, 29 (5), 537–546.
- Holmstrom, B. and J. Tirole**, “Financial Intermediation, Loanable Funds, and the Real Sector,” *Quarterly Journal of Economics*, 1997, 122 (3), 663–691.
- Imbens, G. W.**, “Matching Methods in Practice,” *Journal of Human Resources*, 2004, 50 (2), 373–419.
- and **D. Rubin**, *Causal Inference in Statistics, Social, and Biomedical Sciences*, Cambridge University Press, 2015.
- Jensen, M. C. and W. H. Meckling**, “Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure,” *Journal of Financial Economics*, 1976, 3, 305–360.
- Jiménez, G., S. Ongena, J. L. Peydró, and J. Saurina**, “Hazardous Times for Monetary Policy: What do Twenty-three Million Bank Loans Say About the Effects of Monetary Policy on Credit Risk?,” *Econometrica*, 2014, 82 (2), 463–505.
- Karlan, D. and J. Zinman**, “Observing Unobservables: Identifying Information Asymmetries with a Consumer Credit Field Experiment,” *Econometrica*, 2009, 77 (6), 1993–2008.
- Koijen, R. S. J. and M. Yogo**, “An Equilibrium Model of Institutional Demand and Asset Prices,” 2017. NBER Working Paper #21749.
- Lester, B., A. Shourideh, V. Venkateswaran, and A. Zetlin-Jones**, “Screening and Adverse Selection in Frictional Markets,” *Journal of Political Economy*, 2017. forthcoming.
- Leuven, E. and B. Sianesi**, “PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing,” 2003. <https://ideas.repec.org/c/boc/bocode/s432001.html>.
- Lustig, J.**, “Measuring Welfare Losses from Adverse Selection and Imperfect Competition in Privatized Medicare,” 2011. Boston University Working Paper.
- Mahoney, N. and E. G. Weyl**, “Imperfect Competition in Selection Markets,” *Review of Economics and Statistics*, 2017, 99 (4), 637–651.
- Michelacci, C. and F. Schivardi**, “Does Idiosyncratic Business Risk Matter for Growth?,” *Journal of the European Economic Association*, 2013, 11 (2), 343–368.
- Moskowitz, T. J. and A. Vissing-Jorgensen**, “The Returns of Entrepreneurial Investment: A Private Equity Premium Puzzle?,” *American Economic Review*, 2002, 92 (4), 745.
- Myers, S. C.**, “Determinants of Corporate Borrowing,” *Journal of Financial Economics*, 1977, 5, 147–175.
- Nevo, A.**, “Mergers with Differentiated Products: the Case of the Ready-To-Eat Cereal Industry,” *RAND Journal of Economics*, 2000, 31 (3), 395–421.
- , “A Practitioner’s Guide to Estimation of Random-Coefficients Logit Models of Demand,” *Journal of Economics and Management Strategy*, 2000, 9 (4), 513–548.
- , “Measuring Market Power in the Ready-To-Eat Cereal Industry,” *Econometrica*, 2001, 69 (2), 307–342.
- Panetta, F., F. Schivardi, and M. Shum**, “Do Mergers Improve Information? Evidence from the Loan

- Market,” *Journal of Money, Credit and Banking*, 2009, 41 (4), 673–709.
- Pavanini, N. and F. Schivardi**, “The Value of Information in Relationship Lending,” 2017. Working Paper.
- Petersen, M. A. and R. G. Rajan**, “The Benefits of Lending Relationships: Evidence from Small Business Data,” *Journal of Finance*, 1994, XLIX (1), 3–37.
- and —, “The Effect of Credit Market Competition on Lending Relationships,” *Quarterly Journal of Economics*, 1995, 110, 403–444.
- and —, “Does Distance Still Matter? The Information Revolution in Small Business Lending,” *Journal of Finance*, 2002, LVII (6), 2533–2570.
- Rodano, G., N. Serrano-Velarde, and E. Tarantino**, “Lending Standards over the Credit Cycle,” 2015. Working Paper.
- Rothschild, M. and J. E. Stiglitz**, “Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information,” *Quarterly Journal of Economics*, 1976, 90 (4), 630–649.
- Rubin, D. B.**, “Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation,” *Health Services & Outcomes Research Methodology*, 2001, 2, 169–188.
- Sapienza, P.**, “The Effects of Banking Mergers on Loan Contracts,” *Journal of Finance*, 2002, LVII (1), 329–367.
- Starc, A.**, “Insurer Pricing and Consumer Welfare: Evidence from Medigap,” *RAND Journal of Economics*, 2014, 45 (1), 198–220.
- Stein, J. C.**, “Information Production and Capital Allocation: Decentralized versus Hierarchical Firms,” *Journal of Finance*, 2002, LVII (5), 1891–1921.
- Stiglitz, J. and A. Weiss**, “Credit Rationing in Markets with Imperfect Information,” *American Economic Review*, 1981, 71 (3), 393–410.
- Tirole, J.**, *The Theory of Corporate Finance*, Princeton University Press, 2006.
- Train, K. E.**, *Discrete Choice Methods with Simulation*, Cambridge University Press, 2009.
- and **C. Winston**, “Vehicle Choice Behavior and the Declining Market Share of U.S. Automakers,” *International Economic Review*, 2007, 48 (4), 1469–1496.
- Vives, X.**, *Competition and Stability in Banking: The Role of Regulation and Competition Policy*, Princeton University Press, 2016.
- Wooldridge, J. M.**, *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, 2002.

A Appendix For Online Publication

A.1 Conditional Means Results

In this Appendix, we provide details of the conditional means results discussed in Section 3.3. For simplicity, we focus only on the demand and default equations, and exclude consideration of loan use.

From equation (1), we write the firm's borrowing decision as a binary choice between the outside good and the maximum utility it can receive from one of the J_{mt} inside goods. That is, a firm i will borrow from a bank j in market m in year t if:

$$\text{Max}_{j \in 1, \dots, J_{mt}} \{ \bar{\alpha}_0^D + X_{jmt}^D \beta^D + \xi_{jmt}^D + \alpha^D P_{ijmt} + Y_{ijmt}^D \eta^D + \varepsilon_i^D + \nu_{ijmt} \} \geq \nu_{i0mt}. \quad (\text{A.1})$$

This can be rewritten as:

$$\varepsilon_i^D \geq \underbrace{\nu_{i0mt} - \bar{\alpha}_0^D - \text{Max}_{j \in 1, \dots, J_{mt}} \{ X_{jmt}^D \beta^D + \xi_{jmt}^D + \alpha^D P_{ijmt} + Y_{ijmt}^D \eta^D + \nu_{ijmt} \}}_{T_i}, \quad (\text{A.2})$$

where we can simplify the right-hand side, T_i , into a firm-specific component because each firm i in the sample is active in a single market and single year and, once we take the maximum, j drops out. That is, because unobserved demand for credit is firm- and not bank-specific, the additional Type I Extreme Value errors, ν_{ijmt} , do not complicate the calculation of the relationship between ε_i^D and ε_i^F . As far as adverse selection is concerned, we are interested in whether or not a firm borrows at all, not from which bank it borrows. Given this structure, the results relied upon in the text are straightforward. Because ε_i^D and ε_i^F are jointly normally distributed with variances σ_D^2 and σ_F^2 , and correlation coefficient, ρ_{DF} , then

$$\begin{aligned} E(\varepsilon_i^D | D = 1) &= E(\varepsilon_i^D | \varepsilon_i^D \geq T_i) \\ &= \sigma_D \frac{\phi(T_i/\sigma_D)}{1 - \Phi(T_i/\sigma_D)} > 0, \end{aligned} \quad (\text{A.3})$$

and

$$E(\varepsilon_i^F | D = 1) = \rho_{DF} \frac{\phi(T_i/\sigma_D)}{1 - \Phi(T_i/\sigma_D)}, \quad (\text{A.4})$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the PDF and CDF of the standard normal distribution.

