Asymmetric Information and Imperfect Competition in Lending Markets*

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

June 2014

PRELIMINARY AND INCOMPLETE, PLEASE DON'T CITE WITHOUT PERMISSION

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 adverse selection, and that 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 conduct a policy experiment to double its magnitude. As predicted, in this counterfactual scenario equilibrium prices rise in more competitive markets and decline in more concentrated ones.

*We thank Daniel Ackerberg, Pierre-André Chiappori, Lorenzo Ciari, Valentino Dardanoni, Ramiro de Elejalde, Liran Einav, Moshe Kim, Rocco Macchiavello, Carlos Noton, Tommaso Oliviero, Steven Ongena, Ariel Pakes, Andrea Pozzi, Pasquale Schiraldi, Matt Shum, Michael Waterson, Chris Woodruff, Christine Zulehner and seminar participants at Warwick, PEDL, Barcelona GSE Banking Summer School 2012, EARIE 2012, EUI, Tilburg, Zürich, Bocconi, 2014 Winter Marketing-Economics Summit in Wengen, IO session of the German Economic Association in Hamburg, and St. Gallen 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.

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

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

§LUISS, EIEF and CEPR, fschivardi@luiss.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, mispricing 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. 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 an insurer or a lender, has an information disadvantage with respect to a potential insuree/borrower, it is very unlikely that such a disadvantage can be overcome by the researcher, if not in experimental settings. While one cannot generally construct measures of the ex-ante unobserved characteristics determining riskyness, it is often possible to observe ex-post outcomes, such as filing a claim to an insurance company or 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 (2013)).

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 size, default, and bank pricing based on the insights in Stiglitz and Weiss (1981) that allows us to estimate the extent of adverse selection 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, 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

[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).

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. This implies both that imperfect competition can moderate the welfare losses from asymmetric information and that adverse selection can moderate the welfare losses of market power.

To analyze these questions, we construct a model where banks offer standardized contracts to observationally equivalent firms. Loan contracts 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 set interest rates by competing Bertrand-Nash. 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 and that between unobserved determinants of how much of that loan to use and default. 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.[2] We show with a Monte Carlo simulation that imperfect competition can indeed mitigate the effects of adverse selection.[3] The effects of asymmetric information on prices depends on market power. When markets are competitive, more asymmetric information always leads to higher rates and less credit. As banks' market power increases, this relationship becomes weaker and eventually turns negative.

We estimate the model on highly detailed microdata from the Bank of Italy 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 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

[2] Handel (2013), Lustig (2011), and Starc (2013) 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.

[3] In the Monte Carlo we vary the degree of competition changing the number of banks in the market, as well as varying the price sensitivity of borrowers, which increases/decreases their utility from the outside option of not borrowing.

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

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.

and 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 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 reduces the probability that a firm borrows but, conditional on borrowing, increases the default probability. Among other observables, older firms are both less likely to demand credit, arguably because they have more internally generated funds, and more likely to default. Firms with larger assets demand more credit and default less. In terms of correlation in unobservables, we find a positive correlation between the choice to borrow and default, and between how much loan to use and default. We simulate with a counterfactual experiment the possible consequences of a credit crunch, where risky firms become more exposed to financial distress than safe ones and demand more credit. Our results show that when we increase the correlation in the unobservables (thus increasing the extent of adverse selection), prices in most markets increase, but they fall in some markets. The change in prices is related to different measures of market concentration,[5] supporting the view that market concentration can mitigate the negative effects of asymmetric information. As a consequence of this price decrease, the share of borrowing firms in more concentrated markets increases, and their average default rate falls.

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 evidence that new banks entering local markets perform poorly relative to incumbents, as entrants experience higher default rates and concentration and default rates are positively correlated. 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.

relationships in a companion paper Pavanini and Schivardi (2013).

[5] In the counterfactuals we relate the equilibrium price variation to the estimated markups from the demand model. We also experiment with HHI in terms of branches and loans.

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, Section 4 outlines the structural model, and Section 5 describes the econometric specification of demand, loan size, 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).[6] 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 since 1959.

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 the main source of borrowing of Italian firms. For example, in 1994 they represented 53 percent of the total debts according to the Flow of Funds data. We define the interest rate as the ratio of the payment made in each year by the firm to the bank to the average amount of the loan. The interest payment includes the fixed expenses charged by the bank to the firm (e.g. which encompass the cost of opening the

[6] 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.

[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.

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 over 20,000 potential borrowers, 36% take up a loan in our sample period, and use on average 80% of the amount granted. Of these, around 15% end up being classified as bad loans within the following 3 years.[12] The average amount granted is 350,000 euros, and the average interest rate charged is just below 15%.

Panel B of Table 1 shows summary statistics for the 90 reporting banks. The average total assets level is almost 11 billions, 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 companies; 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 commerce chambers (limited liability companies are obliged to file their balance sheets to the commerce chambers, 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 defined a borrowing firm as one that shows up as a borrower in the CR database. Non borrowing firms

[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] Due to computational constraints, we are able to estimate the model in this version of the paper only on half of the sample. Therefore we randomly pick 50% of the province-year combinations in our sample.

[12] We classify a borrower as defaulter when any of its loans is pass due within the next 3 years from the initial borrowing date.

are defined according to two criteria: (a) they are not in the CR database; (b) they 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. These two groups of firms appear to be fairly similar, except that borrowing firms seems to have more fixed assets and be slightly younger on average. In terms of bank relations, our sample of borrowing firms have on average around 3.4 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 around 70%, and it goes up to 75% 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 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 more than three miles from the branch. 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 almost 6 banks per province-year in our sub-sample, each bank has on average almost 19 branches per province, with a market share of 7% for both branches and loans. On average a bank has been in a province for 22 years.[14]

3 Reduced Form Evidence

We conduct some reduced form analysis to test for evidence of asymmetric information and to justify the use of a structural model. 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. We run these tests on the whole sample and for the first loan ever taken, the set of loans that we will use in the structural estimates.

[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. This might be a possible source of selection bias that we will need to investigate.

[14] 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	20,080	0.36	0.48			
Loan Level	Loan Size	7,170	0.81	1.34			
	Default	7,170	0.15	0.36			
	Amount Granted	7,170	352.85	409.68			
	Interest Rate	7,170	14.67	4.03			
Panel B:	Total Assets	900	10,726.8	16,965.6			
Bank Level	Employees	896	3,179.9	4,582.5			
	Bad Loans	893	6.2	6.3			
	Number of Provinces	861	34.54	30.19			
Panel C:		**Borrowing Firms**			**Non-Borrowing Firms**		
Firm Level	Fixed Assets	7,170	2,631.52	11,136.64	12,876	1,597.84	7,705.57
	Intangible/Tot Assets	7,170	0.19	0.25	12,876	0.19	0.27
	Net Worth	7,170	1,441.18	4,683.39	12,876	1,591.21	6,845.24
	Trade Debit	7,170	1,402.73	4,197.48	12,876	1,358.50	6,723.18
	Profits	7,170	744.90	2,043.97	12,876	517.79	2,636.62
	Cash Flow	7,170	441.37	1,754.81	12,876	474.47	2,362.71
	Firm's Age	7,170	11.68	11.83	12,876	13.76	13.37
	Branch distance (km)	7,170	2.64	6.56			
	Number of Lenders	31,328	3.40	2.36			
	Lines Opened	31,328	1.04	1.55			
	Lines Closed	31,328	0.61	1.23			
	Share of Main Line	26,776	0.70	0.25			
	Share of Main New Line	6,095	0.74	0.25			
Panel D:	Number of Banks	386	5.90	4.11			
Market level	Number of Branches	2,279	18.77	31.34			
	Share of Branches	2,279	0.07	0.09			
	Years in Market	2,279	22.21	13.77			
	Market Shares	2,279	0.07	0.08			

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 size is the share of amount used over granted, default is a dummy for a firm having any of its loans classified as bad at most within 3 years from demanding the loan we consider, 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 8 variables and a firm-year for the others. The balance sheet variables in this panel are winsorized at the 1st and 99th percentile. 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 loans.

3.1 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 only for borrowing firms. We can formalize the problem as a two equations selection model:

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

where d_i is equal to 1 if the firm borrows and f_i is equal to one if the borrower is a defaulter [15] and is observed only if $d_i = 1$. This is similar to the classical selection model analyzed by Heckman (1979), with the only difference that the outcome variable is also binary, rather than being continuous. Adverse selection implies that the correlation between ν and η is positive. If we estimate a linear probability model for default, assuming that ν, η are bivariate normal with correlation coefficient ρ, we can employ the two step procedure of Heckman (1979) by first estimating a probit on d_i, and then constructing the Mills ratio and inserting it in the second equation. A test for a positive correlation between the error terms is a t-test on the coefficient of the Mills ratio in the default equation. As controls in the default equation we use firm level characteristics (total assets, share of intangible assets over total assets, returns on assets, leverage, sales, trade debit, score) as well as sector, year and area dummies. In the selection equation we add the indicators of local financial development in 1936 at the regional level collected by Guiso, Sapienza and Zingales (2004), who show that they are good instruments for financial development today while uncorrelated with current economic performance. They therefore satisfy the condition for a valid exclusion restriction: they affect the probability of obtaining a loan, which varies with the degree of local financial development, but are unlikely to be correlated with the probability of defaulting, conditional on having a loan.[16]

Results reported in Table 2, Panel A, are consistent with the hypothesis that lending is affected by adverse selection. The coefficient of the Mills ratio is positive and statistically significant both when considering first loans and all loans. The magnitude is larger for the second sample, suggesting that adverse selection issues are not confined to the early phase of the firm's borrowing cycle.

