Asymmetric Information and Imperfect Competition in Lending Markets*

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

May 2016

Abstract

We measure the consequences of asymmetric information and imperfect competition in the Italian market for small business lines of credit. We provide evidence that a bank's optimal price response to an increase in adverse selection varies depending on the degree of competition in its local market. More adverse selection causes prices to increase in competitive markets, but can have the opposite effect in more concentrated ones, where banks trade off higher markups and the desire to attract safer borrowers. This implies both that imperfect competition can moderate the welfare losses from an increase in adverse selection, and that an increase in adverse selection can moderate the welfare losses from market power. Exploiting 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, we estimate models of demand for credit, loan pricing, loan use, and firm default to measure the extent and consequences of asymmetric information in this market. While our data include a measure of observable credit risk available to a bank during the application process, we allow firms to have private information about the underlying riskiness of their project. This riskiness influences banks' pricing of loans as higher interest rates attract a riskier pool of borrowers, increasing aggregate default probabilities. We find evidence of adverse selection in the data, and increase it with a policy experiment to evaluate its importance. As predicted, in the counterfactual equilibrium prices rise in more competitive markets and decline in more concentrated ones, where we also observe an increase in access to credit and a reduction in default rates. Thus market power may serve as a shield against the negative effects of an increase in adverse selection.

*We thank Daniel Ackerberg, Jeff Campbell, Pierre-André Chiappori, Lorenzo Ciari, Valentino Dardanoni, Ramiro de Elejalde, Liran Einav, 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 seminar participants at Warwick, PEDL, Barcelona GSE Banking Summer School, EUI, Tilburg, Zürich, Bocconi, 2014 Winter Marketing-Economics Summit in Wengen, IO session of the German Economic Association in Hamburg, St. Gallen, Barcelona 2014 Summer Forum, EARIE 2014, Toulouse, DIW Berlin, NBER 2015 Winter IO Meeting, and UCL 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

‡University of Zürich, nicola.pavanini@econ.uzh.ch

§Bocconi, EIEF and CEPR, fabiano.schivardi@unibocconi.it

1 Introduction

Following the seminal work of Akerlof (1970) and Rothschild and Stiglitz (1976), a large theoretical literature has stressed the key role of asymmetric information in financial markets. This literature has shown that asymmetric information can generate market failures such as credit rationing, inefficient provision, misspricing of risk and, in the limit, market breakdown.[1] Indeed, the recent financial crisis can be seen as an extreme manifestation of the problems that asymmetric information can cause. In fact, following the definition by Mishkin (2012), *a financial crisis is a nonlinear disruption to financial markets in which adverse selection and moral hazard problems become much worse*. Deepening our understanding of the extent and causes of asymmetric information is key for the design of a regulatory framework that limits their negative consequences.

Although the basic theoretical issues are well understood, empirical work is fairly rare. Asymmetric information is by definition hard to measure. If a financial intermediary, such as a lender, has an information disadvantage with respect to a potential borrower, it is very unlikely that such a disadvantage can be overcome by the researcher. While one cannot generally construct measures of the ex-ante unobserved characteristics determining riskiness, it is often possible to observe ex-post outcomes, such as defaulting on a loan. The empirical literature has been built on these facts, analyzing how agents with different ex-post outcomes self select ex-ante into contracts (if any) with different characteristics in terms of price, coverage, deductibles etc. (Chiappori and Salanié (2000), Abbring, Chiappori, Heckman and Pinquet (2003), Lustig (2011), Einav, Jenkins and Levin (2012), Starc (2014)).[2]

We measure the consequences of asymmetric information and imperfect competition in the Italian market for small business lines of credit. We exploit detailed, proprietary data on a representative sample of Italian firms, the population of medium and large Italian banks, individual lines of credit between them, and subsequent individual defaults. While our data include a measure of observable credit risk comparable to that available to a bank during the application process, in our model we 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 demand credit. As shown by Stiglitz and Weiss (1981), in this setting an increase in the interest rate exacerbates adverse selection, inducing a deterioration in the quality of the pool of borrowers. We formulate and structurally estimate a model of credit demand, loan use, default, and bank pricing based on the insights in Stiglitz and Weiss (1981) and Einav et al. (2012) that allows us to estimate the extent of both adverse selection and moral hazard in the market, and to run counterfactuals that approximate economic environments of likely concern to policymakers.

One key contribution of our paper is that we study adverse selection in an imperfectly competitive market. This differs from most of the previous literature, that, due to data limitation or to specific market features, has assumed either perfectly competitive markets, or imperfectly competitive markets subject to significant regulatory oversight. Assuming perfect competition in the market for small business loans is not desirable,

[1] See, for example, Banerjee and Newman (1993), Bernanke and Gertler (1990), DeMeza and Webb (1987), Gale (1990), Hubbard (1998), Mankiw (1986), Mookherjee and Ray (2002).

[2] See Einav and Finkelstein (2011), Einav, Finkelstein and Levin (2010), and Chiappori and Salanié (2013) for extensive surveys of the this literature.

given the local nature of small business lending and the high degree of market concentration at the local level, the latter due to entry barriers in the Italian banking sectors that persisted into the 1990s. We show that the degree of competition can have significant consequences on the equilibrium effects of asymmetric information. Intuitively, with perfect competition banks price at average costs (e.g. Einav and Finkelstein (2011)). When adverse selection increases, the price also rises, as a riskier pool of borrowers implies higher average costs in the form of more defaults. 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 price increases in the standard (e.g. monopoly) pricing calculus. As a consequence, a bank with market power facing an increase in adverse selection will also increase its market share and improve the quality of its borrowers, as a lower price attracts marginal borrowers, which are safer under adverse selection. This implies both that imperfect competition can moderate the welfare losses from an increase in adverse selection and that higher adverse selection can moderate the welfare losses of market power. Lester, Shourideh, Venkateswaran and Zetlin-Jones (2015) and Mahoney and Weyl (2014) provide an intuitive theoretical foundation for this result, consistent with our findings.

To analyze these questions, we construct a model where banks offer standardized contracts to observationally equivalent firms. Loan contracts and banks are differentiated products in terms of, among other characteristics, the amount granted, a bank's network of branches, the years a bank has been in a market, and distance from the 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 private information to the firm. Firms choose the preferred loan, if any, according to a mixed logit demand system. They also choose how much of the credit line to use. Finally, they decide if to repay the loan or default. The degree of adverse selection is determined by two correlations: 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). For a given interest rate, firms' expected profits are increasing with risk due to the insurance effect of loans: banks share a portion of the costs of unsuccessful projects. As a result, higher-risk firms are more willing to demand higher-rate loans. This, in turn, influences the profitability of rate increases by banks.[3] We show with a Monte Carlo simulation that imperfect competition can indeed mitigate the effects of an increase in adverse selection. The effects of asymmetric information on prices depends on market power. When markets are competitive, more adverse selection always leads to higher rates and less credit. As banks' market power increases, this relationship becomes weaker and eventually turns negative. Last, we also show the causal effect of a change in interest rates on default, controlling for selection, and interpret it as moral hazard.

We estimate the model on highly detailed microdata covering individual loans between firms and banks between 1988 and 1998. There are two key elements of this data. The first, from the Italian Central Credit Register (*Centrale dei Rischi*), provides detailed information on all individual loans extended by the 90

[3] Handel (2013), Lustig (2011), and Starc (2014) find similar effects of 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 articulates the non-monotonicity of these effects depending on the strength of competition, an empirically relevant result in our application.

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. Critically, this second dataset includes an observable measure of each firm's default risk (SCORE). Combining them yields a matched panel dataset 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.[4] We define local markets at the level of provinces, administrative units roughly comparable to a US county that, as discussed in detail by Guiso, Pistaferri and Schivardi (2013), constitute a natural geographical unit for small business lending. We estimate individual firms' demand for credit, banks' pricing of these lines, firm's loan use and subsequent default. We extend the econometric approach taken by Einav et al. (2012) to the case of multiple lenders by assuming unobserved tastes for credit independent of the specific bank chosen to supply that credit. We combine this framework with the literature on demand estimation for differentiated products (Berry 1994, Berry, Levinsohn and Pakes 1995, Goolsbee and Petrin 2004). Data on default, loan use, demand, and pricing separately identify the distribution of private riskiness from heterogeneous firm disutility from paying interest.

We provide reduced form evidence of adverse selection along both the intensive and the extensive margin. For the former, we run a positive correlation test as in Chiappori and Salanié (2000). For the latter, we estimate a Heckman selection model. We also provide rough evidence of imperfect competition, showing that interest rates are positively correlated with concentration in local markets. In the structural model, we find that the choice to borrow, the amount used and the decision to default depend on observables as expected. In particular, a higher interest rate and higher distance from branches reduce the probability that a firm borrows. Among other observables, firms with more cash flow are both less likely to demand credit, arguably because they have more internally generated funds, use a smaller share of their loan, and less likely to default. In terms of correlation of unobservables, we find a positive correlation both between the choice to borrow and default, and between how much loan to use and default. We interpret this as evidence of adverse selection. We also find a positive effect of interest rates on default, which we interpret as evidence of moral hazard.

We run a counterfactual to quantify the extent of adverse selection and understand its interaction with imperfect competition. In this policy experiment we increase the degree of adverse selection, identified by the correlation between both demand and default and loan use and default unobservables, and look at how equilibrium prices, demand, and defaults vary in response to this. The economic motivation for this exercise can be thought as the possible consequences of a credit crunch, where risky firms become more exposed to financial distress than safe ones and demand more credit. This counterfactual delivers two important findings. First, there is a heterogeneous response of equilibrium prices, market shares, loan use, and defaults to an increase in adverse selection. Second, these variations are correlated with banks' market power, measured by their estimated markup at the year-province-borrower level. We find that banks with higher markups decrease prices as adverse selection increases, and consequently increase their share of borrowers and de-

[4] A similar approach is followed, among others, by Chiappori and Salanié (2000). We model the dynamics of firm-bank relationships in a companion paper (Pavanini and Schivardi (2016)).

crease their share of defaulters. This implies that banks with higher markups have a counter-cyclical effect on credit supply, responding to an increase in adverse selection with a reduction in prices and an increase in quantity lent. We show that one standard deviation increase in markup reduces the counterfactual variation in bank's prices by 4.3 percentage points, increases its variation in demand probability by 0.2 percentage points and in loan use by around 8,200 euros, and reduces its variation in borrowers' default probability by 2.2 percentage points.

This paper contributes to two main strands of empirical work. The first is the literature on empirical models of asymmetric information, so far mainly focussed on insurance markets. We look at the less developed 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 novelty is to introduce imperfect competition. We show that this is important, as the impact of asymmetric information depends crucially on the nature of competition in the market. The second field we contribute to is the literature on empirical banking, where we are not aware of any structural model that seeks to measure the consequences of asymmetric information and the role competition plays in mediating its effects. Nonetheless, several reduced form papers on Italian banking provide motivation for a model that structurally combines these two effects. For example, Bofondi and Gobbi (2006) show that new banks entering local markets experience higher default rates than incumbents, as the latter have superior information about borrowers and local economic conditions. Gobbi and Lotti (2004) claim that there is a positive correlation between branching and markets with low proprietary information services, and that interest rate spreads are positively related to entry of de novo banks, but not of banks existing in other markets. Finally, Panetta, Schivardi and Shum (2009) show that mergers enhance pricing of observable risk, as merged banks achieve a better match of interest rates and default risk, mainly due to better information processing.

The structure of the paper is the following. In Section 2 we describe the dataset and the market, in Section 3 we present the reduced form tests of adverse selection and imperfect competition, Section 4 outlines the structural model, and Section 5 describes the econometric specification of demand, loan use, default and supply. The estimation and the results are in Section 6, the counterfactuals are in Section 7, Section 8 concludes.

2 Data and Institutional Details

We use a unique dataset of small business credit lines, previously used in Panetta et al. (2009).[5] It is based on three main sources of data. Interest rate data and data on outstanding loans are from the Italian *Centrale dei Rischi*, or Central Credit Register. Firm-level balance sheet data are from the *Centrale dei Bilanci* database. Banks' balance-sheet and income-statement data are from the Banking Supervision Register at the Bank of Italy. By combining these data, we obtain a matched panel dataset of borrowers and lenders extending over an eleven-year period, between 1988 and 1998. We also collected data on bank branches at the local level

[5] For reasons that will be explained below, in this paper we only use on a subset of the original data. This section focusses on the description of this subset, referring the interested reader to Panetta et al. (2009) for descriptive statistics of the full dataset.

since 1959.[6]

The Central Credit Register (hereafter CR) is a database that contains detailed information on individual bank loans extended by Italian banks. Banks must report data at the individual borrower level on the amount granted and effectively utilized for all loans exceeding a given threshold,[7] with a breakdown by type of the loan (credit lines, financial and commercial paper, collateralized loans, medium and long-term loans and personal guarantees). Banks also report if they classify a loan as bad, meaning that they attach a low probability to the event that the firm will be able to repay the loan in full. We define a default as a loan being classified as bad.[8] In addition, a subgroup of around 90 banks (accounting for more than 80 percent of total bank lending) have agreed to file detailed information on the interest rates they charge to individual borrowers on each type of loan.

We restrict our attention to short-term credit lines, which have ideal features for our analysis. First, the bank can change the interest rate at any time, while the borrower can close the credit line without notice. This means that differences between the interest rates on loans are not influenced by differences in the maturity of the loan. Second, the loan contracts included in the CR are homogeneous products, so that they can be meaningfully compared across banks and firms. Third, they are not collateralized, a key feature for our analysis, as adverse selection issues become less relevant for collateralized borrowing. Fourth, short term bank loans are one of the main source of borrowing of Italian firms. According to our data, trade credit represents around 48% of firms' debt, short term bank credit 28%, and long term bank credit 9%. We define the interest rate as the ratio of the payment made in each year by the firm to the bank to the amount of the loan used. The interest payment includes the fixed expenses charged by the bank to the firm (e.g. which encompass the cost of opening the credit line or the cost of mailing the loan statement).

We focus on a subsample of the available data, namely on the main credit line of the first year a firm opens at least one credit line. Considering only the first year is a common assumption in static empirical models of insurance with asymmetric information, starting from Chiappori and Salanié (2000). This is done to avoid modeling heterogenous experience ratings among borrowers and loan renegotiation, challenging topics, and ones that we leave for future research. Moreover, we focus on the main new credit line because it accounts on average for around 75% of the total share of new yearly credit (both usable and used),[9] even if in Italy multiple relationship banking is widely used by firms to reduce liquidity risk (Detragiache, Garella and Guiso (2000)). This means that we restrict our attention only to the first year in which we observe a firm in our data.[10] This reduces the sample size from around 90,000 firms to over 40,000.[11] Table 1, Panel A reports the loan level information that we use in the empirical analysis. Out of around 27,000 firms, 69% take up a loan in our sample period, and use on average 67% of the amount granted. Of these, around 6%

[6] Detailed descriptives on the branch data are in Ciari and Pavanini (2014).

[7] The threshold was 41,000 euros (U.S. $42,000) until December 1995 and 75,000 euros thereafter.

[8] We do not observe if a loan actually reverts to not being bad. However, this seems to be a rather unlikely event. Moreover, classifying a loan as bad has a negative impact on bank accounting ratios, even before the firm formally defaults. So this is clearly a costly event in itself for the bank. See section 2.1 for a complete definition of default.

[9] 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 chosen.

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

[11] We estimate our structural model on a subset of the original dataset, mostly for computational and institutional reasons explained in Section 6. This reduces the sample to around 27,000 firms.

end up being classified as bad loans within our sample.[12] The average amount granted is around 370,000 euros, and the average interest rate charged is 14.5%.

Panel B of Table 1 shows summary statistics for the 90 reporting banks. The average total asset level is almost 11 billion, they employ 3,200 employees and have a share of bad loans over total loans of 6%. The average bank is present in 34 provinces out of 95, but with great variation across banks.