A.2 Numerical Results

A.2.1 Monte Carlo

In this Appendix, we construct a simple Monte Carlo to give intuition for how adverse selection and competition can interact in lending markets. We simulate data for the case of a monopolist bank offering a

loan to $i = 1, \dots, I$ potential borrowers, who are observationally equivalent to the bank and differ only in their unobserved demand for credit (ε_i^D) and their utility from default, ε_i^F , which we call their “riskiness”. For simplicity, we concentrate on the correlation in the unobservables between demand and default (ρ_{DF}), setting borrowers’ loan amount and loan use to 1.

In the Monte Carlo, we keep these data fixed and vary two parameters: borrowers’ price sensitivity, α , as a proxy for the strength of the effects of a competitive fringe on the bank’s residual demand curve,⁵⁰ and the extent of asymmetric information, ρ_{DF} , where $\rho_{DF} < 0$ means advantageous selection and $\rho_{DF} > 0$ means adverse selection. For each of these cases, we compute the bank’s equilibrium price based on the maximization of its expected profit, as described in Section 3.3.

Formally, let borrower i have utility U_i^D from taking credit from the bank, utility U_{i0}^D from not borrowing, and utility U_i^F from defaulting:

$$\begin{aligned} U_i^D &= \bar{\alpha}_0 + \alpha P + \varepsilon_i^D + \nu_i, \\ U_{i0}^D &= \nu_{i0}, \\ U_i^F &= \varepsilon_i^F, \end{aligned} \tag{A.5}$$

where P is the interest rate charged by the bank (common across borrowers since there is no observed heterogeneity), and ν_i, ν_{i0} are distributed as type 1 extreme value. We allow ε_i^D and ε_i^F to be jointly normally distributed, with correlation coefficient $-1 \leq \rho_{DF} \leq 1$, and set the variance of ε_i^D to 4 and the variance of ε_i^F to 1. Last, we set $\bar{\alpha}_0 = 1$. We assume that all borrowers have the same price sensitivity $\alpha < 0$. Our asymmetric information assumption implies that the bank doesn’t observe borrower i ’s individual demand and default unobservables, ε_i^D and ε_i^F , but only their distribution in the population of borrowers.

Given this setup, borrower i ’s demand probability is:

$$\begin{aligned} \Pr_i^D &= \Pr(\bar{\alpha}_0 + \alpha P + \varepsilon_i^D + \nu_i > \nu_{i0}) \\ &= \frac{\exp(\bar{\alpha}_0 + \alpha P + \varepsilon_i^D)}{1 + \exp(\bar{\alpha}_0 + \alpha P + \varepsilon_i^D)} \\ &= \Lambda(\bar{\alpha}_0 + \alpha P + \varepsilon_i^D), \end{aligned} \tag{A.6}$$

where $\Lambda(\cdot)$ is the CDF of the logistic distribution. Given \Pr_i^D , the bank’s expected market share is $Q = \int \Pr_i^D f(\varepsilon_i^D) d\varepsilon_i^D$, where $f(\varepsilon_i^D)$ is the density of ε_i^D . Conditional on demand ($D = 1$), the default probability follows that implied by the joint normality assumption (Wooldridge 2002) and is:

$$\begin{aligned} \Pr_{i,F=1|D=1}^F &= E [\Pr(\varepsilon_i^F > 0 | \varepsilon_i^D, P) | D = 1, P] \\ &= \int \Phi \left(\frac{\frac{\rho_{DF} \varepsilon_i^D}{\sigma^2}}{\sqrt{1 - \frac{\rho_{DF}^2}{\sigma^2}}} \right) f(\varepsilon_i^D | D = 1) d\varepsilon_i^D, \end{aligned} \tag{A.7}$$

⁵⁰ In this example, we capture competition versus the outside option, but have verified that increasing the number of banks in the model gives the same qualitative results.

where $\Phi(\cdot)$ is the CDF of the standard normal distribution, σ is the variance of ε_i^D , and $f(\varepsilon_i^D|D=1)$ is the density of ε_i^D conditional on borrowing. In this simple setting, the bank faces no observable heterogeneity in borrowers' default probability, so its expected share of defaulters is just $F = \Pr_{i,F=1|D=1}^F$. Given these probabilities and our supply-side model described in the paper, expected profit-maximization by the bank delivers the following version of the pricing equation (6):

$$P = \underbrace{\frac{MC}{1 - F - F' \frac{1}{\alpha(1-Q)}}}_{\text{Effective Marginal Cost}} + \underbrace{\frac{-(1-F) \frac{1}{\alpha(1-Q)}}{1 - F - F' \frac{1}{\alpha(1-Q)}}}_{\text{Effective Markup}}, \quad (\text{A.8})$$

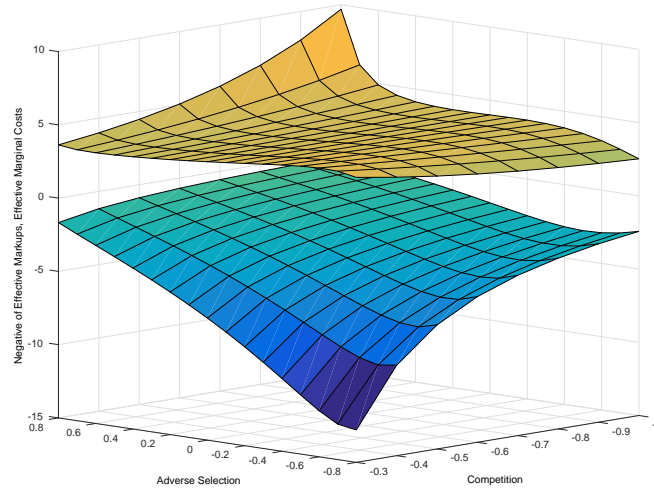
where F' is the derivative of the expected default rate with respect to price, and $\alpha(1-Q)$ is the derivative of the expected market share with respect to price.