3.2 Loan Size and Default

We then consider the relationship between amount of loan used and default probability. 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" and the occurrence of an "accident", conditional on several individual characteristics. We consider two dependent variables for coverage: the absolute amount of credit used as well as the amount of credit used as a share of credit granted. In our lending context we check if firms that use a larger share of their loans are more likely to default on them. Adverse selection should imply that riskier firms use more

[15] As explained in the data section, we define a firm as defaulter if any of its loans are classified as bad up to at most 3 years after borrowing.

[16] This instrument is valid for this simplified setup of the reduced form test, but not for the structural model that we present later, where we need to instrument prices that vary at the bank-market-year level.

Table 2: Positive Correlation Tests

Panel A: Demand and Default		
	First	All
Selection	.131**	.312***
	(.059)	(.023)

Panel B: Loan Size and Default		
	First	All
Used	0.181***	0.170***
	(0.003)	(0.003)
Used/Granted	0.196***	0.186***
	(0.003)	(0.003)

Note: Panel A reports the selection term of a Heckman selection model. The two columns report the coefficient on the Mills ratio in a model where the outcome equation (default or not) is linear. Panel B reports the correlation coefficient of the error terms of a bivariate probit model. Columns labelled "First" only consider the first loan ever, "All" all loans.

credit. We set up the following bivariate probit:

$$\begin{aligned} \ell_i &= \mathbf{1}(X_i\beta + \varepsilon_i > 0) \\ f_i &= \mathbf{1}(X_i\gamma + \eta_i > 0) \end{aligned} \tag{2}$$

where ℓ_i is a dummy equal to one if the loan amount used is above the median, or if the loan amount used over granted is above the median, and f_i takes value of one if the borrower is a defaulter. The vector of controls X_i is composed by year, area, sector, and bank fixed effects, as well as other firm's balance sheet variables, including the score, and the interest rate. We specify the distribution of the residuals ε_i, η_i as jointly normal, with a correlation coefficient ρ. Positive and significant ρ suggests presence of adverse selection. The results of this test are summarized in Table 2 Panel B. The positive correlation is similar for the sample of first loans and for all loans and for both dependent variables.

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 decide which bank $j = 1, ..., J$ to borrow from based on the conditions offered that maximise the expected "profits"[17] of their choice. This determines

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

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.

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.[18]

(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 heterogenous experience ratings among borrowers and loan renegotiation, as the focus of the paper is on first access to credit.[19]

(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. 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 share of new yearly credit (both usable and used), justifying the choice of this simplifying assumption.

(4) Posted Interest Rates: We assume that banks have posted interest rates for types of firms $k = 1, ..., K$ in each market m and period t, depending on the borrowers' characteristics. Following the work by Albareto, Benvenuti, Mocetti, Pagnini and Rossi (2011) on the determinants of interest rates decisions, these types are defined by the amount of credit granted, the firm's sector, the firm's size in terms of sales, and the observable riskiness of the firm defined by the SCORE.[20]

(5) 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 type k in each market m. In a standard insurance or credit market with asymmetric information, firms 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. Empirical evidence of non-exclusivity results also from the pricing regression described in Appendix A, which presents a negative correlation between interest rates and the amount of credit granted.[21] Moreover, the assumption of setting the loan amount as part of the definition of type is also justified by the distribution of amounts granted, characterized by a high

[18] The assumption of asymmetric information in Stiglitz and Weiss (1981) is that lenders observe the mean return of a project, but not its riskiness.

[19] We relax this assumption in a companion paper (Pavanini and Schivardi (2013)).

[20] The construction of these posted interest rates is described in Appendix A.

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

concentration of loans around some specific mass points. 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.

4.1 Demand, Loan Size and Default

Given these assumptions, let there be $i = 1, ..., I$ firms of observable type $k = 1, ..., K$ and $j = 1, ..., J$ banks in $m = 1, ..., M$ markets in period $t = 1, ..., T$. Let firms have the following utility from credit, which determines their demand:

$$U^D_{ikjmt} = \underbrace{\bar{\alpha}^D_0 + \alpha^D_1 P_{jmt} + X'_{jmt}\beta^D + \xi^D_{jmt}}_{\delta^D_{jmt}} + \underbrace{\sigma^D \nu_i + Y'_i \eta^D + \gamma^D_k}_{V^D_i} + \varepsilon^D_{ikjmt}. \tag{3}$$

We normalize to zero 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 share of amount granted to use that maximizes the following utility:

$$U^L_{ikmt} = \underbrace{\alpha^L_0 + \alpha^L_1 P_{jmt} + X'_{jmt}\beta^L + \xi^L_{jmt}}_{\delta^L_{jmt}} + \underbrace{Y'_i \eta^L + \gamma^L_k}_{V^L_i} + \varepsilon^L_{ikmt}. \tag{4}$$

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

$$U^F_{ikmt} = \underbrace{\alpha^F_0 + \alpha^F_1 P_{jmt} + X'_{jmt}\beta^F + \xi^F_{jmt}}_{\delta^F_{jmt}} + \underbrace{Y'_i \eta^F + \gamma^F_k}_{V^F_i} + \varepsilon^F_{ikmt}. \tag{5}$$

Here X_{jmt} are banks' observable attributes, P_{jmt} are the posted interest rates mentioned above,[22] ξ_{jmt} are banks' unobservable (to the econometrician) attributes, Y_i are firms' observable characteristics, and γ_k are types' fixed effects. We assume that ε^D_{ikjmt} 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 $\alpha^D_{0i} = \bar{\alpha}^D_0 + \sigma^D \nu_i$, with $\nu_i \sim N(0, 1)$,[23] to be jointly normally distributed with ε^L_{ikmt}, and ε^F_{ikmt}, such that:

$$\begin{pmatrix} \alpha^D_0 \\ \varepsilon^L \\ \varepsilon^F \end{pmatrix} \sim N\left(\begin{pmatrix} \bar{\alpha}^D_0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^{2D} & \rho_{DL}\sigma^D\sigma^L & \rho_{DF}\sigma^D\sigma^F \\ \rho_{DL}\sigma^D\sigma^L & \sigma^{2L} & \rho_{LF}\sigma^L\sigma^F \\ \rho_{DF}\sigma^D\sigma^F & \rho_{LF}\sigma^L\sigma^F & \sigma^{2F} \end{pmatrix} \right). \tag{6}$$

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

[22] We explain in Appendix A how we separate the bank-market-period price from the type specific price.

[23] 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.

positively correlated with the unobservables that drive default, then riskier firms are more likely to demand. The idea behind the identification of the correlation between α_0^D and ε^F is the following. If we 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_0^D" firm. A positive correlation of α_0^D with ε^F is evidence of adverse selection.

We interpret a positive correlation between the unobservables driving loan size 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.[24] This identification strategy allows us to recover adverse selection parameters that are common across banks and markets, not bank or market specific.[25]

This set up is similar to 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 size 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 size and default equations. We argue that this 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.

4.2 Supply

On the supply side, we let banks set their interest rates competing à la Bertrand Nash. We assume that bank j's profits in market m at time t are given by the sum of the profits made with each subset of its borrowers of types k:

$$\begin{aligned} \Pi_{jkmt} &= (P_{jkmt} - MC_{jmt})Q_{jkmt}(1 - F_{jkmt}) - MC_{jmt}Q_{jkmt}F_{jkmt} \\ &= P_{jkmt}Q_{jkmt}(1 - F_{jkmt}) - MC_{jmt}Q_{jkmt}, \end{aligned} \tag{7}$$

where Q_{jkmt} and F_{jkmt} are bank's expectation of demand and default. In particular, Q_{jkmt} is given by the model's market shares and the expected loan size, and F_{jkmt} is the average default rate for the borrowers of type k that bank j lends to in market m. P_{jkmt} is the price of the loan $(1 + r_j)$. MC_{jmt} are the bank's marginal costs, which we assume to be constant at the bank-market-period level. It is important to note that

[24] 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.

[25] There is not a clear economic interpretation of the correlation between the demand and loan size unobservables, so at the moment we set it to zero for simplicity.

F_{jkmt} depends on price through two channels. First, equation (5) allows for a direct impact of the interest rate on default probability. Second, a higher interest rate also changes the composition of borrowers as stated in Assumption 1: increasing price increases the conditional expectation of α_0^D, as low-utility-from-borrowing firms are more likely to self-select out of the borrowing pool. If $\rho_{DF} \geq 0$, 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:

$$P_{jkmt} = \underbrace{\frac{MC_{jmt}}{1 - F_{jkmt} - F'_{jkmt}\frac{Q_{jkmt}}{Q'_{jkmt}}}}_{\textit{Effective Marginal Costs}} - \underbrace{\frac{(1 - F_{jkmt})\frac{Q_{jkmt}}{Q'_{jkmt}}}{1 - F_{jkmt} - F'_{jkmt}\frac{Q_{jkmt}}{Q'_{jkmt}}}}_{\textit{Markup}}, \tag{8}$$

Note that the equilibrium price depends on what we define as "effective" marginal costs and on a markup term. F'_{jkmt} is the derivative of the expected default rate with respect to prices, and Q'_{jkmt} is the derivative of the market share with respect to prices. $\frac{Q_{jkmt}}{Q'_{jkmt}}$ 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. $F_{jkmt} = F'_{jkmt} = 0$, 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.