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, that sells these data to banks for their lending decisions. The unique feature of the CB data set is that, unlike other widely used data sets on individual companies (such as the Compustat database of US companies), it has wide coverage of small and medium enterprises; moreover, almost all the companies in the CB sample are unlisted. The coverage of these small firms makes the data set 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. In time, the CB has increased the sample size drawing from balance sheets deposited with the chambers of commerce (limited liability companies are obliged to file their balance sheets to the chambers of commerce, that make them available to the public). The database is fairly representative of the Italian non-financial sector. The firms in the CB sample represent about 49.4% of the total sales 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 in Altman (1968) and Altman, Marco and Varetto (1994).

We define a borrowing firm as one that shows up as a borrower in the CR database. Non borrowing firms are defined according to two criteria: they are not in the CR database and report zero bank borrowing in their balance sheets. We use the second definition to exclude firms that are not in our CR database but are still borrowing from banks, either from one of the non-reporting banks or through different loan contracts.[13] Table 1, Panel C reports descriptive statistics for the sample of borrowing and non-borrowing firms. Borrowing firms seem to have larger assets and sales. In terms of bank relations, borrowing firms have on average around 3.5 credit lines active every year. They open one new line every year and close 0.6. Note that these firms are mostly new borrowers, so they are more likely to be in the process of expanding their number of relationships. The share of credit used from the main line is 72%, and it goes up to 76% when a firm borrows for the first year. This shows that focusing on the main line captures most of the credit that firms borrow, especially for new firms.

There is ample evidence that firms, particularly small businesses like the ones in our sample, are tied to their 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. Segmentation of local credit markets is thus very likely to occur. In our

[12] See Section 2.1 for a complete definition of default.

[13] This implies that we exclude from our sample around 27,000 firms that borrow from banks not included in our sample, or borrow from the banks in our sample but using a different type of loan.

market definition we will use provinces as our geographical units. Provinces are administrative unit roughly comparable to a US county. They are a proper measure of local markets in banking for at least 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, the 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 the branch. In our data firms are on average 3.19 km (1.98 miles) far from the branch of their main bank. At the time of our data, there were 95 provinces. We report summary statistics of markets (defined more precisely below) in Panel D of Table 1, which shows that there are around 8 banks per province-year in our sub-sample, each bank has on average almost 14 branches per province, with a market share of 7% for branches and 9% for loans.[14] On average a bank has been in a province for at least 21 years.[15]

Even though our dataset includes both borrowing and non-borrowing firms, we have no information on banks' rejections of applicants. 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 choice set would charge them, and then decide which bank is their best alternative. In our model, a bank that classifies a firm as very risky will not reject it, but will be likely to offer it a very high interest rate. 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, but to the best of our knowledge there is no paper using loan applications for our sample period in Italy. Albertazzi, Bottero and Sene (2014) was one of the first papers making use of loan applications in Italy for the 2003-2012 period. In that data a loan application is identified by an enquiry advanced by a bank to the Credit Register to obtain information on the current credit position of a new potential borrower, not currently borrowing from the bank. The authors classify a loan application to be approved if a new loan is granted within three months since the information request, and rejected if no loan is granted. According to this definition, they find that 21% of the applications result in a new loan within the 3 months window. They also show that each firm in their sample receives on average 0.91 rejections in the 6 months prior to each information request. This definition of rejection doesn't however rule out the case of firms refusing to accept the bank's offer, presented to the firm once the bank has obtained the information from the Credit Register and has decided on the interest rate to charge. Hence, this data doesn't allow to distinguish between banks' rejections and firms' offer refusal, so it is hard to know with certainty how relevant actual rejections are in our context.

2.1 Default Definition

Following Panetta et al. (2009), the definition of default in our data includes firms in liquidation or other bankruptcy proceedings, and those that have not paid repayment installments on loans for at least six months. This corresponds to a default warning that any bank can file to the Italian Credit Register for any of its

[14] The market share of the outside option, defined by the firms that choose not to borrow, is on average around 30%.

[15] We start counting the years from 1959, which is the first year that we observe in the branching data.

Table 1: Summary Statistics

	Variable	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.
Panel A:	Demand	27,128	0.69	0.46			
Loan Level	Loan Use	18,820	246.5	442.6			
	Default	18,820	0.06	0.24			
	Amount Granted	18,820	368.9	474.8			
	Interest Rate	18,820	14.49	4.62			
Panel B:	Total Assets	900	10,727	16,966			
Bank Level	Employees	896	3,180	4,583			
	Bad Loans	893	6.2	6.3			
	Number of Provinces	861	34.54	30.19			
Panel C:		**Borrowing Firms**			**Non-Borrowing Firms**		
Firm Level	Total Assets	18,820	11,244	19,139	8,308	3,553	8,105
	Intangible/Tot Assets	18,820	0.16	0.22	8,308	0.21	0.28
	Profits	18,820	1,028	3,052	8,308	288	1,438
	Cash Flow	18,820	670	2,187	8,308	260	1,221
	Sales	18,820	14,174	22,927	8,308	4,901	11,247
	Trade Debit	18,820	1,676	3,397	8,308	754	3,161
	Firm's Age	18,820	12.75	12.68	8,308	10.69	11.96
	Score	18,820	5.37	1.79	8,308	4.76	2.16
	Branch distance (km)	18,820	3.19	7.22			
	Number of Lenders	72,573	3.45	2.55			
	Lines Opened	72,573	1.15	1.69			
	Lines Closed	72,573	0.64	1.25			
	Share of Main Line	58,848	0.72	0.26			
	Share of Main New Line	14,732	0.76	0.25			
Panel D:	Number of Banks	666	7.93	3.99			
Market level	Number of Branches	5,284	13.82	21.21			
	Share of Branches	5,284	0.07	0.09			
	Years in Market	5,284	21.38	14.23			
	Market Shares	5,284	0.09	0.09			

Note: In Panel A an observation is a firm for the first variable and a loan contract for the others. Demand is a dummy for taking a loan or not, 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, amount granted is in thousands of euros. In Panel B an observation is a bank-year. Employees is the number of employees at the end of the year. Bad loans is a percentage of total loans. In Panel C an observation is a firm for the first 9 variables and a firm-year for the others. The Score is the indicator of the risk of the firm computed each year by the CB (higher values indicate riskier companies). Number of Lenders is the number of banks from which the firm borrows through these credit lines. The last two variables represent the ratio of credit utilized from the main line over total credit utilized, when credit utilized is non-zero. In Panel D an observation is year-province for the Number of Banks, and bank-year-province for the other 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 for since 1959. Market Shares are in terms of number of borrowers.

borrowers. This warning cannot be filed for a single loan overdue, but it's rather the result of a negative evaluation that a bank has of the borrower's overall financial situation, even prior to a legally certified bankruptcy status.[16] This implies that banks classify these firms' loans as irrecuperable, defaulting firms are unable to repay all of their loans and end up exiting the credit market. We find that 80% of the firms that default within our sample exit the sample in the same year, and 16% in the following year.[17] There is institutional and anecdotal evidence[18] that when one bank sends this kind of default warning to the Credit Register it has a "domino effect" on all other loans with any other bank that the defaulting firm has. Most importantly, according to the Italian Civil Code, this default warning remains in the Credit Register as public information available to all banks for the following 10 years, compromising a defaulting firm's access to credit from any bank for that period of time.

Among the new borrowers that we focus on, we find that 54% of the firms that end up defaulting receive a default warning and exit the credit markets within 2 years of their first loan, and another 24% within 4 years. Given the low number of defaulters per year, and the short time period between the first loan and subsequent default, we choose to focus just on the first year of a loan and classify as defaulter a firm that will eventually default within 3 years in our sample. We choose the 3 years limit because we can trace a firm's default until 2001, 3 years after the end of our loans' sample, reducing issues connected with right censoring of our data for firms that start borrowing towards the end of our sample.

2.2 Price Construction

A crucial empirical challenge that we face when connecting the dataset we use to our model set up concerns price prediction. On one hand, we don't observe prices for loans in a firm's choice set that didn't take place, so we need to predict those interest rates based on the observables we have. On the other hand, one of the main determinants of loan prices is borrowers' riskiness perceived by banks, which is predicted by lenders from a combination of hard information, which we observe in the data, and soft information, which we don't observe. As a consequence, we cannot assume with certainty that we have the same information set as each bank about each borrower.

Whether the information gap between us and the lenders becomes a problem for our findings depends on how much soft information matters, relative to hard information, for banks to price risk. We adopt several strategies to limit the extent of this problem. First of all, we just consider the first year in which a firm borrows in the sample, excluding the initial year in our data (1988). The advantage of this approach, introduced in the insurance context by Chiappori and Salanié (2000) among others, is that it limits the information gap between the econometrician and the lender, as we just consider borrowers that approach a bank for the first time. The second point is to select the best model for price prediction among a variety of alternatives,[19] based on institutional and anecdotal evidence, and to test the statistical and economic significance of the residuals of this pricing regression as explanatory variable in a default equation. This allows us both to

[16] Source: Bank of Italy's informative note (*Circolare*) n. 139 of 11/02/1991.

[17] The remaining within 4 years.

[18] Source: www.tuttocentraledeirischi.it, support web page for borrowers dealing with the Credit Register.

[19] For alternative ways of predicting prices see Gerakos and Syverson (2015).

identify the hard information that best predicts prices, and to investigate how much this hard information explains ex post risk compared to soft information, captured by the residuals of the pricing regression. The third point is comparing our findings to the existing literature in corporate finance and empirical banking on loan pricing models. Last, we discuss the possible implications for our results of an inaccurate price prediction.

Before describing the modeling strategy we use to predict prices, it is important to 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' financial situation, whereas the Centrale dei Bilanci provides banks with a detailed archive of firms' balance sheet information. As described in Cerqueiro, Degryse and Ongena (2011), banks use both hard and soft information to determine their lending policies. The authors show that for US data the importance of each factor depends on loan and borrower characteristics, as well as local lending markets, and borrower-lender relationship.

To describe the institutional features of the Italian lending market we rely on the results of a survey conducted by the Bank of Italy of over 300 Italian banks in 2007 about banks' organization of lending, summarized in Albareto, Benvenuti, Mocetti, Pagnini and Rossi (2011). Several features of this survey are relevant for our analysis. First, the survey shows that larger banks, which are the ones we have in our data, are more likely to use hard information and standardized scoring techniques. 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, large banks tend 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 the top management. Multiple layers of hierarchy also imply that large banks allow for shorter terms of office for branch managers, to avoid branch managers to develop relationships with local borrowers and derive 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. Banks' ranking turns out to be the following: (i) Financial statement data (i.e. hard information from 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 data), (v) Availability of guarantees, (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 shows that for small banks instead soft information is much more relevant, even though still less important than the first two forms of hard information.

Two other interesting features for our set up emerge from that survey. First, the importance of credit scoring in banks' lending policies (including pricing), and second, the use by banks of both sales and loan size to segment borrowers into size classes. The survey highlights that 70% of large banks are organized by divisions, with customers segmented by size and typically divided into SMEs and large firms. The variable commonly used for segmenting firms is sales. Therefore, we control for both Score and sales in our pricing and default regressions.

Our model selection for a loan-pricing model is based on OLS pricing regressions where we progressively

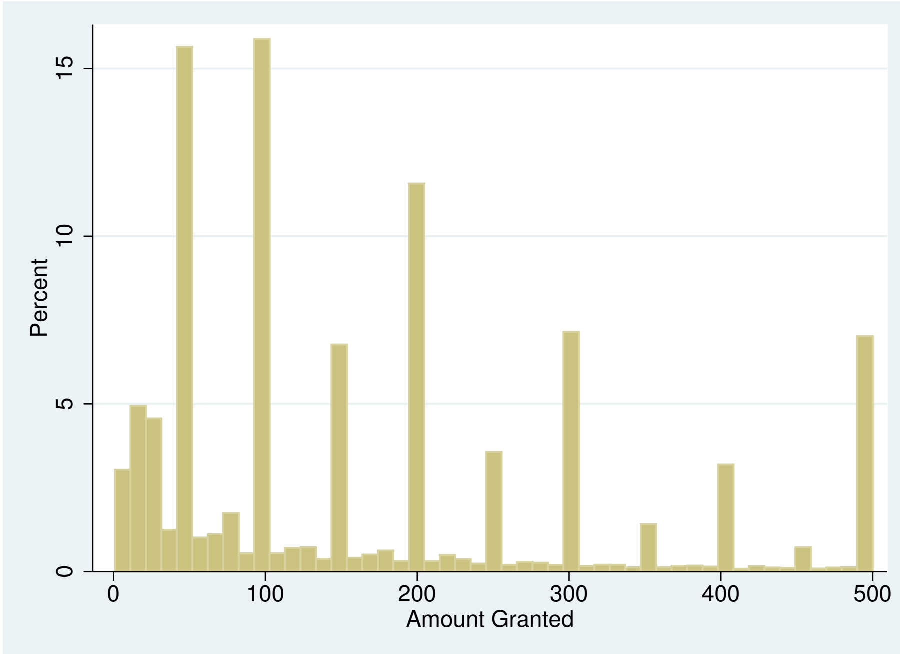
include controls from our dataset, as shown in Table 2.[20] The subset of the data we use only includes the first year in which a firm appears in our sample, and every firm in its first year borrows on average from just below 2.8 banks. In the first 3 columns of Table 2 we show the results from progressively including just year, province, and bank fixed effects. We will eventually also allow for the triple interaction of year-province-bank fixed effects. From column (4) onwards we start including firm and loan specific controls, first linearly as continuous variables and then with fixed effects. The firm-level variables we include in regressions (4) and (5) were the only statistically significant controls that we found. Starting from column (4) we also add the log of deposit costs.[21] We introduce firm fixed effects in the last specification (6).

The only loan level characteristic that we control for is the amount granted, which we assume to be exogenous and determined by the liquidity needs a firm expects to have for that specific year. We will discuss later in greater detail this assumption of exogenous amount granted, justified by the non-exclusive nature of these lending contracts. For now, supporting evidence of this claim is given by the negative relationship between amount granted and interest rates shown in columns (4) to (6) of Table 2, which implies that in absence of exclusivity no convex price schedule can be implemented, because if interest rates rise with the amount borrowed, borrowers can ”linearize” the schedule by opening several credit lines with multiple banks (Chiappori and Salanié (2013)). We don't have other loan level information as these contracts are yearly uncollateralized credit lines, and exhibit therefore no heterogeneity in maturity, collateral, covenants and/or other features. We control for loan amount both linearly and using fixed effects. The decision to discretize the distribution of amounts granted comes from the empirical distribution of these loans shown in Figure 1, as it appears to have a significant number of observations around a few mass points. For example, over 40% of the loans we consider are of either exactly €50,000, €100,000, or €200,000.

[20] We also experimented with using a LASSO regression, but it didn't improve our results as we mostly rely on fixed effects in our preferred specification.

[21] This variable is a proxy for deposit costs that a bank faces in a year-province. It is constructed as a combination of year-region level average deposit rates and the share of branches that each bank has in each province-year.

