Online appendix for: Asymmetric Information and Imperfect Competition in Lending Markets
Published in the American Economic Review

By Gregory S. Crawford, Nicola Pavanini and Fabiano Schivardi

This appendix contains additional material referred to in the published paper. All references beginning with a number (Arabic or Latin) refer to the paper, while those beginning with a letter refer to the appendix itself.

A. Conditional Means Results

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

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

$$\text{(A1)} \qquad \underset{j \in 1,\ldots,J_{mt}}{\text{Max}} \{\bar{\alpha}_0^D + X_{jmt}'^D \beta^D + \xi_{jmt}^D + \alpha^D P_{ijmt} + Y_{ijmt}'^D \eta^D + \varepsilon_i^D + \nu_{ijmt}\} \geq \nu_{i0mt}.$$

This can be rewritten as:

$$\text{(A2)} \qquad \varepsilon_i^D \geq \underbrace{\nu_{i0mt} - \bar{\alpha}_0^D - \underset{j \in 1,\ldots,J_{mt}}{\text{Max}} \{X_{jmt}'^D \beta^D + \xi_{jmt}^D + \alpha^D P_{ijmt} + Y_{ijmt}'^D \eta^D + \nu_{ijmt}\}}_{T_i},$$

where we can simplify the right-hand side, T_i, into a firm-specific component because each firm i in the sample is active in a single market and single year and, once we take the maximum, j drops out. That is, because unobserved demand for credit is firm- and not bank-specific, the additional Type I Extreme Value errors, ν_{ijmt}, do not complicate the calculation of the relationship between ε_i^D and ε_i^F. As far as adverse selection is concerned, we are interested in whether or not a firm borrows at all, not from which bank it borrows.

Given this structure, the results relied upon in the text are straightforward. Because ε_i^D and ε_i^F are jointly normally distributed with variances σ_D^2 and σ_F^2, and correlation coefficient, ρ_{DF}, then

$$\begin{aligned} E(\varepsilon_i^D|D=1) &= E(\varepsilon_i^D|\varepsilon_i^D \geq T_i) \\ &= \sigma_D \frac{\phi(T_i/\sigma_D)}{1-\Phi(T_i/\sigma_D)} > 0, \end{aligned} \tag{A3}$$

and

$$E(\varepsilon_i^F|D=1) = \rho_{DF} \frac{\phi(T_i/\sigma_D)}{1-\Phi(T_i/\sigma_D)}, \tag{A4}$$

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

B. Numerical Results

B1. Monte Carlo

In this Appendix, we construct a simple Monte Carlo to give intuition for how adverse selection and competition can interact in lending markets. We simulate data for the case of a monopolist bank offering a loan to $i = 1, ..., I$ potential borrowers, who are observationally equivalent to the bank and differ only in their unobserved demand for credit (ε_i^D) and their utility from default, ε_i^F, which we call their "riskiness". For simplicity, we concentrate on the correlation in the unobservables between demand and default (ρ_{DF}), setting borrowers' loan amount and loan use to 1.

In the Monte Carlo, we keep these data fixed and vary two parameters: borrowers' price sensitivity, α, as a proxy for the strength of the effects of a competitive fringe on the bank's residual demand curve,[1] and the extent of asymmetric information, ρ_{DF}, where $\rho_{DF} < 0$ means advantageous selection and $\rho_{DF} > 0$ means adverse selection. For each of these cases, we compute the bank's equilibrium price based on the maximization of its expected profit, as described in Section II.C in the paper.

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

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

$$
\begin{aligned}
U_i^D &= \bar{\alpha}_0 + \alpha P + \varepsilon_i^D + \nu_i, \\
U_{i0}^D &= \nu_{i0}, \\
U_i^F &= \varepsilon_i^F,
\end{aligned}
\tag{B1}
$$

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

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

$$
\begin{aligned}
\Pr_i^D &= \Pr(\bar{\alpha}_0 + \alpha P + \varepsilon_i^D + \nu_i > \nu_{i0}) \\
&= \frac{\exp(\bar{\alpha}_0 + \alpha P + \varepsilon_i^D)}{1 + \exp(\bar{\alpha}_0 + \alpha P + \varepsilon_i^D)} \\
&= \Lambda(\bar{\alpha}_0 + \alpha P + \varepsilon_i^D),
\end{aligned}
\tag{B2}
$$