The top graph of Figure A.1 reports effective marginal costs, the bottom graph of Figure A.1 reports the *negative* of effective markups, while Figure A.2 combines these to show equilibrium prices. The figures examine how these three elements vary with different degrees of adverse selection, measured by ρ_{DF} , and “competition” with the outside option, measured by the slope parameter of the monopolist's residual demand curve, α .

Looking at the top graph in Figure A.1, for a high level of competition (i.e. the rightmost point on the top graph) an increase in adverse selection (moving to the northwest) causes effective marginal costs to increase, whereas for low competition (the point closest to the reader on the top graph, again moving northwest) it remains relatively constant. The opposite happens for effective markups as we increase adverse selection. Looking at the bottom graph of Figure A.1, and recalling that it reports the *negative* of effective markups (thus a higher markup is associated with a lower point in the graph), for high levels of competition (the rightmost point on the bottom graph, moving to the northwest), increases in adverse selection decrease effective markups slightly, whereas for low levels of competition, (the closest point to the reader on the bottom graph, again moving northwest), increases in adverse selection decrease effective markups substantially.

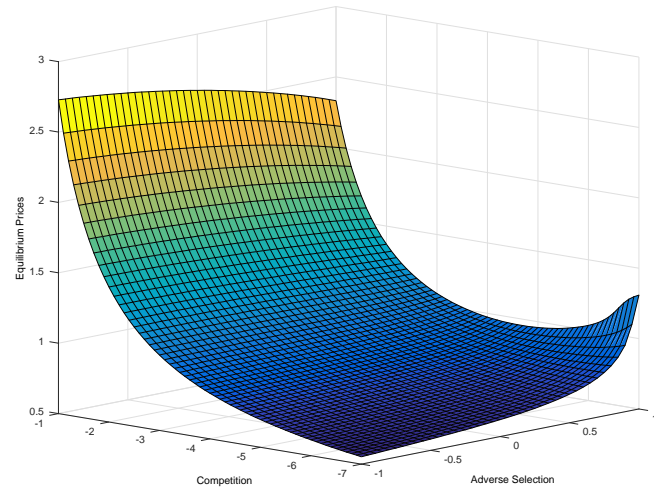
Figure A.2 combines both factors and demonstrates a non-monotonic price response to increases in adverse selection, with the sign of the effect depending on the level of competition. While equilibrium prices rise with adverse selection in a competitive environment (the closest point to the reader, moving to the northeast), the opposite happens in a concentrated market (the leftmost point, moving east). In a competitive environment, markups are low and the average borrower effect dominates, so increasing adverse selection causes prices to rise (driven by the rising effective marginal costs shown in Figure A.1). In a less competitive environment, by contrast, markups are high, enhancing the marginal borrower effect, and increasing adverse selection drives prices down. In competitive markets, banks have small margins and can only increase prices in response to increased average selection, while in less competitive markets, banks with higher markups find it profitable to reduce prices as it allows them to attract relatively safe borrowers.

Figure A.1: Adverse Selection vs Imperfect Competition - Effective Marginal Costs, Negative Effective Markups



Note: The vertical axis shows the value of effective marginal costs and of the negative of the effective markups. The left horizontal axis is level of adverse selection, increasing towards left. The right horizontal axis measures the slope of the residual demand curve (our measure of competition with the outside option), increasing towards the right.

Figure A.2: Adverse Selection vs Imperfect Competition - Equilibrium Prices



Note: The vertical axis shows the level of equilibrium prices. The left horizontal axis measures the slope of the residual demand curve (our measure of competition with the outside option), increasing towards the right. The right horizontal axis is the level of adverse selection, increasing towards right. The horizontal axis definitions and scales in this figure differ from those in Figure A.1 to better display the effects in each.

A.2.2 $\partial F'/\partial \rho_{DF}$ Results

For the simplified setting described in the Monte Carlo, equation (A.7) above presented the formula for the probability firm i defaults given that it has chosen to borrow, $D = 1$. In what follows, we derive this formula and its derivatives in greater detail for use in the discussion in Section 3.3 in the body of the text.

To understand where equation (A.7) came from, note that given the structure of preferences described in equation (A.5) and the joint normality assumption on ε_i^D and ε_i^F , the probability of default, $F = 1$, conditioned on a specific ε_i^D is $\Phi\left(\frac{\rho_{DF}\varepsilon_i^D}{\sigma^2}\right)$. When we calculate the conditional probability of default, we therefore need to take into account that borrowers have selected into the decision to borrow.

In a standard Heckman setting, the selection equation is deterministic: a firm borrows if $\varepsilon_i^D \geq \bar{\alpha}_0 - \alpha P$. The problem in our case is different because our selection equation gives only a probability of borrowing conditional on ε_i^D rather than a deterministic threshold. From equation (A.6), we know that:

$$\Pr_{i|\varepsilon_i^D}^D = \Lambda(\bar{\alpha}_0 + \alpha P + \varepsilon_i^D). \quad (\text{A.9})$$

One can then apply Bayes rule to obtain the conditional distribution of ε_i^D given $D = 1$:

$$\begin{aligned} f(\varepsilon_i^D | D = 1) &= \frac{\Pr_{i|\varepsilon_i^D}^D f(\varepsilon_i^D)}{\Pr_i^D} \\ &= \frac{\Lambda(\bar{\alpha}_0 + \alpha P + \varepsilon_i^D) f(\varepsilon_i^D)}{\int \Lambda(\bar{\alpha}_0 + \alpha P + \varepsilon_i^D) f(\varepsilon_i^D) d\varepsilon_i^D}. \end{aligned} \quad (\text{A.10})$$

Then, the probability of default conditional on borrowing is just the expected value of $\Phi\left(\frac{\rho_{DF}\varepsilon_i^D}{\sigma^2}\right)$ under $f(\varepsilon_i^D | D = 1)$. As in equation (A.7) above,

$$\Pr_{i,F=1|D=1}^F = \int \Phi\left(\frac{\rho_{DF}\varepsilon_i^D}{\sigma^2}\right) f(\varepsilon_i^D | D = 1) d\varepsilon_i^D. \quad (\text{A.11})$$

Let $\Lambda(\bar{\alpha}_0 + \alpha P + \varepsilon_i^D) = \Lambda(\cdot)$ to simplify the notation. Substituting equation (A.10) into equation (A.11) we get:

$$\Pr_{i,F=1|D=1}^F = \underbrace{\frac{1}{\int \Lambda(\cdot) f(\varepsilon_i^D) d\varepsilon_i^D}}_A \underbrace{\int \Phi\left(\frac{\rho_{DF}\varepsilon_i^D}{\sigma^2}\right) \Lambda(\cdot) f(\varepsilon_i^D) d\varepsilon_i^D}_B. \quad (\text{A.12})$$