Figure 1: Distribution of Amount Granted



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

Table 2: Reduced Form Pricing OLS regressions

Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Constant	16.09***	15.39***	15.85***	16.29***	15.39***	16.00***
	(0.07)	(0.15)	(0.28)	(0.17)	(0.13)	(0.10)
Sales	-	-	-	-0.10***	-0.09**	-
				(0.04)	(0.04)	
Total Assets	-	-	-	0.17***	0.19***	-
				(0.05)	(0.06)	
Net Assets	-	-	-	-0.63***	-0.65***	-
				(0.17)	(0.19)	
Short Term Debt	-	-	-	-0.76***	-0.86***	-
				(0.21)	(0.25)	
Profits	-	-	-	-0.14***	-0.11*	-
				(0.06)	(0.06)	
Cash Flow	-	-	-	0.29***	0.26***	-
				(0.07)	(0.08)	
Leverage	-	-	-	-0.04***	-0.04*	-
				(0.02)	(0.02)	
Distance to Branch	-	-	-	0.00	0.00	0.00
				(0.00)	(0.00)	(0.00)
Log of Deposit Costs	-	-	-	0.10***	-	0.11***
				(0.02)	-	(0.02)
Amount Granted	-	-	-	-2.40***	-2.42***	-
				(0.04)	(0.05)	
50,001-100,000	-	-	-	-	-	-0.91***
						(0.09)
100,001-150,000	-	-	-	-	-	-1.55***
						(0.09)
150,001-200,000	-	-	-	-	-	-1.98***
						(0.10)
200,001-300,000	-	-	-	-	-	-2.18***
						(0.10)
300,001-400,000	-	-	-	-	-	-2.63***
						(0.10)
400,001-500,000	-	-	-	-	-	-2.87***
						(0.11)
500,001-1,000,000	-	-	-	-	-	-3.05***
						(0.10)
1,000,001-3,000,000	-	-	-	-	-	-3.46***
						(0.12)
Sector FE	No	No	No	Yes	Yes	No
Score FE	No	No	No	Yes	Yes	No
Year FE	Yes	Yes	Yes	Yes	No	No
Province FE	No	Yes	Yes	Yes	No	No
Bank FE	No	No	Yes	Yes	No	Yes
Bank-Year-Province FE	No	No	No	No	Yes	No
Firm FE	No	No	No	No	No	Yes
R^2	0.2025	0.2275	0.2562	0.3029	0.4045	0.7000
N obs.	92,602	92,602	92,602	92,596	92,596	92,602

Note: An observation is a firm-bank. This sample only includes the first year that a firm appears in our sample, excluding the first year 1988. Standard errors are clustered at the bank-province-year level. Firm controls for regressions (4) and (5) are rescaled to interpret the coefficients more easily: the linear term for Amount Granted is in €10,000, Sales, Total Assets, Net Assets and Short Term Debt are in millions of €, Profits and Cash Flow are in €100,000. Based on statistical significance and sub-sector homogeneity, we construct the Sector fixed effects grouping sectors into 3 categories: *Primary* for primary, minerals' extraction, chemicals, metals, energy; *Manufacturing* for food and beverages, textile and clothing, wood and paper and publishing, mechanical and electronic machines, production of transport vehicles, other manufacturing, constructions; *Commerce and Services* for commerce of transport vehicles, other commerce, hotels and restaurants, transports and storing and communications, real estate, financial intermediaries, public administration.

Based on the price regressions presented above, we now investigate whether the unexplained variation in prices is a predictor of subsequent firms' default. We want to choose the pricing model that minimizes the part of unexplained price variation which correlates with ex-post risk. For this reason we predict the residuals from each of the regressions above, and use them as explanatory variable in a default regression. We use a linear probability model for ease of interpretation, but estimates from a discrete choice regression yield similar results. For each specification we use the same controls as in each pricing equation, apart from (6) as firms default on all lines almost simultaneously, so we have one observation per firm and cannot use firm fixed effects anymore. As shown in Table 3, we find that in all but one specifications the residuals have a positive and significant effect on default, however this effect is economically very small. In specifications (1) to (5) we find that 1 standard deviation increase in the residuals increases default by a range between 7.5% and 4.2% of its standard deviation. These results provide evidence that only the pricing residuals from the last specification (6) are uncorrelated with firms' defaults.

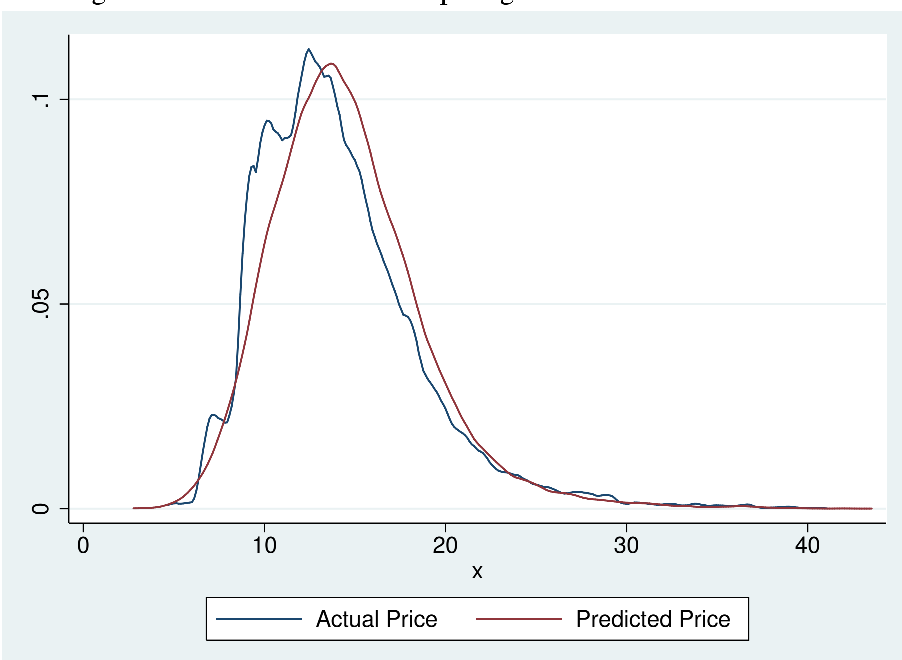
Table 3: Reduced Form Default OLS regressions

Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Residual	0.004***	0.003***	0.003***	0.003***	0.003***	0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Residual t-stat	14.23	12.55	12.41	9.79	7.90	0.94
Residual Mean	0.16	0.12	0.11	0.14	0.13	-0.02
Residual SD	4.45	4.39	4.32	4.20	3.86	2.17
Default Mean	0.06	0.06	0.06	0.06	0.06	0.06
Default SD	0.24	0.24	0.24	0.24	0.24	0.24
1 Residual SD vs % of 1 Default SD	7.4%	5.5%	5.4%	4.6%	4.2%	0.1%
Amount Granted FE	No	No	No	No	No	Yes
Sector FE	No	No	No	Yes	Yes	Yes
Score FE	No	No	No	Yes	Yes	Yes
Firm Controls	No	No	No	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	No	No
Province FE	No	Yes	Yes	Yes	No	No
Bank FE	No	No	Yes	Yes	No	No
Bank-Year-Province FE	No	No	No	No	Yes	Yes
R^2	0.0285	0.0504	0.0564	0.0959	0.3219	0.3210
N obs.	35,319	35,319	35,319	35,316	35,316	35,316

Note: An observation is a firm-bank. Standard errors are clustered at the bank-province-year level. All the specifications are the same as in Table 2, apart from column (6).

We decide to use regression (6) as our baseline pricing model for three main reasons. First, it has the best predictive power in terms of R^2 among all the models we experimented with. Second, its pricing residuals are uncorrelated with firms' default, implying that the price variation unexplained by our model doesn't convey any soft information that the bank uses to price risk. Third, the firm fixed effects in the pricing regression control for any firm-level unobservables that would otherwise cause selection bias. We show the overlap of the kernel densities of actual and predicted prices for borrowing firms in Figure 2. Another important dimension to consider to evaluate our price prediction is the comparison with the existing literature. The dispersion of loan interest rates offered by banks to small and medium enterprises has been documented in various papers in the empirical banking literature (Petersen and Rajan (1994), Berger and Udell (1995), Degryse and Ongena (2005)). According to Cerqueiro et al. (2011) it is actually an empirical regularity that contracted loan rates are typically difficult to predict. The authors estimate a loan-pricing model and compare their model fit to various papers that attempted to construct similar models using lender, borrower and contract information. They find an R^2 of 25%, whereas Petersen and Rajan (1994) obtained an R^2 of 14.5%. Degryse and Ongena (2005) get an R^2 of 22%, but this increases to 67% when they focus on larger loans (above \$ 50,000), and decreases to 1% for smaller loans (below \$ 5,000). We obtain an R^2 of 70%.

Figure 2: Kernel Densities Comparing Actual and Predicted Prices



The pricing regression we are running for a firm i borrowing from bank j at price P_{ij} takes the following

form:

$$P_{ij} = \underbrace{\widehat{\gamma}_0 + \widehat{\gamma}_1 Z'_{ij} + \widehat{\lambda}_j + \widehat{\omega}_i}_{\widetilde{P}_{ij}} + \epsilon_{ij}, \tag{1}$$

where $\widehat{\omega}_i$ and $\widehat{\lambda}_j$ are firm and bank fixed effects, and Z'_{ij} includes amount granted, distance between firms and banks, and deposit costs. We omit market m and time t subscripts, as each firm appears only in one market at one point in time. Using combinations of $\widehat{\gamma}_0$, $\widehat{\gamma}_1$, $\widehat{\lambda}_j$, and $\widehat{\omega}_i$ we are able to predict prices $\widetilde{P}_{ij}$ offered to borrowing firms from banks they ended up not choosing. To predict prices offered to non-borrowing firms we use propensity score matching, that is we match several borrowing firms to non-borrowing firms that are similar in observables' space, and then randomly assign a borrowing firm's $\widehat{\omega}_i$ to a matched non-borrowing firm. We assign the amount granted to non-borrowing firms using the same approach.[22] A similar approach has also been used in Adams, Einav and Levin (2009).

This method to predict prices and match borrowers to non-borrowers has two potential drawbacks. First, our pricing model is able to explain around 70% of the variation in loan interest rates. Second, 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. These drawbacks can generate measurement error in the price, so we want to investigate the potential direction of the bias in our main coefficients of interest, outlying here a stylized version of our structural model. We define a firm i's utility from demanding (superscript D) from a bank j and defaulting (superscript F) that just depend on price and unobservables. We decompose prices into what we can and cannot predict, where ϵ_{ij} is the price variation unexplained by our regression, and ν_{ij} is the measurement error from a potential wrong assignment of firm fixed effects. We identify adverse selection as a positive correlation between the unobservables of these two equations:

$$\begin{aligned}
U^D_{ij} &= \alpha^D P_{ij} + \varepsilon^D_{ij} \\
&= \alpha^D \left[\widetilde{P}_{ij} + \epsilon_{ij} + \nu_{ij}\right] + \varepsilon^D_{ij} \\
&= \underbrace{\alpha^D \widetilde{P}_{ij}}_{\substack{\text{Hard information}\\ \text{observed by bank}\\ \text{and econometrician}}} + \underbrace{\alpha^D(\epsilon_{ij}+\nu_{ij})}_{\substack{\text{Soft information}\\ \text{observed by bank}\\ \text{but not econometrician}}} + \underbrace{\varepsilon^D_{ij}}_{\substack{\text{Soft information}\\ \text{unobserved by bank}\\ \text{and econometrician}}} \\
U^F_{ij} &= \overbrace{\alpha^F \widetilde{P}_{ij}} + \overbrace{\alpha^F(\epsilon_{ij}+\nu_{ij})} + \overbrace{\varepsilon^F_{ij}} \quad .
\end{aligned} \tag{2}$$

One of the features of this set up is a non-classical form of measurement error in prices identified by $\epsilon_{ij}+\nu_{ij}$. We observe in our data that $\epsilon_{ij}+\nu_{ij}$ is uncorrelated with $\widetilde{P}_{ij}$, so following Pischke (2007) this yields to consistent OLS estimates and correct standard errors. There is another more substantial concern related to this measurement error in prices, that is the identification of adverse selection. Following the literature, we identify adverse selection with a positive correlation between the unobservables that drive demand and default. In this set up these unobservables are $\alpha^D(\epsilon_{ij}+\nu_{ij})+\varepsilon^D_{ij}$ for demand and $\alpha^F(\epsilon_{ij}+\nu_{ij})+\varepsilon^F_{ij}$ for default. If there was no measurement error, we would interpret a positive correlation between ε^D_{ij} and

[22] A detailed description of the matching model is in the Appendix.

ε_{ij}^F as adverse selection. As expected, what we find in our structural model is that $\alpha^D < 0$ and $\alpha^F > 0$, which means a negative correlation between $\alpha^D(\epsilon_{ij} + \nu_{ij})$ and $\alpha^F(\epsilon_{ij} + \nu_{ij})$. This implies that finding a positive correlation between unobservables in the presence of this kind of measurement error, as we find, is actually underestimating the true extent of adverse selection. Our positive correlation result can therefore be interpreted as a lower bound to the true value of adverse selection.

3 Reduced Form Evidence

3.1 Asymmetric Information

We conduct some reduced form analysis to test for evidence of asymmetric information and to justify the use of a structural model. To do so we follow the early empirical literature on positive correlation tests introduced by Chiappori and Salanié (2000). We propose two tests, one based on the choice to take up a loan and another based on the choice of how much to draw on the credit 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 will rely on in the structural model, explained in Section 4.

3.1.1 Demand and Default

We start by investigating whether firms that are more likely to demand credit are also more likely to default. The CB dataset includes both firms borrowing and not borrowing, while we only observe default on the loan for borrowing firms. We can formalize the problem as a two equations selection model:

$$\begin{aligned} d_i &= \mathbf{1}(X_i^d \beta + \nu_i > 0) \\ f_i &= \begin{cases} X_i^f \gamma + \eta_i & \text{if } d_i = 1 \\ - & \text{if } d_i = 0 \end{cases} \end{aligned} \tag{3}$$

where d_i is equal to 1 if the firm borrows and f_i is equal to one if the borrower is a defaulter. 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 and η.[23] Results of this Heckman selection model are reported in the first two columns (*Extensive Margin*) of Table 4, where the decisions to borrow and default are regressed on year, province, score, amount granted, and sector fixed effects, as well as on a set of relevant forms' balance sheet variables. We use as instruments in the selection (i.e. borrowing) equation variables that satisfy the exclusion restriction, having a statistically significant effect on demand but not on default: the number of banks in a firm's market, as a proxy for banks' competition, and age of

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

the firm, capturing a firm's reputation and ability to negotiate with banks. We find a positive and significant correlation coefficient between the unobservables driving demand and default, which we interpret as preliminary evidence of adverse selection on the extensive margin.

3.1.2 Loan Use and Default

We then consider the relationship between loan use and default. Differently from the previous subsection, 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 "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 regressions:

$$\begin{aligned} \ell_i &= X_i\beta + \varepsilon_i \\ f_i &= X_i\gamma + \eta_i \end{aligned} \tag{4}$$

where ℓ_i is the amount of loan used, and f_i takes value of one if the borrower is a defaulter. The vector of controls X_i is composed of year, province, bank, score, amount granted, and sector fixed effects, as well as on a set of relevant firms' balance sheet variables. We specify the distribution of the residuals ε_i, η_i as joint normal, with a correlation coefficient ρ. Positive and significant ρ suggests the presence of adverse selection. The results of this test are summarized in the last two columns (*Intensive Margin*) of Table 4. We again find a positive correlation, consistent with adverse selection on the intensive margin.

Based on these suggestive results, we estimate a structural model to measure the extent of adverse selection in this market and its consequences for market outcomes. The structural framework has two main advantages compared to these reduced form tests. First, it has a more flexible residuals' correlation structure that allows us to estimate them jointly. Second, jointly with a supply side model we can use it to run counterfactual policy experiments to measure the consequences of adverse selection and imperfect competition. We introduce this model in Section 4.