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

$$
\begin{aligned}
\Pr_{i,F=1|D=1}^F &= E\left[\Pr(\varepsilon_i^F > 0 | \varepsilon_i^D, P) | D = 1, P\right] \\
&= \int \Phi\left(\frac{\frac{\rho_{DF}\varepsilon_i^D}{\sigma^2}}{\sqrt{1 - \frac{\rho_{DF}^2}{\sigma^2}}}\right) f(\varepsilon_i^D | D = 1) d\varepsilon_i^D,
\end{aligned}
\tag{B3}
$$

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

$$P = \underbrace{\frac{MC}{1 - F - F' \frac{1}{\alpha(1-Q)}}}_{\text{Effective Marginal Cost}} + \underbrace{\frac{-(1-F)\frac{1}{\alpha(1-Q)}}{1 - F - F' \frac{1}{\alpha(1-Q)}}}_{\text{Effective Markup}}, \tag{B4}$$

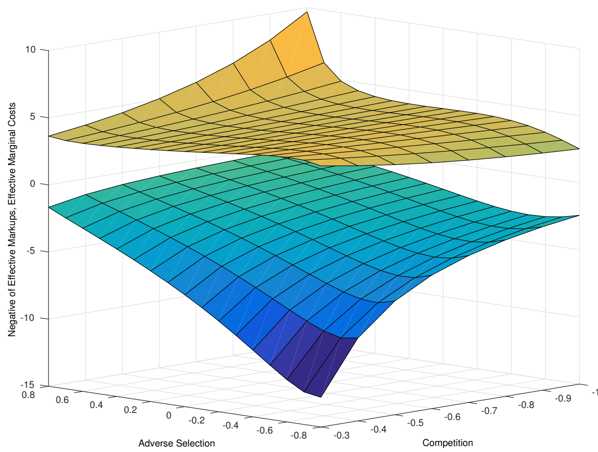
where F' is the derivative of the expected default rate with respect to price, and $\alpha(1-Q)$ is the derivative of the expected market share with respect to price.

The top graph of Figure B1 reports effective marginal costs, the bottom graph of Figure B1 reports the *negative* of effective markups, while Figure B2 combines these to show equilibrium prices. The figures examine how these three elements vary with different degrees of adverse selection, measured by ρ_{DF}, and "competition" with the outside option, measured by the slope parameter of the monopolist's residual demand curve, α.

Looking at the top graph in Figure B1, for a high level of competition (i.e. the rightmost point on the top graph) an increase in adverse selection (moving to the northwest) causes effective marginal costs to increase, whereas for low competition (the point closest to the reader on the top graph, again moving northwest) it remains relatively constant. The opposite happens for effective markups as we increase adverse selection. Looking at the bottom graph of Figure B1, and recalling that it reports the *negative* of effective markups (thus a higher markup is associated with a lower point in the graph), for high levels of competition (the rightmost point on the bottom graph, moving to the northwest), increases in adverse selection decrease effective markups slightly, whereas for low levels of competition, (the closest point to the reader on the bottom graph, again moving northwest), increases in adverse selection decrease effective markups substantially.

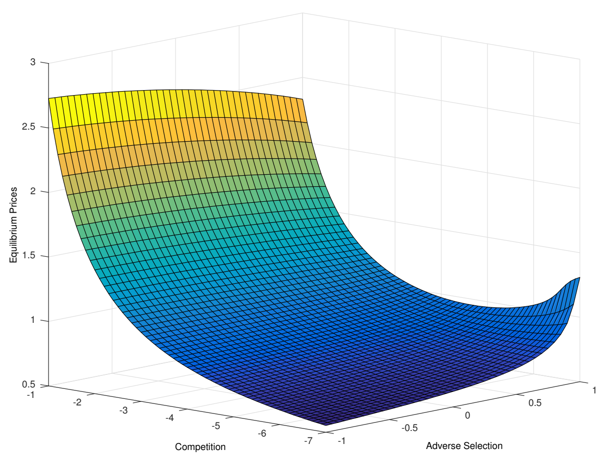
Figure B2 combines both factors and demonstrates a non-monotonic price response to increases in adverse selection, with the sign of the effect depending on the level of competition. While equilibrium prices rise with adverse selection in a competitive environment (the closest point to the reader, moving to the northeast), the opposite happens in a concentrated market (the leftmost point, moving east). In a competitive environment, markups are low and the average borrower effect dominates, so increasing adverse selection causes prices to rise (driven by the rising effective marginal costs shown in Figure B1). In a less competitive environment, by contrast, markups are high, enhancing the marginal borrower effect, and increasing adverse selection drives prices down. In competitive markets, banks have small margins and can only increase prices in response to increased average selection, while in less competitive markets, banks with higher markups find it profitable to reduce prices as it allows them to attract relatively safe borrowers.

