THE UNIVERSITY OF WARWICK

Sep 2013 No.167

Asymmetric Information and Imperfect Competition in the Loan Market

Gregory S.Crawford, Nicola Pavanini
and Fabiano Schivardi

WORKING PAPER SERIES

Centre for Competitive Advantage in the Global Economy

Department of Economics

CAGE
Competitive Advantage in the Global Economy

E·S·R·C
ECONOMIC & SOCIAL RESEARCH COUNCIL

Asymmetric Information and Imperfect Competition in the Loan Market*

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

September 2013

Abstract

We measure the consequences of asymmetric information in the Italian market for small business lines of credit. Exploiting detailed, proprietary data on a random sample of Italian firms, the population of medium and large Italian banks, individual lines of credit between them, and subsequent individual defaults, we estimate models of demand for credit, loan pricing, loan use, and firm default based on the seminal work of Stiglitz and Weiss (1981) to measure the extent and consequences of asymmetric information in this market. While our data include a measure of observable credit risk comparable to that 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. Data on default, loan size, demand, and pricing separately identify the distribution of private riskiness from heterogeneous firm disutility from paying interest. Preliminary results suggest evidence of asymmetric information, separately identifying adverse selection and moral hazard. We use our results to quantify the impact of asymmetric information on pricing and welfare, and the role imperfect competition plays in mediating these effects.

*We thank Daniel Ackerberg, Pierre-André Chiappori, Lorenzo Ciari, Valentino Dardanoni, Ramiro de Elejalde, Liran Einav, Moshe Kim, Rocco Macchiavello, Carlos Noton, Ariel Pakes, Andrea Pozzi, Pasquale Schiraldi, Matt Shum, Michael Waterson, Chris Woodruff and seminar participants at Warwick, PEDL, Barcelona GSE Banking Summer School 2012 and EARIE 2012 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 Zurich, CEPR and CAGE, gregory.crawford@econ.uzh.ch

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

§LUISS, EIEF and CEPR, fschivardi@luiss.it

1 Introduction

The presence and consequences of asymmetric information in lending markets and the development of policies aimed at reducing these inefficiencies are of crucial importance for credit allocation and financial development.[1] A primary source of asymmetric information arises from banks' uncertainty about borrowers' creditworthiness. This can generate two types of barriers to efficient credit allocation in the loan market: adverse selection in the likelihood of repayment and moral hazard in the riskiness of firms' business decisions, also affecting repayment. While widely recognised to be a central feature of insurance and loan markets, measures of the costs of adverse selection and moral hazard are rare.

In this paper, we measure the consequences of asymmetric information and imperfect competition in the Italian market for small business lines of credit. To do so, we exploit detailed, proprietary data on a random sample of Italian firms, the population of medium and large Italian banks, individual lines of credit between them, and subsequent individual defaults previously analysed in Panetta, Schivardi and Shum (2009). While our data include a measure of observable credit risk comparable to that 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. To measure the distribution of asymmetric firm riskiness, we estimate models of credit, loan size, default, and bank pricing. Data on default, loan use, demand, and pricing separately identify the distribution of private riskiness from heterogeneous firm disutility from paying interest. Preliminary results suggest evidence of asymmetric information, separately identifying adverse selection and moral hazard. We then use our results to simulate counterfactual outcomes varying both asymmetric information and competition in local banking markets. We do this to measure the consequences of adverse selection and moral hazard, and to investigate how competition can mitigate or exacerbate these effects.

There are two types of economic agents in our empirical model: firms and banks. Following Stiglitz and Weiss (1981), we assume 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. 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. We assume banks are differentiated providers of credit that compete by setting prices (interest rates). In addition to the tradeoff of higher infra-marginal profits against lost marginal customers associated with standard pricing decisions, higher rates for any bank also worsen the risk composition of its accepted loans. This increases its aggregate default rates, lowering its profitability and mitigating against rate increases. Handel (2011), Lustig (2011), and Starc (2012) find similar effects of adverse selection and imperfect competition in US health insurance markets.

We estimate the model on highly confidential microdata from the Bank of Italy covering individual loans

[1] There is a vast literature on credit market failures due to information frictions. Among others Banerjee and Newman (1993), Bernanke and Gertler (1990), DeMeza and Webb (1987), Gale (1990), Hubbard (1998), Mankiw (1986), Mookherjee and Ray (2002), Stiglitz and Weiss (1981).

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 Italian banks above a certain size, 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 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.[2] In the final analysis, we estimate individual firms' demand for credit, banks' pricing of these lines, firm's loan use and subsequent default following Einav, Jenkins and Levin (2011) 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.

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. (2011)) approaches. Our novelty is to introduce 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're 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 et al. (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.

We also provide reduced form evidence to motivate our structural model. Following Chiappori and Salanié (2000)'s positive correlation test, we find evidence of adverse selection and moral hazard for new borrowers. Preliminary structural results show, instead, that the Italian market for small business credit lines is affected by adverse selection, but not by moral hazard. We experiment with various subsets of our sample. We provide some intuitive Monte Carlo simulations to understand the counterfactual policy experiments we're working on.

The structure of the paper is the following. In Section 2 we describe the dataset and the market, in Section 3 we present the reduced form tests of adverse selection and moral hazard, Section 4 outlines the structural

[2] A similar approach is followed, among others, by Chiappori and Salanié (2000). We model the dynamics of firm-bank relationships in a companion paper Pavanini and Schivardi (2013).

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 have access to a unique dataset of small business credit lines, previously used in Panetta et al. (2009). We use 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.

The Central Credit Register (hereafter CR) is a database that contains detailed information on all 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 [3], with a breakdown by type of the loan (credit lines, financial and commercial paper, collateralized loans, medium and long-term loans and personal guarantees). 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. Summary statistics for these banks are reported in Panel A of Table 1.

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 (for example, they are not collateralized), so that they can be meaningfully compared across banks and firms. Third, 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 credit line or the cost of mailing the loan statement).

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 a consortium of banks interested in pooling information about their customers. A firm is included in the CB sample if it borrows from at least one of the banks in the consortium. 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. Table 1, Panel B reports descriptive statistics for the sample of borrowing and non-borrowing firms. These two groups of firms appear to be fairly similar in terms of size, leverage and riskiness, but as expected borrowing firms have a higher share of short term debt compared to non-borrowing ones. 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.