4.3 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 monopoly bank facing $i = 1, ..., N$ heterogeneous borrowers. For simplicity, we concentrate on adverse selection between demand and default (ρ_{DF}), setting loan size to 1 and $\rho_{DL} = \rho_{LF} = 0$. We keep this data fixed and vary the number of banks, borrowers' price sensitivity, 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 banks' equilibrium prices based on our model. Let borrower i have U_{ij}^D utility from taking credit from bank j, U_{i0}^D utility from not borrowing, and U_i^F utility from defaulting:

$$\begin{aligned} U_{ij}^D &= \alpha_{0i} + \alpha_1 P_j + \epsilon_{ij}, \\ &= \bar{\alpha}_0 + \sigma\nu_i + \alpha_1 P_j + \epsilon_{ij}, \\ U_{i0}^D &= \epsilon_{i0}, \\ U_i^F &= \varepsilon_i, \end{aligned} \tag{9}$$

where P_j is the interest rate charged by bank j, ϵ_{ij}, ϵ_{i0} are distributed as type 1 extreme value, and $\nu_i \sim N(0, 1)$. We set $\sigma = 2$ and $\bar{\alpha}_0 = 1$, and allow α_i 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 $\alpha_1 < 0$. 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_j for everyone. Given this setup, the demand probability will be given by:

$$\begin{aligned} \Pr_{ij}^{D} &= \Pr(\alpha_{0i} + \alpha_1 P_j + \epsilon_{ij} > \alpha_{0i} + \alpha_1 P_h + \epsilon_{ih} \; \forall h \neq j) \\ &= \frac{\exp(\alpha_{0i}+\alpha_1 P_j)}{1+\sum_\ell \exp(\alpha_{0i}+\alpha_1 P_\ell)} \\ &= \Lambda(\alpha_{0i} + \alpha_1 P_j), \end{aligned} \tag{10}$$

and we will construct banks' market shares as $S_j = \frac{1}{N}\sum_i \Pr_{ij}^{D}$. Conditional on demand, default probability will follow from Wooldridge (2002) as:

$$\begin{aligned} \Pr_{ij}^{F|D=j} &= E\big[\Pr(F=1|\nu, P_j)|D=j, P_j\big] \\ &= \frac{1}{\Lambda(\alpha_{0i}+\alpha_1 P_j)} \int_{-(\alpha_{0i}+\alpha_1 P_j)}^{\infty} \Phi\left(\frac{\frac{\rho\nu}{\sigma^2}}{\sqrt{1-\frac{\rho^2}{\sigma^2}}}\right)\phi(\nu)d\nu, \end{aligned} \tag{11}$$

and we will construct banks' share of defaulters as $F_j = \frac{1}{N_j}\sum_i \Pr_{ij}^{F}$, where N_j is the number of borrowers of bank j. Given these probabilities and our supply side model described in equations (7) and (8), the first order conditions will deliver the following equilibrium pricing equation for each bank:

$$P_j^* = \underbrace{\frac{MC}{1 - F_j - F_j' \frac{1}{\alpha_1(1-S_j)}}}_{\textit{Effective Marginal Costs}} - \underbrace{\frac{(1-F_j)\frac{1}{\alpha_1(1-S_j)}}{1 - F_j - F_j' \frac{1}{\alpha_1(1-S_j)}}}_{\textit{Markup}}, \tag{12}$$

where the first term on the right hand side represents what we define as "effective" marginal costs (EffMC), and the second term represents the markup (MKP). F_j' is the derivative of the expected default rate with respect to prices, and $\alpha_1(1-S_j)$ is the derivative of the market share with respect to prices. For $\alpha_1 < 0$ it can be shown that the EffMC term is always positive and the markup term is always negative.

The different effects of these two factors on equilibrium prices is crucial to understand the interaction between asymmetric information and imperfect competition. This is displayed in Figures 1 and 2, 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_1. This means that for the moment we are capturing competition versus the outside option, but we have verified that increasing the number of banks gives the same result.

Looking at Figure 1, 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 (leftmost point, moving to the northeast), but it decreases substantially

for a low level of competition (closest point to the reader, moving to the northeast). What the graph shows in fact 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 2, the combination of these two factors results in a non-monotonic equilibrium price response to an increase in adverse selection. If on one hand equilibrium prices rise 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 increasing EffMC drive prices up, whereas in the second case the declining markup drives prices down. More intuitively, in a highly competitive market where banks have small price-cost margins, higher prices is the only possible response to an increase in adverse selection. However, banks with a higher price-cost margin will find it more profitable to reduce prices, as this will allow them to lower the average riskiness to their pool of borrowers.

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 that borrows, and $j = 1, ..., J_{mt}$ be the bank/loan identifier in market m at time t. Moreover, let $k = 1, ..., K$ identify the type of firm that is borrowing. The k index further segments the market, as banks can lend across all types of firms within the same market, but firms can only borrow at the interest rate offered to their own type. Let Y_i 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), X_{jmt} a vector of bank-province-year specific attributes (number of branches in the market, years of presence in the market, bank fixed effects), and γ_k types' fixed effects.

We estimate a system of three equations: demand for credit, amount of loan used, and default. We use a 2-step method based on maximum simulated likelihood and instrumental variables (Train (2009)). In the first step we estimate the firm-level parameters $\eta = \{\eta^D, \eta^L, \eta^F\}$, the types' fixed effects $\gamma_k = \{\gamma_k^D, \gamma_k^L, \gamma_k^F\}$, the correlation coefficients $\rho = \{\rho_{DF}, \rho_{DL}, \rho_{LF}\}$, and the covariances σ^D and σ^L from the firms' choice probabilities.[26] We follow Einav et al. (2012), but differ 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 $\widehat{\delta}_{jmt} = \{\widehat{\delta}_{jmt}^D, \widehat{\delta}_{jmt}^L, \widehat{\delta}_{jmt}^F\}$ using the contraction method introduced by Berry (1994).

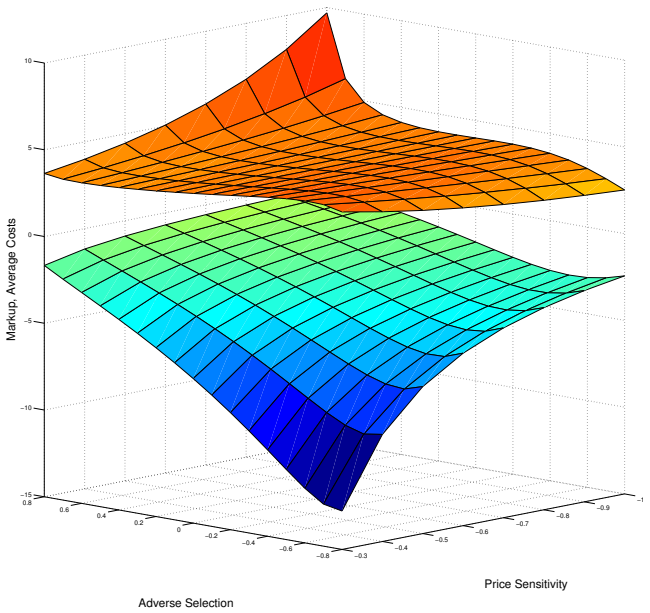
The probability that borrower i of type k in market m at time t chooses lender j is given by:

$$\Pr_{ikjmt}^D = \int \left[\frac{\exp(\widehat{\delta}_{jmt}^D(X_{jmt}, P_{jmt}, \xi_{jmt}^D, \beta^D) + V_i^D(Y_i, \eta^D, \gamma_k^D))}{1 + \sum_\ell \exp(\widehat{\delta}_{\ell mt}^D(X_{\ell mt}, P_{\ell mt}, \xi_{\ell mt}^D, \beta^D) + V_i^D(Y_i, \eta^D, \gamma_k^D))} \right] f(\alpha_{0i}^D|\theta) d\alpha_{0i}^D, \tag{13}$$

where $f(\alpha_{0i}^D|\theta)$ is the density of α_{0i}^D, and θ are the parameters of its distribution that we want to estimate. The estimation of this choice model only provides the estimates of $\eta^D, \gamma_k^D, \sigma^D$, but not of the parameters

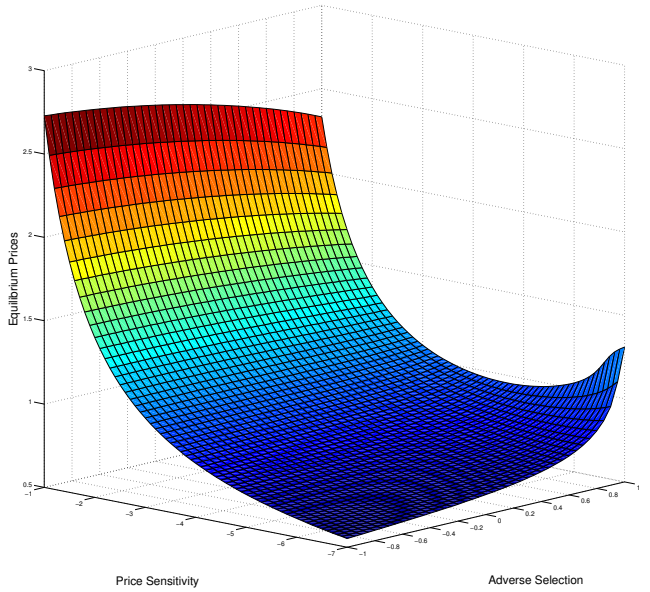
[26] In this version of the paper we are still not estimating ρ_{DL}, which we set to zero, and σ^D, which we set to 1. For the second, it is due to the well known identification problem of the standard deviations of random coefficients in Berry et al. (1995), explained in Berry, Levinsohn and Pakes (2004) and Train and Winston (2007). We are working on incorporating second preferred choices into the model to guarantee better identification and to be able to estimate this parameter.