Figure B1. Adverse Selection vs Imperfect Competition - Effective Marginal Costs, Negative Effective Markups



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

Figure B2. Adverse Selection vs Imperfect Competition - Equilibrium Prices



Note: The vertical axis shows the level of equilibrium prices. The left horizontal axis measures the slope of the residual demand curve (our measure of competition with the outside option), increasing towards the right. The right horizontal axis is the level of adverse selection, increasing towards right. The horizontal axis definitions and scales in this figure differ from those in Figure B1 to better display the effects in each.

B2. $\partial F'/\partial \rho_{DF}$ Results

For the simplified setting described in the Monte Carlo, equation (B3) above presented the formula for the probability firm i defaults given that it has chosen to borrow, $D = 1$. In what follows, we derive this formula and its derivatives in greater detail for use in the discussion in Section II.C in the body of the text.

To understand where equation (B3) came from, note that given the structure of preferences described in equation (B1) and the joint normality assumption on ε_i^D and ε_i^F, the probability of default, $F = 1$, conditioned on a specific ε_i^D is $\Phi\left(\frac{\frac{\rho_{DF}\varepsilon_i^D}{\sigma^2}}{\sqrt{1-\frac{\rho_{DF}^2}{\sigma^2}}}\right)$. When we calculate the conditional probability of default, we therefore need to take into account that borrowers have selected into the decision to borrow.

In a standard Heckman setting, the selection equation is deterministic: a firm borrows if $\varepsilon_i^D \geq \bar{\alpha}_0 - \alpha P$. The problem in our case is different because our selection equation gives only a probability of borrowing conditional on ε_i^D rather than a deterministic threshold. From equation (B2), we know that:

$$\text{(B5)} \qquad \Pr{}_{i|\varepsilon_i^D}^D = \Lambda(\overline{\alpha}_0 + \alpha P + \varepsilon_i^D).$$

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

$$\text{(B6)} \qquad \begin{aligned} f(\varepsilon_i^D|D=1) &= \frac{\Pr_{i|\varepsilon_i^D}^D f(\varepsilon_i^D)}{\Pr_i^D} \\ &= \frac{\Lambda(\overline{\alpha}_0 + \alpha P + \varepsilon_i^D) f(\varepsilon_i^D)}{\int \Lambda(\overline{\alpha}_0 + \alpha P + \varepsilon_i^D) f(\varepsilon_i^D) d\varepsilon_i^D}. \end{aligned}$$

Then, the probability of default conditional on borrowing is just the expected value of $\Phi\left(\frac{\frac{\rho_{DF}\varepsilon_i^D}{\sigma^2}}{\sqrt{1-\frac{\rho_{DF}^2}{\sigma^2}}}\right)$ under $f(\varepsilon_i^D|D=1)$. As in equation (B3) above,

$$\text{(B7)} \qquad \Pr{}_{i,F=1|D=1}^F = \int \Phi\left(\frac{\frac{\rho_{DF}\varepsilon_i^D}{\sigma^2}}{\sqrt{1-\frac{\rho_{DF}^2}{\sigma^2}}}\right) f(\varepsilon_i^D|D=1) d\varepsilon_i^D.$$

Let $\Lambda(\overline{\alpha}_0 + \alpha P + \varepsilon_i^D) = \Lambda(.)$ to simplify the notation. Substituting equation (B6) into equation (B7) we get:

$$\text{(B8)} \qquad \Pr^F_{i,F=1|D=1} = \underbrace{\frac{1}{\int \Lambda(.) f(\varepsilon_i^D) d\varepsilon_i^D}}_{\mathcal{A}} \underbrace{\int \Phi\left(\frac{\frac{\rho_{DF}\varepsilon_i^D}{\sigma^2}}{\sqrt{1-\frac{\rho_{DF}^2}{\sigma^2}}}\right) \Lambda(.) f(\varepsilon_i^D) d\varepsilon_i^D}_{\mathcal{B}}.$$