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

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).

Table 1: Summary statistics: Banks and Firms

Variable	Obs.	Mean	Stand. Dev.	5^{th} pctile	Median	95^{th} pctile
			Panel A: The Bank Sample			
Total Assets	900	10,726.8	16,965.6	481.3	3,709	54,354.1
Employees	896	3,179.9	4,582.5	206	1,137	14,038
Bad Loans	893	6.2	6.3	1.9	4.9	15.8
Costs-Income ratio	893	34.5	6.1	25.4	33.1	43.2
			Panel B.1: The Borrowing Firm Sample			
Total Assets	302,747	8.1	12.2	0.9	4.2	29.1
Employees	272,816	54.4	75.6	3	30	195
Leverage	305,151	0.57	0.28	0	0.62	0.95
Return on Sales	301,821	1.1	7.5	-9.7	1.2	11.1
Short Term Debt	305,752	33	22.9	0	32	70.7
SCORE	307,532	5.2	1.8	2	5	8
No. of Lenders	329,623	4.4	3.3	1	4	11
Utilized Credit	319,792	50.2	54.3	0	38.2	138.4
			Panel B.2: The Non-Borrowing Firm Sample			
Total Assets	209,754	8.8	20.1	0	2.8	38.8
Employees	176,248	60.6	124.4	0	20	269
Leverage	208,441	0.49	0.36	0	0.52	1
Return on Sales	197,624	2.2	19.6	-17.1	1.2	22.4
Short Term Debt	195,663	23.9	25.7	0	15.7	73.9
SCORE	206,378	4.8	2.1	1	5	8

Note: An observation is the number of bank-years with non-missing records in Panel A, and firm-years in Panel B. Total assets are in millions of euros. Employees is the number of employees at the end of the year. Bad loans is a percentage of total loans. Cost-income ratio is the ratio of overhead to gross income (in %). Return on sales is calculated as the percentage ratio of current profits over total sales. Short term debt is expressed as a proportion of total debt. The SCORE is the indicator of the risk of the company computed each year by the Centrale dei Bilanci (higher values indicate riskier companies). Number of lenders is the number of banks from which the company borrows. Utilized credit is expressed as a proportion of credit granted. The first five variables in each of the firm's sample are winsorized at the 1st and 99th percentile.

2.1 Main new credit lines

The focus of this paper is on a subsample of the available dataset, namely on the first year of the main new credit line that each firm ever opens (within our sample). Considering only the first year is a common assumption in empirical models of insurance with asymmetric information, starting from Chiappori and Salanié (2000). This is done to avoid modeling heterogenous experience ratings among borrowers and loan renegotiation, 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 70% of the total share of new yearly credit (both usable and used), even if in Italy multiple relationship banking is widely used by firms to reduce liquidity risk (Detragiache and Guiso (2000)).

This means that we restrict our attention only on the first year we observe a firm in our data[4]. We also just analyze the main line that a firm uses. This restrict the sample size from around 1,200,000 to around 23,600 observations. The main features of this subsample are presented in Table 2. If on average firms have around 4.5 credit lines active every year, they end up borrowing only from 3.5. Firms open and close almost one line per year. The mean length of relationships that start and terminate within our sample is around 2.5 years.[5] The share of credit used from the main line is around 70%, and it goes up to 80% 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.

Table 2: Summary statistics on credit lines

Variable	Mean	Stand. Dev.	5^{th} pctile	Median	95^{th} pctile
Number of Lines	4.54	3.16	1	4	11
Number of Lines Used	3.48	2.71	1	3	9
Lines Opened	0.99	1.59	0	0	4
Lines Closed	0.92	1.59	0	0	4
Line Length	2.55	1.71	1	2	6
Share of Main Line	0.68	0.27	0.26	0.67	1
Share of Main New Line	0.79	0.24	0.34	0.93	1

Note: An observation is the number of firm-years with non-missing records for all variables apart from line length, whose observation is a firm-bank relationship that begins and finishes within our sample.

[4]To avoid left censoring issues we drop the first year of our sample (1988) and just look at new relationships starting from 1989.
[5]We avoid right censoring not considering relationships still active during the last year of our sample (1998).

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 present separate tests for adverse selection and moral hazard. In the first test, we look at the correlation between the probability of taking a loan and the probability of defaulting. In the second test, we look for a positive correlation between the unobservables driving the choice of how much credit to use and the unobservables influencing default. The choice of these tests gives a flavor of the identification strategy for adverse selection and moral hazard 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, to check if asymmetric information is more severe for the newly started relationships that we're focusing on.

3.1 Adverse Selection

We investigate whether firms that are more likely to demand credit are also more likely to default. Given that the CB dataset includes both firms borrowing and not borrowing, we predict the probability of borrowing based on all the observables that we have: year, province, sector fixed effects, as well as firms' balance sheet information. However, as we don't observe default for non-borrowing firms, we just take the firms that borrow, and use this predicted probability of taking a loan as a regressor for the probability of default. We run the following probit:

$$z_i = \mathbf{1}(X_i\gamma + \eta_i > 0) \tag{1}$$

where z_i is equal to one if the borrower is a defaulter, and X_i is a vector of controls including year, province, sector, and bank fixed effects, as well as other firm's balance sheet variables, amount granted and interest rate. We define as defaulter a firm that defaults at any point in time within our sample. We find a positive and significant correlation between demand probability and default only for the first loan ever. Marginal effects are reported in Table 3, showing that one unit increase in the probability of taking a loan will increase default probability by 4.6 percentage points, where the average default probability is 6.5%.

3.2 Moral Hazard

Following the literature on positive correlation tests introduced by Chiappori and Salanié (2000), we test for moral hazard specifying a bivariate probit model that identifies the relationship between amount of loan used and default probability. The idea of this approach, typically used in various insurance contexts, is to 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. In our lending context we check if firms that demand larger loans are more likely to default on them. Moral hazard should imply that riskier firms use more credit. We set up the following bivariate probit:

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

where y_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 z_i takes value of one if the borrower defaults on that loan. The vector of controls X_i is composed by year, region, 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 moral hazard. The results of this test are summarised in Table 3, where we also provide some additional evidence. The positive correlation is stronger for the first main credit line, compared to all loans with different maturities, showing that as the firm-bank relationship evolves over time there is some learning that reduces asymmetric information.