The derivative of default conditional on borrowing with respect to P , that is F'_{ijmt} in Section 3.3, is given

by:

$$\frac{\partial \Pr_{i,F=1|D=1}^F}{\partial P} = \frac{\partial \mathcal{A}}{\partial P} \mathcal{B} + \mathcal{A} \frac{\partial \mathcal{B}}{\partial P}, \quad (\text{A.13})$$

with the derivative components in this equation given by the following two terms:

$$\begin{aligned} \frac{\partial \mathcal{A}}{\partial P} &= -\alpha \frac{\int \Lambda(\cdot)(1 - \Lambda(\cdot))f(\varepsilon_i^D)d\varepsilon_i^D}{\left(\int \Lambda(\cdot)f(\varepsilon_i^D)d\varepsilon_i^D\right)^2}, \\ \frac{\partial \mathcal{B}}{\partial P} &= \int \Phi\left(\frac{\frac{\rho_{DF}\varepsilon_i^D}{\sigma^2}}{\sqrt{1 - \frac{\rho_{DF}^2}{\sigma^2}}}\right) \alpha \Lambda(\cdot)(1 - \Lambda(\cdot))f(\varepsilon_i^D)d\varepsilon_i^D. \end{aligned} \quad (\text{A.14})$$

The second derivative of default conditional on borrowing with respect to P and ρ_{DF} , referred to in the body of the text as $\partial F'_{ijmt}/\partial \rho_{DF}$, is given by:

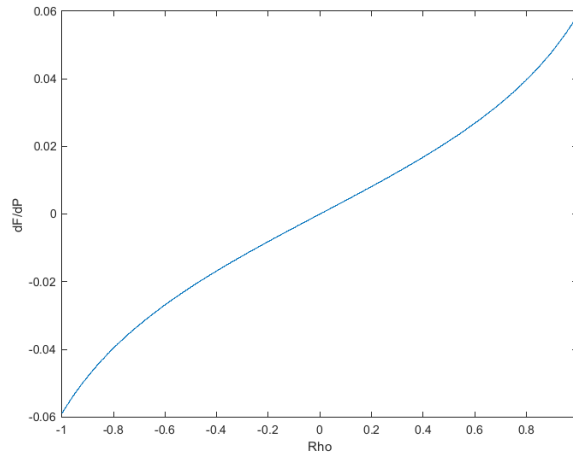
$$\frac{\partial^2 \Pr_{i,F=1|D=1}^F}{\partial P \partial \rho_{DF}} = \frac{\partial \mathcal{A}}{\partial P} \frac{\partial \mathcal{B}}{\partial \rho_{DF}} + \mathcal{A} \frac{\partial^2 \mathcal{B}}{\partial P \partial \rho_{DF}}, \quad (\text{A.15})$$

with the two new derivative components in this equation given by the following terms:

$$\begin{aligned} \frac{\partial \mathcal{B}}{\partial \rho_{DF}} &= \int \phi\left(\frac{\frac{\rho_{DF}\varepsilon_i^D}{\sigma^2}}{\sqrt{1 - \frac{\rho_{DF}^2}{\sigma^2}}}\right) \frac{\frac{\varepsilon_i^D}{\sigma^2}}{\left(1 - \frac{\rho_{DF}^2}{\sigma^2}\right)^{\frac{3}{2}}} \Lambda(\cdot)f(\varepsilon_i^D)d\varepsilon_i^D, \\ \frac{\partial^2 \mathcal{B}}{\partial P \partial \rho_{DF}} &= \int \phi\left(\frac{\frac{\rho_{DF}\varepsilon_i^D}{\sigma^2}}{\sqrt{1 - \frac{\rho_{DF}^2}{\sigma^2}}}\right) \frac{\frac{\varepsilon_i^D}{\sigma^2}}{\left(1 - \frac{\rho_{DF}^2}{\sigma^2}\right)^{\frac{3}{2}}} \alpha \Lambda(\cdot)(1 - \Lambda(\cdot))f(\varepsilon_i^D)d\varepsilon_i^D, \end{aligned} \quad (\text{A.16})$$

where ϕ is the PDF of a standard normal distribution. Even in this simple setting, it is difficult to sign these derivatives analytically. Instead, we simulate them for $\rho \in [-1, 1]$, $\bar{\alpha}_0 = 0$, $\alpha = -1$, and $\sigma = 1$ to obtain the following figure:

Figure A.3: $\partial \Pr_{i,F=1|D=1}^F / \partial P$



As can be seen there, the derivative of the default probability with respect to price is the same sign as ρ_{DF} and the slope of this line, measuring $\partial^2 \Pr_{i,F=1|D=1}^F / \partial P \partial \rho_{DF}$, is everywhere increasing when $\rho_{DF} > 0$.

A.3 Constructing the Dataset

We have assembled data from the following sources:

- **Firm Data:** Dataset from *Centrale dei Bilanci* with yearly (1988-1998) balance sheet data for each firm, including data both for firms that take credit and those that do not. This data also includes the year of birth for each firm, its location at the city council level, and what we call in Section 2.2 each firm’s “Score.” The Score represents our measure of a firm’s observable default risk.
- **Loan Data:** Dataset from *Centrale dei Rischi* with yearly (1988-1998) firm-bank loan contracts, including the amount of granted credit, the amount of this used by the firm, the loan’s interest rate, and whether or not the firm has defaulted on this loan (Section 2.3 in the paper describes our definition of default). As discussed in Section 2.1, this data is only available for 94 large banks (representing more than 80% of total bank lending) and for short term lines of credit.
- **Bank Data:** Dataset with yearly (1988-2002) balance sheet data for each bank, including yearly total loans that each bank gives in each province, and its share of the total loans granted in each province.
- **Branch Data:** Dataset with yearly (1959-2005) branches for the population ($\sim 1,500$) of banks at the city council level.
- **Coordinates Data:** Based on the National Institute for Statistics *ISTAT* city council classification, we assign to each city council the geographic coordinates that will allow us to calculate firm-branch distances.

We first merge the firm data with the loan data, in order to combine all the borrowing and non-borrowing firms. We then take all the banks actively lending in each province and assume that those represent the choice set for each firm, regardless of whether they have a branch in that province or not.⁵¹ We assume that each firm chooses one main credit line from among those offered by the banks active in its province or chooses not to take any line of credit (the outside good). The main line is defined as the line for which the amount used, regardless of the amount granted, is the highest. For cases in which multiple lines have the same amount used, then the one with the lowest price is defined as the main line. We calculate the distance in kilometers between the city council of each firm and the closest city council where each bank from the choice set has a branch using the geographic coordinates.

A.4 Reduced Form Evidence

In this section we fully describe the reduced form evidence summarized in Section 2.4 in the paper.

⁵¹ There is evidence in other papers (Bofondi and Gobbi 2006), as well as in our data, that a few banks lend in some provinces even if they don’t have a branch there.