Figure 1: 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 2: 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 1 to better display the effects in each.

in δ^D. Looking at the second equation, the share of credit used over granted conditional on borrowing, the probability of observing a utilization of L_{ikmt} is given by:

$$\begin{aligned}\Pr^L_{ikmt,L=L^*|D=1,\alpha^D_{0i}} &= \Pr(L_{ikmt} = \delta^L_{jmt} + V^L_i + \varepsilon^L_{ikmt}|\alpha^D_{0i}) \\ &= \int \frac{1}{\tilde{\sigma}_{\varepsilon^L_{ikmt}|\alpha^D_{0i}}} \phi_{\varepsilon^L_{ikmt}|\alpha^D_{0i}} \left(\frac{L_{ikmt} - \widehat{\delta}^L_{jmt}(X_{jmt}, P_{jmt}, \xi^L_{jmt}, \beta^L) - V^L_i(Y_i, \eta^L, \gamma^L_k) - \tilde{\mu}_{\varepsilon^L_{ikmt}|\alpha^D_{0i}}}{\tilde{\sigma}_{\varepsilon^L_{ikmt}|\alpha^D_{0i}}} \right) f(\alpha^D_{0i}|\theta) d\alpha^D_{0i},\end{aligned} \tag{14}$$

$$\text{where} \qquad \varepsilon^L_{ikmt}|\alpha^D_{0i} \sim N\big(\underbrace{\sigma^L \rho_{DL} \nu_i}_{\tilde{\mu}_{\varepsilon^L_{ikmt}|\alpha^D_{0i}}}, \underbrace{\sigma^{2L}(1-\rho^2_{DL})}_{\tilde{\sigma}^2_{\varepsilon^L_{ikmt}|\alpha^D_{0i}}}\big)$$

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

$$\Pr^F_{ikmt,F=1|D=1,\alpha^D_{0i},\varepsilon^L_{ikmt}} = \int \Phi_{\varepsilon^F_{ikmt}|\alpha^D_{0i},\varepsilon^L_{ikmt}} \left[\frac{\widehat{\delta}^F_{jmt}(X_{jmt}, P_{jmt}, \xi^F_{jmt}, \beta^F) + V^F_i(Y_i, \eta^F, \gamma^F_k) - \tilde{\mu}_{\varepsilon^F_{ikmt}|\alpha^D_{0i},\varepsilon^L_{ikmt}}}{\tilde{\sigma}_{\varepsilon^F_{ikmt}|\alpha^D_{0i},\varepsilon^L_{ikmt}}} \right] f(\alpha^D_{0i}|\theta) d\alpha^D_{0i}, \tag{15}$$

$$\text{where} \qquad \varepsilon^F_{ikmt}|\alpha^D_{0i}, \varepsilon^L_{ikmt} \sim N\big(\underbrace{A\sigma^D \nu_i + B\varepsilon^L_{ikmt}}_{\tilde{\mu}_{\varepsilon^F_{ikmt}|\alpha^D_{0i},\varepsilon^L_{ikmt}}}, \underbrace{\sigma^{2F} - (A\rho_{DF} + B\rho_{LF})}_{\tilde{\sigma}^2_{\varepsilon^F_{ikmt}|\alpha^D_{0i},\varepsilon^L_{ikmt}}}\big)$$

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

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

where the residuals ε^F_{ikmt} are conditional on demand and loan amount unobservables. Similarly to the demand side, the estimation of these two choice equations, jointly with the demand one, only delivers the parameters $\eta^L, \eta^F, \gamma^L_k, \gamma^F_k, \rho, \sigma^L$.

In the second step, the estimated constants $\widehat{\delta}_{jmt}$ are the dependent variables of instrumental variable regressions that recover the parameters $\bar{\alpha}^D_0, \alpha^D_1, \alpha^L_0, \alpha^L_1, \alpha^F_0, \alpha^F_1, \beta^D, \beta^L, \beta^F$ of the bank specific attributes X_{jmt} and prices P_{jmt}. This second step also controls for the potential endogeneity bias caused by the correlation between prices and unobserved (to the econometrician) bank attributes $\xi_{jmt} = \{\xi^D_{jmt}, \xi^L_{jmt}, \xi^F_{jmt}\}$. Following Berry (1994), 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:

$$\delta^{D,r+1}_{jmt} = \delta^{D,r}_{jmt} + \ln\left(\frac{S^D_{jmt}}{\widehat{S}^D_{jmt}(\delta^{D,r}_{jmt})} \right). \tag{16}$$

The predicted market shares are defined as $\widehat{S}^D_{jmt} = \sum_i \Pr^D_{ikjmt}/N_{mt}$, where N_{mt} are the number of borrowers in market m at time t. Given the value of these constant terms, the parameters $\bar{\alpha}^D_0, \alpha^D_1, \beta^D$ are estimated

using instrumental variables:

$$\delta^D_{jmt} = \bar{\alpha}^D_0 + \alpha^D_1 P_{jmt} + X'_{jmt}\beta^D + \xi^D_{jmt}, \tag{17}$$

with ξ^D_{jmt} being the mean zero structural econometric error term. Similarly, the lender-market constants for loan size δ^L_{jmt} and default δ^F_{jmt} are estimated using a nonlinear least squares search routine as in Goolsbee and Petrin (2004), which solves for:

$$\delta^L_{jmt} = \arg\min_{\delta} \sum_j \left(\hat{S}^L_{jmt}(\eta^L, \delta^L) - S^L_{jmt}\right)^2, \tag{18}$$

$$\delta^F_{jmt} = \arg\min_{\delta} \sum_j \left(\hat{S}^L_{jmt}(\eta^F, \delta^F) - S^L_{jmt}\right)^2, \tag{19}$$

where $\widehat{S}^L_{jmt}, \widehat{S}^F_{jmt}$ and S^L_{jmt}, S^F_{jmt} are the predicted and actual shares of loan sizes and defaults for lender j in market m at time t. Given the value of these constant terms, the parameters $\alpha^L_0, \alpha^L_1, \beta^L$ and $\alpha^F_0, \alpha^F_1, \beta^F$ are estimated using instrumental variables:

$$\delta^L_{jmt} = \alpha^L_0 + \alpha^L_1 P_{jmt} + X'_{jmt}\beta^L + \xi^L_{jmt}, \tag{20}$$

$$\delta^F_{jmt} = \alpha^F_0 + \alpha^F_1 P_{jmt} + X'_{jmt}\beta^F + \xi^F_{jmt}. \tag{21}$$

6 Estimation

Following from section 5, we use the demand, loan size and default probabilities to construct the simulated maximum likelihood that allows us to recover the parameters in $\eta, \gamma_k, \sigma^D, \sigma^L, \rho$:

$$\log L = \sum_i \log(\text{Pr}^D_{ikjmt}) d_{ikjmt} + \sum_{i \in D} \Big[\log(\text{Pr}^L_{ikmt}) + \log(\text{Pr}^F_{ikmt}) f_{ikmt} + \log(1 - \text{Pr}^F_{ikmt})(1 - f_{ikmt}) \Big], \tag{22}$$

where d_{ikjmt} is the dummy for the choice by firm i of type k of bank j in market m at time t, and f_{ikmt} is the dummy identifying its default. In order to estimate the remaining parameters we need an additional step explained below.

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). 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.[27] [28] Following other papers on Italian local credit markets (Felici and Pagnini (2008), Bofondi and Gobbi (2006), 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.[29] Our markets are then constructed as province-year combinations. We define the loan size variable as the share of loan used over loan granted, and define default as an ever default variable, as explained before.

The observable explanatory variables that determine firm's demand, loan size and default choices are firm and bank characteristics, summarized in Table 1. In the first set of regressors we include firms' fixed assets, the ratio of intangible over total assets, net worth, trade debit, profits, cash flow, and age, where trade debit is the debit that the firm has with its suppliers or clients. We also include types' fixed effects, where a type is defined as a combination of amount granted, sector, size, and score.[30] In the second group we use prices, bank's share of branches in the province, number of years the bank had at least one branch in the province, and bank dummies. We also control for the distance between each firm and the closest branch of each bank. We provide details on these variables in the appendix. We motivate the choice of these explanatory variables in Section 6.3.

6.2 Identification

The use of instrumental variables in the second step of the estimation aims at correcting the potential endogeneity bias in the price coefficient for the three equations. The bias derives from the possible correlation between prices P_{jmt} and unobserved (to the econometrician) bank-market level characteristics ξ_{jmt}. These unobserved attributes can be thought as the borrowers' valuation of a banks' brand, quality, and credibility, which are assumed to influence borrowers' demand, loan size, and default decisions, but are also very likely to be correlated with banks' interest rates. Think for example of ξ_{jmt} as a banks' reputation for offering valuable and helpful assistance to its borrowers in their business projects, which is unobserved to the econometrician. Borrowers will value this quality when deciding which bank to get credit from, and they will also be affected in their likelihood of using more or less credit and of defaulting. Consequently, the bank will be likely to charge a higher interest rate, given the potentially higher markup that this attribute can provide. Moreover, assuming default is increasing in interest rates, a good assistance can lower the borrower's default probability, allowing banks to charge a higher rate.