Table 3: Positive correlation tests

	First Loan Ever	**Whole Sample**
	Adverse Selection	
Demand Prob	0.046***	0.008
	(0.014)	(0.014)
	Moral Hazard	
Used	0.181***	0.170***
	(0.003)	(0.003)
Used/Granted	0.196***	0.186***
	(0.003)	(0.003)

4 The Model

The framework we construct aims at quantifying the effects of adverse selection and moral hazard on demand 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 of their choice. This determines demand for credit. Conditional on demand, firms decide the amount of credit to use and whether to default or not. Supply of credit results from banks' static Bertrand-Nash competition on interest rates r_j.

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 average 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.[6]

(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.[7] As in the reduced form tests, we define as defaulter a firm that defaults at any point in time within our sample.

(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 and Guiso (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 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 $k = 1, ..., K$, 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 sales (firm size), sector and observable riskiness of the firm.[8]

(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 B_{ij} 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 exclusivity, banks can offer menus of rate-amount to reduce the extend 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,

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

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

[8] The construction of this posted interest rates is described in the appendix.

which is not a feature of our setting, where borrowers can open multiple credit lines with different lenders. Empirical evidence of non-exclusivity is shown in Table 6 in the Appendix, which presents a negative correlation between interest rates and amount granted.[9] Moreover, the assumption of setting the loan size just as j-specific is also justified by the distribution of amounts granted, where we observe a high 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 endogenous amount of loan used.

Assume there are $i = 1, ..., I$ firms of type $k = 1, ..., K$ and $j = 1, ..., J$ banks in $m = 1, ..., M$ markets. We omit the k subscript for simplicity. Let firms have the following utility from credit, which determines their demand:

$$U_{ijm}^D = \alpha_i + \delta_{jm}^D(X_{jm}, P_{jm}, \beta^D) + V_{ijm}^D(Y_{ijm}, \eta^D) + \varepsilon_{ijm}^D. \tag{3}$$

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:

$$U_{ijm}^L = \delta_{jm}^L(X_{jm}, P_{jm}, \beta^L) + V_{ijm}^L(Y_{ijm}, \eta^L) + \varepsilon_{im}^L. \tag{4}$$

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

$$U_{ijm}^F = \delta_{jm}^F(X_{jm}, P_{jm}, \beta^F) + V_{ijm}^F(Y_{ijm}, \eta^F) + \varepsilon_{im}^F. \tag{5}$$

Here X_{jm} are banks' observable attributes, P_{jm} are the posted interest rates mentioned above, and Y_{ijm} are firms' observable characteristics. We assume that ε_{ijm}^D 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 $\alpha_i = \bar{\alpha} + \sigma_D \nu_i$, with $\nu_i \sim N(0, 1)$, to be jointly normally distributed with ε_{im}^L, and ε_{im}^F, such that:

$$\begin{pmatrix} \alpha \\ \varepsilon^L \\ \varepsilon^F \end{pmatrix} \sim N\left(\begin{pmatrix} \bar{\alpha} \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_D^2 & 0 & \rho_{DF} \\ 0 & \sigma_L^2 & \rho_{LF} \\ \rho_{DF} & \rho_{LF} & 1 \end{pmatrix} \right). \tag{6}$$

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 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_i and ε_{im}^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_i" firm. A positive correlation of α_i with ε_{im}^F is evidence of adverse selection.

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

We interpret a positive correlation between the unobservables driving loan size and default (ρ_{LF}) as evidence of moral hazard. 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. We define this as moral hazard because the decision on how much loan to use is an action taken after the borrower and lender have agreed on the contract terms. With this definition of moral hazard we are trying to capture the case in which a risky firm (high ε^F_{im}), before signing the contract, already knows that due to its high ε^L_{im} 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_{im} for the firm that was hit, forcing it to use more of the loan, but not due to moral hazard. [10] This identification strategy allows us to recover adverse selection and moral hazard parameters that are common across banks and markets, not bank or market specific.

On the supply side, we let banks set their interest rates competing à la Bertrand Nash. We assume that bank j's profits are given by the sum of profits made with each subset of borrowers' types k:

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

where Q_{jkm} and F_{jkm} are bank's expectation of demand and default. In particular, Q_{jkm} is given by the model's market shares and the expected loan size, and F_{jkm} is the average default rate for the borrowers of type k that bank j lends to in market m, following Assumption 1. P_{jkm} is the price of the loan $(1 + r_j)$. MC_{jm} are the bank's marginal costs, which we assume to be constant at the bank-market level. The first order conditions of this profit function deliver the following pricing equation:

$$\begin{aligned} P_{jkm} &= MC_{jm} - \tfrac{Q_{jkm}}{Q'_{jkm}} + AIC_{jkm}, \\ with \quad AIC_{jkm} &= \frac{MC_{jm}F_{jkm} + MC_{jm}F'_{jkm}\frac{Q_{jkm}}{Q'_{jkm}} - F'_{jkm}\left(\frac{Q_{jkm}}{Q'_{jkm}}\right)^2}{1 - F_{jkm} - F'_{jkm}\frac{Q_{jkm}}{Q'_{jkm}}}. \end{aligned} \tag{8}$$

Note that the equilibrium price depends on marginal costs and markup $\frac{Q_{jkm}}{Q'_{jkm}}$, as in a standard Bertrand-Nash model with differentiated products, but also on a term defined as the Asymmetric Information Cost (AIC_{jkm}). This term is a function of default probability F_{jkm}, derivative of default with respect to prices F'_{jkm}, bank's markup and marginal costs. Both marginal revenues and the default probability determine the shape of the banks' profit function, driving it in different directions. The effect of an increase in interest rates increases on one hand the marginal revenues from borrowers that don't drop out, but on the other hand increases also costs from defaults, in the presence of adverse selection.

[10] We don't have a clear economic interpretation of the correlation between demand and loan size unobservables, so at the moment we are setting the correlation between them to zero for simplicity. We are planning to estimate it in future versions of the model.

