Markups, Productivity and the Financial Capability of Firms*

Carlo Altomonte† Domenico Favoino‡ Tommaso Sonno§

This version: September 2016

Abstract

In this paper we introduce financial frictions in a framework of monopolistically competitive firms with endogenous markups and heterogeneous productivity, as in Melitz and Ottaviano (2008). Before producing, firms need to invest part of their fixed costs in start-up capital to be used as collateral in order to obtain a loan necessary to finance production costs. In addition to productivity, firms are also heterogeneous in their financial capability: some firms obtain start-up capital at better conditions, thus decreasing their cost of collateral. As a result, on top of productivity, financial capability and collateral requirements also enter in the expression of the equilibrium firm-level markup. At the aggregate level, the model shows that tighter credit constraints in the form of higher collateral requirements mitigate the pro-competitive effect of trade. Theoretical results are structurally tested employing a refined measure of productivity purged from the differential access of firms to start-up capital, and capitalizing on a representative sample of manufacturing firms covering a subset of European countries during the financial crisis.

*The research for this paper was financially supported by the Baffi Carefin Centre.
†Bocconi University and Bruegel
‡Tinbergen Institute
§Université Catholique de Louvain and London School of Economics

1 Introduction

A large and growing literature shows how financial market imperfections affect country and industry-specific economic outcomes not only per se, but also through their interplay with firms' characteristics. In particular, access to external finance has been recognized as an important determinant of export and innovation activity of firms, on top of productivity.[1] In addition, there seems to be high within-industry heterogeneity of firms with respect to credit constraints, even after controlling for firm characteristics, such as size and productivity.[2]

Motivated by this evidence, this paper extends a framework of monopolistically competitive firms heterogeneous in productivity and with endogenous markups (as in Melitz and Ottaviano, 2008) to incorporate the presence of financial frictions. Firms need to allocate part of their initial fixed entry cost in tangible fixed assets that have to be used as collateral in order to obtain a loan necessary to cover part of their production costs (allocation problem). Once the loan is obtained, then firms maximize profits, given productivity (production problem). To account for the heterogeneous access of firms to external finance, firms in our framework also differ in their ability to raise collateral at lower financial costs. As a result, financial capability and collateral requirements drive, together with productivity, firm level equilibrium markups.

The implication of this finding is that, in our industry equilibrium, higher collateral requirements mitigate the pro-competitive effects (lower markups) observed in standard models of trade and firm heterogeneity.

All the results derived from the theoretical model are tested empirically on a representative sample of manufacturing firms covering a subset of European countries during the financial crisis (the EFIGE dataset).[3] EFIGE data have been integrated with balance-sheet data drawn from the Amadeus database managed by Bureau van Dijck, retrieving usable balance-sheet information for each surveyed firm from 2002 to 2013, thus enabling the calculation of time-varying firm-specific measures of productivity and markups.

We exploit these data and our theoretical model to back out a structural non-parametric

[1] See among others Minetti and Zhu (2011); Gorodnichenko and Schnitzer (2013); Manova (2013); Peters and Schnitzer (2015); Muuls (2015).

[2] Irlacher and Unger (2016) use World Bank firm-level data across countries to decompose the total variation in a measure for credit access (tangible over total assets) into within- and between-industry variation, finding that roughly 80% of the variation is within (narrowly defined) industries, also after controlling for firm-level characteristics, a feature that we also retrieve in our data.

[3] The European Firms in the Global Economy (EFIGE) dataset is a harmonized cross-country dataset containing quantitative as well as qualitative information on around 150 items for a representative sample of some 15,000 manufacturing firms in the following countries: Austria, France, Germany, Hungary, Italy, Spain, and the United Kingdom.

measure of the (unobserved) firm-specific financial capability, starting from simple balance sheet data. We then estimate a measure of firm-specific productivity purged from the effect of financial capability. In fact, to the extent that firms face a differential access to external finance, on top of the standard simultaneity bias TFP estimation could suffer from the potential endogeneity between (unobserved) financial capability and the amount of total fixed assets (capital) used by the firm in production, since more financially capable firms can more easily obtain fixed assets. If financial capability is not controlled for in the estimation, the latter ends up in the error term, leading to a positive correlation between the same error term and capital. This in turn leads to potentially biased production function coefficients affecting (downward) the estimated TFP. Hence, we have modified the Woolridge (2009) algorithm for productivity estimation in order to include financial capability as an additional control. We then use the corrected TFP measure to retrieve firm-level markups (as in DeLoecker and Warzynski, 2012), and test our theoretical model.

Our paper mainly contributes to the recent literature on international trade under credit constraints. Peter and Schnitzer (2015) incorporate credit constraints in Melitz and Ottaviano (2008) through a model of endogenous technology adoption, in which the cost of purchasing the advanced technology has to be financed externally. In their framework, however, they assume that technology adoption results in an increase of the price margin by a fixed amount, and do not work out the implications of credit constraints for markups.