Table 4: Reduce Form Evidence for Adverse Selection

	Extensive Margin		Intensive Margin	
Variables	**Demand**	**Default**	**Loan Use**	**Default**
Correlation between Unobservables	0.09^{**}		0.032^{***}	
	(0.01)		(0.000)	
Constant	-0.99^{***}	-0.01	-0.11	-0.06^{**}
	(0.14)	(0.02)	(0.40)	(0.03)
Intangible Assets	-1.44^{***}	-0.04^{***}	-0.20	-0.04^{***}
	(0.07)	(0.01)	(0.14)	(0.01)
Ratio of Intangible Assets	-0.40^{***}	0.02^{**}	-0.12	0.02^{**}
	(0.04)	(0.01)	(0.12)	(0.01)
Total Assets	0.66^{***}	0.01^{***}	0.32^{***}	0.00^{**}
	(0.02)	(0.00)	(0.03)	(0.00)
Sales	0.48^{***}	-0.00^{**}	-0.09^{***}	-0.00^{***}
	(0.01)	(0.00)	(0.02)	(0.00)
Profits	0.45^{***}	0.02^{*}	0.58^{***}	0.01
	(0.05)	(0.01)	(0.12)	(0.01)
Cashflow	-0.56^{***}	-0.03^{*}	-2.24^{***}	-0.02
	(0.08)	(0.01)	(0.18)	(0.01)
Trade Credit	-0.18^{***}	-0.00	-0.05^{***}	-0.00
	(0.01)	(0.00)	(0.01)	(0.00)
Age of Firm	0.48^{***}	0.02	-0.58^{***}	0.01
	(0.08)	(0.01)	(0.20)	(0.01)
N. of Banks in Market	0.05^{***}	-	-0.00	0.00
	(0.01)		(0.02)	(0.00)
Price	-	-	-0.02^{***}	0.00^{***}
			(0.01)	(0.00)
Distance to Branch	-	-	-0.00	-0.00^{***}
			(0.00)	(0.00)
Year FE	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes
Bank FE	No	No	Yes	Yes
Score FE	Yes	Yes	Yes	Yes
Amount Granted FE	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes
R^2	0.2601	0.1046	0.4345	0.1009
N Obs.	25,101	18,820	18,820	18,820

Note: In each regression an observation is a firm. We rescale some variables for ease of interpretation. 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.

3.2 Imperfect Competition

We also provide some descriptive statistics on imperfect competition, the other friction in this market that we focus on. We construct a concentration index (HHI - Herfindahl-Hirschman Index) based on each bank's

share of credit used within a province-year, and investigate its correlation with interest rates, conditional on various sets of observables. We use as interest rates the predicted prices described in Section 2.2. We show with various specifications, summarized in Table 5, that concentration is positively and significantly correlated with interest rates, as expected. In our preferred specification in column 4, we find that a 10% increase in concentration is associated with a 0.05% increase in prices. In the first column we show that there is a positive and significant correlation between concentration and price within a year-province, and that this holds also when controlling for bank fixed effects, as well as for bank-year and bank-province fixed effects. Finally, the correlation becomes stronger once we control for loan amount and distance between firm and banks, and for firm fixed effects.

Table 5: Reduced Form Evidence of Imperfect Competition

Variables	**(1)**	**(2)**	**(3)**	**(4)**
Log HHI	0.004^{**}	0.004^{**}	0.004^{**}	0.005^{***}
	(0.002)	(0.002)	(0.002)	(0.002)
Year FE	Yes	No	No	No
Province FE	Yes	No	No	No
Bank FE	Yes	No	No	No
Bank-Year FE	No	Yes	Yes	Yes
Bank-Province FE	No	Yes	Yes	Yes
Loan Controls	No	No	Yes	Yes
Firm FE	No	No	No	Yes
R^2	0.3140	0.3198	0.4207	0.9416
N obs.	285,734	285,734	285,734	285,734

Note: An observation is a firm-year-province-bank. Loan Controls include fixed effects for amount granted and distance between firm and bank.

4 The Model

The framework we construct aims at quantifying the effects of asymmetric information on the demand for and supply of credit for Italian firms. In order to test for this, we assume that each firm $i = 1, ..., I$ is willing to invest in a project and is looking for credit to finance it. Firms in each market m and period t decide which bank $j = 1, ..., J_{mt}$ to borrow from, based on the conditions offered that maximize the expected "profits"[24] of their choice. This determines the demand for credit. Conditional on demand, firms decide the amount of credit to use and whether to default or not. The supply of credit results from banks' static Bertrand-Nash competition on interest rates, an assumption we motivate later in this section.

[24] We will define these profits as utilities later on, to distinguish them from banks' profits.

The theoretical model we develop is based on the following assumptions:

(1) Asymmetric Information: Following Stiglitz and Weiss (1981), we assume that the asymmetry of information is on the riskiness of the firm, known by the firm but not by the bank, whereas the distribution of riskiness among all firms is known by both. We identify this riskiness with the firm's probability of default. We let borrowers and lenders be risk neutral.

(2) First Year of New Loans: We limit our analysis to the first year of newly granted loans. This is a common assumption in empirical models of insurance with asymmetric information, starting from Chiappori and Salanié (2000). This is done to avoid modeling heterogenous experience ratings among borrowers and loan renegotiation, as the focus of the paper is on first access to credit.

(3) Main New Credit Line: We just consider the choice of the main new credit line that firms open for the first time within our sample. The main line is defined as the one from which the firm borrows the most. As shown by Detragiache et al. (2000), in Italy, multiple relationship banking is widely used by firms to reduce liquidity risk. However, the share of the main credit line opened accounts on average for over 70% of the total amount of new yearly credit (both usable and used), justifying the choice of this simplifying assumption.[25]

(4) Exogenous Amount of Credit: We limit our analysis to the interest rate as the only screening device, as in Stiglitz and Weiss (1981). Therefore, we 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 just offers a posted interest rate for that specific amount to each firm i in each market m. In a standard insurance or credit market with asymmetric information, insurers or banks are likely to compete not only on prices, but on other clauses of the contract as well. In our context, the amount granted could be another dimension over which banks compete. In a world with lending exclusivity, banks can offer menus of amounts granted with matched interest rates to reduce the extent of asymmetric information, for example charging rates that increase more than proportionally with the amount granted. However, this is the case only with contract exclusivity, which is not a feature of our setting, where borrowers can open multiple credit lines with different lenders. As explained in Chiappori and Salanié (2013), in absence of exclusivity no convex price schedule can be implemented, because if interest rates rise with the amount borrowed, borrowers can "linearize" the schedule by opening several credit lines with multiple banks. Empirical evidence of non-exclusivity results also from the pricing regression in Table 2, which presents a negative correlation between interest rates and the amount of credit granted.[26] We also assume no collateral, as the type of loans we analyze are uncollateralized. We do however allow for an endogenous amount of loan used.

---

[25] We tried to make use of a borrower's ranking of its lenders, in terms of amount used, for identification purposes. However, only a subset of the firms in our sample borrows from multiple banks, so we ended up not using this information.

[26] We thank Pierre-André Chiappori for his suggestions on this point.

4.1 Demand, Loan Use and Default

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

$$U^D_{ijmt} = \underbrace{\bar{\alpha}^D_1 + X'_{jmt}\beta^D + \xi^D_{jmt}}_{\delta^D_{jmt}} + \underbrace{\sigma^D\nu_i - \alpha^D_2\widetilde{P}_{ijmt} + Y'_{ijmt}\eta^D}_{V^D_{ijmt}} + \varepsilon^D_{ijmt}. \tag{5}$$

We let $U^D_{i0mt} = \varepsilon^D_{i0mt}$ be the utility from the outside option, which is not borrowing. Firms will choose the bank that maximizes their utility, or will choose not to borrow. Then, conditional on borrowing, they will choose the amount of credit to use that maximizes the following utility:

$$U^L_{ijmt} = \underbrace{\alpha^L_1 + X'_{jmt}\beta^L + \xi^L_{jmt}}_{\delta^L_{jmt}} - \underbrace{\alpha^L_2 P_{ijmt} + Y'_{ijmt}\eta^L}_{V^L_{ijmt}} + \varepsilon^L_{ijmt}. \tag{6}$$

Finally, conditional on borrowing, they will choose to default if the following utility is greater than zero:

$$U^F_{ijmt} = \underbrace{\alpha^F_1 + X'_{jmt}\beta^F + \xi^F_{jmt}}_{\delta^F_{jmt}} + \underbrace{\alpha^F_2 P_{ijmt} + Y'_{ijmt}\eta^F}_{V^F_{ijmt}} + \varepsilon^F_{ijmt}. \tag{7}$$

Here X_{jmt} are banks' observable attributes, $\widetilde{P}_{ijmt}$ are the predicted interest rates described in Section 2.2, P_{ijmt} are actual interest rates, which we observe for firms that demand, ξ_{jmt} are banks' unobservable (to the econometrician) attributes, and Y_{ijmt} are firm specific and firm-bank specific observable characteristics. We assume that ε^D_{ijmt} is distributed as a type 1 extreme value, following the literature on demand estimation for differentiated products (Berry (1994), Berry et al. (1995)). We let the random coefficient of the demand's constant term $\alpha^D_{1i} = \bar{\alpha}^D_1 + \sigma^D\nu_i$,[27] with $\nu_i \sim N(0, 1)$,[28] to be jointly normally distributed with ε^L_{ijmt}, and ε^F_{ijmt}, such that:

$$\begin{pmatrix} \alpha^D_1 \\ \varepsilon^L \\ \varepsilon^F \end{pmatrix} \sim N\left(\begin{pmatrix} \bar{\alpha}^D_1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^{D2} & \rho_{DL}\sigma^D\sigma^L & \rho_{DF}\sigma^D \\ \rho_{DL}\sigma^D\sigma^L & \sigma^{L2} & \rho_{LF}\sigma^L \\ \rho_{DF}\sigma^D & \rho_{LF}\sigma^L & 1 \end{pmatrix} \right). \tag{8}$$

We interpret a positive correlation between the firm specific unobservables driving demand and default (ρ_{DF}) as evidence of adverse selection. The intuition is that if the unobservables that drive demand are positively correlated with the unobservables that drive default, then riskier firms are more likely to demand loans. The idea behind the identification of the correlation between α^D_1 and ε^F is the following. If we

[27] Following Nevo (2000b), we interpret $-\alpha^D_2\widetilde{P}_{ijmt} + Y'_{ijmt}\eta^D$ as observed heterogeneity in the constant random coefficient. Given that the constant is normalized to zero for the outside option, also $-\alpha^D_2\widetilde{P}_{ijmt} + Y'_{ijmt}\eta^D$ will be zero for the outside option in order for α^D_2 and η^D to be indentified. These demographics help us to control for the observable sources of the borrower's taste for credit (regardless of which bank it chooses), leaving ν_i as the unobservables taste for credit.

[28] We use 100 Halton draws for simulation. According to Train and Winston (2007), 100 Halton draws achieve greater accuracy in mixed logit estimations than 1,000 pseudo-random draws.

observe a firm taking out a loan, while the model tells us that this firm should be unlikely to take the loan, then this is a "high α_1^D" firm. A positive correlation of α_1^D with ε^F is evidence of adverse selection.

We interpret a positive correlation between the unobservables driving loan usage and default (ρ_{LF}) as other possible evidence of adverse selection. The intuition is that if the unobservables that drive the choice of how much credit to use are positively correlated with the unobservables that drive default, then riskier firms will use more credit. With this definition of adverse selection we are trying to capture the case in which a risky firm (high ε^F), before signing the contract, already knows that due to its high ε^L it will use a higher share of the loan. However, our definition cannot rule out the case in which two ex-ante equally risky firms take the same loan, and one of them is hit by a negative shock after the contract has been signed. This shock increases ε^L for the firm that was hit, forcing it to use more of the loan.[29] This is however not a major concern in our case, as we just focus on the first year of the firm-bank relationship, when firms are less leveraged as we show in the next section. The correlation between unobservables driving demand and loan use (ρ_{DL}) doesn't have a clear economic interpretation in terms of asymmetric information, but it's important to estimate it jointly with the other elements of the variance-covariance matrix, to avoid capturing with ρ_{DF} and ρ_{LF} any possible spurious correlation. The joint estimation of these parameters guarantees a better identification of adverse selection compared to the reduced form estimates, where each correlation coefficient was estimated separately. Note that this identification strategy allows us to recover adverse selection parameters that are common across banks and markets, not bank or market specific.[30]

This set up builds on Einav et al. (2012), but differs in the specification of the demand utility. In our case, borrowers' choices follow a multinomial distribution, instead of a binomial. This raises the issue of correlating residuals from the demand model, which vary across borrowers and alternatives (i.e. lenders), to the residuals from the loan use and default models, which instead vary only across borrowers. We follow the approach of Ackerberg and Botticini (2002) and allow the normally distributed random coefficient on the constant term to be correlated with the residuals from the loan use and default equations. We argue that this is a practical and intuitive solution, as it simplifies the problem and allows for a correlation between unobservables only at the level of the borrower. This implies that in the presence of adverse selection a riskier firm is more likely to demand from any lender, and not differently across different lenders.

We interpret as possible evidence of moral hazard a positive effect of interest rates on default ($\alpha_2^F > 0$), which implies that an increase in this relevant term of the loan contract makes it more likely for a borrower to default. We provide this interpretation for α_2^F because we use a control function approach that allows us to identify the causal effect of interest rates on default,[31] following Adams et al. (2009), and also because this effect is conditional on adverse selection, modeled through the correlation between unobservables.

Last, as usual in structural estimation, our framework is a compromise between tractability and modeling complexity. We attempted to introduce more structure into the model, following more closely Stiglitz and Weiss (1981)'s set up, but faced substantial identification problems. We therefore decided to rely more closely on the empirical literature on credit markets with selection (Einav et al. (2012)), and on demand for

[29] In this case, ρ_{LF} could be interpreted as evidence of either adverse selection or moral hazard. See Abbring et al. (2003) for distinguishing between those sources of asymmetric information.

[30] Extending the model to allow for heterogeneity across banks is scope for future research.

[31] See Section 6.2 for a detailed discussion on identification.

differentiated products (Berry (1994), Berry et al. (1995)), which also has its recent applications in banking and finance (Ho and Ishii (2011), Koijen and Yogo (2015), Egan, Hortaçsu and Matvos (2015)). If on one hand our approach is closer to a more reduced form selection model, on the other it offers a good tractability and allows us to incorporate both imperfect competition and selection effects in a structural pricing equation. Based on this model we can implement several counterfactual exercises, that can offer relevant guidance for policy interventions.

4.2 Alternative Frictions

We interpret a positive correlation between demand for credit and firms' riskiness as evidence of adverse selection. An alternative explanation is that such correlation 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 sub-optimal decisions to serve the shareholders' interests. In our context this can imply risk shifting by firms with more debt, which in turn decreases the quality of firms' projects and increases their default probability. Another related plausible explanation for our results follows from Myers (1977), who argues that firms with more debt are more likely to run into a debt overhang situation, declining to fund good projects and increasing their default probability. Although theoretically possible, we believe that these alternative explanations are unlikely to hold in our data for several reasons.

First, we rely on a sample of SMEs, most of which are owned and controlled by an individual or a family.[32] In these firms, agency conflicts can only arise between the firm and the bank, as ownership is concentrated and bank debt is the main source of finance (together with trade credit).[33] Typically, owners of family firms tend to immobilize a large portion of their overall wealth in the firm itself (Moskowitz and Vissing-Jorgensen (2002)) and, as a consequence, they undertake less risky, more conservative projects than a well diversified owner or an external CEO (Michelacci and Schivardi (2013)). Excessive risk taking is therefore unlikely to arise from the financial structure. Second, we focus on the main lender, which therefore has strong incentives to exert monitoring and spot suboptimal investment choices. Third, we look at the first line of credit. As shown in Table 6, we consider firms with a level of debt still relatively low, while the theories above apply to firms with a high level of leverage. Fourth, in all our regressions we control for various indicators of incentives to engage in risk shifting, such as leverage (a measure of financial vulnerability), net worth, cash flows, profits and trade credit. Finally, Myers (1977) also suggests an easy way out from the suboptimal investments caused by agency costs, which is shortening debt maturity. He states that permanent debt capital is best obtained rolling over short maturity debt claims, with continuous and gradual renegotiation that allows the firm to switch to other sources of debt at any time. In our setting we do focus on a very short maturity form of debt. In fact, credit lines can be closed at any time by the bank, so the specific contract we consider is less likely to give rise to the agency issues discussed above. Indeed, in our data, firms rely mostly on short term bank debt and trade credit, which are both short term debts, thus reducing agency

[32] Bugamelli, Cannari, Lotti and Magri (2012) show that 85.6% of Italian firms are family businesses, and in 66.3% of these the family also manages the firm, compared to the 25.8% in France, 28% in Germany, 35.5% in Spain, and 10.4% in the UK.