4.1 Monte Carlo

We construct a simple numerical example to give an idea of the model's predictions. We simulate data for a monopolist bank facing heterogeneous borrowers. Just to give proof of concept, we concentrate on adverse selection, setting loan size to 1 and $\rho_{LF} = 0$. We keep this data fixed, varying prices and adverse selection, where $\rho_{DF} = 0$ means no adverse selection and $\rho_{DF} = 1$ is maximum adverse selection. For each value of prices and adverse selection, we compute firms' demand and default, and bank's profits based on our model. We also vary the degree of markup that the monopolist has, to mimic the effects of more or less competition (in this case with the outside option). Let firm i have U_{i1}^D utility from borrowing, U_{i0}^D utility from not borrowing, and U_i^F utility from defaulting:

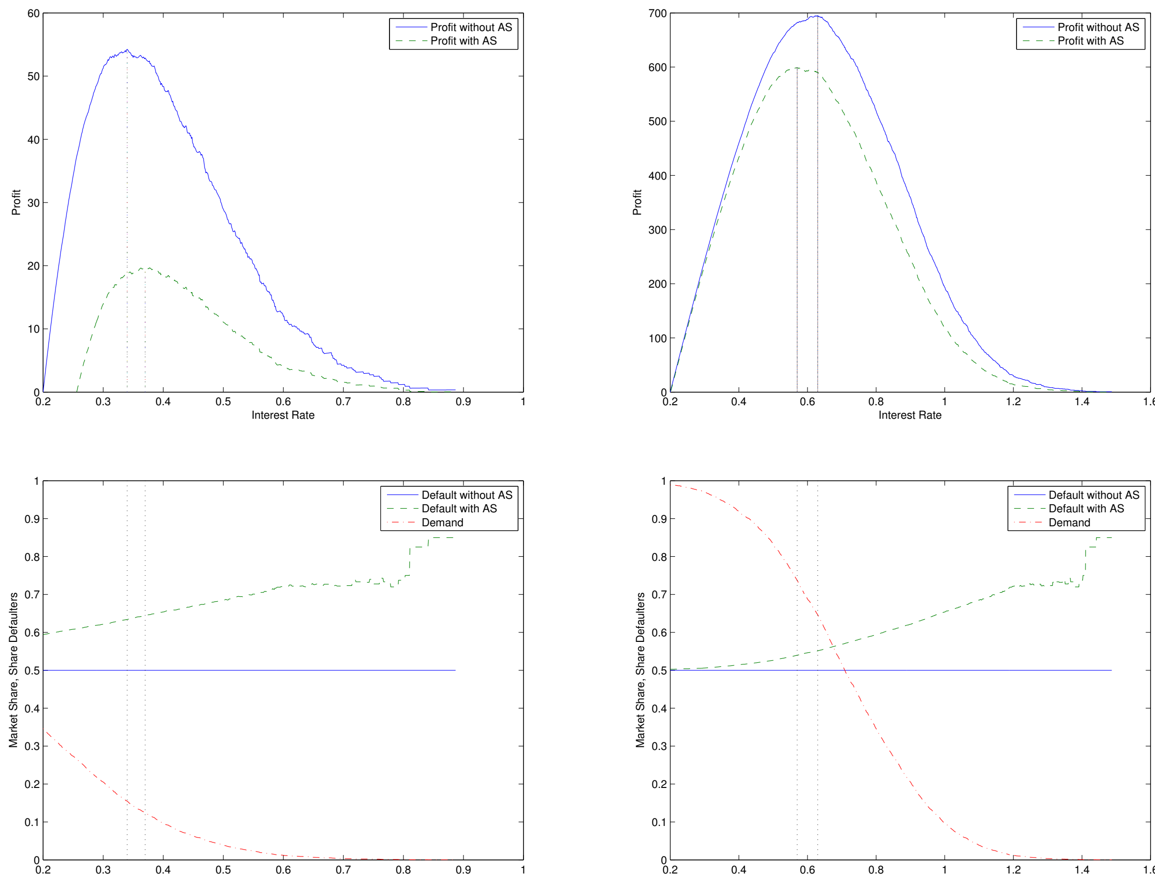
$$
\begin{aligned}
U_{i1}^D &= \bar{\alpha} + \sigma\nu_i - \beta P, \\
U_{i0}^D &= 0, \\
U_i^F &= \varepsilon_i,
\end{aligned} \tag{9}
$$

where P is the interest rate charged. We set $\sigma = 1$ and $\beta = 1$, and allow for a positive or zero correlation between $\alpha_i = \bar{\alpha} + \sigma\nu_i$ and ε_i, jointly normally distributed. We use two different values of $\bar{\alpha} = \{0.3, 0.9\}$, which means shifting borrowers' demand for credit up or down, increasing or decreasing the monopolist's markup. We do this for the two cases of no adverse selection ($\rho_{DF} = 0$), and full adverse selection ($\rho_{DF} = 1$).

The left side of Figure 1 shows the case of low monopolist's markup $\bar{\alpha} = 0.3$, whereas the right side shows the case of higher markup $\bar{\alpha} = 0.9$. The top figures display how the bank's profits without adverse selection (blue solid line) and with adverse selection (green dashed line) vary with prices (on the horizontal axis). The bottom figures show how the average borrowers' default rate without adverse selection (blue solid line) and with adverse selection (green dashed line) vary with prices (on the horizontal axis), as well as the downward sloping demand curve (red dash-dot line). Default rates are the share of borrowing firms that default. In both cases, low and high markup, adverse selection gives rise to a riskier pool of borrowers, as displayed in the bottom figures where the green dashed line is always above the blue solid line.

As emerges from the two top figures, in the low markup case adverse selection increases equilibrium prices, whereas in the high markup cases it reduces them. Once we introduce adverse selection, the monopolist bank can decide to rise or lower prices in response to that (or leave them unchanged). In the first case, increasing prices means increasing default rates and reducing quantities, as shown in the left graphs. Higher defaults lower profits, but the effect of reduced quantities on profits can go either way, depending on demand elasticity. If demand is sufficiently inelastic, as it's the case for the bottom-left figure, then few borrowers will drop out and bank's profits will increase. Moreover, this gain in revenues will compensate the profit loss due to higher defaults. In the second case, reducing prices means reducing default rates and increasing quantities, as shown in the right graphs. Lower defaults foster profits, but again the quantity effect on profits depends on demand elasticity. A very elastic demand, like for the bottom-right figure, implies that a small reduction in prices increases quantities substantially. Therefore, with a sufficiently elastic demand curve, the combination of lower defaults and higher quantities makes it optimal for the bank to lower prices.

Figure 1: Bank's Profits, Prices and Defaults with and without Asymmetric Information, and with low markup ($\bar{\alpha} = 0.3$) and high markup ($\bar{\alpha} = 0.9$).