To address the simultaneity problem, following Nevo (2001), we include bank dummies to capture the bank characteristics that do not vary by market (year-province). This means that the correlation between prices and banks' nationwide-level unobserved characteristics is fully accounted for with these fixed effects, and does not require any instruments. Hence, we can rewrite equation (17), and similarly equations (20) and (21), as:

$$\delta^D_{jmt} = \bar{\alpha}^D_0 + \alpha^D_1 P_{jmt} + X'_{jmt}\beta^D + \xi^D_j + \Delta\xi^D_{jmt}, \tag{23}$$

where ξ^D_j are banks' fixed effects and $\Delta\xi^D_{jmt}$ are bank-market-time specific deviations from the national

[27] We do this in a companion paper Pavanini and Schivardi (2013).

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

[29] See Pavanini and Schivardi (2013) and Ciari and Pavanini (2013) for a detailed discussion on the definition of local banking markets in Italy.

[30] See the Appendix A for a detailed description of the types.

mean valuation of the bank. Therefore, we need to use instrumental variables to account for the potential correlation between interest rates and these bank-market-time specific deviations. We argue that a valid instrument is represented by the share of branches in a specific market of merging rival banks.[31] Since mergers only happen in a single year, this accounts to relying on the across time correlation of prices with changes in concentration among branches at the market level. The first stage regression shows that this correlation is positive and significant, implying that greater rivals' concentration leads to higher interest rates. We verify empirically the rank condition for instruments' validity with the first stage estimates[32], showing that the instruments are good predictors of interest rates. We compare OLS and IV second stages, to show how the instruments lessen the simultaneity bias.[33] Last, for the exclusion restriction to hold, we assume that bank-market-time specific deviations $\Delta\xi^D_{jmt}$ are uncorrelated with the share of branches of merging rival banks in a market-time combination. We interpret these deviations, for example, as market specific differences in a bank's quality with respect to its national average quality. These can be thought as differences in local managers' capacities, or in a bank's management connections with the local industries and authorities. These factors are likely to influence a bank's prices in that local market, but not the merging decision of rivals, which are usually taken at a national level and are effective across various markets.

6.3 Results

The estimates of the structural model are presented in Table 3. The three columns of results refer respectively to the demand, loan size 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 correlation coefficients of interest, i.e. the correlation between unobservables of demand and default (ρ_{DF}) and the correlation between unobservables of loan size and default (ρ_{LF}). We decided to include those specific firm characteristics to control for different measures of firms' assets, profitability, debt, age, and distance, and for our definition of observable type. We chose among the wide set of balance sheet variables running various reduced form regressions for demand, loan size, and default. We wanted to control for different measures of firm size, in the form of assets[34], but also for some measures of firms' current performance, in terms of profits and cash flow. We also tried to control for other specific forms of finance that firms have access to, such as debt from suppliers[35]. Finally, 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.[36] As introduced in the previous sections, we include fixed effects for the type of the firm as

[31] We experimented also with other instruments, with similar results. In one case we followed the approach of Nevo (2001) and Hausman and Taylor (1981), which implies instrumenting the prices charged by a bank j in a market m with the average of the prices that the same bank charges in all the other markets. We also tried with banks' expenditure in software per employee, weighted by the number of branches in a market, and with the sum of rivals' characteristics, as in Berry et al. (1995).

[32] First stage estimates are reported in Appendix B.

[33] OLS and IV second stage estimates are reported in Appendix B.

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

[35] 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.

[36] It is important to include distance as Degryse and Ongena (2005) show empirical evidence, using Belgian data, that in lending relationship transportation costs cause spatial price discrimination. They find that loan rates decrease with the distance between the

a determinant of the posted prices.[37] Following the survey of Albareto et al. (2011), these types are constructed as the combination of the firm's sector (primary, secondary, tertiary), size (sales above or below the median), riskiness (three risk categories based on the SCORE), and amount granted (five categories between 0 and 3,000,000 €). 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 asymmetric information, both in terms of the correlation between demand and default unobservables and loan size 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. In general, it seems that firms with more net worth, cash flow and trade debit are less likely to demand credit, but firms with more fixed assets and profits are more likely to borrow. Older firms are also less likely to demand. Firms seem to favor banks with a higher share of branches, but are less likely to demand form banks with longer experience in a market. This might be because the sample we are considering is of new borrowers, which might be perceived as more risky by experienced banks. Hence, these firms are more likely to get better conditions from less experienced banks. The share of loan used seems to follow the same logic as demand for fixed assets and ratio of intangible assets, as well as profits, cash flow, interest rates and share of branches. Differently from demand, the share of loan used over granted is increasing in the distance from the branch firms are borrowing from. For what concerns the default probability, this is negatively influenced by more assets, cash flow and trade debit, but positively affected by net worth, profits, and firm's age. As expected, higher interest rates increase default probability.

The mean of own and cross price elasticities for the main 5 banks in the sample are reported in Table 4. We find that on average a 1% increase in interest rate reduces a bank's own market share by over 2%, and increases competitor banks' shares by about 0.2%.

borrower and the lender, and increase with the distance between the borrower and the competing lenders.

[37] See Appendix A for a detailed description of how we construct types.

Table 3: Structural Estimates

Variables			Demand	Loan Size	Default
1st Stage Firm Level	**Assets**	Fixed Assets	0.197***	0.018**	-0.026**
			(0.017)	(0.007)	(0.009)
		Intangible/Total Assets	1.352***	2.001***	-1.189***
			(0.429)	(0.123)	(0.211)
		Net Worth	-0.370***	0.021	0.116***
			(0.024)	(0.016)	(0.016)
	Profitability	Profits	0.623***	0.040***	0.167***
			(0.036)	(0.016)	(0.014)
		Cash Flow	-0.327***	-0.049***	-0.451***
			(0.028)	(0.015)	(0.027)
	Debt	Trade Debit	-0.692***	0.003	-0.116***
			(0.044)	(0.018)	(0.023)
	Others	Firm's Age	-0.798***	-0.036	0.131***
			(0.050)	(0.029)	(0.014)
		Distance	-5.481***	0.088**	-0.020
			(0.257)	(0.037)	(0.019)
		Type FE	Yes	Yes	Yes
2nd Stage Bank Level		Interest Rate	-3.669***	-0.295***	2.387***
			(0.348)	(0.097)	(0.389)
		Number of Branches	-2.746***	-0.102	-0.236
			(0.269)	(0.075)	(0.300)
		Share of Branches	12.646***	0.429**	0.098
			(0.721)	(0.201)	(0.805)
		Years in Market	-1.001***	-0.032	-0.124
			(0.172)	(0.048)	(0.193)
		Bank FE	Yes	Yes	Yes
Adverse Selection	**Demand-Default**	ρ_{DF}			0.304***
					(0.006)
	Loan Size-Default	ρ_{LF}			0.159***
					(0.001)
Obs			301,334	7,170	7,170

Note: Standard errors in brackets. $*$ is significant at the 10% level, $**$ at the 5% level, and $***$ at the 1% level.

7 Counterfactuals (to be completed)

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 increase in adverse selection, and analyze the consequence of this change on equilibrium prices, quantities, and defaults. An increase in adverse selection captures the idea that during a financial crisis investment opportunities contract for all firms, but risky firms will be more exposed than safer ones, demanding more credit. We simulate this scenario doubling the estimated correlation coefficients. Once we recover the new equilibrium outcomes of interest in the new scenario, we investigate whether the variations that we observe from the baseline model are correlated with various measures of competition in the different local markets.

In this counterfactual exercise we follow the example of Nevo (2000) and recover each bank's marginal costs using the pricing equation (8):

$$\widehat{MC}_{jmt} = P_{jmt}\Big[1 - F_{jmt} - F'_{jmt}\frac{Q_{jmt}}{Q'_{jmt}}\Big] + \frac{(1 - F_{jmt})\frac{Q_{jmt}}{Q'_{jmt}}}{1 - F_{jmt} - F'_{jmt}\frac{Q_{jmt}}{Q'_{jmt}}} \tag{24}$$

Under the assumption of marginal costs being the same in each scenario, we re-calculate banks' market shares, loan sizes and defaults with the counterfactual level of adverse selection, and derive the new equilibrium prices as:

$$\widetilde{P}_{jmt} = \frac{\widehat{MC}_{jmt}}{1 - \widetilde{F}_{jmt} - \widetilde{F}'_{jmt}\frac{\widetilde{Q}_{jmt}}{\widetilde{Q}'_{jmt}}} - \frac{(1 - \widetilde{F}_{jmt})\frac{\widetilde{Q}_{jmt}}{\widetilde{Q}'_{jmt}}}{1 - \widetilde{F}_{jmt} - \widetilde{F}'_{jmt}\frac{\widetilde{Q}_{jmt}}{\widetilde{Q}'_{jmt}}}, \tag{25}$$

where $\widetilde{Q}_{jmt}$ and $\widetilde{F}_{jmt}$ are the new equilibrium quantities and defaults under the counterfactual scenario. Following the non-monotonic price response predicted in the Monte Carlo experiment, we investigate what happens to equilibrium prices in this counterfactual scenario with respect to the actual prices. As shown in Figure 3, we find that almost all the prices vary, with the majority increasing by up to 5% (with some outliers not included in this figure), but some of them decreasing by at most 5%. We relate this price variation[38] to a measure of bank-province-year specific market power, which is the predicted markup derived from the demand model, i.e. the last term in equation (24).[39] We present this relationship in Figure 4, where we show evidence of a negative and statistically significant correlation between market power and price variation.[40] This means that prices increase in more competitive markets and decrease in more concentrated ones, confirming the predictions of the model.