The derivative of default conditional on borrowing with respect to P, that is F'_{ijmt} in Section II.C, is given by:

$$\text{(B9)} \qquad \frac{\partial \Pr^F_{i,F=1|D=1}}{\partial P} = \frac{\partial \mathcal{A}}{\partial P}\mathcal{B} + \mathcal{A}\frac{\partial \mathcal{B}}{\partial P},$$

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

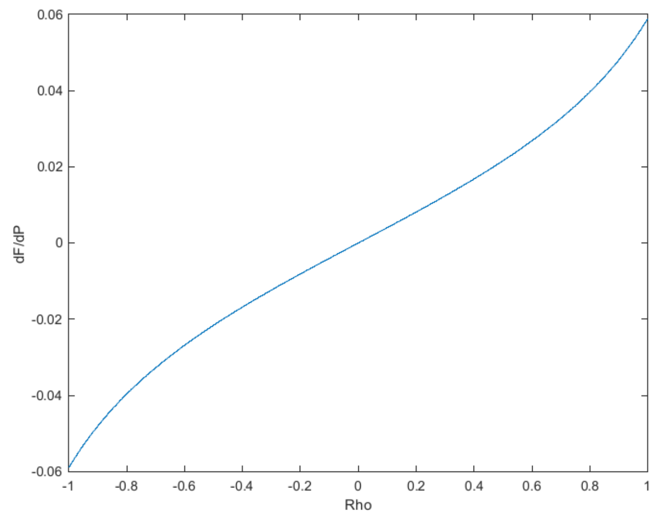
$$\text{(B10)} \qquad \begin{aligned} \frac{\partial \mathcal{A}}{\partial P} &= -\alpha \frac{\int \Lambda(.)(1-\Lambda(.)) f(\varepsilon_i^D) d\varepsilon_i^D}{\left(\int \Lambda(.) f(\varepsilon_i^D) d\varepsilon_i^D\right)^2}, \\ \frac{\partial \mathcal{B}}{\partial P} &= \int \Phi\left(\frac{\frac{\rho_{DF}\varepsilon_i^D}{\sigma^2}}{\sqrt{1-\frac{\rho_{DF}^2}{\sigma^2}}}\right) \alpha\Lambda(.)\left(1-\Lambda(.)\right) f(\varepsilon_i^D) d\varepsilon_i^D. \end{aligned}$$

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

$$\text{(B11)} \qquad \frac{\partial^2 \Pr^F_{i,F=1|D=1}}{\partial P \partial \rho_{DF}} = \frac{\partial \mathcal{A}}{\partial P}\frac{\partial \mathcal{B}}{\partial \rho_{DF}} + \mathcal{A}\frac{\partial^2 \mathcal{B}}{\partial P \partial \rho_{DF}},$$

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

Figure B3. $\partial \mathrm{Pr}^F_{i,F=1|D=1}/\partial P$



$$\text{(B12)}\quad \begin{aligned} \frac{\partial \mathcal{B}}{\partial \rho_{DF}} &= \int \phi\left(\frac{\frac{\rho_{DF}\varepsilon_i^D}{\sigma^2}}{\sqrt{1-\frac{\rho_{DF}^2}{\sigma^2}}}\right)\frac{\frac{\varepsilon_i^D}{\sigma^2}}{\left(1-\frac{\rho_{DF}^2}{\sigma^2}\right)^{\frac{3}{2}}}\Lambda(.)f(\varepsilon_i^D)d\varepsilon_i^D, \\ \frac{\partial^2 \mathcal{B}}{\partial P \partial \rho_{DF}} &= \int \phi\left(\frac{\frac{\rho_{DF}\varepsilon_i^D}{\sigma^2}}{\sqrt{1-\frac{\rho_{DF}^2}{\sigma^2}}}\right)\frac{\frac{\varepsilon_i^D}{\sigma^2}}{\left(1-\frac{\rho_{DF}^2}{\sigma^2}\right)^{\frac{3}{2}}}\alpha\Lambda(.)\left(1-\Lambda(.)\right)f(\varepsilon_i^D)d\varepsilon_i^D, \end{aligned}$$

where ϕ is the PDF of a standard normal distribution. Even in this simple setting, it is difficult to sign these derivatives analytically. Instead, we simulate them for $\rho \in [-1,1]$, $\overline{\alpha}_0 = 0$, $\alpha = -1$, and $\sigma = 1$ to obtain Figure B3. As can be seen there, the derivative of the default probability with respect to price is the same sign as ρ_{DF} and the slope of this line, measuring $\partial^2 \mathrm{Pr}^F_{i,F=1|D=1}/\partial P \partial \rho_{DF}$, is everywhere increasing when $\rho_{DF} > 0$.