5 Econometric Specification

5.1 Demand, Loan Size and Default

Following the model presented above, let $m = 1, ..., M$ index the year-geographical market combination, $i = 1, ..., I$ the firm that borrows, and $j = 1, ..., J_m$ be the bank/loan identifier in market m. 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. From now on we will omit the k index for convenience, and assume that the type also defines the market. Let Y_{ijm} be a vector of firm and firm-bank specific characteristics (firm's balance sheet data, firm's distance to closest bank's branch), and X_{jm} a vector of bank specific attributes (number of branches in the market, years of presence in the market, bank 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 individual-level parameters $\eta = \{\eta^D, \eta^L, \eta^F\}$, and the correlation coefficients ρ_{DF} and ρ_{LF} from the firms' choice probabilities. We follow Einav et al. (2011), but differ from them as we estimate demand using a mixed logit with random coefficients, rather than a probit. We also recover the lender-market specific constants $\widehat{\delta}_{jm} = \{\widehat{\delta}^D_{jm}, \widehat{\delta}^L_{jm}, \widehat{\delta}^F_{jm}\}$ using the contraction method introduced by Berry (1994).

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

$$\text{Pr}^D_{ijm} = \int \left[\frac{\exp(\alpha_i + \widehat{\delta}^D_{jm}(X_{jm}, P_{jm}, \beta^D) + V^D_{ijm}(Y_{ijm}, \eta^D))}{1 + \sum_\ell \exp(\alpha_i + \widehat{\delta}^D_{\ell m}(X_{\ell m}, P_{\ell m}, \beta^D) + V^D_{i\ell m}(Y_{ijm}, \eta^D))} \right] f(\alpha_i|\theta) d\alpha_i, \tag{10}$$

where $f(\alpha_i|\theta)$ is the density of α_i, and θ are the parameters of its distribution that we want to estimate. The estimation of this choice model only provides the estimates of η^D and θ, but not of the parameters in δ^D. Looking at the second equation, the amount of credit used conditional on borrowing, the probability of observing a utilization of L_{ijm} is given by:

$$\text{Pr}^L_{ijm} = \phi \left[\frac{L_i - \widehat{\delta}^L_{jm}(X_{jm}, P_{jm}, \beta^L) - V^L_{ijm}(Y_{ijm}, \eta^L)}{\sigma_L} \right], \tag{11}$$

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

$$\text{Pr}^F_{ijm,F=1|D=1,\alpha^D_i,\varepsilon^L_i} = \Phi_{\varepsilon^F_i|\alpha^D_i,\varepsilon^L_i} \Big[- \widehat{\delta}^F_{jm}(X_{jm}, P_{jm}, \beta^F) - V^F_{ijm}(Y_{ijm}, \eta^F) \Big], \tag{12}$$

where the residuals ε^F_{ijm} are conditional on demand and loan amount unobservables[11]. Similarly to the demand side, the estimation of these two choice equations only delivers the η^L and η^F parameters.

[11] The conditional probability in equation 12 is constructed as $\Phi_{\varepsilon^F_i} \left[\frac{-\delta^F_{jm} - V^F_{ijm} - \frac{\rho_{DF}\nu_i}{\sigma^2_D} - \frac{\rho_{LF}\varepsilon^L_i}{\sigma^2_L}}{\sqrt{1 - \frac{\rho^2_{DF}}{\sigma^2_D} - \frac{\rho^2_{LF}}{\sigma^2_L}}} \right]$

In the second step, the estimated constants $\widehat{\delta}_{jm}$ are the dependent variable of instrumental variable regressions that recovers the parameters $\beta = \{\beta^D, \beta^L, \beta^F\}$ of the bank specific attributes X_{jm} and prices P_{jm}. This second step also controls for the potential endogeneity bias caused by the correlation between prices and unobserved (by the econometrician) bank attributes $\xi_{jm} = \{\xi^D_{jm}, \xi^L_{jm}, \xi^F_{jm}\}$. Following Berry (1994), the contraction method on the demand side finds the δ^D that equate predicted market shares $\hat{S}_{jm}$ to actual market shares S_{jm}. This iterative process is defined by:

$$\delta^{D,t+1}_{jm} = \delta^{D,t}_{jm} + \ln\left(\frac{S_{jm}}{\hat{S}_{jm}(\delta^{D,t}_{jm})}\right). \tag{13}$$

The predicted market shares are defined as $\hat{S}_{jm} = \sum_i \mathrm{Pr}^D_{ijm}/N_m$, where N_m are the number of borrowers in market m. Given the value of these constant terms, the β^D parameters are estimated using instrumental variables:

$$\delta^D_{jm} = (X_{jm}, P_{jm})'\beta^D + \xi^D_{jm}, \tag{14}$$

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

$$\delta^L_{jm} = \arg\min_{\delta} \sum_j \left(\hat{s}_{jm}(\eta^L, \delta^L) - s_{jm}\right)^2, \tag{15}$$

$$\delta^F_{jm} = \arg\min_{\delta} \sum_j \left(\hat{s}_{jm}(\eta^F, \delta^F) - s_{jm}\right)^2, \tag{16}$$

where $\hat{s}_{jm}$ and s_{jm} are the predicted and actual shares of loan sizes and defaults for lender j in market m. Given the value of these constant terms, the β^L and β^F parameters are estimated using instrumental variables:

$$\delta^L_{jm} = (X_{jm}, P_{jm})'\beta^L + \xi^L_{jm}, \tag{17}$$

$$\delta^F_{jm} = (X_{jm}, P_{jm})'\beta^F + \xi^F_{jm}, \tag{18}$$

6 Estimation

Following from section 5.1, we use the demand, loan size and default probabilities to construct the simulated maximum likelihood that allows us to recover the parameters in $\eta = \{\eta^D, \eta^L, \eta^F\}$, and the correlation coefficients ρ_{DF} and ρ_{LF}:

$$\log L = \sum_i \log(\mathrm{Pr}^D_{ijm}) D_{ijm} + \sum_{i \in D} \left[\log(\mathrm{Pr}^L_{ijm}) + \log(\mathrm{Pr}^F_{ijm}) F_{ijm} + \log(1 - \mathrm{Pr}^F_{ijm})(1 - F_{ijm}) \right]. \quad (19)$$

In order to estimate the parameters $\Theta = \{\beta^D, \beta^L, \beta^F, \gamma\}$ we need an additional step. Given some instruments Z_{jm}, we recover Θ using instrumental variables.