Table A.1: Reduced Form Evidence of Imperfect Competition

Variables	(1)	(2)	(3)	(4)
Log HHI Loans	0.02 (0.00)			
Log HHI Firms		0.00 (0.00)		
Log CR3 Loans			0.01 (0.01)	
Log CR3 Firms				0.02 (0.00)
R ²	0.923	0.923	0.923	0.923
N obs.	469,633	469,677	469,666	469,633

Note: The table shows the results of OLS regressions of the predicted interest rate (in percentage points) on lending market concentration measures at the market-year level. An observation is a firm-bank-year. HHI Loans is the Herfindahl-Hirschman Index based on the share of credit. HHI Firms is the Herfindahl-Hirschman Index based on the share of firms. CR3 Loans is the concentration ratio for the 3 banks with the highest market share in terms of loans in each market-year. CR3 Firms is the concentration ratio for the 3 banks with the highest market share in terms of borrowers in each market-year. All regressions include bank-year fixed effects, bank-market fixed effects, firm fixed effects, amount granted fixed effects and the distance between the firm and the bank. Standard errors are clustered at the bank-year level.

A.4.1 Imperfect Competition

In this subsection, we provide some descriptive evidence on imperfect competition in the Italian market for small business lines of credit. We construct four alternative measures of concentration and investigate their correlation with interest rates conditional on various sets of observables. First, we calculate the Herfindahl-Hirschman Index (HHI) of market concentration based on each bank's share of used credit within a market-year (hereafter HHI Loans). Second, we define the same index based instead on each bank's share of borrowers within a market-year (HHI Firms). Third, we construct the 3-bank concentration ratio in terms of used credit in each market-year (CR3 Loans). Finally, we construct the 3-bank concentration ratio in terms of borrowers in each market-year (CR3 Firms). We then use each of these as regressors in a regression of the log of predicted interest rates.⁵² We show in Table A.1 that concentration is positively correlated with interest rates in all specifications, as expected, and is statistically significant for two out of the four measures we use. In the first and fourth specifications, respectively in columns (1) and (4), we find that a 10% increase in concentration is associated with a 0.2% increase in interest rates.

A.4.2 Asymmetric Information

In this subsection, we conduct reduced-form tests for the presence of asymmetric information. To do so, we follow the early empirical literature on positive correlation tests introduced by Chiappori and Salanié (2000).

⁵² We use predicted interest rates because it's the price variable we use in our structural demand model. See Section 4.1 for a detailed description on how we construct predicted interest rates.

We propose two tests, one based on the choice to take up a line of credit and another based on the choice of how much credit to draw on that line. Both tests are based on the correlation between the unobservables driving these choices and the unobservables influencing default. The choice of these tests gives a flavor of the identification strategy that we rely on in the structural model, as explained in greater detail in Section 4.3 in the body of the text.

Demand and Default: We start by investigating whether firms that are more likely to demand credit are also more likely to default. Our data include both firms that borrow and those that do not, while we only observe default on loans for borrowing firms. We can formalize this problem as a selection model with two equations:

$$\begin{aligned} d_i &= \mathbf{1}(X_i^d \beta + \nu_i > 0), \\ f_i &= \begin{cases} X_i^f \gamma + \eta_i & \text{if } d_i = 1 \\ - & \text{if } d_i = 0, \end{cases} \end{aligned} \quad (\text{A.17})$$

where d_i is equal to one if the firm borrows and f_i is equal to one if the borrower defaults. f_i is observed only if $d_i = 1$. This is similar to the classical selection model analyzed by Heckman (1979), where we interpret as adverse selection a positive correlation between ν_i and η_i .⁵³ Results of this Heckman selection model are reported in the first two columns (“Extensive Margin”) of Table A.2, where the decisions to borrow and default are regressed on year, market, firms’ Score, amount of granted credit, and sector fixed effects, as well as on a set of firms’ balance sheet variables. We use as an instrument in the selection (i.e. borrowing) equation the number of banks in a firm’s market, which we interpret as a proxy for the competitiveness of banks’ local markets.⁵⁴ We find a positive and significant correlation coefficient of 0.09 between the unobservables driving demand and default, which we interpret as preliminary evidence of asymmetric information on the extensive margin.

Loan Use and Default: We then consider the relationship between loan use and default. Unlike the previous test, we are not in a selection framework as the same firms are observed in both equations. Still, the idea is the same, as we test for a positive correlation between the unobservables that determine the choice of “extent of coverage” (loan use) and the occurrence of an “accident” (default), conditional on several firm characteristics. Following the intuition of the previous test, adverse selection should imply that riskier firms use more credit. We set up the following seemingly unrelated regression (SUR):

$$\begin{aligned} \ell_i &= X_i \beta + \varepsilon_i, \\ f_i &= X_i \gamma + \eta_i, \end{aligned} \quad (\text{A.18})$$

where ℓ_i is the amount of its loan used by firm i , and, as before, f_i takes value of one if the borrower defaults. The vector of controls X_i is composed of year, market, bank, firms’ Score, amount of granted credit, and sector fixed effects, as well as on a set of firms’ balance sheet variables. We specify the distribution of the residuals ε_i, η_i as joint normal, with a correlation coefficient ρ . A positive and significant estimate of

⁵³ We estimate default as a linear probability model for ease of interpretation, but estimates from a discrete choice regression yield similar results.

⁵⁴ This is likely to satisfy the exclusion restriction that it influences demand for credit (via interest rates), but is uncorrelated with a firm’s idiosyncratic decision to default. We find a first stage F-statistic of 75.94.

ρ suggests the presence of asymmetric information. The results of this test are summarized in the last two columns (“Intensive Margin”) of Table A.2. We again find a positive correlation, consistent with asymmetric information on the intensive margin.

A.5 Matching Model

As discussed in Section 4.1, we must predict the prices charged on loans from banks from whom firms chose not to borrow. We also need to predict both prices and the amount of granted credit a firm that chooses not to borrow at all would require should it have chosen to borrow. As summarized in Section 4.1.3, we use propensity score matching to estimate the prices and amounts of granted credit for non-borrowing firms. This subsection describes how we do so.

Following Imbens (2004), Imbens and Rubin (2015), and Caliendo and Kopeinig (2008), we construct an iterative process to appropriately select the relevant variables determining the propensity score and obtain the best possible match. Our choice of covariates for the matching, that is the variables determining whether a firm borrows or not, is guided by economic theory and knowledge of the institutional setting, as well as by the overlap in variables’ distributions and statistics from the matching results.

The final set of variables that we use are all fixed effects and include fixed effects for the year, firms’ Score, firms’ geographical area, sales, and assets. In line with Caliendo and Kopeinig (2008), we apply the following specific criteria in our selection in order to determine both which variables to include and the degree of discretization for the fixed effects. First, we only include controls that influence simultaneously the participation (borrowing vs non-borrowing) and the outcome variables (interest rates or amount of granted credit, depending on the variable we are predicting). Second, variables must be unaffected by participation, or the anticipation of it, so should be either fixed over time or measured before participation. The Score respects this rule, and we assume that the value of assets and sales are persistent over time. We choose to only control for a Score above or below 6, as Rodano, Serrano-Velarde and Tarantino (2015) showed that this is the most relevant threshold level for lending standards. Third, for the common support assumption to hold, some randomness is needed, so some firms with identical characteristics should be observed in each state (borrowing versus not). For this reason, we choose a parsimonious set of variables and seek to avoid over-parametrization.