C. Constructing the Dataset

We have assembled data from the following sources:

- **Firm Data:** Dataset from *Centrale dei Bilanci* with yearly (1988-1998) balance sheet data for each firm, including data both for firms that take credit and those that do not. This data also includes the year of birth for each firm, its location at the city council level, and what we call in Section I.B each firm's "Score." The Score represents our measure of a firm's observable default risk.

- **Loan Data:** Dataset from *Centrale dei Rischi* with yearly (1988-1998) firm-bank loan contracts, including the amount of granted credit, the amount of this used by the firm, the loan's interest rate, and whether or not the firm has defaulted on this loan (Section I.C in the paper describes our definition of default). As discussed in Section I.A, this data is only available for 94 large banks (representing more than 80% of total bank lending) and for short term lines of credit.
- **Bank Data:** Dataset with yearly (1988-2002) balance sheet data for each bank, including yearly total loans that each bank gives in each province, and its share of the total loans granted in each province.
- **Branch Data:** Dataset with yearly (1959-2005) branches for the population ($\sim$ 1,500) of banks at the city council level.
- **Coordinates Data:** Based on the National Institute for Statistics *ISTAT* city council classification, we assign to each city council the geographic coordinates that will allow us to calculate firm-branch distances.

We first merge the firm data with the loan data, in order to combine all the borrowing and non-borrowing firms. We then take all the banks actively lending in each province and assume that those represent the choice set for each firm, regardless of whether they have a branch in that province or not.[2] We assume that each firm chooses one main credit line from among those offered by the banks active in its province or chooses not to take any line of credit (the outside good). The main line is defined as the line for which the amount used, regardless of the amount granted, is the highest. For cases in which multiple lines have the same amount used, then the one with the lowest price is defined as the main line. We calculate the distance in kilometers between the city council of each firm and the closest city council where each bank from the choice set has a branch using the geographic coordinates.

D. Reduced Form Evidence

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

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

Table D1—Reduced Form Evidence of Imperfect Competition

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

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

D1. Imperfect Competition

In this subsection, we provide some descriptive evidence on imperfect competition in the Italian market for small business lines of credit. We construct four alternative measures of concentration and investigate their correlation with interest rates conditional on various sets of observables. First, we calculate the Herfindahl-Hirschman Index (HHI) of market concentration based on each bank's share of used credit within a market-year (hereafter HHI Loans). Second, we define the same index based instead on each bank's share of borrowers within a market-year (HHI Firms). Third, we construct the 3-bank concentration ratio in terms of used credit in each market-year (CR3 Loans). Finally, we construct the 3-bank concentration ratio in terms of borrowers in each market-year (CR3 Firms). We then use each of these as regressors in a regression of the log of predicted interest rates.[3] We show in Table D1 that concentration is positively correlated with interest rates in all specifications, as expected, and is statistically significant for two out of the four measures we use. In the first and fourth specifications, respectively in columns (1) and (4), we find that a 10% increase in concentration is associated with a 0.2% increase in interest rates.

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

D2. Asymmetric Information

In this subsection, we conduct reduced-form tests for the presence of asymmetric information. To do so, we follow the early empirical literature on positive correlation tests introduced by P. A. Chiappori and B. Salanié (2000). We propose two tests, one based on the choice to take up a line of credit and another based on the choice of how much credit to draw on that line. Both tests are based on the correlation between the unobservables driving these choices and the unobservables influencing default. The choice of these tests gives a flavor of the identification strategy that we rely on in the structural model, as explained in greater detail in Section III.C in the body of the text.