Egger and Seidel (2012) is the paper closest to our approach. In their framework they follow Manova (2013) and let credit constraints emerge from an interplay of collateral assets, the share of costs that have to be financed externally, and the development of financial markets affecting the probability of repayment, as we do. Manova (2013) incorporates financial frictions in a CES framework with constant markups la Melitz (2003), while Egger and Seidel (2012) use a setting of linear demand and monopolistically competitive firms heterogeneous in productivity (Melitz and Ottaviano, 2008) in order to discuss the implications of credit constraints for prices and markups, thus similar to our research question. Differently from our approach, however, they do not take into account the heterogeneity of firms in access to external finance. As a result, in their framework, and contrary to ours, profit-maximizing quantities and prices are unaffected by credit constraints, with credit constraints playing a role only through the cost cutoff parameter.

Our paper also talks to other strands of the finance literature. From Graham (1998), Vig (2013) and Brumm et al. (2015) we take the idea that the amount and quality of tangible assets collected by firms typically influence the availability of collateral that banks require as a guarantee against loans. We also exploit evidence that larger firm size is typically associated to higher (need of) loans and thus collateral, as shown e.g. by Rampini and Viswanathan (2013). Capitalizing on these findings, we can control for the variation in the size of collateral across firms, and assume that the amount of collateral per unit of output is dictated by banks across firms in a given industry. The latter allows us to solve the allocation problem independently from the optimal firm size and thus productivity.

Another key ingredient of the model borrowed from the financial literature is the relationship between the tangible assets owned by the firm and the collateral she can pledge. Tangible assets differ in terms of "redeployability" (see Campello and Giambona, 2012; Carlson et al. 2004; Zhang, 2005; Cooper, 2006; Berger et al., 2011; Geraldo et al., forthcoming). Redeployable tangible assets (eg. land) are less firm-specific, but can be more easily sold and thus are more easily accepted as collateral. Because of their higher liquidity, the price of redeployable assets is the result of a bargaining process on the market between the supplier of the same asset and the firm. In our model, firms have a different ability in bargaining, and thus can obtain redeployable assets at different prices: as a result, more financially capable firms will benefit from a lower cost of collateral. On the other hand, non-redeployable assets (e.g. machinery) are more firm-specific, and thus their price can be ultimately subsumed by each firm's marginal cost.

The paper is organized as follows. We present our theoretical framework in Section II. Section III describes our data and introduces our estimation routines for financial capability, productivity and markups. In section IV we discuss the empirical strategy used to test our predictions and present our main results, together with robustness checks. The final section concludes.

2 Theoretical Model

2.1 Setup

In order to produce, liquidity constrained firms need to finance a share of their production costs through loans from a perfectly competitive banking sector. To provide a loan, banks require a sector-specific amount of tangible assets to be used as collateral. Production entails fixed costs to enter the market. Hence, before producing, firms have to allocate a share of the fixed costs to obtain assets used for production and as collateral (startup capital). Specifically, startup capital is composed of two types of tangible assets: redeployable (land, buildings) and non-redeployable (machinery). Redeployable assets are easier to collateralize.

Firms are heterogeneous in their productivity (marginal cost of production), as well as in financial capability: more financially capable firms have a lower cost of obtaining redeployable assets, and thus startup capital.

The two sources of heterogeneity are ex-ante uncorrelated, as marginal costs of production and financial capability are drawn from two independent probability distributions. They are also considered sequentially by firms: first, given her financial capability, a firm minimizes the cost of generating startup capital combining redeployable and non-redeployable assets (allocation problem); second, once endowed with startup capital, the firm obtains the loan and maximize profits, given her productivity (production problem). Still, the two sources of heterogeneity, even if ex-ante independent and sequentially considered, jointly influence the firm behavior. Specifically, more productive firms will end up in equilibrium with higher production volumes, and thus will be requested a higher amount of startup capital. At the same time, a firm with a higher level of financial capability will face lower costs for obtaining the required amount of startup capital, which in turn will influence her optimal production levels.

To disentangle the effects of these two sources of heterogeneity in a tractable way we exploit the fact that, while larger firms have larger amount of startup capital, the share of tangible assets used as collateral is exogenously dictated by banks, and thus equal in terms of unit of output for each firm in a given industry. We can therefore solve the allocation problem independently from the optimal firm size, deriving an expression for the cost advantage that a firm characterized by a given level of financial capability will have in creating the required amount of startup capital. We can then rewrite the production problem augmented with this cost advantage term and solve the model.