We summarize in Table A.3 the normalized differences in means between treatment (non-borrowers) and control (borrowers) groups.⁵⁵ Imbens (2004) defines as “modest” normalized differences below 0.3 in absolute value, and all of our differences for the continuous variables used are below that threshold. We implement several matching methods and find similar results across them. We choose to focus on k-nearest neighbor matching, as it allows us to assign several untreated (borrowing) firms to each treated (non-borrowing) one.

We follow the standard literature in performing several statistical tests to assess the quality of the matching. Variable selection is based on statistical significance, the “hit or miss” method (Heckman, Ichimura and

⁵⁵ The normalized difference for a variable with mean μ and variance σ^2 is given by $\frac{\mu_T - \mu_C}{\sqrt{\sigma_T^2 + \sigma_C^2}/2}$, where T stands for treated (non-borrowing) and C stands for control (borrowing) groups.

Table A.2: Reduce Form Evidence for Adverse Selection

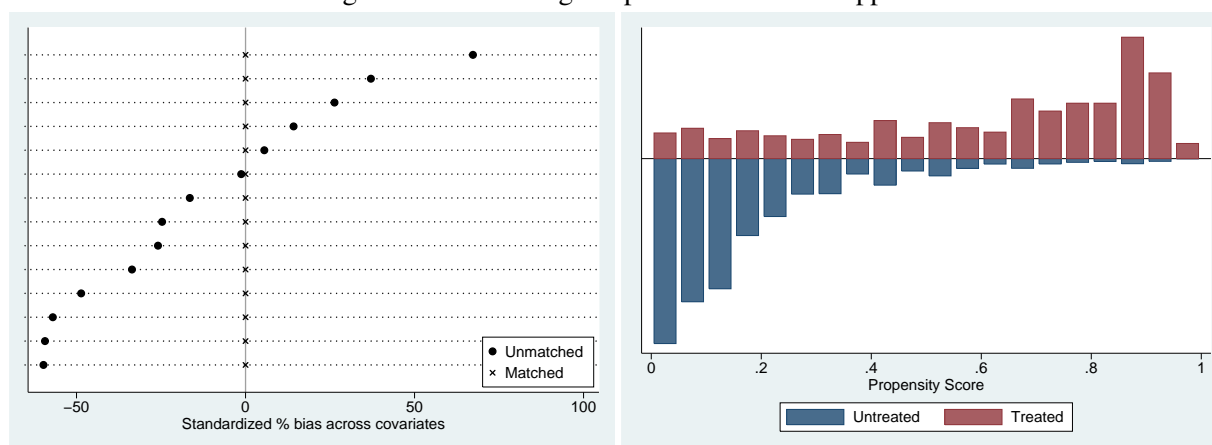
Variables	Extensive Margin		Intensive Margin	
	Demand	Default	Loan Use	Default
Correlation between Unobservables	0.09 (0.01)		0.03 (0.00)	
Total Assets	0.55 (0.02)	0.00 (0.00)	0.34 (0.02)	0.00 (0.00)
Intangible Assets	-0.95 (0.07)	-0.04 (0.01)	-0.39 (0.12)	-0.03 (0.01)
Intangible/Total Assets	-0.39 (0.03)	0.01 (0.01)	-0.08 (0.10)	0.01 (0.01)
Sales	0.46 (0.01)	-0.00 (0.00)	-0.05 (0.01)	-0.00 (0.00)
Profits	0.53 (0.05)	0.02 (0.01)	0.48 (0.12)	0.01 (0.01)
Cash Flow	-0.43 (0.07)	-0.03 (0.01)	-2.29 (0.16)	-0.02 (0.01)
Trade Debit	-0.17 (0.00)	0.00 (0.00)	-0.08 (0.01)	0.00 (0.00)
Firm's Age	0.54 (0.07)	0.01 (0.01)	-0.07 (0.17)	0.01 (0.01)
Number of Banks in Market	0.04 (0.00)	-	-0.00 (0.01)	0.00 (0.00)
Distance to Branch	-	-	-0.00 (0.00)	-0.00 (0.00)
Interest Rate	-	-	-0.02 (0.01)	0.00 (0.00)
Bank FE	No	No	Yes	Yes
Market FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Score FE	Yes	Yes	Yes	Yes
Amount Granted FE	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes
R ²	0.222	0.091	0.437	0.094
N Obs.	36,520	25,351	25,351	25,351

Note: The table shows the results of a Heckman selection model for Demand and Default (Extensive Margin) and of a SUR model for Loan Use and Default (Intensive Margin). The dependent variable is a dummy for credit demand in Column 1, a dummy for default in column 2 and 4, and the amount of credit used in column 3. See Table 1 for variables' definition. The excluded instrument for the selection model is the number of banks in a firm's market. In each regression an observation is a firm. We rescale some variables to interpret the coefficients more easily: Intangible Assets, Sales, Total Assets, Profits, and Cashflow are in €100,000. Trade Debit is in €1,000,000. Age of Firm is in 100 years.

Table A.3: Normalized Differences

Variable	Obs	Normalized Difference
Score	52,310	-0.294
Sales	52,310	-0.076
Total Assets	52,310	-0.066

Figure A.4: Matching Graph and Common Support



Todd 1997),⁵⁶ and comparisons of several statistics before and after the matching. These include the Pseudo-R-squared, the Likelihood Ratio, the mean and median bias, and Rubin’s B and R .⁵⁷ Caliendo and Kopeinig (2008) explain that a rule of thumb for a good match is to have mean and median biases below 3% to 5%. According to Leuven and Sianesi (2003), Rubin’s B should be below 25% and Rubin’s R should be between 0.5 and 2. Finally, a good matching outcome should deliver Pseudo R-squared and Likelihood Ratio tests high in the unmatched case, and very low in the matched case. Our results pass all these statistical tests, as shown in Tables A.4 and A.5. Last, we show a graph of the bias reduction and test the common support of the propensity score between treated and untreated in Figure A.4. Even though there is a large mass at each tail, these figures show that the values for both groups span the full range of propensity scores, implying that we have enough overlap as long as we allow for replacement.

⁵⁶ Variables are chosen to maximize within-sample prediction rates, i.e. maximizing the cases in which the estimated propensity score for each observation is greater than the sample proportion of firms taking the treatment (in our case not borrowing).

⁵⁷ B is the number of standard deviations between the means of the groups, and R is the ratio of treatment variance to control variance (Rubin 2001).