6.1 Constructing the Sample

We construct a subsample of the data to show preliminary structural results from the estimation of our model. For computational reasons, we reduce the sample along different dimensions. 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 modelling the evolution of the borrower-lender relationship is beyond the scope of this paper.[12] Additionally, we select only two of the main sizes of credit line, in terms of amount granted, which are €50,000 and €200,000. These represent about 43% of the loans. We keep most of the sectors, excluding five small ones that count only for 5% of the observations. We also exclude some of the smallest provinces for lack of observations. Finally, we concentrate on the 1989-1995 period.[13]

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. 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. In fact, even though we focus on a firm's first borrowing decision, we qualify it as a defaulter if it defaults anytime within the following years in our sample. We do this because we identify the default decision as revealing the firm's unobserved type, which we assume doesn't change over time. The definition of an observable firm type k is based on amount granted, score and sector.

The observable explanatory variables that determine firm's demand, loan size and default choices are firm and bank characteristics. In the first set of regressors we include firm's assets, total active, trade debit, sales, cashflow and leverage, where trade debit is the debit that the firm has with its suppliers or clients, and leverage is the ratio of firm's debt over the sum of debt and capital. In the second group we use prices, bank's share of branches in the province, and number of years the bank had at least one branch in the province. 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.

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

[13] A more extensive description of the construction of the sample is in the appendix.

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_{jm} and unobserved (to the econometrician) bank-market level characteristics ξ_{jm}. 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_{jm} as a banks' reputation for offering valuable and helpful assistance to its borrowers in their business projects. Borrowers will value this quality when deciding which bank to get credit from, and they will also be affected in their likelihood of borrowing more or less 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 don't vary by market. This means that the correlation between prices and banks' nationwide-level unobserved characteristics is fully accounted for with these fixed effects, and doesn't require any instruments. Hence, we can rewrite equation 14, and similarly equations 17 and 18, as:

$$\delta_{jm}^{D} = (X_{jm}, P_{jm})'\beta^{D} + \xi_{j}^{D} + \Delta\xi_{jm}^{D}, \tag{20}$$

where ξ_j^D are banks' fixed effects and $\Delta\xi_{jm}^D$ are bank-market specific deviations from the national 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 specific deviations. We follow 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 verify empirically the rank condition for instruments' validity with the first stage estimates[14], showing that the instruments are good predictors of interest rates. We compare OLS and IV second stages, to show how the instruments contribute to attenuate the simultaneity bias.[15] Last, for the exclusion restriction to hold, we assume that bank-market specific deviations $\Delta\xi_{jm}^D$ are uncorrelated across cities. 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. These factors are likely to influence a bank's prices in that local market, but not a bank's prices in other markets, which will be instead correlated with other local managers' qualities. The correlation between a bank's prices across markets will be mostly driven by common costs, like the interbank rate, or deposit interest rates, which doesn't affect our exclusion restriction given that we're controlling for bank fixed effects.

[14]First stage estimates are reported in the Appendix B.

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

6.3 Results

The estimates of the structural model are presented in Table 4. The three columns 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 demand and default (ρ_{DF}) and the correlation between loan size and default (ρ_{LF}). We decide to include those specific firm characteristics as a result of various reduced form regressions for demand, loan size, and default. We wanted to control for different measures of firm size, in the form of tangible and intangible assets[16], but also for some measures of firms' current performance, in terms of sales and cash-flow. We accounted for how leveraged a firm is, but also tried to control for other specific forms of finance that firms have access to, as debit from suppliers.[17] Finally, we computed 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.[18] We also include the share of branches that a bank has in a market, as well as the number of years that it has been in the market for. 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 confirm evidence of asymmetric information, but just in terms of adverse selection. We find no evidence of moral hazard, as opposed to the reduced form test that we presented earlier. One of the reasons for this different outcome is that in the reduced form test the loan size dependent variable was binary (loan size above or below the median of its distribution), whereas now we are using a continuous variable, the share of used over granted. 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 assets are less likely to demand credit, but more leveraged firms are more likely to borrow. 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.

Similarly to the demand equation, firms with more assets are likely to use a smaller share of their credit lines. Also higher cash-flow, as expected, has a negative effect on the size of the loan used. More access to credit from suppliers reduces the need to use the credit line, but more leveraged firms will be more likely to use a greater portion. Interestingly, higher distance from the bank's branch seems to have a positive effect on the amount of loan used. Finally, firms' default decisions are negatively correlated with trade debit, sales and cash-flow, but positively affected by leverage.

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

[17]Petersen and Rajan (1995) use the amount 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.

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

Table 4: Structural results for all sectors

Variable	**Demand**	**Loan Size**	**Default**
	Firm level variables		
Intangible Assets	-0.171***	-1.521***	-0.073
	(0.062)	(0.056)	(0.296)
Total Active	-0.158***	0.453***	0.045**
	(0.005)	(0.004)	(0.019)
Net Asset	-0.196***	-0.204***	0.090
	(0.021)	(0.020)	(0.079)
Trade Debit	-0.003	-0.430***	-0.258***
	(0.005)	(0.010)	(0.060)
Sales	0.012***	0.050***	-0.194***
	(0.003)	(0.003)	(0.019)
Cashflow	-0.055	-1.348***	-1.463***
	(0.015)	(0.048)	(0.470)
Leverage	1.730***	7.809***	-2.162**
	(0.345)	(0.279)	(1.163)
Distance	-0.522***	0.074***	0.072
	(0.024)	(0.019)	(0.111)
	Bank level variables		
Intercept	2.733***	0.575	-2.324***
	(0.464)	(1.080)	(0.642)
Interest Rate	-0.173***	0.029	-0.039
	(0.025)	(0.058)	(0.034)
Share of Branches	1.018*	-0.278	0.338
	(0.592)	(1.380)	(0.820)
Years in Market	-0.030***	-0.032***	0.001
	(0.005)	(0.011)	(0.007)
Bank FE	Yes	Yes	Yes
ρ_{DF}			0.186***
			(0.098)
ρ_{LF}			0.035
			(0.031)
Obs	56,581	2,763	2,763

7 Counterfactuals

We run two counterfactual policy experiments to quantify the welfare effects of asymmetric information, as well as to understand the relationship between asymmetric information and imperfect competition. We construct a standard measure of welfare as the sum of consumer surplus and producer profits, and compare total amount of credit given, prices, default rates, and welfare under different scenarios. In the first policy experiment, we simulate the case of no asymmetric information, setting the correlation coefficients to zero. Given our estimation results, we just focus on shutting down adverse selection, as the moral hazard coefficient was already not statistically different from zero. This allows us to measure the welfare cost of asymmetric information. In the second experiment, we simulate the effects of a credit crunch, making it more costly for banks to finance credit. We do this increasing banks' marginal costs by 10%.