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

$$\begin{aligned} d_i &= \mathbf{1}(X_i^d \beta + \nu_i > 0), \\ f_i &= \begin{cases} X_i^f \gamma + \eta_i & \text{if } d_i = 1 \\ - & \text{if } d_i = 0, \end{cases} \end{aligned} \tag{D1}$$

where d_i is equal to one if the firm borrows and f_i is equal to one if the borrower defaults. f_i is observed only if $d_i = 1$. This is similar to the classical selection model analyzed by J. J. Heckman (1979), where we interpret as adverse selection a positive correlation between ν_i and η_i.[4] Results of this Heckman selection model are reported in the first two columns ("Extensive Margin") of Table D2, where the decisions to borrow and default are regressed on year, market, firms' Score, amount of granted credit, and sector fixed effects, as well as on a set of firms' balance sheet variables. We use as an instrument in the selection (i.e. borrowing) equation the number of banks in a firm's market, which we interpret as a proxy for the competitiveness of banks' local markets.[5] We find a positive and significant correlation coefficient of 0.09 between the unobservables driving demand and default, which we interpret as preliminary evidence of asymmetric information on the extensive margin.

Loan Use and Default: We then consider the relationship between loan use and default. Unlike the previous test, we are not in a selection framework as the same firms are observed in both equations. Still, the idea is the same, as we test for a positive correlation between the unobservables that determine the choice of "extent of coverage" (loan use) and the occurrence of an "accident" (default), conditional on several firm characteristics. Following

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

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

the intuition of the previous test, adverse selection should imply that riskier firms use more credit. We set up the following seemingly unrelated regression (SUR):

$$\text{(D2)} \qquad \begin{aligned} \ell_i &= X_i\beta + \varepsilon_i, \\ f_i &= X_i\gamma + \eta_i, \end{aligned}$$

where ℓ_i is the amount of its loan used by firm i, and, as before, f_i takes value of one if the borrower defaults. The vector of controls X_i is composed of year, market, bank, firms' Score, amount of granted credit, and sector fixed effects, as well as on a set of firms' balance sheet variables. We specify the distribution of the residuals ε_i, η_i as joint normal, with a correlation coefficient ρ. A positive and significant estimate of ρ suggests the presence of asymmetric information. The results of this test are summarized in the last two columns ("Intensive Margin") of Table D2. We again find a positive correlation, consistent with asymmetric information on the intensive margin.

E. Matching Model

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

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

The final set of variables that we use are all fixed effects and include fixed effects for the year, firms' Score, firms' geographical area, sales, and assets. In line with Caliendo and Kopeinig (2008), we apply the following specific criteria in our selection in order to determine both which variables to include and the degree of discretization for the fixed effects. First, we only include controls that influence simultaneously the participation (borrowing vs non-borrowing) and the outcome variables (interest rates or amount of granted credit, depending on the variable we are predicting). Second, variables must be unaffected by participation, or the anticipation of it, so should be either fixed over time or measured before participation. The Score respects this rule, and we assume that the value of assets

Table D2—Reduced Form Evidence for Adverse Selection

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

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

and sales are persistent over time. We choose to only control for a Score above or below 6, as G. Rodano, N. Serrano-Velarde and E. Tarantino (2015) showed that this is the most relevant threshold level for lending standards. Third, for the common support assumption to hold, some randomness is needed, so some firms with identical characteristics should be observed in each state (borrowing versus not). For this reason, we choose a parsimonious set of variables and seek to avoid over-parametrization.

We summarize in Table E1 the normalized differences in means between treatment (non-borrowers) and control (borrowers) groups.[6] Imbens (2004) defines as "modest" normalized differences below 0.3 in absolute value, and all of our differences for the continuous variables used are below that threshold. We implement several matching methods and find similar results across them. We choose to focus on k-nearest neighbor matching, as it allows us to assign several untreated (borrowing) firms to each treated (non-borrowing) one.

We follow the standard literature in performing several statistical tests to assess the quality of the matching. Variable selection is based on statistical significance, the "hit or miss" method (J. J. Heckman, H. Ichimura and P. E. Todd 1997),[7] and comparisons of several statistics before and after the matching. These include the Pseudo-R-squared, the Likelihood Ratio, the mean and median bias, and Rubin's B and R.[8] Caliendo and Kopeinig (2008) explain that a rule of thumb for a good match is to have mean and median biases below 3% to 5%. According to E. Leuven and B. Sianesi (2003), Rubin's B should be below 25% and Rubin's R should be between 0.5 and 2. Finally, a good matching outcome should deliver Pseudo R-squared and Likelihood Ratio tests high in the unmatched case, and very low in the matched case. Our results pass all these statistical tests, as shown in Tables E2 and E3. Last, we show a graph of the bias reduction and test the common support of the propensity score between treated and untreated in Figure E1. Even though there is a large mass at each tail, these figures show that the values for both groups span the full range of propensity scores, implying that we have enough overlap as long as we allow for replacement.