Table A.4: Matching Results 1

Variable	Unmatched vs. Matched	Mean		% Bias	% Bias Reduction	t-Test	
		Treated	Control			t	p> t
1991-1992	U	0.392	0.132	-33.6		-32.57	0.000
	M	0.392	0.392	0.0	100.0	0.00	1.000
1993-1994	U	0.333	0.308	5.6		5.90	0.000
	M	0.333	0.333	0.0	100.0	-0.00	1.000
1995-1996	U	0.249	0.110	37.1		41.82	0.000
	M	0.249	0.249	0.0	99.9	0.03	0.980
1997-1998	U	0.309	0.245	14.2		15.25	0.000
	M	0.309	0.309	-0.0	99.8	-0.02	0.981
Score>6	U	0.281	0.358	-16.5		-17.24	0.000
	M	0.281	0.281	0.0	100.0	-0.00	1.000
North Area	U	0.656	0.662	-1.3		-1.35	0.176
	M	0.656	0.656	0.0	100.0	-0.00	1.000
Sales Category 2	U	0.391	0.114	67.3		77.37	0.000
	M	0.391	0.391	0.0	100.0	-0.00	1.000
Sales Category 3	U	0.135	0.229	-24.7		-25.17	0.000
	M	0.135	0.135	0.0	100.0	0.00	1.000
Sales Category 4	U	0.055	0.265	-59.8		-57.30	0.000
	M	0.055	0.055	0.0	100.0	0.00	1.000
Sales Category 5	U	0.056	0.265	-59.3		-56.92	0.000
	M	0.056	0.056	0.0	100.0	-0.00	1.000
Assets Category 2	U	0.275	0.166	26.3		28.85	0.000
	M	0.275	0.275	0.0	100.0	0.00	1.000
Assets Category 3	U	0.132	0.231	-25.9		-26.36	0.000
	M	0.132	0.132	0.0	100.0	-0.00	1.000
Assets Category 4	U	0.079	0.255	-48.6		-47.59	0.000
	M	0.079	0.079	0.0	100.0	0.00	1.000
Assets Category 5	U	0.061	0.263	-57.0		-54.94	0.000
	M	0.061	0.061	0.0	100.0	0.00	1.000

Table A.5: Matching Results 2

Sample	Pseudo- R^2	LR χ^2	$p > \chi^2$	Mean Bias	Median Bias	Rubin's B	Rubin's R
Unmatched	0.367	23,790.6	0.000	34.1	30.0	170.0	1.26
Matched	-0.000	-0.00	1.000	0.0	0.0	0.0	1.00

Note: A rule of thumb for a good match is to have mean and median biases below 3% to 5%, Rubin's B below 25% and Rubin's R between 0.5 and 2.

A.6 The Effects of Measurement Error in Predicted Prices

As discussed in footnotes 21 (in Section 4.1.1), 28 (in Section 4.1.3), and 31 (in Section 4.2.1), we assume that our model of price prediction does not induce conventional econometric measurement error problems. In this subsection, we demonstrate that as long as the residuals in the pricing regression are uncorrelated with default unobservables, any measurement error due to our use of predicted prices can at most result in a conservative estimate of the degree of adverse selection.

To do so, we outline a stylized version of our structural model to investigate the potential direction of the bias due to measurement error in prices in the correlation coefficients that identify adverse selection, our main coefficients of interest. We define firm i 's utility from demanding (superscript D) from bank j , and its utility from defaulting (superscript F), to depend only on price and unobservables. We decompose prices into what we can and cannot predict as follows:

$$P_{ij} = \tilde{P}_{ij} + \tilde{\tau}_{ij}, \quad (\text{A.19})$$

where P_{ij} is the true price, \tilde{P}_{ij} is the predicted price, and $\tilde{\tau}_{ij}$ is the measurement error. We identify adverse selection as a positive correlation between the unobservables of the demand and default equations, ε_i^D and ε_i^F :

$$U_{ij}^D = \alpha^D P_{ij} + \varepsilon_i^D, \quad (\text{A.20})$$

$$U_{ij}^F = \alpha^F P_{ij} + \varepsilon_i^F. \quad (\text{A.21})$$

As described at length in Section 4.1.3, our preferred model of price prediction uses data on all the loans taken by each firm, allowing us to include firm fixed effects in its estimation. The fourth column of Table 2 shows that this significantly increases the fit of the regression (as measured by its R^2) and the fourth column of Table 3 shows that the residuals from this pricing regression are uncorrelated with default. In this stylized model, we therefore claim that the measurement error, $\tilde{\tau}_{ij}$, is uncorrelated with ε_i^F .

Recall also that in our structural model, we only use predicted prices in the demand estimation and not in the estimation of the loan use and default equations. In this stylized model, this means we only need to account for measurement error in equation (A.20) and not equation (A.21). Substituting (A.19) in (A.20), we get:

$$U_{ij}^D = \alpha^D (\tilde{P}_{ij} + \tilde{\tau}_{ij}) + \varepsilon_i^D = \alpha^D \tilde{P}_{ij} + \zeta_{ij}, \quad (\text{A.22})$$

where $\zeta_{ij} \equiv \alpha^D \tilde{\tau}_{ij} + \varepsilon_i^D$. The true correlation coefficient between the propensity to borrow and to default is defined by:

$$\rho_{DF} = \frac{\text{Cov}(\varepsilon_i^D, \varepsilon_i^F)}{\sigma_{\varepsilon_i^D} \sigma_{\varepsilon_i^F}}. \quad (\text{A.23})$$

In the presence of measurement error, we would estimate:

$$\tilde{\rho}_{DF} = \frac{\text{Cov}(\zeta_{ij}, \varepsilon_i^F)}{\sigma_{\zeta_{ij}} \sigma_{\varepsilon_i^F}} = \frac{\text{Cov}(\varepsilon_i^D, \varepsilon_i^F)}{(\sigma_{\varepsilon_i^D} + \alpha^D \sigma_{\tilde{\tau}_{ij}}) \sigma_{\varepsilon_i^F}} = \rho_{DF} \frac{\sigma_{\varepsilon_i^D}}{\sigma_{\varepsilon_i^D} + \alpha^D \sigma_{\tilde{\tau}_{ij}}}, \quad (\text{A.24})$$

where the second equality follows from the assumed lack of correlation between $\tilde{\tau}_{ij}$ and ε_i^F . One can see that the only effect of measurement error in (A.24) is the additional variance in the composite demand error caused by the measurement error.

Equation (A.24) shows that, typical of classical measurement error problems, in the presence of measurement error our estimate of ρ_{DF} would be biased towards zero. The size of any bias would depend on both α^D and the standard deviation of $\tilde{\tau}_{ij}$ relative to that of ε_i^D , that is, the firm's private information. We believe that our price prediction reduces measurement error to a minimum. Indeed, as we argue in the text, firms too are likely to predict banks' prices, rather than getting a price quotation from each of them, suggesting the variance of the demand error (which we normalize in the text to that of the standard Type I Extreme Value) using our predicted prices should be comparable to that of firms. Even if this weren't the case, equation (A.24) shows that the degree of adverse selection we estimate is a lower bound on the true degree of adverse selection. Measurement error by itself, therefore, cannot explain the correlation we find between the propensity to demand credit and to default.