2.2 Demand Side

We consider an economy with L consumers, each supplying one unit of labour. Consumers can allocate their income over two goods: a homogeneous good, supplied by perfectly competitive firms, and a differentiated good. The market for the latter is characterized by monopolistic competition, with consumers exhibiting love for variety and horizontal product differentiation. Preferences are quasi-linear in order to allow for variable markups across firms, as in Melitz and Ottaviano (2008) among others:

$$U = q_0 + \alpha \int_{i \in \Omega} q_i^c \, di - \frac{1}{2}\gamma \int_{i \in \Omega} (q_i^c)^2 \, di - \frac{1}{2}\eta \left[\int_{i \in \Omega} q_i^c \, di \right]^2 \tag{1}$$

where the set Ω contains a continuum of differentiated varieties, each of which is indexed by i. q_0 represents the demand for the homogeneous good, taken as numeraire, while q_i^c corresponds to the individual consumption of variety i of the differentiated good. α and η are utility function parameters indexing the substitution pattern between the homogeneous and the differentiated good; γ represents the degree of differentiation of varieties $i \in \Omega$. Conditional on the demand for the homogeneous good being positive, i.e. $q_0 > 0$, and solving the utility maximization problem of the individual consumer, it is possible to derive the inverse demand for each variety:

$$p_i = \alpha - \gamma q_i^c - \eta \int_{i \in \Omega} q_i^c \, di \text{ }, \forall i \in \Omega \tag{2}$$

By inverting (2) we obtain the individual demand for variety i in the set of consumed varieties Ω^*, where the latter is a subset of Ω for which $q_i^c > 0$ and retrieve the following linear market demand system:

$$q_i = Lq_i^c = \frac{\alpha L}{\gamma + \eta N} - \frac{L}{\gamma} p_i + \frac{\eta N \overline{p} L}{\gamma(\gamma + \eta N)} \text{ }, \forall i \in \Omega^* \tag{3}$$

N represents the number of consumed varieties, which also corresponds to the number of firms in the market since each firm is a monopolist in the production of its own variety; $\overline{p} = \frac{1}{N} \int_{i \in \Omega^*} p_i di$ is the average price charged by firms in the differentiated sector.
In order to obtain an expression for the maximum price that a consumer is willing to pay, we set $q_i = 0$ in the demand for variety i and obtain the following:

$$p_{max} = \frac{\alpha\gamma + \eta N \overline{p}}{\gamma + \eta N} \tag{4}$$

Therefore, prices for varieties of the differentiated good must be such that $p_i \leq p_{max}$, $\forall i \in \Omega^*$, which implies that Ω^* is the largest subset of Ω that satisfies the price condition above.

2.3 Technology

Firms use one factor of production, labour, inelastically supplied in a competitive market. The production of the homogeneous good requires one unit of labour, which implies a wage normalized to one. Both the differentiated and the homogeneous good are produced under constant returns to scale, but the entry in the former industry involves a sunk cost f_E, representing start-up investments which constitute the initial endowment of each firm.

Firms are heterogeneous in productivity, having a firm-specific marginal cost of production $c \in [0, c_M]$ randomly drawn from a given distribution. Based on observation of their marginal production costs, firms then decide whether to stay in the market and produce a quantity $q(c)$ at a total production costs $cq(c)$, or exit.

2.4 Financing of firms and collateral

In our framework, liquidity constrained firms need to borrow money from banks in order to finance a share of their production costs $cq(c)$.[4] Banks, which operate in a perfectly competitive banking sector, define contract details for loans and make a take-it or leave-it offer to firms, including the collateral needed against the loan.

Firms with larger output $q(c)$, having to finance a higher production cost, will require a larger volume of credit and thus would need more collateral, which is an empirical regularity detected in the literature (Rampini and Viswanathan, 2013). Specifically, firms will be required by banks to pledge an amount of collateral equal to $\beta f_E q(c) \leq f_E$, with $\beta \in [0, 1]$ being industry-specific.[5] In other words, firms will have to use part of their fixed entry cost f_E to generate an amount of collateral proportional to their scale of production. This is the startup capital required by each firm. As it can be seen, startup capital is ultimately a function of a firm's productivity, as the latter determines the firm optimal output. However, the amount of collateral to be raised per unit of output, βf_E, is constant across firms in a given industry, and independent from productivity.

The required amount of startup capital is allocated by firms across different types of tangible fixed assets.[6] In particular the literature has identified two different categories of tangible assets that firms can use as collateral (Campello and Giambona, 2012): redeployable assets (Re) constituted by land, plants and buildings; and non-redeployable assets (NRe), i.e. machinery and equipment. Redeployable assets are easier to resell on organized markets: being more liquid, they can be easily used as collateral and thus facilitate firms' borrowing.

[4] In the trade literature with credit constraints, Manova (2013) among others assumes that loans obtained by firms are required to finance fixed export costs; in the innovation literature, external finance is typically needed for investments that increase productivity/lower marginal costs (e.g. Peters and Schnitzer, 2015; Eckel and Unger, 2015; Gorodnichenko and Schnitzer, 2014; Mayneris, 2012).

[5] As in Manova (2013) and Peters and Schnitzer (2015), β is the industry-specific share of assets required as collateral by the bank and seized if the firm is not able to repay the debt.

[6] The use of tangibles as collateral for loans is a standard practice for firms asking for loans and a common feature of the finance literature, as discussed among others by Graham (1998), Vig (2013) or Brumm et al. (2015). The fact that larger firms also have more tangible assets is another well known stylized fact. Manova (2013) assumes that fixed entry cost already constitute part of the collateral that firms can use, although she does not exclude that firms might invest in assets to increase their capacity for raising outside finance.

Non-redeployable assets are more firm-specific and with a value that deteriorates over time (because of technological obsolescence): as such, they are less easy to be used as collateral.