We also look at the variation in quantities in the counterfactual scenario. We focus on the change in the share

[38] We measure price variation as: $\Delta P_{jmt} = \frac{\widetilde{P}_{jmt} - P_{jmt}}{P_{jmt}}$.

[39] We use the markup estimated in the baseline model. We tried also with other measures of competition at the local market level, like HHI of branches and loans, with similar results.

[40] We run a regression of bank-province-year level price variation on bank-province-year level markup, controlling for province-year fixed effects, and find a negative and significant coefficient.

of borrowing firms[41] due to an increase in adverse selection. As shown in Figure 5, similarly to the price variation, we find that the share of borrowing firms varies in both direction, both increasing and decreasing. When we regress this share variation on the average markup at the province-year level, we find a positive and significant coefficient, presented in Figure 6. This suggests that the share of borrowing firms increases in more concentrated markets, as a natural consequence of the price reduction.

Last, we consider the effect of an increase in adverse selection on the average default rates[42] that a bank faces in a province-year. Figure 7 confirms that again most of the defaults are unchanged, but that a fraction of the banks in some provinces-years experiences either an increase or a reduction in borrowers' defaults. When we investigate the relationship between changes in default rates and markups, shown in Figure 8, we find that banks with a higher markup tend to reduce their default rates. This is the effect of the reduction in prices, which on one side attracts safer borrowers, and on the other reduces the probability of default of borrowers, given the estimated positive and significant effect of interest rates on defaults in Table 3.

[41] We measure this as the variation in the sum of the shares of the inside goods: $\Delta \sum_j S_{jmt} = \sum_j \widetilde{S}_{jmt} - \sum_j S_{jmt}$

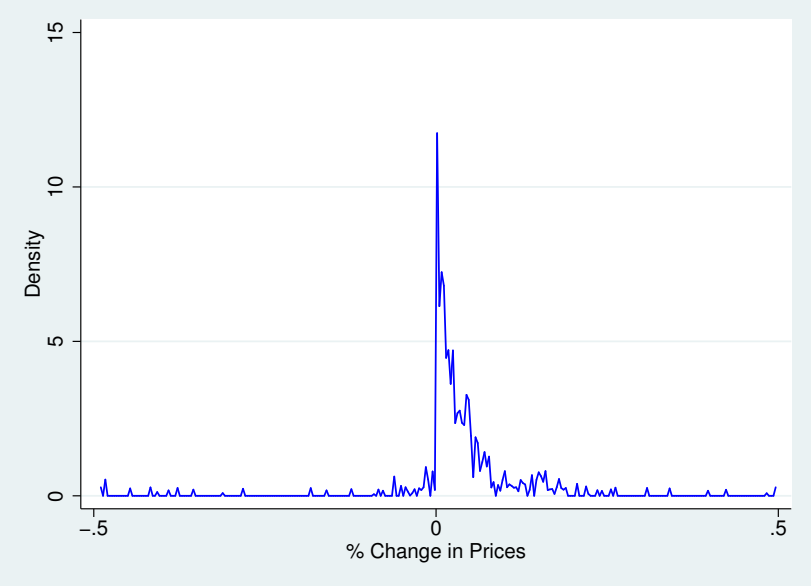
[42] We measure default rates' variation as: $\Delta F_{jmt} = \widetilde{F}_{jmt} - F_{jmt}$

Table 4: 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	-2.635	0.133	0.206	0.168	0.136
Bank 2	0.164	-2.394	0.163	0.164	0.154
Bank 3	0.261	0.258	-2.662	0.276	0.257
Bank 4	0.202	0.171	0.234	-2.562	0.162
Bank 5	0.184	0.161	0.183	0.174	-2.734

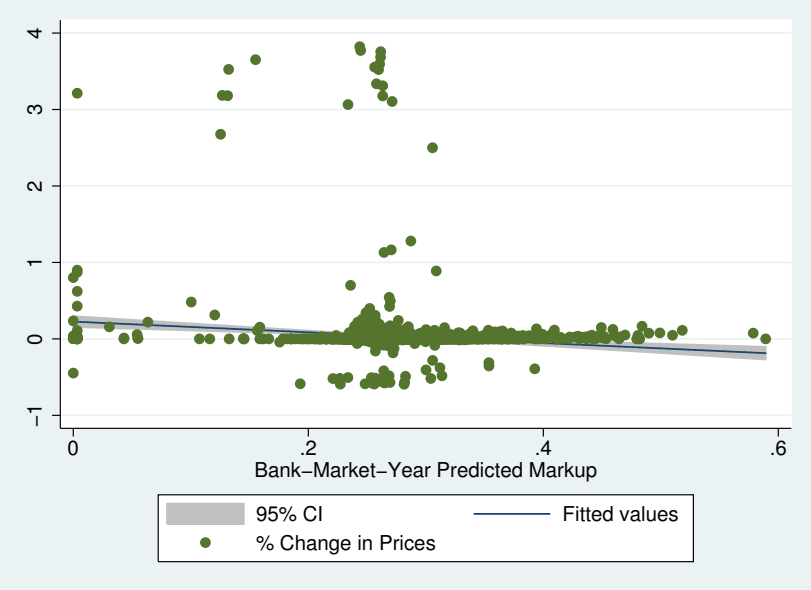
Note: These are the first 5 banks in terms of national market shares. Elasticities are interpreted as the percentage change in market shares in response to a 1% increase in prices.

Figure 3: Kernel Density of Price Variations



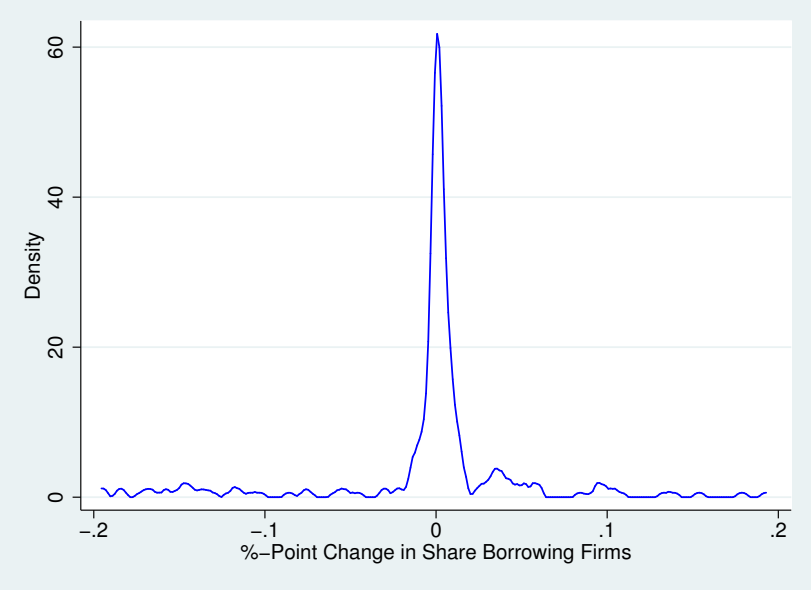
Note: The vertical axis is the density. The horizontal axis is the percentage variation between actual and counterfactual prices.

Figure 4: Regression of Price Variation on Markup



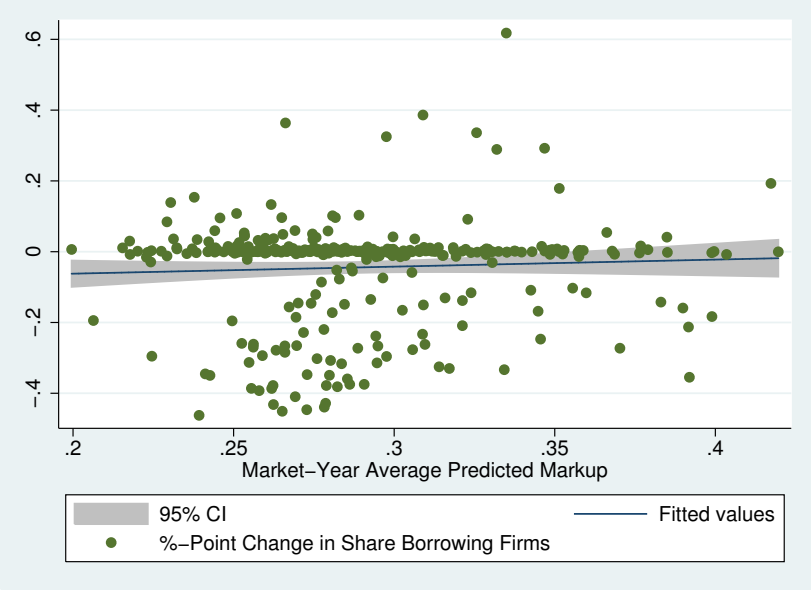
Note: The vertical axis is the percentage variation between actual and counterfactual prices. The horizontal axis is the bank-province-year level measure of markup.