[33] Barba Navaretti, Bugamelli, Schivardi, Altomonte, Horgos and Maggioni (2011) show that the share of venture capital in Italian firms' external financing is between 0.35% and 0.52%.

concerns.

Table 6: Leverage for New vs Old Borrowers

Sample	Obs	Mean	Std. Dev.	5^{th} Pctile	Median	95^{th} Pctile
New Borrowers	54,014	0.503	0.914	0	0.554	0.997
Old Borrowers	400,008	0.548	4.670	0	0.596	0.987

Note: An observation is a firm-year. Leverage is defined as a firm's debt over its liabilities. New borrowers are firms that borrow for the first year in our sample, old borrowers are firms borrowing from the second year onwards.

Another important aspect that might affect our results is the possibility that firms and banks adjust their structure of financing as asymmetric information varies. Again, although theoretically sound, this is unlikely to play an important role in our data. As already discussed, SMEs finance their operations mostly through bank loans. 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 markets and SMEs where very unlikely to list. Private equity financing was also very rare and the market of bonds for small firms non existent, also due to legal restrictions to bond issuance for SMEs. As such, firms had little alternative to bank debt, so that it is unlikely that changes in the degree of asymmetric information could lead to substantial changes in firms' financial choices.

4.3 Supply

On the supply side, we let banks set their interest rates competing à la Bertrand Nash. We assume that bank j's expected profits from firm i in market m at time t are given by:

$$\begin{aligned} \Pi_{ijmt} &= (\widetilde{P}_{ijmt} - MC_{ijmt})Q_{ijmt}(1 - F_{ijmt}) - MC_{ijmt}Q_{ijmt}F_{ijmt} \\ &= \widetilde{P}_{ijmt}Q_{ijmt}(1 - F_{ijmt}) - MC_{ijmt}Q_{ijmt}, \end{aligned} \tag{9}$$

where Q_{ijmt} and F_{ijmt} are banks' expectation of each firm's demand and default. In particular, Q_{ijmt} is given by the model's demand probability and the expected loan use, and F_{ijmt} is the expected default rate for each borrower. $\widetilde{P}_{ijmt}$ is the posted price of the loan, and MC_{ijmt} are the bank's marginal costs. It is important to note that F_{ijmt} depends on price through two channels. First, equation (7) allows for a direct impact of the interest rate on firms' default probabilities. Second, a higher interest rate also changes the composition of borrowers as stated in Assumption 1: increasing price increases the conditional expectation of α_1^D, as low-utility-from-borrowing firms are more likely to self-select out of the borrowing pool. If $\rho_{DF} > 0$, this implies that an increase in prices increases the probability of default of the pool of borrowers.

The first order conditions of this profit function deliver the following pricing equation:

$$\widetilde{P}_{ijmt} = \underbrace{\frac{MC_{ijmt}}{1 - F_{ijmt} - F'_{ijmt}\frac{Q_{ijmt}}{Q'_{ijmt}}}}_{\textit{Effective Marginal Costs}} + \underbrace{\frac{-(1 - F_{ijmt})\frac{Q_{ijmt}}{Q'_{ijmt}}}{1 - F_{ijmt} - F'_{ijmt}\frac{Q_{ijmt}}{Q'_{ijmt}}}}_{\textit{Markup}}, \tag{10}$$

Note that the equilibrium price depends on what we define as "effective" marginal costs and on a markup term. F'_{ijmt} is the derivative of the expected default with respect to the price, and Q'_{ijmt} is the derivative of the demand probability with respect to the price. $\frac{Q_{ijmt}}{Q'_{ijmt}}$ would be the markup in a Bertrand-Nash model with differentiated products and no asymmetric information. In fact if there was no default, i.e. $F_{ijmt} = F'_{ijmt} = 0$, we would be back to a standard equilibrium pricing equation for differentiated firms competing à la Bertrand-Nash as in Berry et al. (1995). We will analyze this equilibrium pricing equation in greater detail in the next section.

4.4 Monte Carlo

We construct a simple numerical example to give the intuition underlying the model's predictions. We simulate data for the case of a monopolist bank facing $i = 1, ..., I$ heterogeneous borrowers, observationally equivalent to the bank. For simplicity, we concentrate on adverse selection between demand and default (ρ_{DF}), setting loan use to 1 and $\rho_{DL} = \rho_{LF} = 0$. We keep this data fixed and vary both borrowers' price sensitivity, as a proxy for the strength of the effects of a competitive fringe on the bank's (residual) demand curve, and the extent of asymmetric information, 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 prices based on our model. Let borrower i have U_i^D utility from taking credit from the bank, U_{i0}^D utility from not borrowing, and U_i^F utility from defaulting:

$$\begin{aligned} U_i^D &= \alpha_{1i} + \alpha_2 P + \epsilon_i, \\ &= \bar{\alpha}_1 + \sigma\nu_i + \alpha_2 P + \epsilon_i, \\ U_{i0}^D &= \epsilon_{i0}, \\ U_i^F &= \varepsilon_i, \end{aligned} \tag{11}$$

where P is the interest rate charged by the bank, ϵ_i, ϵ_{i0} are distributed as type 1 extreme value, and $\nu_i \sim N(0, 1)$. We set $\sigma = 2$ and $\bar{\alpha}_1 = 1$, and allow α_{1i} and ε_i to be jointly normally distributed, with correlation coefficient $-1 \leq \rho \leq 1$. We assume that all the borrowers have the same price sensitivity $\alpha_2 < 0$. Our asymmetric information assumption implies that a bank doesn't observe its borrowers' individual default probability, but only its distribution. As a consequence, it only offers one pooling price P for everyone.

Given this setup, the demand probability will be given by:

$$
\begin{aligned}
\Pr_i^D &= \Pr(\alpha_{1i} + \alpha_2 P + \epsilon_i > \epsilon_{i0}) \\
&= \tfrac{\exp(\alpha_{1i}+\alpha_2 P)}{1+\exp(\alpha_{1i}+\alpha_2 P)} \\
&= \Lambda(\alpha_{1i} + \alpha_2 P),
\end{aligned} \tag{12}
$$

and we will construct the bank's market share as $S = \frac{1}{N}\sum_i \Pr_i^D$, where N is the number of borrowers of the bank. Conditional on demand, the default probability is (Wooldridge (2002)):

$$
\begin{aligned}
\Pr_i^{F|D=1} &= E\big[\Pr(F=1|\nu, P)|D=1, P\big] \\
&= \tfrac{1}{\Lambda(\alpha_{1i}+\alpha_2 P)} \int_{-(\alpha_{1i}+\alpha_2 P)}^{\infty} \Phi\left(\frac{\frac{\rho\nu}{\sigma^2}}{\sqrt{1-\frac{\rho^2}{\sigma^2}}}\right)\phi(\nu)d\nu,
\end{aligned} \tag{13}
$$

and we will construct the bank's share of defaulters as $F = \frac{1}{N}\sum_i \Pr_i^F$. Given these probabilities and our supply side model described in equations (9) and (10), the first order condition will deliver the following equilibrium pricing equation:

$$
P^* = \underbrace{\frac{MC}{1 - F - F'\frac{1}{\alpha_2(1-S)}}}_{\textit{Effective Marginal Costs}} + \underbrace{\frac{-(1-F)\frac{1}{\alpha_2(1-S)}}{1 - F - F'\frac{1}{\alpha_2(1-S)}}}_{\textit{Markup}}, \tag{14}
$$

where the first term on the right hand side represents what we define as "effective" marginal cost (EffMC), and the second term represents the markup (MKP). F' is the derivative of the expected default rate with respect to price, and $\alpha_2(1-S)$ is the derivative of the market share with respect to price.

The different effects of EffMC and MKP on the equilibrium price is crucial to understanding the interaction between asymmetric information and imperfect competition. This is displayed in Figures 3 and 4, where the top graph represents EffMC above and negative of the markup below, and the bottom graph shows equilibrium prices for a monopolist bank. We let these three elements vary across different degrees of adverse selection, measured by ρ, and "competition", measured by α_2. In this example we are capturing competition versus the outside option, but we have verified that increasing the number of banks in the model gives the same qualitative result.

Looking at Figure 3, for a high level of competition (i.e. rightmost point on the figure) an increase in adverse selection (moving to the northwest) causes EffMC to increase, whereas for low competition (point closest to the reader, again moving northwest) they remain relatively constant. The intuition for this result is the following. Higher adverse selection implies higher correlation between borrowers' willingness to pay (WTP) and their riskiness. Hence, with strong competition only firms with high WTP will borrow, whereas with less competition even firms with low WTP will take credit, lowering the average riskiness of the pool of borrowers. The opposite happens for the markup curve as we increase adverse selection, because it remains nearly constant for high competition (rightmost point, moving to the northwest), but it decreases substantially for a low level of competition (closest point to the reader, moving to the northwest). The reason for this sharp reduction in markup with low competition as adverse selection increases is that the marginal

borrowers become the safest ones, implying a reduction in the banks' market power to keep them. Finally, what the graph shows is that both an increase in adverse selection and an increase in competition reduces a bank's markup, implying that adverse selection has a mitigating effect on market power.

As shown in Figure 4, the combination of these two factors results in a non-monotonic equilibrium price response to an increase in adverse selection. If on one hand the equilibrium price rises in a very competitive environment (closest point to the reader, moving to the northeast), the opposite happens in a concentrated market (leftmost point, moving to the northeast). This is because in the first case the price depends more on EffMC than markup and the increasing EffMC drives the price up. In a highly competitive market where banks have a small price-cost margin, a higher price is the only possible response to an increase in adverse selection. In the second case, the price depends more on markup and the declining markup drives the price down. Banks with a higher price-cost margin will find it more profitable to reduce the price, as this will allow them to lower the average riskiness to their pool of borrowers.

This effect can be better understood focusing on how the denominator of EffMC and MKP varies differently in response to an increase in adverse selection, depending on the measure of competition α_2. Higher adverse selection means higher default rates, therefore an increase in both the second (F) and third ($F' \frac{1}{\alpha_2(1-S)}$) terms of the denominator. On one hand a higher F reduces the denominator, pushing for an increase in prices, but on the other hand a higher F' raises the denominator, because $\alpha_2 < 0$. Whether the net effect of higher adverse selection is a higher or lower denominator depends on the relative importance of a bank's market power, determined by α_2 in our setting. Low market power, or very strong competition with the outside option ($\alpha_2 \rightarrow \infty$), means that the effect of F dominates, so prices will increase in response to an increase in adverse selection. High market power, or very weak competition with the outside option ($\alpha_2 \rightarrow 0$), means that the effect of F' dominates, so prices will decrease in response to an increase in adverse selection. In other words, market power weights the relative importance that banks' give in their equilibrium pricing to the marginal default response to an increase in prices.

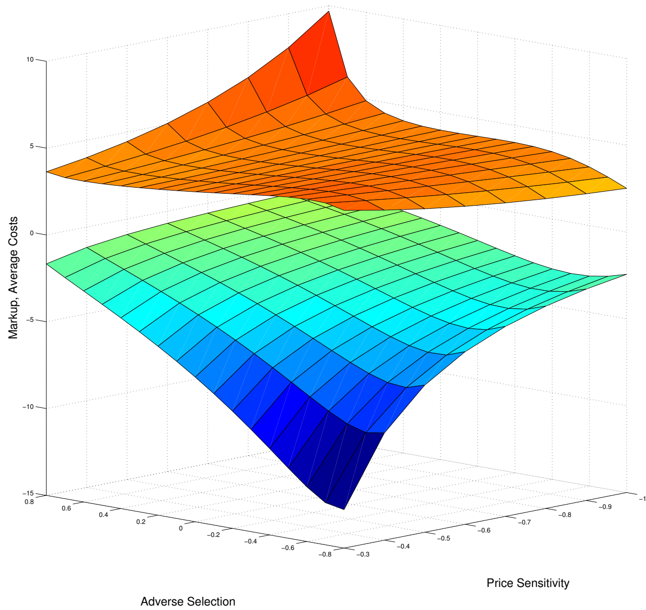
In this simplified example the increase in adverse selection affects the derivative of default with respect to prices (F') only through the change in the correlation coefficient. However, in the structural model we will also allow for another channel to affect F', that is the direct effect of prices on default. This effect can be interpreted as moral hazard, as it measures how interest rates affect the default probability conditional on the selection effect.

5 Econometric Specification

Following the model presented above, let $m = 1, ..., M$ index a province, $t = 1, ..., T$ a year, $i = 1, ..., I$ the firm, and $j = 1, ..., J_{mt}$ be the bank/loan identifier in market m at time t. Let Y_{ijmt} be a vector of firm and firm-bank specific characteristics (firm's balance sheet data, firm's age, and firm's distance to the closest branch of each bank), and X_{jmt} a vector of bank-province-year specific attributes (number of branches in the market, years of presence in the market, bank, year, and province fixed effects).

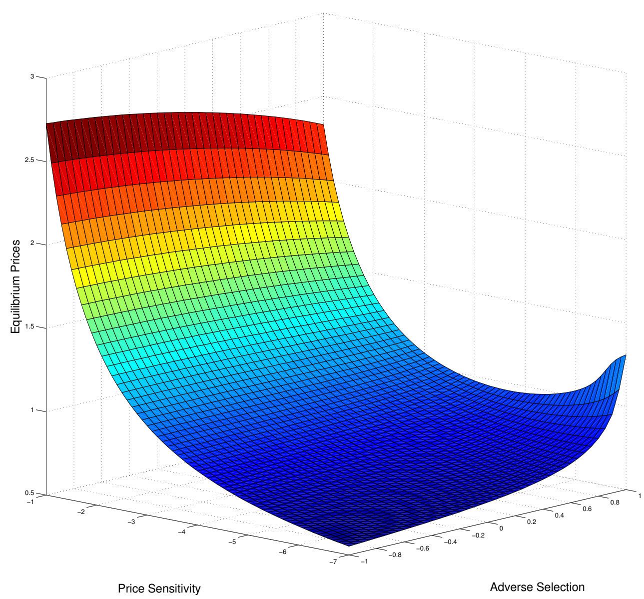
We estimate a system of three equations: demand for credit, amount of loan used, and default. For the

Figure 3: Adverse Selection vs Imperfect Competition - Effective Marginal Costs, negative Markups



Note: The vertical axis shows the value of effective marginal costs and of the negative of the markup. The left horizontal axis is level of adverse selection, increasing towards left. The right horizontal axis is the level of price sensitivity (our measure of competition with the outside option), increasing towards the right.