Because of their higher liquidity, the price of redeployable assets is the result of a bargaining process on the market between the supplier of the same asset and the firm. In particular, we assume that firms are heterogeneous in their financial capability of negotiating the price of redeployable assets. Specifically, firms are characterized by a specific level of financial capability $\tau \in [0, 1]$ randomly drawn from a probability distribution right after the entry of the firm.[7] The price of redeployable assets Re is thus $1 - \epsilon(\tau)$, with $\epsilon(\tau) \geq 0$ and increasing in τ.

The intuition here is that firms with better financial expertise can fetch a lower price on the market for their redeployable assets. This is in line with evidence provided by Guner et al. (2008), showing how e.g. the financial expertise of directors plays a positive role in finance and investment policies adopted by the firm. Glode et al. (2012) model instead the financial expertise of firms as the ability in estimating the value of securities, and show how these characteristic increase the ability of firms of raising capital.[8]

The price of non-redeployable assets instead can be normalized at unity. In fact, as these assets tend to be firm-specific in their use, one can assume that part of their price is formed on perfectly competitive markets plus an idiosyncratic component characterizing each firm, ultimately subsumed in the marginal production cost c.

The optimal allocation between redeployable and non-redeployable assets, given the constraint on the amount of startup capital, is then the result of the following minimization problem:

$$\min C(Re, NRe) = (1 - \epsilon(\tau))\, Re + NRe \tag{5}$$

subject to the constraint:

$$\left(Re^{\frac{\sigma-1}{\sigma}} + NRe^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} = \beta f_E$$

The term $C(Re, NRe)$ represents the cost of tangible asset per unit of output that the firm spends when allocating its endowment f_E in redeployable (Re) and non-redeployable

[7] The probability distributions $\tau \in [0, 1]$ and of $c \in [0, c_M]$ are assumed to be independent.

[8] Alternatively, one can think at the relationship lending literature, studying how the relationship of managers with banks increase funds availability and reduce loan rates (see Elyasiani and Goldberg, 2004, for a review of the recent literature). More financially capable managers might be more skilled in bargaining with banks, thus reducing the overall cost of collateral needed to obtain a given loan. The channel through which these effects can have an impact on firms lending are both fund availability and quantity (see Boot and Thakor, 1994; Berger and Udell, 1996; Cole et al., 2004), or prices and collateral (see Berger and Udell, 1998; Petersen and Rajan, 1995).

(NRe) assets, given the price of the same assets, with $\sigma > 1$. The constraint βf_E is the amount of collateral (per unit of output) required by banks.

From the minimization of the cost function 5 we obtain

$$C(\tau) = \frac{\beta f_E \left[(1-\epsilon(\tau)) + (1-\epsilon(\tau))^{\sigma}\right]}{\left[1 + (1-\epsilon(\tau))^{\sigma-1}\right]^{\frac{\sigma}{\sigma-1}}} \tag{6}$$

in which $C(\tau)$ is strictly decreasing in the financial capability of the firm. Equation 6 also allows us to define the financial capability cutoff, i.e. the level of financial capability that makes firms indifferent (in terms of costs) in allocating their endowment between Re vs. NRe assets. This corresponds to $\widetilde{\tau}$ such that $\epsilon(\widetilde{\tau}) = 0$, i.e. a firm characterized by the cutoff financial capability would not obtain any type of advantage in the price of redeployable assets. As a consequence, the cost of the τ-cutoff firm is equal to:

$$C(\widetilde{\tau}) = \beta f_E 2^{\frac{1}{1-\sigma}} \tag{7}$$

The latter represents the cost upper bound that the less financially capable firm will face in order to produce the required amount of tangible assets per unit of output.

The implications of heterogeneity in financial capability can be seen considering the case of all firms having the same financial expertise $\bar{\tau}$. As firms in the industry have the same fixed entry cost f_E, in our setting they will end up with the same cost to produce the required amount of tangible asset per unit of output $C(\bar{\tau})$. In this case, the total cost of producing tangible assets for any firm $\overline{TC}(c) = C(\bar{\tau})q(c)$ will just be a function of the firm's size, i.e. ultimately of its marginal costs. In other words, even introducing a financial sector in our framework, without heterogeneity in financial capability productivity will remain the only endogenous variable needed to characterize the entire equilibrium of the industry: a given marginal cost c would in fact determine the firm's size $q(c)$ and from here the volume of the loan as a share of production costs $cq(c)$, as well as the total cost $\overline{TC}(c)$ of producing the tangible assets required to be able to pledge some collateral against these loans. Introducing heterogeneity also on financial capability τ, on top of productivity, allows instead to derive non-trivial implications for firms' behavior, especially when studying the implications of financial shocks.