Figure 5: Kernel Density of Quantity Variations



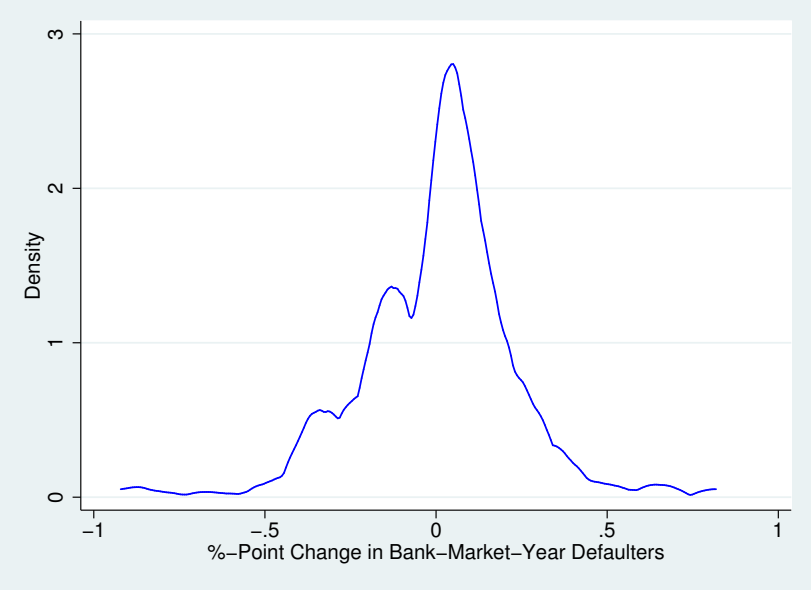
Note: The vertical axis is the density. The horizontal axis is the variation between actual and counterfactual share of the borrowing firms in each province-year in percentage points.

Figure 6: Regression of Quantity Variation on Markup



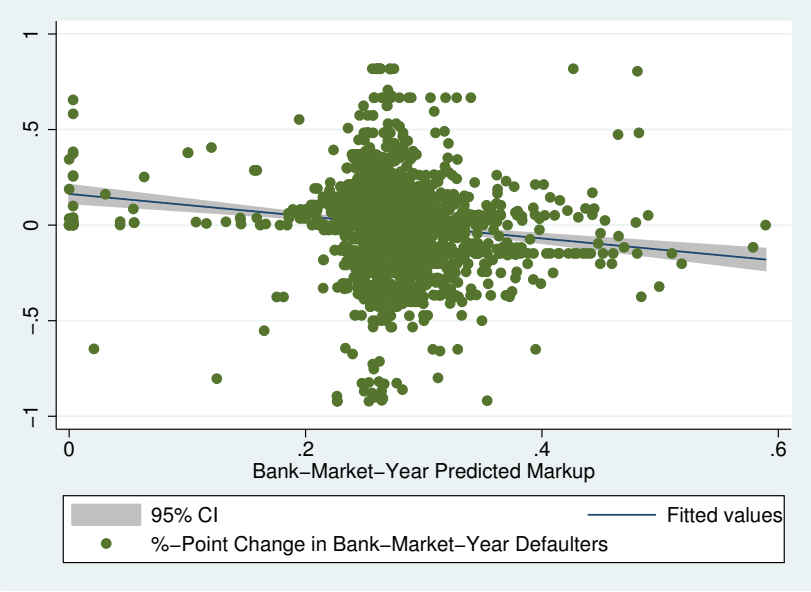
Note: The vertical axis is the variation between actual and counterfactual share of the borrowing firms in each province-year in percentage points. The horizontal axis is the percentage variation between actual and counterfactual prices.

Figure 7: Kernel Density of Default Variations



Note: The vertical axis is the density. The horizontal axis is the variation between actual and counterfactual default rates in percentage points.

Figure 8: Regression of Default Variation on Markup



Note: The vertical axis is the variation between actual and counterfactual default rates in percentage points. The horizontal axis is the bank-province-year level measure of markup.

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 this data, we provide reduced form evidence of adverse selection in the spirit of the positive correlation test on unobservables by Chiappori and Salanié (2000). We find stronger presence of asymmetric information for new borrowers.

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 average 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, 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, both in the form of a positive correlation between unobservables determining demand and default, and unobservables affecting the size of the loan and 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, doubling adverse selection. As predicted, in this counterfactual scenario equilibrium prices rise for more competitive markets and decline for more concentrated ones. Moreover, 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.

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.

Altman, E., "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.

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.

____ , ____ , and ____ , "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Vehicle Market," *Journal of Political Economy*, 2004, *112* (1), 68–104.

Bofondi, M. and G. Gobbi, "Informational Barriers to Entry into Credit Markets," *Review of Finance*, 2006, *10*, 39–67.

Chiappori, P. A. and B. Salanié, "Testing for Asymmetric Information in Insurance Markets," *Journal of Political Economy*, 2000, *108* (1), 56–78.

Ciari, L. and N. Pavanini, "Market Structure and Multi Market Contact," 2013. Working Paper.

Degryse, H. and S. Ongena, "Distance, Lending Relationships, and Competition," *Journal of Finance*, 2005, *LX* (1), 231–266.

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.

Einav, L., M. Jenkins, and J. Levin, "Contract Pricing in Consumer Credit Markets," *Econometrica*, 2012, *80* (4), 1387Ð1432.

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.

Gerakos, J. and C. Syverson, "Competition in the Audit Market: Policy Implications," 2014. Working Paper.

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.

____ , P. Sapienza, and L. Zingales, "Does Local Financial Development Matter?," *Quarterly Journal of Economics*, 2004, *119* (3), 929–969.

Handel, B., "Adverse Selection and Switching Costs in Health Insurance Markets: When Nudging Hurts," 2013. Forthcoming *American Economic Review*.

Hausman, J. A. and W. E. Taylor, "Panel Data and Unobservable Individual Effects," *Econometrica*, 1981, *49* (6), 1377–1398.

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

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

Karlan, D. and J. Zinman, "Observing Unobservables: Identifying Information Asymmetries with a Consumer Credit Field Experiment," *Econometrica*, 2009, *77* (6), 1993–2008.

Lustig, J., "Measuring Welfare Losses from Adverse Selection and Imperfect Competition in Privatized Mediare," 2011. Boston University Working Paper.

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

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

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.

____ , "Measuring Market Power in the Ready-To-Eat Cereal Industry," *Econometrica*, 2001, *69* (2), 307–342.

Panetta, F., F. Schivardi, and M. Shum, "Do Mergers Improve Information? Evidence from the Loan Market," *Journal of Money, Credit and Banking*, 2009, *41* (4), 673–709.

Pavanini, N. and F. Schivardi, "Dynamic Entry and Exit with Learning by Branching," 2013. University of Warwick Working Paper.

Petersen, M. A. and R. G. Rajan, "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.

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.

Starc, A., "Insurer Pricing and Consumer Welfare: Evidence from Medigap," 2013. RAND Journal of Economics forthcoming.

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 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 [43]. 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 Predicting Prices and Amounts Granted

We only consider the first year in which a firm appears in our sample. We assume that banks have a posted price for each observable type of firm in each market, defined as a year-province combination. We recover this synthetic price using regression analysis based on the actual prices that we observe.[44] We need to do this to predict the price that would have been offered to firms not borrowing in the data, as well as the price

[43] 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.

[44] We are working on alternative ways of predicting prices, following Gerakos and Syverson (2014).

that would have been offered to borrowing firms by banks other than the chosen one. For this reason, we use not only the interest rate charged for the main credit line, but also the rates for the other lines that a firm opens in its first year. We run the following OLS regression:

$$P_{ikjmt} = \underbrace{\widehat{\alpha} + \widehat{\lambda}_{jmt}}_{\widetilde{P}_{jmt}} + \underbrace{\widehat{\omega}_k}_{\widetilde{\Delta P}_k} + \varepsilon_{ikjmt}, \tag{26}$$

where P_{ikjmt} is the interest rate that bank j charges to firm i of type k in market m at time t, λ_{jmt} is a bank-market-time interactive fixed effect and ω_k are interactive type dummies. We define an observable type based on the amount granted, the sector, the size in terms of sales, and the observable riskiness (SCORE). The underlying assumption here is that the effect on prices of the observable type's characteristics is additively separable with respect to the bank-market effect. Table 5 summarizes all the categories that define a type, as well as some interest rate descriptives for each value of these categories. This regression allows us to recover the bank-market-time specific average price $\widetilde{P}_{jmt}$, as well as the type specific deviation from this average $\widetilde{\Delta P}_k$. Given the large number of variables estimated, we don't report the results of the regression.[45] However, given that in the estimation we will use only the bank-market-time specific average prices, we present some descriptive statistics comparing the predicted prices $\widetilde{P}_{jmt}$ with the actual prices P_{jmt} in the data in Table 6, as well as two overlapping kernel densities in Figure 9, to show the goodness of fit of the model.