Figure 4: Adverse Selection vs Imperfect Competition - Equilibrium Prices



Note: The vertical axis shows the level of equilibrium prices. The left horizontal axis is level of price sensitivity (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 axis definitions in this figure differ from those in Figure 3 to better display the effects in each.

demand model we use a 2-step method based on maximum simulated likelihood and OLS, as in Train (2009). In the first step we estimate the firm-level parameters $\eta = \{\alpha_2^D, \alpha_2^L, \alpha_2^F, \eta^D, \eta^L, \eta^F\}$, bank-level parameters for loan use and default $\beta^{LF} = \{\alpha_1^L, \alpha_1^F, \beta^L, \beta^F\}$ and the elements of the variance-covariance matrix $\Sigma = \{\sigma^D, \sigma^L, \rho_{DF}, \rho_{DL}, \rho_{LF}\}$ from the firms' choice probabilities. This specification builds on Einav et al. (2012), but differs from them as we estimate demand using a mixed logit with random coefficients, rather than a probit. We also recover the lender-province-year specific constants for the demand model $\widehat{\delta}_{jmt}^D$ using the contraction method introduced by Berry et al. (1995).[34]

The probability that borrower i in market m at time t chooses lender j is given by:

$$\text{Pr}_{ijmt}^D = \int \left[\frac{\exp(\widehat{\delta}_{jmt}^D(X_{jmt}, \xi_{jmt}^D, \bar{\alpha}_1^D, \beta^D) + V_{ijmt}^D(\widetilde{P}_{ijmt}, Y_{ijmt}, \sigma^D, \alpha_2^D, \eta^D))}{1 + \sum_\ell \exp(\widehat{\delta}_{\ell mt}^D(X_{\ell mt}, \xi_{\ell mt}^D, \bar{\alpha}_1^D, \beta^D) + V_{i\ell mt}^D(\widetilde{P}_{i\ell mt}, Y_{i\ell mt}, \sigma^D, \alpha_2^D, \eta^D))} \right] f(\alpha_{1i}^D|\theta) d\alpha_{1i}^D, \tag{15}$$

where $f(\alpha_{1i}^D|\theta)$ is the density of α_{1i}^D, and θ are the parameters of its distribution that we want to estimate. Looking at the second equation, the amount of credit used conditional on borrowing, the probability of observing a utilization of L_{ijmt} is given by:

$$\begin{aligned} \text{Pr}_{ijmt, L=L^*|D=1}^L &= E\big[\text{Pr}(L_{ijmt} = \delta_{jmt}^L + V_{ijmt}^L + \varepsilon_{ijmt}^L | \alpha_{1i}^D) | D = 1\big] \\ &= \int \frac{\phi_{\varepsilon_{ijmt}^L|\alpha_{1i}^D} \left(\frac{L_{ijmt} - \widehat{\delta}_{jmt}^L(X_{jmt}, \xi_{jmt}^L, \alpha_1^L, \beta^L) - V_{ijmt}^L(P_{ijmt}, Y_{ijmt}, \alpha_2^D, \eta^L) - \tilde{\mu}_{\varepsilon_{ijmt}^L|\alpha_{1i}^D}}{\tilde{\sigma}_{\varepsilon_{ijmt}^L|\alpha_{1i}^D}} \right)}{\tilde{\sigma}_{\varepsilon_{ijmt}^L|\alpha_{1i}^D}} f(\alpha_{1i}^D|\theta) d\alpha_{1i}^D, \end{aligned} \tag{16}$$

$$\text{where} \qquad \varepsilon_{ijmt}^L | \alpha_{1i}^D \sim N\big(\underbrace{\sigma^L \rho_{DL} \nu_i}_{\tilde{\mu}_{\varepsilon_{ijmt}^L|\alpha_{1i}^D}}, \underbrace{\sigma^{2L}(1 - \rho_{DL}^2)}_{\tilde{\sigma}^2_{\varepsilon_{ijmt}^L|\alpha_{1i}^D}} \big)$$

where ϕ is a standard normal pdf. Finally, the probability of default conditional on taking a loan is:

$$\text{Pr}_{ijmt, F=1|D=1, \varepsilon_{ijmt}^L}^F = \int \Phi_{\varepsilon_{ijmt}^F|\alpha_{1i}^D, \varepsilon_{ijmt}^L} \left(\frac{\widehat{\delta}_{jmt}^F(X_{jmt}, \xi_{jmt}^F, \alpha_1^F, \beta^F) + V_{ijmt}^F(P_{ijmt}, Y_{ijmt}, \alpha_2^F, \eta^F) - \tilde{\mu}_{\varepsilon_{ijmt}^F|\alpha_{1i}^D, \varepsilon_{ijmt}^L}}{\tilde{\sigma}_{\varepsilon_{ijmt}^F|\alpha_{1i}^D, \varepsilon_{ijmt}^L}} \right) f(\alpha_{1i}^D|\theta) d\alpha_{1i}^D, \tag{17}$$

[34] We are unable to use the contraction for loan use and default as we have many zeros in loan use and default lender-province-year specific shares. For this reason, and due to the smaller number of observations in loan use and default compared to demand, we let $\widehat{\delta}_{jmt}^L = \alpha_1^L + X'_{jmt}\beta^L + \xi_j^L + \tau_{mt}^L$ and $\widehat{\delta}_{jmt}^F = \alpha_1^F + X'_{jmt}\beta^F + \xi_j^F + \tau_{mt}^F$, where ξ_j^L, ξ_j^F are bank fixed effects and τ_{mt}^L, τ_{mt}^F are province-year fixed effects.

$$\text{where} \qquad \varepsilon^F_{ijmt}|\alpha^D_{1i},\varepsilon^L_{ijmt} \sim N\big(\underbrace{A\sigma^D\nu_i + B\varepsilon^L_{ijmt}}_{\tilde{\mu}_{\varepsilon^F_{ijmt}|\alpha^D_{1i},\varepsilon^L_{ijmt}}}, \underbrace{\sigma^{2F} - (A\rho_{DF} + B\rho_{LF})}_{\tilde{\sigma}^2_{\varepsilon^F_{ijmt}|\alpha^D_{1i},\varepsilon^L_{ijmt}}}\big)$$

$$A = \tfrac{\rho_{DF}\sigma^{2L} - \rho_{LF}\rho_{DL}}{\sigma^{2D}\sigma^{2L} - \rho^2_{DL}}$$
$$B = \tfrac{-\rho_{DF}\rho_{DL} + \rho_{LF}\sigma^{2D}}{\sigma^{2D}\sigma^{2L} - \rho^2_{DL}}$$

where the residuals ε^F_{ijmt} are conditional on demand and loan amount unobservables. The joint estimation of these three choice equations through maximum simulated likelihood only delivers the parameters in η, β^{LF} and Σ, based on the following log-likelihood function:

$$\log L = \sum_i d_{ijmt}\{\log(\text{Pr}^D_{ijmt}) + \log(\text{Pr}^L_{ijmt}) + f_{ijmt}\log(\text{Pr}^F_{ijmt}) + (1 - f_{ijmt})\log(\text{Pr}^F_{ijmt})\} \tag{18}$$

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

In the second step, the estimated constants $\widehat{\delta}^D_{jmt}$ are the dependent variables of OLS regressions that recover the parameters $\beta^D = \{\bar{\alpha}^D_1, \beta^D\}$ of the bank specific attributes X_{jmt}. Following Berry et al. (1995), the contraction method on the demand side finds the δ^D that equate predicted market shares $\widehat{S}^D_{jmt}$ to actual market shares S^D_{jmt}. This iterative process is defined by:

$$\delta^{D,r+1}_{jmt} = \delta^{D,r}_{jmt} + \ln\left(\frac{S^D_{jmt}}{\widehat{S}^D_{jmt}(\delta^{D,r}_{jmt})}\right). \tag{19}$$

The predicted market shares are defined as $\widehat{S}^D_{jmt} = \sum_i \text{Pr}^D_{ijmt}/N_{mt}$, where N_{mt} are the number of firms in market m at time t. Given the value of these constant terms, the parameters $\bar{\alpha}^D_1, \beta^D$ are estimated using OLS:

$$\delta^D_{jmt} = \bar{\alpha}^D_1 + X'_{jmt}\beta^D + \xi^D_{jmt}, \tag{20}$$

with ξ^D_{jmt} being the mean zero structural econometric error term.

6 Estimation

6.1 Constructing the Sample

As already mentioned, we focus on the first line of credit that a firm opens (at least within our dataset), excluding the first year (1988) to avoid left censoring. We do this to concentrate on new borrowers, where

we expect to find stronger asymmetric information, and because modeling the evolution of the borrower-lender relationship is beyond the scope of this paper.[35] Following other papers on Italian local credit markets, like Felici and Pagnini (2008), Bofondi and Gobbi (2006), and Gobbi and Lotti (2004), we identify banking markets as the Italian provinces, also used by Italian supervisory authorities as proxies for the local markets for deposits and loans to SMEs.[36] Our markets are then constructed as province-year combinations. We assume that each firm's choice set is defined by all the banks that are actively lending in its province.[37]

The observable explanatory variables that determine firm's demand, loan use and default choices are firm and bank characteristics, previously summarized in Table 1. In the first set of regressors we include firms' total assets, the ratio of intangible over total assets, profits, cash flow, sales, trade debit, firm's age, distance from the closest branch, and interest rates. Trade debit is the debit that the firm has with its suppliers or clients. We also include score, sector, and loan amount fixed effects. In the demand model we estimate bank-year-province fixed effects with the contraction method explained above. In the second group we use bank's share of branches in the province, number of years the bank had at least one branch in the province. For the second step of the demand model we control for bank, year, and province dummies, whereas in the loan use and default models we allow for bank and year-province fixed effects. We motivate the choice of these explanatory variables in Section 6.3.

We estimate our structural model on a subset of the original dataset, mostly for computational and institutional reasons. Cohen and Mazzeo (2007) define a local banking market as a geographic area where consumers don't choose lenders outside their province, that means that the market cannot be too small, and that doesn't include distinct or overlapping markets within, implying that the market cannot 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 province, out of an initial sample of 977 year-province combinations we drop approximately the first and last decile of their size distribution, in terms of number of observations.[38] This leaves us with 666 markets, around 66% of the original number of markets.

6.2 Identification

As described in Section 2.2 and in the Appendix, we impute the prices that we don't observe in the data using propensity score matching (PSM), where we match borrowers to non-borrowers based on the probability of borrowing, defined as:

$$\Pr[D_{imt} = 1 | M_{imt}] = \Phi(M_{imt}\tau), \tag{21}$$

[35] A more extensive description of the construction of the sample is in Appendix A.

[36] See Ciari and Pavanini (2014) and Guiso et al. (2013) for a detailed discussion on the definition of local banking markets in Italy.

[37] As many other papers estimating demand for differentiated products, we face the problem of zero market shares causing the second stage IV estimation to deliver inconsistent estimates, as described in Gandhi, Lu and Shi (2013). We deal with this excluding from the borrowers' choice set banks with non-zero branches but with zero market shares in a local market.

[38] To improve the convergence of the contraction mapping we also eliminate markets with zero or close to zero market share of the outside option.

where $M_{imt} \subset Y_{ijmt}$ is the subset of the firm observables Y_{ijmt} that pass the PSM tests for a good match. We then run the following regression to predict prices for firm i with bank j in market m at time t:

$$P_{ijmt} = \underbrace{\gamma_0 + \gamma_1 d_{ij} + \gamma_2 L_{ij} + \gamma_3 Z_{jmt} + \lambda_j + \omega_i}_{\widetilde{P}_{ijmt}} + \epsilon_{ijmt}, \tag{22}$$

where P_{ijmt} are actual prices, $\widetilde{P}_{ijmt}$ are predicted prices, d_{ij} is distance from the branch, L_{ij} are dummies for the size of the amount granted, Z_{jmt} are deposit costs, λ_j are bank fixed effects, ω_i are firm fixed effects, and ϵ_{ijmt} are residuals. We use $\Phi(M_{imt}\tau)$ to assign borrowers' $\widehat{\omega}_i$ to non-borrowers. We then use predicted prices $\widetilde{P}_{ijmt}$ to estimate demand based on the following utility function:

$$U^D_{ijmt} = \alpha^D_{1i} - \alpha^D_2 \widetilde{P}_{ijmt} + Y'_{ijmt}\eta^D + \delta^D_{jmt} + \varepsilon^D_{ijmt}, \tag{23}$$

where $\alpha^D_{1i} = \overline{\alpha}^D_1 + \sigma^D \nu_i$, with $\nu \sim N(0,1)$, is a random coefficient on the constant and Y_{ijmt} includes firms' balance sheet information, distance from bank branches, and size of amount granted. In this demand model the only identifying variation in $\widetilde{P}_{ijmt}$ is given by $\widehat{\omega}_i$, because within each firm all of the variation in $\widehat{P}_{ijmt}$ is absorbed by $d_{ij}, L_{ij}, \delta^D_{jmt}$, leaving α^D_2 to be only identified by variation in $\widehat{\omega}_i$ across firms with the same observables. A potential endongeneity concern is given by the correlation between $\widehat{\omega}_i$ and ε^D_{ijmt}. We overcome this relying on the standard assumption in PSM of conditional independence between treatment and outcome conditional on the propensity score (Caliendo and Kopeinig (2008)), which in our case means assuming that $\widehat{\omega}_i$ and D_{imt} are independent conditional on $\Phi(M_{imt}\tau)$.

We also worry that actual interest rates, which we observe and use for the loan use and default equations, might be correlated with loan use and (in particular) default unobservables. To address this potential endogeneity concern we use a control function approach, following Train (2009) and Adams et al. (2009). We use deposit costs as instruments, as they are correlated with interest rates (see Table 2), but are unlikely to be correlated with loan use and default unobservables. We proceed regressing interest rates P_{ijmt} on the same observables that we use for loan use and default, plus the instruments. We then use the residuals from this pricing regression $\widehat{u}_{ijmt}$ as controls in the following utility from choosing how much loan to use and utility from defaulting:

$$\begin{aligned} U^L_{ijmt} &= \alpha^L_1 + X'_{jmt}\beta^L + \alpha^L_2 P_{ijmt} + \alpha^L_3 \widehat{u}_{ijmt} + Y'_{ijmt}\eta^L + \delta^L_j + \delta^L_{mt} + \varepsilon^L_{ijmt} \\ U^F_{ijmt} &= \alpha^F_1 + X'_{jmt}\beta^F + \alpha^F_2 P_{ijmt} + \alpha^F_3 \widehat{u}_{ijmt} + Y'_{ijmt}\eta^F + \delta^F_j + \delta^F_{mt} + \varepsilon^F_{ijmt}. \end{aligned} \tag{24}$$

Once we control for the residuals $\widehat{u}_{ijmt}$, and conditional on the other controls, the remaining price variability is attributable to differences in deposit rates, that are independent from the individual decision on how much to draw and to default on the credit line. Following Adams et al. (2009), the price coefficient in the default equation α^F_2 can be interpreted in terms of moral hazard: it indicates how much the probability of default increases for an increase in price unrelated to firms' characteristics, including riskiness observed by the bank and priced accordingly.

6.3 Results

The estimates of the structural model are presented in Table 7. The three columns of results refer respectively to the demand, loan use and default equations. The top part of the table shows the effect of firm characteristics, the middle one the effect of bank characteristics, and the bottom one shows the covariance matrix, where we interpret as adverse selection the correlation between unobservables of demand and default (ρ_{DF}) and the correlation between unobservables of loan use and default (ρ_{LF}). We decided to include specific firm characteristics to control for different measures of firms' assets, profitability, and debt. We chose among the wide set of balance sheet variables running various reduced form regressions for demand, loan use, and default. We wanted to control for different measures of firm size, in the form of assets and ratio of intangible assets,[39] but also for some measures of firms' current performance, in terms of profits, cash flow, and sales. We also tried to control for other specific forms of finance that firms have access to, such as credit from suppliers.[40] We computed 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.[41] We also include fixed effects for the Score of the firm, its sector (primary, secondary, tertiary), the loan amount, and various combinations of bank, year and province dummies.[42] Another important variable we include are the predicted interest rates. Finally, we also included the number and the share of branches that a bank has in a market (province-year), as well as the number of years that it has been in the market. We have data on branches from 1959, so we can observe banks' presence in each council for the 30 years before the beginning of our loan sample. These variables aim at capturing the level of experience that a bank has in a market, as well as the density of its network of branches with respect to its competitors, which can both be relevant features influencing firms' decisions.

The estimates present evidence of adverse selection both in the correlation between demand and default unobservables, and in the correlation between loan use and default unobservables. This confirms the results of the the reduced form test that we presented earlier. Looking at the demand side, we find that distance and prices have a negative impact on demand, as expected. Firms with more cash flow and trade debit are less likely to demand credit, but firms with more assets, profits, and sales are more likely to borrow. Older firms are also more likely to demand. Firms tend to favor banks with a higher share and number of branches, and are more likely to demand form banks with longer experience in a market. The loan used seems to follow the same logic as demand for most of the relevant variables, such as interest rates, trade debit, cash flow, profits and assets. For what concerns the default probability, this is increasing in interest rates as expected, and decreasing in profits and trade debit. We interpret the positive effect of interest rates on default as evidence

[39] Albareto et al. (2011) describe the importance of firms' size in the organization of lending in the Italian banking sector.