2.5 Banking sector

Following Egger and Seidel (2012), firms need to externally fund a share ρ of their total production costs $cq(c)$ and have to repay $R(c)$ to banks. Repayment occurs with exogenous probability λ, with $\lambda \in (0, 1]$, which is determined by the strength of financial institutions, while with probability $(1 - \lambda)$ the financial contract is not enforced, the firm defaults, and the creditor seizes the startup capital $\beta f_E q(c)$.[9]

To close the deal, the participation constraint of a bank is then:

$$-\rho c q(c) + \lambda R(c) + (1 - \lambda)\beta f_E q(c) \geq 0 \tag{8}$$

As we can easily see, no interest rate is charged by banks because of perfect competition in the banking sector. For the same reason, the participation constraint holds with equality for all banks. Hence, it is possible to derive an expression for the repayment function:

$$R(c) = \frac{1}{\lambda}[\rho c - (1 - \lambda)\beta f_E]q(c) \tag{9}$$

Moreover, although all firms with a financial capability larger than $\widetilde{\tau}$ can in principle obtain a loan, firms will apply only if a liquidity constraint is satisfied, such that net revenues are at least equal to the repayment of the loan $R(c)$ to the bank (see e.g. Manova, 2013). In evaluating this constraint in our setting, we should also consider the heterogeneity of firms in terms of the costs incurred to raise the required startup capital.

Hence, we derive an expression for the cost advantage that a firm characterized by financial capability τ will have in creating the required amount of startup capital with respect to the cutoff firm. By substracting (6) from (7) we get

$$\theta(\tau) = C(\widetilde{\tau}) - C(\tau) = \beta f_E \left[2^{\frac{1}{1-\sigma}} - \frac{(1 - \epsilon(\tau))}{\left[1 + (1 - \epsilon(\tau))^{\sigma - 1}\right]^{\frac{1}{\sigma - 1}}} \right] \tag{10}$$

The last expression, increasing in τ, describes the cost advantage of a firm with financial capability τ. Consistently, the cost advantage of the financial capability cutoff firm will be zero: $\theta(\widetilde{\tau}) = 0$.

From here, we can write the liquidity constraint of the firm, now incorporating the two sources of firm heterogeneity in marginal costs and financial capability (c, τ)

[9] Note that banks would supply loans to all firms characterized by a financial capability above the cutoff, as all these firms are able to raise startup capital.

$$p(c,\tau)q(c,\tau) - (1-\rho)cq(c,\tau) + \theta(\tau)q(c,\tau) \geq R(c,\tau) \tag{11}$$

A firm for which the above inequality does not hold would not be able to obtain the loan because of its inability to reimburse the debt to the borrower. This firm would exit the market right after the entry, i.e. after the random draw of its τ and marginal cost of production c.

2.6 Profit maximization

Each firm in the differentiated sector maximizes the following profit function

$$\Pi(c,\tau) = p(c,\tau)q(c,\tau) - (1-\rho)cq(c,\tau) - \lambda R(c,\tau) - (1-\lambda)\beta f_E q(c,\tau) - C(\tau)q(c,\tau)$$

As there are two sources of heterogeneity (c and τ), to solve the model we have to consider the cutoff level of marginal costs at which profits are zero (a free exit condition, as in Melitz and Ottaviano, 2008), given the cost advantage in generating startup capital obtained by a firm with financial capability τ with respect to the cutoff $\widetilde{\tau}$, and under the participation constraint (8), the liquidity constraint (11) and the demand for the supplied variety (3).

By plugging the expression for repayment (9) in the profit function, and substituting the generic expression of costs $C(\tau)$ with the cost advantage $\theta(\tau)$, we obtain a much simpler form for firm's profits, in which the financial capability cutoff is already incorporated:

$$\pi(c,\tau) = p(c,\tau)q(c,\tau) - cq(c,\tau) + \theta(\tau)q(c,\tau) \tag{12}$$

Solving the profit maximization problem and using the demand constraint (3) to derive $\frac{\partial p}{\partial q} = -\frac{\gamma}{L}$ yields the FOC:

$$p(c,\tau) - \frac{\gamma}{L}q(c,\tau) - c + \theta(\tau) = 0$$

By rearranging the terms in the above equation, we obtain the supply condition:

$$q(c,\tau) = \frac{L}{\gamma}\left[p(c,\tau) - c + \theta(\tau)\right] \tag{13}$$

We can now use the liquidity constraint (11) in order to impose a free entry condition and derive the marginal cost cutoff c_D. Knowing that firms that would not be able to repay the debt will directly exit the market, the liquidity constraint must hold with equality for the cutoff firm. Moreover, since the cutoff firm corresponds to that firm that sets $p_i = p_{max}$, we can rewrite (11) as follows:

$$p_{max}q(c_D,\tau) - (1-\rho)c_Dq(c_D,\tau) + \theta(\tau)q(c_D,\tau) = R(c_D,\tau)$$

Rearranging the terms in the equation above yields a simple expression for p_{max} as a function of the cost cutoff c_D:

$$p_{max} = \omega c_D - \phi - \theta(\tau) \tag{14}$$

where $\omega = \frac{\rho}{\lambda} + 1 - \rho$ and $\phi = \frac{1-\lambda}{\lambda}\beta f_E$ are constants.