We are mainly interested in prices at the bank-market level because we use this variable only in the second stage of our estimation. We provide the intuition of our approach with a simplified version of the demand model. Let the utility of firm i of type k in market m at time t to choose bank j be:

$$U_{ikjmt} = \delta_{jmt} + \gamma_k + \varepsilon_{ikjmt}, \tag{27}$$

in the first stage of our estimation we recover $\widehat{\delta}_{jmt}$ and $\widehat{\gamma}_k$, that are bank-market fixed effects and types' fixed effects, which in the second stage are used to derive the price coefficients. We are just interested in the price coefficient at the bank-market-time level, and the price variation at the type level will be captured by γ_k, so we will only run the following second stage:

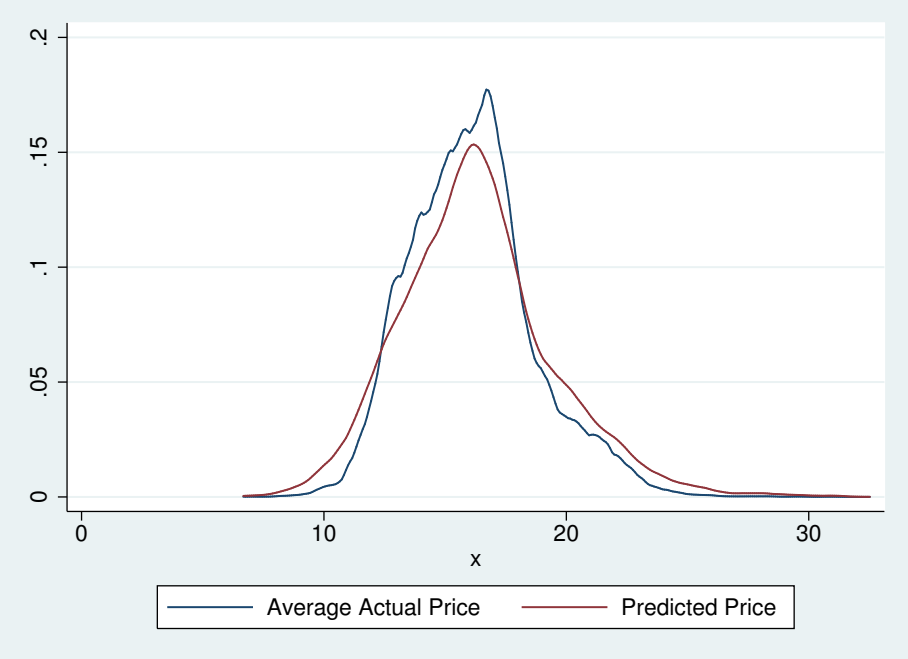
$$\widehat{\delta}_{jmt} = \alpha_0 + \alpha_1 \widetilde{P}_{jmt} + \beta X_{jmt} + \xi_{jmt}. \tag{28}$$

Similarly to the price, we also need to predict what the amount granted would be for firms that don't borrow. We do so using regression analysis from the borrowing firms. This is simplified by the fact that the distribution of amounts granted among the borrowing firms shows evident mass points corresponding to round numbers (mostly between 50 and 500 thousands euros), which are strongly correlated with several firm characteristics (for example, bigger firms get a greater amount). Given that we just need one amount granted for each non-borrowing firm, we calculate the median amount granted to each firm in our data, and group the resulting amounts in the 5 categories listed in Table 5. We regress them against several firm level

[45] We find that the bank-market-time fixed effects as well as the types' fixed effects are jointly significant, and the R^2 is 0.5092.

controls[46] and a province-year-sector interactive fixed effect. The model predictions compared to the actual amounts are shown in Table 7. The model performs relatively well for the most demanded amounts (between 50 and 500 thousand euros), but performs poorly for the least demanded ones (below 50 and above 500 thousands).

Figure 9: Kernel Densities Comparing Actual and Predicted Prices



[46] Tangible and intangible assets, total assets, net assets, short term debt, sales, profits, cashflow, SCORE, long term and short term total bank debt, returns on assets, age of the firm.

Table 5: Types' Summary Statistics

Category				Interest Rate		
		Obs	Percent	Mean	Median	Std. Dev.
Amount Granted	0 - 50,000	12,135	12.16	16.75	16.34	4.70
	50,001 - 100,000	17,014	17.05	15.60	15.19	4.24
	100,001 - 200,000	22,823	22.87	14.65	14.26	3.92
	200,001 - 500,000	27,440	27.50	13.81	13.39	3.67
	500,001 - 3,000,000	20,363	20.41	12.45	12.15	3.35
Sector	Primary	17,076	17.11	14.30	13.70	4.29
	Secondary	42,775	42.87	14.53	14.01	4.17
	Tertiary	39,924	40.01	14.27	13.73	4.01
Size	Small	40,223	40.31	15.45	14.95	3.93
	Large	59,552	59.69	13.67	13.07	4.11
SCORE	Low Risk	25,684	25.74	13.95	13.32	4.28
	Medium Risk	33,659	33.73	14.20	13.72	4.09
	High Risk	40,432	40.52	14.82	14.3	4.02
Total		99,775	100.00	14.39	13.82	4.13

Note: Interest rates is winsorized for the top and bottom 1% of its distribution. We exclude loans above 3,000,000 euros, which represent 2.5% of the loans in our sample. Primary sector includes primary, minerals' extraction, chemicals, metals, energy. Secondary sector includes food and beverages, textile and clothing, wood, paper and publishing, mechanical and electronic machines, production of transport vehicles, other manufacturing, and constructions. Tertiary sector includes commerce of transport vehicles, other commerce, hotels and restaurants, transport, storing and communications, real estate, financial intermediaries, and public administration. Size is defined as firms above or below the median of the distribution of yearly sales, which is around 10 million euros. Low risk is for SCORE values between 1 and 4, medium risk between 5 and 6, and high risk 7 to 9.

Table 6: Descriptives Comparing Actual and Predicted Prices

Statistics	**Actual Price**	**Predicted Price**
Mean	16.594	16.592
Standard Deviation	2.579	2.579
10^{th} Percentile	13.400	13.399
50^{th} Percentile	16.500	16.500
90^{th} Percentile	19.954	19.954
Correlation Coefficient		1.000***
P-Value		0.000

Note: These are average prices at the year-province-bank level for the types' default categories: 0-50,000 euros, primary sector, small size, low risk.

Table 7: Percentage of Predictions and Actual Amounts Granted in Thousands of Euros

			Actual			
Predicted	0 - 50	51 - 100	101 - 200	201 - 500	501 - 3,000	**Total**
0 - 50	2.6	1.7	0.4	0.1	0.0	0.72
51 - 100	23.7	18.6	10.9	4.5	1.2	10.19
101 - 200	63.3	65.8	66.6	59.6	40.0	59.90
201 - 500	10.4	13.9	22.1	35.8	58.2	29.08
501 - 3,000	0.0	0.0	0.0	0.0	0.6	0.11
Total	5,667	11,222	15,580	17,301	9,268	59,038

Note: Each column sums up to 100%. The last column on the right represents the predicted total number of observations for each mass point, whereas the last row represents the actual total number of observations for each mass point.

B Appendix B - IV First Stage and OLS vs IV Second Stage

Table 8: IV First Stage and OLS vs IV Second Stage for Demand

	First Stage	**Second Stage**	
Variable	**Interest Rate**	**OLS**	**IV**
Share of Branches of Merging Rivals	0.215***	-	-
	(0.021)		
Interest Rate	-	4.506***	-3.669***
		(0.303)	(0.348)
Number of Branches	-0.014	-2.766***	-2.746***
	(0.016)	(0.233)	(0.269)
Share of Branches	0.200***	11.497***	12.646***
	(0.043)	(0.627)	(0.721)
Years in Market	-0.030***	-0.799***	-1.001***
	(0.010)	(0.150)	(0.172)
Constant	0.801***	-2.111***	4.580***
	(0.012)	(0.306)	(0.352)
Bank FE	Yes	Yes	Yes
Obs	2,279	2,279	2,279
R^2	0.2028	0.2943	-
F-Stat	97.606	-	-

Note: Standard errors in brackets. $*$ is significant at the 10% level, $**$ at the 5% level, and $***$ at the 1% level.

Table 9: IV First Stage and OLS vs IV Second Stage for Loan Size

	First Stage	**Second Stage**	
Variable	**Interest Rate**	**OLS**	**IV**
Share of Branches of Merging Rivals	0.215***	-	-
	(0.021)		
Interest Rate	-	0.030	-0.295***
		(0.097)	(0.097)
Number of Branches	-0.014	-0.103	-0.102
	(0.016)	(0.075)	(0.075)
Share of Branches	0.200***	0.383*	0.429**
	(0.043)	(0.200)	(0.201)
Years in Market	-0.030***	-0.024	-0.032
	(0.010)	(0.048)	(0.048)
Constant	0.801***	0.070	0.337***
	(0.012)	(0.098)	(0.098)
Bank FE	Yes	Yes	Yes
Obs	2,279	2,279	2,279
R^2	0.2028	0.0425	-
F-Stat	97.606	-	-

Note: Standard errors in brackets. $*$ is significant at the 10% level, $**$ at the 5% level, and $***$ at the 1% level.

Table 10: IV First Stage and OLS vs IV Second Stage for Default

	First Stage	**Second Stage**	
Variable	**Interest Rate**	**OLS**	**IV**
Share of Branches of Merging Rivals	0.215***	-	-
	(0.021)		
Interest Rate	-	-0.283	2.387***
		(0.384)	(0.389)
Number of Branches	-0.014	-0.230	-0.236
	(0.016)	(0.297)	(0.300)
Share of Branches	0.200***	0.473	0.098
	(0.043)	(0.796)	(0.805)
Years in Market	-0.030***	-0.190	-0.124
	(0.010)	(0.191)	(0.193)
Constant	0.801***	-0.727*	-2.913***
	(0.012)	(0.389)	(0.393)
Bank FE	Yes	Yes	Yes
Obs	2,279	2,279	2,279
R^2	0.2028	0.1740	-
F-Stat	97.606	-	-

Note: Standard errors in brackets. $*$ is significant at the 10% level, $**$ at the 5% level, and $***$ at the 1% level.