[40] Petersen and Rajan (1995) use the amount of trade credit as a key variable to determine if borrowers are credit constrained, as it's typically a more expensive form of credit than banks' credit lines.

[41] It is important to include distance as Degryse and Ongena (2005) show evidence of spatial price discrimination in lending relationship using Belgian data. They find that loan rates decrease with the distance between the borrower and the lender, and increase with the distance between the borrower and the competing lenders.

[42] As explained in Section 5, we estimate the demand model in two steps, and the loan use and default models in one step. This implies that for the demand model we control for bank-year-province fixed effects in the first stage, and then for separate bank, year and province fixed effects in the second stage. For loan use and default instead we are just able to control for bank and year-province fixed effects.

of moral hazard.

The mean of own and cross price elasticities for the main 5 banks in the sample are reported in Table 8. We find that on average a 1% increase in interest rate reduces a bank's own market share by around 4%, and increases competitor banks' shares by about 0.3%.

Table 7: Structural Estimates

Variables			Demand	Loan Use	Default
1^{st} Stage Firm Level	Assets	Total Assets	5.95***	0.08***	0.07***
			(0.09)	(0.00)	(0.02)
		Intangible/Total Assets	-0.85***	-0.01*	0.06*
			(0.06)	(0.01)	(0.04)
	Profitability	Profits	1.14***	0.01***	-0.32***
			(0.06)	(0.00)	(0.04)
		Cash Flow	-0.93***	-0.05***	0.03
			(0.06)	(0.00)	(0.02)
		Sales	6.97***	-0.03***	0.07*
			(0.08)	(0.00)	(0.03)
	Debt	Trade Debit	-3.21***	-0.02***	-0.11***
			(0.05)	(0.00)	(0.02)
	Others	Firm's Age	0.21***	-0.02***	0.03
			(0.04)	(0.00)	(0.03)
		Distance	-1.19***	-0.05	-0.06
			(0.04)	(0.10)	(0.07)
		Interest Rate	-0.26**	-0.06***	0.40***
			(0.02)	(0.00)	(0.00)
		Score FE	Yes	Yes	Yes
		Sector FE	Yes	Yes	Yes
		Loan Amount FE	Yes	Yes	Yes
		Bank-Year-Province FE	Yes	No	No
2^{nd} Stage Bank Level		Number of Branches	0.72***	0.01**	-0.35***
			(0.06)	(0.00)	(0.06)
		Share of Branches	1.61***	-0.09***	-0.00
			(0.26)	(0.01)	(0.02)
		Years in Market	0.47***	0.02***	-0.29***
			(0.05)	(0.00)	(0.11)
		Bank FE	Yes	Yes	Yes
		Year FE	Yes	No	No
		Province FE	Yes	No	No
		Year-Province FE	No	Yes	Yes
Observations			312,906	18,820	18,820
Covariance matrix			$\sigma^D = 0.11^{***}$		
			(0.00)		
			$\rho_{DL} = 0.02^{***}$	$\sigma^L = 0.30^{***}$	
			(0.00)	(0.00)	
		Adverse Selection	$\rho_{DF} = 0.06^{***}$	$\rho_{LF} = 0.09^{***}$	$\sigma^F = 1$
			(0.00)	(0.00)	

Note: The top panel shows the estimates of the first stage η, γ_k. The middle panel shows the estimates of the second stage $\alpha, \beta, \tau_t, \xi_j$. The bottom panel shows the estimates of the variance-covariance Σ. The first column of results refers to the demand model, the second to the loan use, and the third to the default. Standard errors are in brackets. $*$ is significant at the 10% level, $**$ at the 5% level, and $***$ at the 1% level. 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 iteration of parametric bootstrap.

Table 8: Mean across Markets of Own and Cross Price Elasticities for Main Banks

Banks	**Bank 1**	**Bank 2**	**Bank 3**	**Bank 4**	**Bank 5**
Bank 1	-4.06	0.19	0.22	0.20	0.22
Bank 2	0.26	-3.77	0.26	0.26	0.25
Bank 3	0.31	0.32	-4.31	0.33	0.36
Bank 4	0.24	0.23	0.26	-4.20	0.23
Bank 5	0.27	0.26	0.30	0.25	-3.81

Note: These are the first 5 banks in terms of national market shares. Elasticities are interpreted as the percentage change in market shares in response to a 1% increase in prices.

6.4 Fit of the Model

We provide some descriptives on the fit of the model. We choose to focus on the main objects of interest of the model, that is demand probabilities, amount of loan used, default probabilities, and equilibrium prices. We compare actual and predicted probabilities, as well as actual prices to those that maximize our profit function in equation 9. Summary statistics are presented in Table 9. We follow Nevo (2000a) and recover each bank's borrower-specific marginal costs using the pricing equation (10):

$$\widehat{MC}_{ijmt} = \widetilde{P}_{ijmt}\Big[1 - F_{ijmt} - F'_{ijmt}\frac{Q_{ijmt}}{Q'_{ijmt}}\Big] + \frac{(1-F_{ijmt})\frac{Q_{ijmt}}{Q'_{ijmt}}}{1 - F_{ijmt} - F'_{ijmt}\frac{Q_{ijmt}}{Q'_{ijmt}}}. \tag{25}$$

Based on the model fit, we decide to exclude some firm-bank combinations from the sample for the counterfactuals. Even though our model predicts equilibrium prices well,[43] there are very few outliers for which the model gives a poor prediction. These wrong predictions correspond to just around 2% of the prices and can be easily interpreted based on our model set up. They represent those few cases where a bank's profit from a loan, based on our profit function assumptions and estimated parameters, are negative under the observed prices. For this reason our profit optimization routine delivers new equilibrium prices that need to be high enough to both reduce the probability of a firm taking out a loan, and to reduce to zero the predicted loan use. We decide to exclude these few firm-bank observations from our sample for the counterfactuals. Note that this slight reduction in sample size has almost no effect on the counterfactuals' results.

The model shows a good fit, as the mean of the four variables of interest predicted by the model is always very close to the data. However, as expected, the standard deviation of the model's predictions for demand, loan use, and default is lower than the actual data. The model's price prediction is instead very accurate once we exclude the 2% of outliers. Last, we find that effective marginal costs represent on average 78%

[43] The median deviation of model predicted prices from actual prices is 0, but we have some outliers. While not perfect, we note that we do not impose the pricing moment conditions in estimation (as is often done in the empirical literature), which would necessarily improve the model fit at the expense of having an impact on the estimated adverse selection parameters which are our primary point of interest.

of the price, and effective markups about 22%. Note that the 5^{th} percentile of predicted effective markups is zero because for around 20% of the credit lines we observe a zero loan use, which in turn implies a zero markup for those loans.

In order to validate our model estimates, we investigate whether our predicted marginal costs correlate with banks' cost information we have access to. Presumably the main factors that affect a bank's marginal costs are the Libor,[44] wages and property values in each local market, and deposit rates. We run an OLS regression focussing on the relationship between predicted marginal costs and deposit interest rates, and control for the combined effects of the other factors using province-year fixed effects and bank fixed effects. We concentrate on this element because we only have yearly data, so cannot capture most of daily variation in the Libor, and because banks in the same province-year are likely to face the same labor and property costs. We have information instead on deposit interest rates at the bank-year-province level, and find that conditional on bank and province-year fixed effects they have a positive and significant correlation with our predicted marginal costs. According to our estimates, 5 standard deviation increases in deposit rates are associated with around 1 percentage point increase in predicted marginal costs.

Table 9: Descriptives on Model Fit

Variables	**Nobs**	**Mean**	**Std. Dev.**	**5^{th} Pctile**	**50^{th} Pctile**	**95^{th} Pctile**
Actual Demand	312,906	0.087	0.281	0	0	1
Predicted Demand	312,906	0.087	0.163	0.003	0.035	0.418
Actual Loan Use	18,820	246.5	442.6	0	96	1,018
Predicted Loan Use	18,820	245.6	295.7	-56.9	160.3	998.7
Actual Default	18,820	0.063	0.244	0	0	1
Predicted Default	18,820	0.063	0.106	0	0.016	0.284
Actual Prices	276,205	14.86	4.41	8.83	14.30	22.66
Predicted Prices	276,205	14.96	4.42	8.91	14.39	22.79
Predicted MC	276,205	12.86	5.83	5.12	11.99	23.79
Predicted Effective MC	276,205	11.56	4.73	5.00	11.00	19.90
Predicted Effective Markups	276,205	3.31	1.30	0.00	3.77	4.52

Note: Prices are average prices at the year-province-bank-firm level.

7 Counterfactuals

We run a counterfactual policy experiment to quantify the effects of asymmetric information, as well as to understand the relationship between asymmetric information and imperfect competition. We simulate an

[44] London Interbank Offered Rate, the most relevant European interest rate for interbank lending in the 1990s, with daily variation.

increase in adverse selection, identified by the correlation between unobservables driving both demand and default (ρ_{DF}) and loan size and default (ρ_{LF}). We analyze the consequence of this change on equilibrium prices, quantities, and defaults. The rationale for this counterfactual exercise is twofold. First, we want to give an economic interpretation to the effect of the estimated correlation coefficients in terms of relevant outcome variables. This means looking at how prices, demand, and defaults vary as we increase the correlation coefficients, just like we did in the Monte Carlo exercise. Second, an increase in adverse selection is a distinctive feature of a financial crisis (Mishkin (2012)). During a period of crisis firms' investment opportunities contract, affecting differently the borrowing behavior of sound and risky firms. While the former will borrow less, as they don't invest but still expect to have a positive cash flow to use as operating liquidity, the latter will keep borrowing as they expect their cash flow to be an insufficient source of working capital. This can be thought as an increase in adverse selection, as it implies a higher correlation between borrowers' willingness to pay and their probability of default.

We simulate a scenario in which we double the estimated correlation coefficients ρ_{DF} and ρ_{LF}.[45] Once we recover the equilibrium outcomes of interest in the new scenario, we investigate whether the variations that we observe from the baseline model are correlated with a measure of market power that every bank has with each of its potential borrowers. We follow Nevo (2000a) assuming that marginal costs remain the same in the counterfactual scenario. We re-calculate firms-banks' demand probabilities, loan use and defaults with the counterfactual level of adverse selection, and derive the new equilibrium prices as:

$$\overline{P}_{ijmt} = \frac{\widehat{MC}_{ijmt}}{1 - \overline{F}_{ijmt} - \overline{F}'_{ijmt}\frac{\overline{Q}_{ijmt}}{\overline{Q}'_{ijmt}}} + \frac{-(1-\overline{F}_{ijmt})\frac{\overline{Q}_{ijmt}}{\overline{Q}'_{ijmt}}}{1 - \overline{F}_{ijmt} - \overline{F}'_{ijmt}\frac{\overline{Q}_{ijmt}}{\overline{Q}'_{ijmt}}}, \tag{26}$$

where $\overline{Q}_{ijmt}$ and $\overline{F}_{ijmt}$ are the new equilibrium quantities and defaults in the counterfactual setting. We define quantities as demand probabilities ($\overline{Q}^{D}_{ijmt}$) multiplied by the predicted amount of loan used ($\overline{Q}^{L}_{ijmt}$), and defaults as the firm's predicted default probability. Following the non-monotonic price response predicted in the Monte Carlo experiment, we investigate how equilibrium prices, quantities, and defaults vary with respect to the the baseline case.[46]

We measure these changes at the firm-bank-year-province level and show them in Table 10, where we report both percentage variation (%) and percentage point variation (P.P.) for prices (i.e. interest rates), demand probabilities, and default probabilities.[47] Kernel densities of the distributions of these variations are plotted in Figures 5.

[45] We have experimented with greater or smaller changes in the correlation coefficients, obtaining similar results scaled up or down depending on the size of the variation. We increase both correlation coefficients as they are both measures of adverse selection, and from a policy perspective we are more interested in the combined effect rather than in the relative importance of each in driving the outcomes that we are finding. To make sure that the variance of $\varepsilon^{F}_{ijmt}|\alpha^{D}_{1i}, \varepsilon^{L}_{ijmt}$ remains positive in the counterfactual we increase ρ_{LF} to 0.22, slightly more than doubling it.

[46] We predict prices based on our supply model also for the baseline case. This means that we compare predicted prices under the estimated correlation coefficients and predicted prices under the counterfactual correlation coefficients.

[47] Percentage variation in prices is measured as $\Delta P_{ijmt} = 100 * \frac{\overline{P}_{ijmt} - \widetilde{P}_{ijmt}}{\widetilde{P}_{ijmt}}$, and similarly for demand probabilities, loan use, and default. Percentage point variation in prices is measured as $\Delta P_{ijmt} = \overline{P}_{ijmt} - \widetilde{P}_{ijmt}$, whereas for demand probabilities is measured as $\Delta Q^{D}_{ijmt} = 100 * (\overline{Q}^{D}_{ijmt} - Q^{D}_{ijmt})$, and similarly for default.

The rise in adverse selection causes a median price increase of 26.5%, corresponding to a 3.7 percentage points raise in interest rates. Looking at both the descriptive statistics in Table 10 and at the top left panel in Figure 5 we note that there is substantial heterogeneity in these price responses, with some interest rates raising by 70 percentage points and above, and a few actually declining. We don't find the very stark increase in some of the prices too surprising, as it represent an implicit way for the model to generate a rejection of a risky borrower, making sure that a very high interest rates can both minimize the demand probability and the loan use. As expected, most of the demand probabilities decline, as well as most of the loan uses. Higher adverse selection tends to worsen banks' pool of borrowers, despite some firms experiencing a very marginal reduction in default probability.

Based on these preliminary patters, we want to investigate how these variations relate to the level of market power that each bank has. We use the estimated effective markup term in the baseline case to identify banks' market power, that is the last term on the right hand side of equation 10, taken at the firm-bank-year-province level. We run OLS regressions of variations in interest rate, demand probability, loan use, and default at the bank-firm-year-province level on markups and firm fixed effects. Table 11 shows that the markup term negatively affects the price variation, positively affects the change in demand probability and loan use, and negatively impacts the default probability. This confirms the intuition of the Monte Carlo exercise presented earlier, because banks with higher market power respond to an increase in adverse selection lowering their prices and expanding their credit supply, or increasing their price by less than banks with low market power. Following Stiglitz and Weiss (1981), the reduction in prices attracts the marginal borrowers, which under adverse selection are safer than the infra-marginal ones, lowering the bank's share of defaulters. Looking at the estimated coefficients for the markup term, we find that one standard deviation increase in effective markup reduces the price variation ΔP_{ijmt} by around 4.3 percentage points, increases the variation in demand probabilities ΔQ^D_{ijmt} by 0.2 percentage points and in loan use ΔQ^L_{ijmt} by around 8,200 euros, and reduces variation in defaults ΔF_{ijmt} by 2.2 percentage points.