Note that, since $\theta(\tau)$ is increasing in τ, the maximum price charged by a firm corresponds to the price made by the least financially capable firm, since $\theta(\widetilde{\tau})$ is the lower bound of $\theta(\tau)$. For this reason, in correspondence of p_{max} we have that $C(\tau) = C(\widetilde{\tau})$.

2.7 Equilibrium

At equilibrium, the demand for each variety equals supply:

$$\left[\frac{\alpha\gamma}{\gamma+\eta N} + \frac{\eta N\overline{p}}{\gamma+\eta N} - p(c,\tau)\right]\frac{L}{\gamma} = \frac{L}{\gamma}[p(c,\tau) - c + \theta(\tau)]$$

As the two terms on the left hand side are equal to p_{max}, by substituting it with its expression in (14) and rearranging we obtain the equilibrium price charged by a firm characterized by a given set of (c,τ)

$$p(c,\tau) = \frac{1}{2}\left[\omega c_D + c - \phi - \theta(\tau)\right] \tag{15}$$

From here, we can derive an expression for the equilibrium markup of a (c,τ)-firm by subtracting the marginal cost from the equilibrium price:

$$\mu(c,\tau) = p(c,\tau) - MC(c,\tau) = \frac{1}{2}\left[\omega c_D - c - \phi + \theta(\tau)\right] \tag{16}$$

By looking at expression (16), it is easy to note that, as in Melitz and Ottaviano (2008), the equilibrium markup charged by a (c,τ)-firm is increasing in the production cost cutoff c_D and decreasing in the firm-specific marginal cost of production c. Hence, the more productive a firm, the higher would be its markup (holding constant the effects on the equilibrium cost cut-off c_D of the industry, herein discussed).

Differently from Melitz and Ottaviano (2008) however the introduction of credit constraints, as well as a second source of firm heterogeneity, namely financial capability in raising startup capital, affect the expression of the markup. First, from eq. (16) is clear that a higher financial capability translates, *ceteris paribus*, into a higher markup, via the effect of the cost advantage $\theta(\tau)$. The intuition is that a higher financial capability leads to a higher cost advantage in generating the required amount of startup capital: similar to productivity, more financially capable firms then transfer this advantage into a markup premium. As a result, conditional on the existence of credit constraints / collateral requirements, the dispersion of firm-level markups around a productivity level normally observed in the data could be explained by this second source of firm heterogeneity.[10] To validate this hypothesis, in the second part of the paper we will empirically test on firm-level data the following

Proposition #1. *The equilibrium markup $\mu(c,\tau)$ of a firm characterized by a pair (c,τ) is ceteris paribus an increasing function of financial capability τ.*

The second change brought about by our model to the expression of the firm-level markup is the role of credit constraints, as subsumed in the constant parameters ω and ϕ. To explore how a change in credit constraints affects the behavior of firms, we need however to solve for the industry equilibrium, as the cost cut-off c_D is endogenous to these variables.

2.8 Parameterization

To fully characterize the industry equilibrium, we have to solve for the value of the cost cut-offs c_D and $\widetilde{\tau}$. As in Melitz and Ottaviano (2008), we assume that the marginal cost of production c follows an inverse Pareto distribution with a shape parameter $k \geq 1$ over the support $[0, c_M]$. As we have no ex-ante prior on the distribution of financial capability of firms, we assume that τ follows a uniform distribution in the interval [0,1]. The cumulative density functions of c and τ can then be written as:

$$G(c) = \left(\frac{c}{c_M}\right)^k \quad \text{with} \quad c \in [0, c_M]$$

$$F(\tau) = \tau \quad \text{with} \quad \tau \in [0,1]$$

respectively. The density functions therefore are $g(c) = \frac{kc^{k-1}}{c_M^k}$ and $f(\tau) = \frac{1}{1-a}$.

Under these assumptions, it is possible to solve for the financial capability cutoff $\widetilde{\tau}$, independently from c_D. Recall that the cut-off of τ is defined as $\epsilon(\widetilde{\tau}) = 0$. We thus need to

[10] Given the independence of the two distribution of costs and financial capability, we can consider the cost cut-off c_D as given in analyzing the effect of τ on the markup.

specify a functional form of $\epsilon(\tau)$, i.e. the price advantage enjoyed by the τ firm in the purchase of the redeployable asset. We assume that $\epsilon(\tau) = \tau - a$, with $a \in [0, 1)$ being a constant. It is easy to note that $\epsilon(\tau)$ increases in τ and the function equals 0 in correspondence of a, therefore implying that the financial capability cutoff is $\widetilde{\tau} = a$.[11]

To solve for the cost cut-off c_D we need to apply the free-entry condition, defined over both sources of heterogeneity. From the equality of demand and supply we can derive an expression for a firm's profits in equilibrium:

$$\pi(c,\tau) = \frac{L}{4\gamma}\left[\omega c_D - c - \phi + \theta(\tau)\right]^2 \tag{17}$$

Firms would be willing to enter the market until expected profits are equal to the fixed cost of entry f_E, i.e.:

$$\pi^e = \int_0^{c_D}\int_a^1 \frac{L}{4\gamma}\left[\omega c_D - c - \phi + \theta(\tau)\right]^2 \; dF(\tau)\; dG(c) = f_E \tag{18}$$