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

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

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

Table E1—Normalized Differences

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

Table E2—Matching Results 1

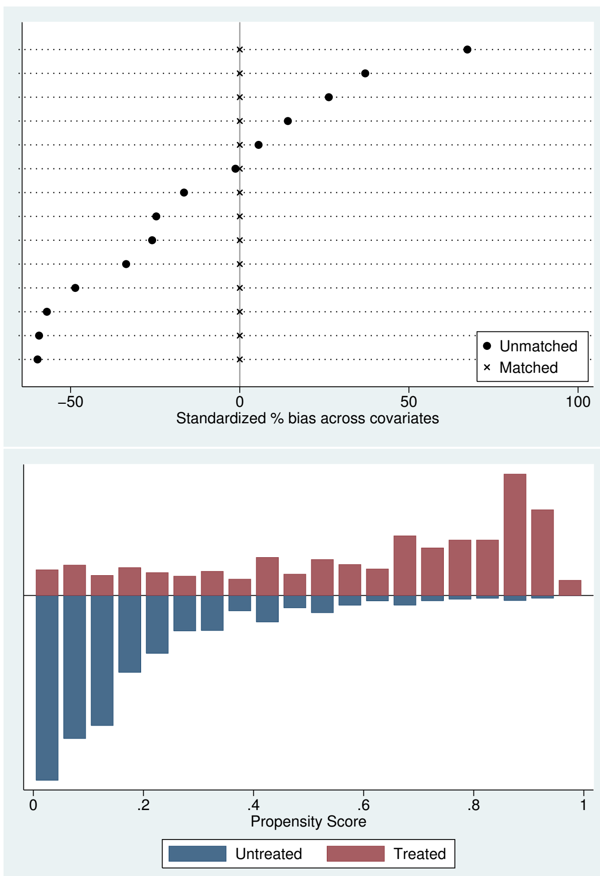
	Unmatched	**Mean**			**% Bias**	**t-Test**	
Variable	**vs. Matched**	**Treated**	**Control**	**% Bias**	**Reduction**	**t**	**p > |t|**
1991-1992	U	0.392	0.132	-33.6		-32.57	0.000
	M	0.392	0.392	0.0	100.0	0.00	1.000
1993-1994	U	0.333	0.308	5.6		5.90	0.000
	M	0.333	0.333	0.0	100.0	-0.00	1.000
1995-1996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

Table E3—Matching Results 2

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

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

Figure E1. Matching Graph and Common Support



F. The Effects of Measurement Error in Predicted Prices

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

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

$$P_{ij} = \widetilde{P}_{ij} + \widetilde{\tau}_{ij}, \tag{F1}$$

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

$$U_{ij}^D = \alpha^D P_{ij} + \varepsilon_i^D, \tag{F2}$$

$$U_{ij}^F = \alpha^F P_{ij} + \varepsilon_i^F. \tag{F3}$$

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

Recall also that in our structural model, we only use predicted prices in the demand estimation and not in the estimation of the loan use and default equations. In this stylized model, this means we only need to account for measurement error in equation (F2) and not equation (F3). Substituting (F1) in (F2), we get:

$$U_{ij}^D = \alpha^D(\widetilde{P}_{ij} + \widetilde{\tau}_{ij}) + \varepsilon_i^D = \alpha^D \widetilde{P}_{ij} + \zeta_{ij}, \tag{F4}$$

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

$$\rho_{DF} = \frac{\text{cov}(\varepsilon_i^D, \varepsilon_i^F)}{\sigma_{\varepsilon_i^D} \sigma_{\varepsilon_i^F}}. \tag{F5}$$

In the presence of measurement error, we would estimate:

$$\widetilde{\rho}_{DF} = \frac{\text{cov}(\zeta_{ij}, \varepsilon_i^F)}{\sigma_{\zeta_{ij}} \sigma_{\varepsilon_i^F}} = \frac{\text{cov}(\varepsilon_i^D, \varepsilon_i^F)}{(\sigma_{\varepsilon_i^D} + \alpha^D \sigma_{\widetilde{\tau}_{ij}})\sigma_{\varepsilon_i^F}} = \rho_{DF} \frac{\sigma_{\varepsilon_i^D}}{\sigma_{\varepsilon_i^D} + \alpha^D \sigma_{\widetilde{\tau}_{ij}}}, \tag{F6}$$

where the second equality follows from the assumed lack of correlation between $\widetilde{\tau}_{ij}$ and ε_i^F. One can see that the only effect of measurement error in (F6) is the additional variance in the composite demand error caused by the measurement error.