A.7 Additional Tables and Figures

Table A.6: IV First Stage and OLS vs IV Second Stage for Demand

Variable	First Stage	Second Stage	
	Interest Rate	OLS	IV
Number Accounts 2 nd quintile	0.05 (0.01)	-	-
Number Accounts 3 rd quintile	0.07 (0.01)	-	-
Number Accounts 4 th quintile	0.03 (0.01)	-	-
Number Accounts 5 th quintile	0.02 (0.01)	-	-
Deposit Rate 2 nd quintile	0.05 (0.01)	-	-
Deposit Rate 3 rd quintile	0.08 (0.01)	-	-
Deposit Rate 4 th quintile	0.08 (0.02)	-	-
Deposit Rate 5 th quintile	0.11 (0.02)	-	-
Log of Deposit Amount	-0.06 (0.03)	-	-
Interest Rate	-	0.16 (0.22)	-1.45 (0.62)
Number of Branches	0.00 (0.02)	4.37 (0.21)	4.38 (0.21)
Share of Branches	0.00 (0.00)	0.53 (0.05)	0.53 (0.05)
Years in Market	0.03 (0.00)	0.01 (0.01)	0.06 (0.03)
Constant	0.95 (0.46)	-3.05 (0.24)	-4.14 (3.66)
Bank FE	Yes	Yes	Yes
Market FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Obs	6,036	6,036	6,036
R ²	0.670	0.766	0.766
F-Stat	25.67	-	-

Note: Standard errors in brackets. See Table 1 for variables' definition. Standard errors are clustered at the bank-year level.

Table A.7: IV First Stage for Loan Use and Default

Variable	First Stage Interest Rate
Interest Rates in Other Markets	0.56 (0.04)
Constant	8.37 (3.14)
Firm Controls	Yes
Bank Controls	Yes
Score FE	Yes
Sector FE	Yes
Loan Amount FE	Yes
Bank FE	Yes
Market-Year FE	Yes
Obs	35,173
R^2	0.376
F-Stat	184.24

Note: Standard errors in brackets. Firm and bank controls include respectively all the firm level and bank level controls in Table 4. Standard errors are clustered at the market-year level.

Table A.8: Structural Estimates with Random Coefficient on Prices

Variables		Demand	Loan Use	Default	
Firm Level	Price	Interest Rate	-1.40 (0.62)	-0.03 (0.00)	0.92 (0.02)
		Assets	Total Assets	5.68 (0.08)	0.09 (0.00)
	Intangible/Total Assets		-0.79 (0.05)	-0.01 (0.01)	0.09 (0.05)
	Profitability	Profits	1.08 (0.05)	0.01 (0.00)	0.03 (0.02)
		Cash Flow	-0.92 (0.05)	-0.05 (0.00)	-0.12 (0.02)
		Sales	6.97 (0.07)	-0.01 (0.00)	-0.36 (0.04)
	Debt	Trade Debit	-3.34 (0.05)	-0.04 (0.00)	0.12 (0.02)
		Others	Firm's Age	0.22 (0.03)	0.00 (0.00)
	Distance to Branch		-1.24 (0.03)	-0.01 (0.01)	-0.04 (0.06)
	Bank Level	Number of Branches	4.35 (0.15)	0.01 (0.00)	0.02 (0.02)
Share of Branches			0.53 (0.04)	-0.07 (0.01)	-0.22 (0.10)
Years in Market		0.06 (0.03)	0.01 (0.00)	-0.10 (0.02)	
First-stage fixed effects	Score	Yes	Yes	Yes	
	Sector	Yes	Yes	Yes	
	Loan Amount	Yes	Yes	Yes	
	Bank-Market-Year	Yes	No	No	
	Bank	No	Yes	Yes	
	Market-Year	No	Yes	Yes	
Observations		506,230	25,351	25,351	
Covariance matrix (Σ)	Adverse Selection	$\sigma_D = 0.19$ (0.00)	$\sigma_L = 0.30$ (0.00)	$\sigma_F = 1$	
		$\rho_{DL} = 0.04$ (0.00)	$\rho_{LF} = 0.09$ (0.00)		

Note: All coefficients are estimated in the first stage, with the exception of the Interest Rate, the Number of Branches, the Share of Branches and the Years in Market for the demand equation, that are estimated in the second stage. Second stage fixed effects, only for the demand equation, are at the bank, market and year level. See Table 1 for variables' definition. Standard errors are in brackets. First stage standard errors are calculated by the inverse of the Information matrix, obtained providing the solver with analytical gradient and hessian. Second stage standard errors are computed with 200 bootstrap replications.

Figure A.5: Default Probability Distribution for Baseline Scenario and for Higher Adverse Selection Counterfactual Scenario

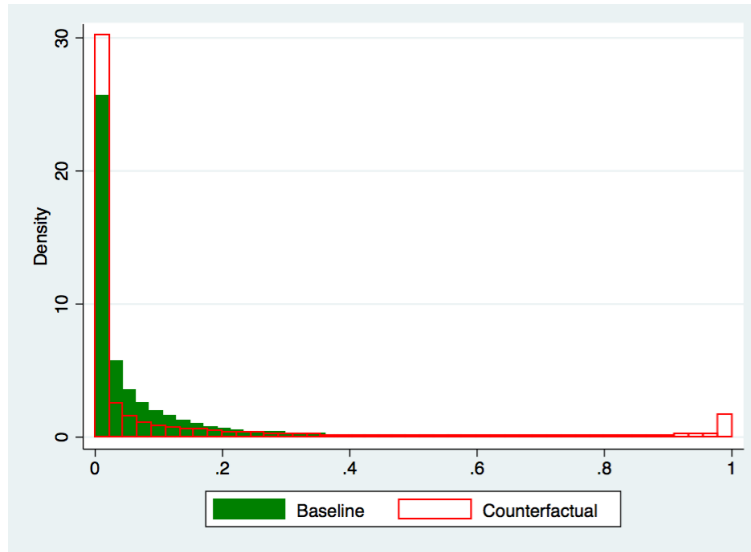


Table A.9: Regressions of Counterfactual Outcomes' Changes on Markups for Higher Adverse Selection and Higher Marginal Costs

Variables	ΔP_{ijmt}	ΔQ_{ijmt}^D	ΔQ_{ijmt}^L	ΔF_{ijmt}
Higher Adverse Selection				
Effective Markup	-0.25 (0.03)	0.14 (0.01)	0.02 (0.00)	-0.12 (0.07)
Bank-Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
R ²	0.763	0.787	0.125	0.896
N obs.	434,490	418,667	434,490	421,407
Higher Marginal Costs				
Effective Markup	-0.30 (0.02)	0.21 (0.01)	0.01 (0.00)	-0.27 (0.01)
Bank-Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
R ²	0.750	0.682	0.130	0.830
N obs.	434,490	418,740	434,490	420,122

Note: An observation is a firm-bank-market-year. Price, demand probabilities, and loan use changes are measured in percentages. Default changes are measured in percentage points. See footnote 45 for dependent variables' definition. Effective Markup is constructed as the negative of the second term on the right hand side of equation (6). Standard errors are clustered at the firm level. Note that we omit the top 3 percentiles of the markup distribution, to avoid our results being driven by outliers. Given the presence of outliers also in the demand and default percentage variations, we omit the top 3 percentiles from those distributions as well.