Since $dG(c) = g(c)dc$ and $dF(\tau) = f(\tau)d\tau$, we can rewrite the integral as:

$$\pi^e = \frac{Lk}{4\gamma c_M^k (1-a)}\int_0^{c_D}\int_a^1 \left[\omega c_D - c - \phi + \theta(\tau)\right]^2 c^{k-1}\; d\tau\; dc = f_E \tag{19}$$

As shown in Appendix A, it is not possible to find an explicit solution for c_D in 19. However, as shown in the same Appendix, one can prove that a positive solution always exists and it is unique conditional on a choice of c_M. The parameterization then allows us to find an expression for the equilibrium number of firms in the market and derive average performance measures as a function of c_D.[12]

By equating both expressions (4) and (14) for p_{max} and solving N, we obtain

$$N = \frac{\gamma}{\eta}\frac{(\alpha - \omega c_D + \phi)}{(\omega c_D - \phi - \overline{p})} \tag{20}$$

[11] The assumption of a uniform distribution of τ together with its independence from marginal costs c (as the allocation problem of tangible assets is not related to productivity) implies that the financial capability cutoff $\widetilde{\tau}$ does not depend on market characteristics but rather is a constant, whereas the cost cutoff c_D is endogenous like in Melitz and Ottaviano (2008). This simplification allows to control for a second source of heterogeneity in the firm-level equilibrium equations, while maintaining the model tractable at the level of industry aggregates. The latter does not entail however a loss of generality, as it can be shown that for relevant shocks (e.g. market size) the effects of the cost and financial capability cutoffs go in the same direction.

[12] The limitation of the setup is that the effect of the variable of interest (say market size) on a given performance measure (say markup) cannot be solved numerically, but will have to be assessed by taking into account the sign of the effect of the variable of interest on c_D.

which corresponds to the number of firms, and therefore varieties of the differentiated good, active in the market in equilibrium. From here we can derive an expression for the average price for the differentiated good and obtain a parameterized expression also for N.

Since the average price is a function of the the average marginal cost and financial capability, we first derive expressions for these measures. Following Melitz and Ottaviano (2008), we define the average marginal cost of production as:

$$\bar{c} = \frac{\int_0^{c_D} cg(c)dc}{G(c_D)} = \frac{kc_D}{k+1} \tag{21}$$

Since the financial capability τ is distributed as a uniform over the interval $(0, 1)$, we also have that:

$$\bar{\tau} = \frac{\int_a^1 \tau f(\tau)d\tau}{F(1-a)} = \frac{1+a}{2} \tag{22}$$

Now we can derive an expression for the average price charged by firms active in the market, which corresponds to:

$$\bar{p} = \frac{1}{2}\frac{\int_0^{c_D}\int_a^1 \left[\omega c_D + c - \phi - \theta(\tau)\right] f(\tau)g(c)dcd\tau}{G(c_D)F(1-a)}$$

Solving the integral yields:

$$\bar{p} = \frac{1}{2}\left[\frac{\omega k + \omega + k}{k+1}c_D - \phi - \frac{1-a}{2}\right] \tag{23}$$

By substituting equation (23) in the expression for the number of firms derived above (20), we have that, at the equilibrium, the number of firms in the market for the differentiated good equals

$$N = \frac{2\gamma}{\eta}\frac{(\alpha - \omega c_D + \phi)}{\left(\frac{\omega k+\omega+k}{k+1}c_D - \phi - \frac{1-a}{2}\right)}$$

Finally, in the same way in which we obtained an expression for the average price $\bar{p}$, we can derive the expression for the average markup

$$\bar{\mu} = \frac{1}{2}\left[\frac{\omega k + \omega - k}{k+1}c_D - \Theta\right] \tag{24}$$

where $\Theta = \phi - \bar{\tau} + a$ is a constant depending on the density distribution of financial capability and the exogenous parameters of the model.

2.9 Trade shock and the role of financial constraints

Our setup allows us to analyze the effects of an increase in the market size L, which is analogous to a symmetric opening of the economy to trade as in Melitz and Ottaviano (2008). Differentiating eq. (24) yields

$$\frac{\partial \overline{\mu}}{\partial L} = \frac{1}{2}\left[\frac{\omega k + \omega - k}{k+1}\frac{\partial c_D}{\partial L}\right] < 0$$

To explore the implications of a trade shock we thus have to look at the effect of an increase in L on the cost cutoff c_D. As shown in Appendix B, we have that $\frac{\partial c_D}{\partial L} < 0$, i.e. an increase in market size tends to reduce the average industry markup by lowering the cost cutoff, in line with the pro-competitive effect of trade identified in the literature.[13]

Still, a closer look at the expression for the average markups reveals that financial constraints can play a role in the reaction of the economy to a trade shock. In fact, the magnitude of the derivative of the cost cutoff with respect to L depends, among others, on the amount of collateral requirements β (see Appendix B for a discussion). In particular, when β is relatively large, i.e. when banks require more collateral for the same loan, the effect of a change in L on the cost cutoff is relatively low.[14] We can formalize this finding with the following