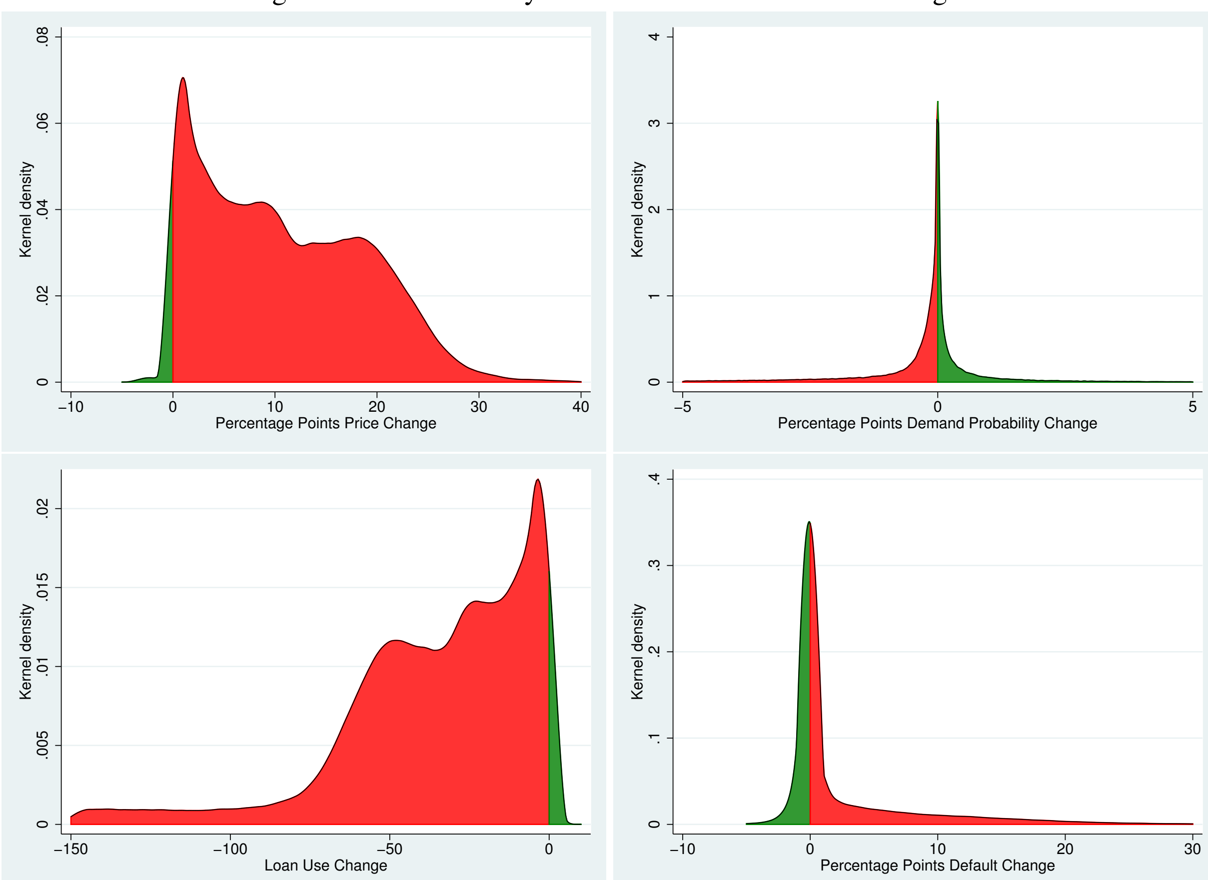
These results highlight several important policy implications. First, an increase in adverse selection causes most of the prices in our sample to increase, most of the quantities to fall and most of the defaults to rise. This implies that, consistent with the theoretical literature on the adverse effects of asymmetric information, such asymmetries can severely worsen lending conditions in this market, and suggests that additional policies to mitigate this market failure would be beneficial. In other words, while market power can soften the adverse effects of asymmetric information, these effects are not sufficient *on average*. That being said, our second implication is that some markets are different: there is substantial heterogeneity in price, quantity and default responses to this rise in adverse selection across banks and markets. It is of crucial importance for policymakers to understand how some banks and/or markets can absorb these shocks and some others cannot. We offer one possible explanation for this heterogeneity, which is market power. We show that banks with higher markups have a counter-cyclical effect on credit supply, responding to an increase in adverse selection with a reduction in prices and an increase in quantity lent. Hence, if on one hand competition in lending markets is beneficial for borrowers as it can reduce interest rates, on the other hand it forces banks to follow the business cycle and increase rates as adverse selection rises, making borrowing firms more likely to be credit rationed during this kind of shocks.

Table 10: Descriptives Counterfactual % and Percentage Points Variations in Outcomes

Variables	**Nobs**	**10^{th} Pctile**	**50^{th} Pctile**	**90^{th} Pctile**
% Price Variation	276,205	0.0	26.5	422.4
P.P. Price Variation	276,205	0.00	3.71	72.83
% Demand Variation	276,205	-30.3	0.0	6.0
P.P. Demand Variation	276,205	-1.28	0.00	0.33
% Loan Use Variation	275,887	-100.0	-6.8	0.0
Loan Use Variation	276,205	-129.4	-9.8	0.0
% Default Variation	276,205	-100.0	-6.2	410.5
P.P. Default Variation	276,205	-0.53	0.00	70.66

Note: These variations are at the firm-bank-year-province. Loan Use variations are in thousands of euros. % Loan Use Variation has a smaller number of observations corresponding to the few cases in which loan use is zero in the baseline case and non-zero in the counterfactual case.

Figure 5: Kernel Density of Counterfactual Outcome Changes



Note: An observation is a firm-year-province-bank. The vertical axis is the density. The horizontal axis is the percentage point outcome variation between the baseline case and the counterfactual scenario for prices, demand probabilities, and default, and it's the loan use change in thousands of euros. The graph is trimmed to avoid outliers and excludes the cases of no variation in the outcomes.

Table 11: Regressions of Counterfactual Outcomes' Variations on Markups

Variables	ΔP_{ijmt}	ΔQ^D_{ijmt}	ΔQ^L_{ijmt}	ΔF_{ijmt}
Effective Markup	-3.32***	0.18***	6.30***	-1.66***
	(0.08)	(0.01)	(0.16)	(0.03)
Firm FE	Yes	Yes	Yes	Yes
R^2	0.761	0.232	0.773	0.781
N obs.	276,205	276,205	275,887	276,205

Note: An observation is a firm-year-province-bank. Price, demand probabilities, and default variations are measured in percentage points. Loan use is measured in thousands of euros.

8 Conclusion

In this paper we analyzed the interaction between imperfect competition and asymmetric information in the Italian market for small business lines of credit. We have access to a rich dataset with detailed information about credit line contracts between firms and banks, including all the main Italian credit institutions and a highly representative sample of firms. Using these data, we provide reduced form evidence of adverse selection in the spirit of the positive correlation test on unobservables by Chiappori and Salanié (2000), as well as reduced form evidence consistent with imperfect competition.

Based on this evidence, we propose a structural model of firms' demand for credit, loan use, and default, as well as of banks' pricing. We let differentiated banks compete à la Bertrand-Nash on interest rates in local credit markets, but also use interest rates as a screening device, as in Stiglitz and Weiss (1981). The model allows for imperfect competition in the lending market, accounting for asymmetric information between borrowers and lenders. We assume in fact that firms know the riskiness of their own project, but banks can only observe the distribution of riskiness of their borrowers, conditional on observable firm characteristics. When we introduce asymmetric information, our model of oligopolistic competition predicts different banks' interest rate reactions to an increase in adverse selection, depending on the level of competition. We provide Monte Carlo evidence of a non-monotonic optimal bank's price response to an increase in adverse selection, depending on different measures of competition. More adverse selection causes prices to increase in competitive markets, but can have the opposite effect in more concentrated ones, where banks can leverage over their markup to lower prices and attract safer borrowers.

We find evidence of adverse selection in the data, in the form of a positive correlation between unobservables determining demand and default and loan use and default. We provide evidence both with reduced form tests and estimating a structural model. We also provide evidence of moral hazard identified with the causal effect of interest rates on borrowers' default. We conduct a policy experiment to simulate the effects of a credit crunch, in which risky firms experience a more severe financial distress and demand more credit, increasing

the extent of adverse selection. As predicted, in this counterfactual scenario equilibrium prices rise for more competitive markets and decline for more concentrated ones. As a consequence, the share of borrowing firms increases in more concentrated markets, and default rates fall.

References

Abbring, J. H., P. A. Chiappori, J. J. Heckman, and J. Pinquet, “Adverse Selection and Moral Hazard in Insurance: Can Dynamic Data Help to Distinguish?,” *Journal of the European Economic Association*, 2003, *1*, 512–521.

Ackerberg, D. A. and M. Botticini, “Endogenous Matching and the Empirical Determinants of Contract Form,” *Journal of Political Economy*, 2002, *110* (3), 564–591.

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., M. Bottero, and G. Sene, “Sharing Information on Lending Decisions: An Empirical Assessment,” 2014. Temi di discussione (Economic working papers) 980, Bank of Italy.

Altman, E. I., “Financial ratios, discriminant analysis and the prediction of corporate bankruptcy,” *Journal of Finance*, 1968, *21*, 589–609.

——, **G. Marco, and F. Varetto**, “Comporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience),” *Journal of Banking and Finance*, 1994, *18*, 505–529.

Banerjee, A. and A. Newman, “Occupational Choice and the Process of Development,” *Journal of Political Economy*, 1993, *101* (2), 274–298.

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*.

Berger, A. N. and G. F. Udell, “Relationship Lending and Lines of Credit in Small Firm Finance,” *Journal of Business*, 1995, *68* (3), 351–381.

Bernanke, A. and M. Gertler, “Financial Fragility and Economic Performance,” *Quarterly Journal of Economics*, 1990, *105* (1), 87–114.

Berry, S., “Estimating Discrete-Choice Models of Product Differentiation,” *The 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," *The 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.

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.

DeMeza, D. and D. C. Webb, "Too Much Investment: A Problem of Asymmetric Information," *The Quarterly Journal of Economics*, 1987, *102* (2), 281–292.

Detragiache, E., P. Garella, and L. Guiso, "Multiple versus Single Banking Relationships: Theory and Evidence," *The 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," 2015. Working Paper.

Einav, L., A. Finkelstein, and J. Levin, "Beyond Testing: Empirical Models of Insurance Markets," *Annual Review of Economics*, 2010, *2*, 311–336.

____ 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.

Felici, R. and R. Pagnini, "Distance, Bank Heterogeneity and Entry in Local Banking Markets," *Journal of Industrial Economics*, 2008, *53* (3), 500–534.

Gale, W. G., "Federal Lending and the Market for Credit," *Journal of Public Economics*, 1990, *42* (2), 177–193.

Gandhi, A., Z. Lu, and X. Shi, "Estimating Demand for Differentiated Products with Error in Market Shares," 2013. CeMMAP Working Paper CWP03/13.

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.

Goolsbee, A. and A. Petrin, "The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV," *Econometrica*, 2004, *72* (2), 351–381.

Guiso, L., L. Pistaferri, and F. Schivardi, "Credit within the Firm," *The Review of Economic Studies*, 2013, *80* (1), 211–247.

Handel, B., "Adverse Selection and Switching Costs in Health Insurance Markets: When Nudging Hurts," *American Economic Review*, 2013, *103* (7), 2643–2682.

Heckman, J. J., "Sample Selection Bias as a Specification Error," *Econometrica*, 1979, *47* (1), 153–161.

Ho, K. and J. Ishii, "Location and Competition in Retail Banking," *International Journal of Industrial Organization*, 2011, *29* (5), 537–546.

Hubbard, R. G., "Capital Market Imperfections and Investment," *Journal of Economic Literature*, 1998, *36* (1), 193–225.

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," 2015. Working Paper.

Lester, B., A. Shourideh, V. Venkateswaran, and A. Zetlin-Jones, "Screening and Adverse Selection in Frictional Markets," 2015. Working Paper.

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 Mediare," 2011. Boston University Working Paper.

Mahoney, N. and E. G. Weyl, "Imperfect Competition in Selection Markets," 2014. NBER Working Paper 20411.

Mankiw, N. G., "The Allocation of Credit and Financial Collapse," *Quarterly Journal of Economics*, 1986, *101* (3), 455–470.

Michelacci, C. and F. Schivardi, "Does Idiosyncratic Business Risk Matter for Growth?," *Journal of the*

European Economic Association, 2013, *11* (2), 343–368.

Mishkin, F. S., *The Economics of Money, Banking, and Financial Markets*, Pearson Education, 2012.

Mookherjee, D. and D. Ray, "Contractual Structure and Wealth Accumulation," *American Economic Review*, 2002, *92* (4), 818–849.

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," *The 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.

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," 2016. Working Paper.

Petersen, M. A. and R. G. Rajan, "The Benefits of Lending Relationships: Evidence from Small Business Data," *Journal of Finance*, 1994, *49* (1), 3–37.

____ **and** ____ , "The Effect of Credit Market Competition on Lending Relationships," *The Quarterly Journal of Economics*, 1995, *110*, 403–444.

____ **and** ____ , "Does Distance Still Matter? The Information Revolution in Small Business Lending," *Journal of Finance*, 2002, *57* (6), 2533–2570.

Pischke, S., "Lecture Notes on Measurement Error," 2007. Spring 2007 Lecture Notes.

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.

Starc, A., "Insurer Pricing and Consumer Welfare: Evidence from Medigap," *RAND Journal of Economics*, 2014, *45* (1), 198–220.

Stiglitz, J. and A. Weiss, "Credit Rationing in Markets with Imperfect Information," *American Economic Review*, 1981, *71* (3), 393–410.

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.

Wooldridge, J. M., *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, 2002.

A For Online Publication - Appendix A - Constructing the Dataset

We have assembled various datasets from different sources, which are the following:

- **Firm Data:** Dataset from *Centrale dei Bilanci* with yearly (1988-1998) balance sheet data for each firm, including both firms that take credit and don't (outside option). This also includes the year of birth of each firm and its location at the city council level.
- **Score Data:** Dataset for each firm with yearly (1982-1998) score data, with also the 6 years preceding 1988. We retain from this data the 1982-1987 average, standard deviation, and weighted average (more weight to more recent years) of the score.
- **Loan Data:** Dataset from *Centrale dei Rischi* with yearly (1988-1998) firm-bank loan contracts, including amount granted, amount used, interest rate, firm's default. This is only for the main 94 banks and for short term credit lines.
- **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 each bank at the city council level. This includes the population of banks ($\sim$ 1,500 banks).
- **Coordinates Data:** Based on the *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 and score datasets with the loan data, in order to have all the borrowing and not borrowing firms together. 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 [48]. We assume that each firm chooses one main credit line among all the banks available in its province. 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 chosen. We calculate the distance in km between the city council of each firm and the city council where each bank from the choice set has a branch using the geographic coordinates. For each firm-bank pair, we only keep the branch that is closest to the firm.

A.1 Matching Model

We recover prices and amounts granted for non-borrowing firms using propensity score matching between borrowers and non-borrowers, 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

[48] 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.

the variables determining whether a firm borrows or not, is guided by economic theory and knowledge of the institutional setting, as well as on the overlap in variables' distributions, and statistics from the matching results.

The final set of variables that we use are all fixed effects for: year, score, geographical area, sales and assets. In line with Caliendo and Kopeinig (2008), we apply the following specific criteria in our selection, 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 (interest rates). Second, variables must be unaffected by participation, or anticipation of it, so should be either fixed over time or measured before participation. The score respects this rule, the value of assets is assumed to be persistent overtime, and sales capture demand effects. 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 states. For this reason we choose a parsimonious set of variables and avoid any over-parametrization.

We summarize in Table 12 the normalized differences in means between treatment (non-borrowers) and control (borrowers) groups.[49] 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 the 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. Variables' selection is based on statistical significance, the "hit or miss" method (Heckman et al. (1997)),[50] and comparisons of several statistics before and after the matching. These includes the Pseudo-R^2, the Likelihood Ratio, the mean and median bias, and Rubin's B and R.[51] 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^2 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 13 and 14. Last, we show a graph of the bias reduction and test the common support of the propensity score between treated and untreated in Figure 6. 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.

[49] 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.

[50] Variables are chosen to maximize the 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).

[51] 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 12: Normalized Differences

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

Table 13: Matching Results 1

	Unmatched	**Mean**			**% Bias**	**t-Test**	
Variable	**vs. Matched**	**Treated**	**Control**	**% Bias**	**Reduction**	**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 14: 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.

Figure 6: Matching Graph and Common Support