7.1 Policy Experiment 1: Welfare Cost of Asymmetric Information

In this counterfactual exercise we vary the correlation coefficients identifying asymmetric information to calculate what is the cost in terms of welfare of this inefficiency. We also look at the new equilibrium quantities, prices, and default rates. Given that in our results we only find evidence of adverse selection, we will limit our analysis to the case of $\rho_{DF} = 0$. Following the example of Nevo (2000), we recover each bank's marginal costs using the pricing equation:

$$\widehat{MC}_{jkm} = P_{jkm} + \frac{Q_{jkm}}{Q'_{jkm}} - AIC_{jkm} \tag{21}$$

Under the assumption of marginal costs being the same with or without asymmetric information, we recalculate banks' market shares, loan sizes and defaults with $\rho_{DF} = 0$, and derive the new equilibrium prices as:

$$P^*_{jkm} = \widehat{MC}_{jkm} - \frac{Q_{jkm}}{Q'_{jkm}}(P^*_{jkm}) + AIC_{jkm}(P^*_{jkm}) \tag{22}$$

Once we derived the new equilibrium outcomes, we look at the change in consumer welfare calculating the compensating variation for each borrower as:

$$CV_i = \frac{\ln\left[\sum_j^J \exp(IV_{ijkm}^{\rho_{DF}=0})\right] - \ln\left[\sum_j^J \exp(IV_{ijkm}^{\widehat{\rho}_{DF}})\right]}{\widehat{\beta}_p^D} \tag{23}$$

where the denominator is the price coefficient in the demand equation, and the numerator is the difference between inclusive values $IV_{ijkm} = \alpha_i + \delta^D_{jm} + V^D_{ijm}$ without ($\rho_{DF} = 0$) and with ($\widehat{\rho}_{DF}$) asymmetric information. Counterfactual results are presented in Table 5.

7.2 Policy Experiment 2: Credit Crunch

This experiment looks at the effects of an increase of banks' cost of borrowing, in the presence of asymmetric information and imperfect competition, on welfare, quantities, prices, and borrowers' default rates. We simulate a 10% increase in banks' marginal costs to mimic the effects of a credit crunch on retail banking. Counterfactual results are presented in Table 5.

Table 5: Counterfactual Experiments

Variables	**Baseline**		**No Asymmetric Info**		**Credit Crunch**	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Market Shares	21.9%	16.6%	12.8%	21.9%	11.9%	21.2%
Prices	16.01	2.01	26.18	15.73	27.39	15.62
Defaults	3.3%	1.2%	6.2%	23.7%	6.2%	23.7%
Banks' Profits	272.4	1,236.1				
Total Banks' Profits	752,610					
Consumers' Surplus	190,960					
Welfare	943,570					

Note: An observation for market shares, prices, defaults and banks' profits is a bank-type-market-year combination. Profits and surplus are in millions of euros.

8 Conclusion

In this paper we analyzed the interaction between imperfect competition and asymmetric information in the Italian market for small business lines of credit. We have access to a rich dataset with detailed information about credit line contracts between firms and banks, including all the main Italian credit institutions and a highly representative sample of firms. Using this data, we provide reduced form evidence of adverse selection and moral hazard, 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 banks' markups and demand elasticity. We show that banks' market power can mitigate the effect of higher costs from asymmetric information. In our setting, low bank markup and inelastic demand cause interest rates to increase, whereas high bank markup and elastic demand lead to lower rates, with respect to the benchmark case of no asymmetric information.

Our structural results provide evidence of adverse selection for new borrowers, but not of moral hazard. We propose two counterfactual policy experiments. The first aims at quantifying the welfare costs of asymmetric information, comparing welfare with and without it. The second investigates the effects of a credit crunch, in the presence of asymmetric information and imperfect competition, on equilibrium prices, amount lent, borrowers' default rates and welfare.

References

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," *RAND Journal of Economics*, 1994, *25* (2), 242–262.

____, J. Levinsohn, and A. Pakes, "Automobile Prices in Market Equilibrium," *Econometrica*, 1995, *63* (4), 841–890.

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

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

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. and P. Garella 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," 2011. Econometrica forthcoming.

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.

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.

Handel, B., "Adverse Selection and Switching Costs in Health Insurance Markets: When Nudging Hurts," 2011. American Economic Review Revise and Resubmit.

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

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 Lending," 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.

Starc, A., "Insurer Pricing and Consumer Welfare: Evidence from Medigap," 2012. The Wharton School, University of Pennsylvania Working Paper.

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.

A Appendix - 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 [19]. We assume that each firm chooses one main credit line among all the banks available in its province. 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 Constructing the Posted Prices

We only consider the first main loan ever taken by a firm. We assume that banks have a posted price for each type of firm in each year-province combination. We recover this synthetic price using regression analysis based on actual prices. 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 that would have been offered to borrowing firms by banks other than the chosen one. Table 6 shows how we construct these prices. We present some descriptive statistics comparing the posted prices to the actual prices in the data in Table 7, as well as two overlapping kernel densities in Figure 2, to show the goodness of fit of the model. We are working on a non-parametric approach

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

to obtain a better fit.

As resulting from the posted price regression, the amount granted is an important driver of price. Therefore, we need to construct also 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 50, 100 and 200 thousands euros), which are strongly correlated with several firm characteristics (for example, bigger firms get a greater amount). Table 8 shows the distribution of amount granted. We predict the amount that not borrowing firms would demand, given their characteristics, keeping only firms that borrow up to 1 million euros. Differently from the price regression, we also include firms that have already been borrowing in the previous years.