Equation (F6) shows that, typical of classical measurement error problems, in the presence of measurement error our estimate of ρ_{DF} would be biased towards zero. The size of any bias would depend on both α^D and the standard deviation of $\widetilde{\tau}_{ij}$ relative to that of ε_i^D, that is, the firm's private information. We believe that our price prediction reduces measurement error to a minimum. Indeed, as we argue in the text, firms too are likely to predict banks' prices, rather than getting a price quotation from each of them, suggesting thei variance of the demand error (which we normalize in the text to that of the standard Type I Extreme Value) using our predicted prices should be comparable to that of firms. Even if this weren't the case, equation (F6) shows that the degree of adverse selection we estimate is a lower bound on the true degree of adverse selection. Measurement error by itself, therefore, cannot explain the correlation we find between the propensity to demand credit and to default.

G. Additional Tables and Figures

Table G1—IV First Stage and OLS vs IV Second Stage for Demand

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

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

Table G2—IV First Stage for Loan Use and Default

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

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

Table G3—Structural Estimates with Random Coefficient on Prices

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

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

FIGURE G1. DEFAULT PROBABILITY DISTRIBUTION FOR BASELINE SCENARIO AND FOR HIGHER ADVERSE SELECTION COUNTERFACTUAL SCENARIO

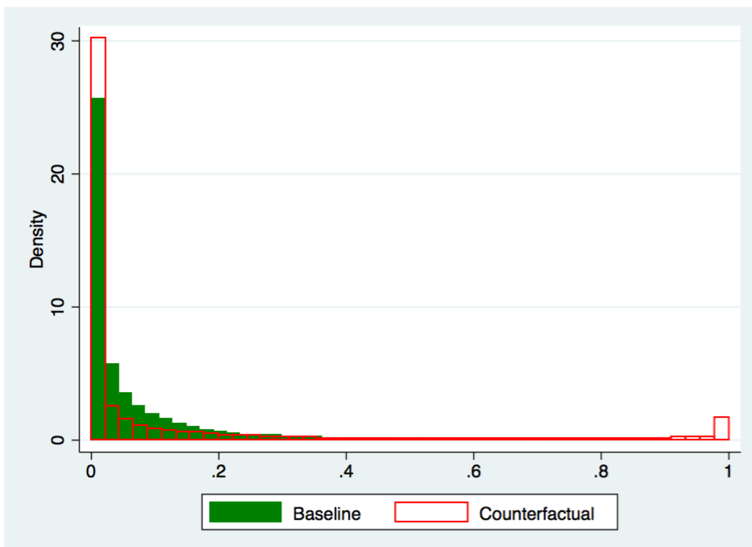


Table G4—Regressions of Counterfactual Outcomes' Changes on Markups for Higher Adverse Selection and Higher Marginal Costs

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

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

*

REFERENCES

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

Caliendo, M., and S. Kopeinig. 2008. "Some Practical Guidance for the Implementation of Propensity Score Matching." *Journal of Economic Surveys*, 22(1): 31–72.

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

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

Heckman, J. J., H. Ichimura, and P. E. Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *Review of Economic Studies*, 64(4): 605–654.

Imbens, G. W. 2004. "Matching Methods in Practice." *Journal of Human Resources*, 50(2): 373–419.

Imbens, G. W., and D. Rubin. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences.* Cambridge University Press.

Leuven, E., and B. Sianesi. 2003. "PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing." `https://ideas.repec.org/c/boc/bocode/s432001.html`.

Rodano, G., N. Serrano-Velarde, and E. Tarantino. 2015. "Lending Standards over the Credit Cycle." Working Paper.

Rubin, D. B. 2001. "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation." *Health Services & Outcomes Research Methodology*, 2: 169–188.

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