Given the 9 mass points in the distribution of amount granted, we use a multinomial logit regression to determine the probability of the firm's choice of one of these amounts. We condition on firm characteristics, as well as year, region, sector and bank fixed effects. We use the estimated coefficients to predict the amount that not borrowing firms would have demanded. We run the multinomial logit only for the amount of the main line that each firm uses in each year. The model predictions compared to the actual amounts are shown in Table 9. The model performs relatively well for the most demanded amounts, but performs poorly for the least demanded ones. In particular, 25% of the predictions are correct, another 24% is just one mass point above or below the actual one, and another 17% is two mass points above or below the correct one. The predicted amounts granted for each firm are then used to predict the posted prices that these firms would be charged if they were to borrow, using the estimated coefficients from Table 6. We also predict the price that borrowing firms would be offered by rival banks, keeping fixed the amount granted by the actual lending bank.

Figure 2: Kernel Densities Comparing Actual and Predicted Prices

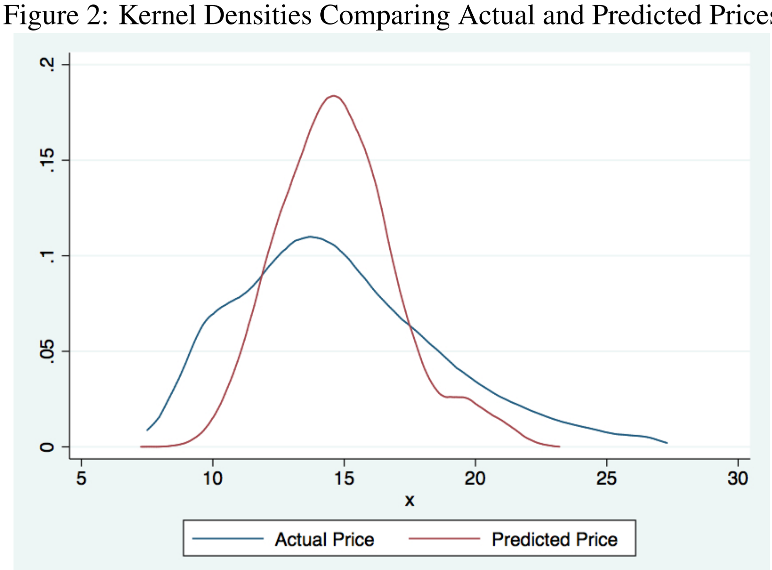


Table 6: Posted Prices for First Loan Ever

Variable	**Interest Rate**
Score	-0.073
	(0.067)
Score^2	0.026***
	(0.006)
Amount	-13.022***
	(0.684)
Amount^2	0.020***
	(0.002)
Amount^3	-0.00001***
	(0.000)
Intercept	16.986***
	(0.177)
Year FE	Yes
Banks FE	Yes
Province FE	Yes
Sector FE	Yes
Obs	23,616
R^2	0.3593

Table 7: Descriptives Comparing Actual and Predicted Prices

Statistics	**Actual Price**	**Predicted Price**
Mean	14.76	14.77
Standard Deviation	3.86	2.30
10th Percentile	10.01	11.91
50th Percentile	14.30	14.63
90th Percentile	20.08	17.69
Regression Coefficient		0.9996***
Standard Error		(0.009)
Correlation Coefficient		0.5969***
P-Value		0.000

Table 8: Distribution of Amounts Granted in Euros

Amount Granted	Observations	%	Cumulate %
50,000	304,394	25.30	25.30
100,000	236,387	19.65	44.94
200,000	217,771	18.10	63.04
300,000	125,266	10.41	73.45
400,000	57,930	4.81	78.27
500,000	81,370	6.76	85.03
600,000	30,203	2.51	87.54
800,000	30,300	2.52	90.06
1,000,000	49,890	4.15	94.21
More than 1,000,000	69,701	5.79	100.00
Total	1,203,212	100.00	-

Table 9: Percentage of Predictions and Actual Amounts Granted in Thousands of Euros

					Actual					
Predicted	50	100	200	300	400	500	600	800	1,000	**Total**
50	**33**	20	18	9	5	4	2	2	2	98,153
100	14	**18**	18	13	7	9	5	5	7	35,673
200	9	13	**20**	15	9	10	6	6	7	103,005
300	1	5	12	**15**	9	18	8	12	16	1,401
400	4	0	11	12	**13**	12	11	15	16	144
500	4	10	12	11	5	**19**	6	8	21	7,888
600	66	0	0	0	0	0	**0**	0	33	3
800	0	20	0	0	0	0	0	**40**	40	5
1,000	3	7	10	9	6	17	6	9	**28**	8,840
Total	48,784	41,901	47,911	32,858	19,101	22,227	12,619	12,481	17,230	255,112

Note: Each horizontal line 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 - First stage of IV regressions and OLS regression

We report the first stage of the IV regressions for demand, loan size and default to verify instruments' validity. An observation is a bank-type-province-year combination. The instrument appears to be a very strong predictor of prices. The F-test for weak instruments is well above the rule of thumb value of 10.

Table 10: First stage of IV regression

Variable	**Interest Rate**
Avg Price in Other Mtks	0.954***
	(0.006)
Share of Branches	0.139
	(0.161)
Years in Market	0.004***
	(0.001)
Intercept	0.988***
	(0.120)
Banks FE	Yes
Obs	2,763
Adj R^2	0.9256

We also report the OLS second stage, to be compared with the IV second stage and compare the price coefficient with and without instruments.

Table 11: OLS vs IV second stage

Variable	**Demand**		**Loan Size**		**Default**	
	OLS	**IV**	**OLS**	**IV**	**OLS**	**IV**
Intercept	2.951***	2.733***	-0.491	0.575	-2.734***	-2.324***
	(0.442)	(0.464)	(1.033)	(1.080)	(0.614)	(0.642)
Interest Rate	-0.187***	-0.173***	0.094	0.029	-0.015	-0.039
	(0.023)	(0.025)	(0.055)	(0.058)	(0.032)	(0.034)
Share of Branches	1.029*	1.018*	-0.284	-0.278	0.311	0.338
	(0.591)	(0.592)	(1.381)	(1.380)	(0.820)	(0.820)
Years in Market	-0.030***	-0.030***	-0.032***	-0.032***	0.002	0.001
	(0.005)	(0.005)	(0.011)	(0.011)	(0.007)	(0.007)
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs	2,763	2,763	2,763	2,763	2,763	2,763