Proposition #2. *An increase in the market size L lowers the average markup $\overline{\mu}$. Tighter credit constraints in the form of higher collateral requirements tend however to mitigate the pro-competitive effect of trade.*

3 Data and estimations

We first describe the data we use, focusing on those unique characteristics that allow us to disentangle the different sources of firm heterogeneity (productivity vs. financial capability). We then provide a structural estimation of the (unobserved) firm-specific financial capability. Finally, we exploit the latter to estimate a measure of firm-specific productivity purged from the effect of financial capability. From here, we back out firm-level markups, and proceed to

[13] Given the assumption of $\widetilde{\tau} = \alpha$ being independent from market characteristics, financial capability does not affect the reaction of average markups to a trade shock. Extending the model to the case of an endogenous financial capability cutoff, i.e. solving the free entry condition also for $\widetilde{\tau}$ would still yield a pro-competitive effect of trade on average markups also through the financial capability channel.

[14] The ECB Bank Lending Survey shows how in the years 2008 and 2009 collateral requirements by banks have tightened threefold in the euro area

test our Propositions.

3.1 Firm-level data

Our firm-level data derive from the first survey on European Firms in a Global Economy (Efige), a research project funded by the European Community's Seventh Framework Programme (FP7/2007-2013). The project aims at analyzing the competitive performance of European firms in a comparative perspective. This dataset is the first harmonized cross-country dataset containing quantitative as well as qualitative information on around 150 items for a representative sample of some 15,000 manufacturing firms in the following countries: Austria, France, Germany, Hungary, Italy, Spain, and the United Kingdom. These items cover international strategies, R&D, innovation, employment, financing and organizational activities of firms, before and after the financial crisis. The complete questionnaire is available on the Efige web page, www.efige.org. A discussion of the dataset as well as its validation is available in Altomonte et al (2012), while Bekes et al. (2011) discuss explicitly the reaction of firms to the crisis as measured in the survey.

An interesting characteristic of the Efige dataset is that, on top of the unique and comparable cross-country firm-level information contained in the survey, data can be matched with balance sheet figures. More precisely, we have been able to integrate Efige data with balance-sheet information drawn from the Amadeus database managed by Bureau van Dijck, retrieving twelve years of usable balance-sheet information for each surveyed firm, from 2001 to 2013. Overall, the dataset includes about 3,000 firms operating in Germany, France, Italy and Spain, some 2,200 firms in the United Kingdom, and about 500 firms for Austria and Hungary, as reported in Table 1. Descriptive statistics are reported in Appendix C.

Table 1: Efige sample size, by country

Country	Number of firms
Austria	443
France	2,973
Germany	2,935
Hungary	488
Italy	3,021
Spain	2,832
UK	2,067
Total	14,759

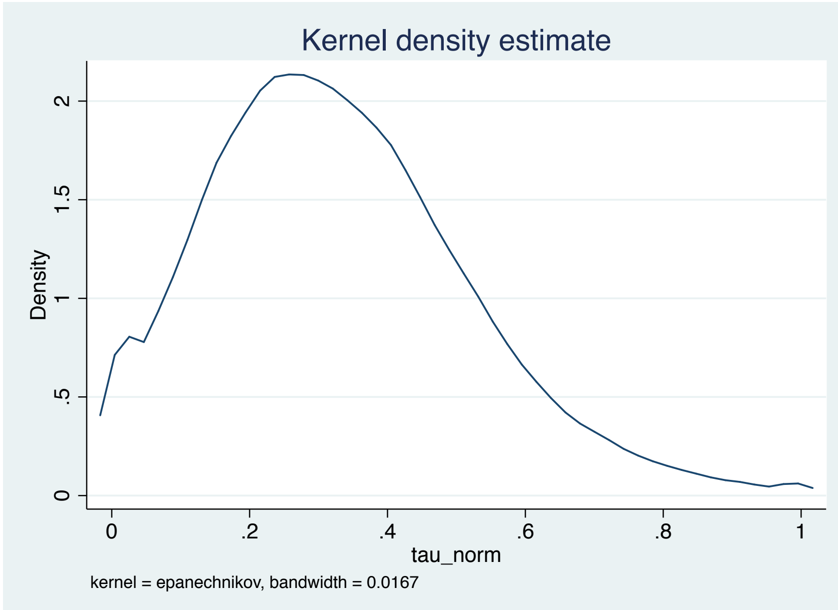
The sampling design follows a stratification by industry, region and firm size structure. Firms with less than 10 employees have been excluded from the survey, that instead presents an oversampling of larger firms with more than 250 employees to allow for adequate statistical inference for this size class.[15]

3.2 Estimation of financial capability

The theoretical model allows to back out an estimate of financial capability at the firm level starting from balance sheet information. The starting point is to write an expression of startup capital (the total amount of tangible assets used by each firm as collateral) in nominal terms. Recalling that $C(\tau)$ in equation (6) represents the cost of tangible asset per unit of output that the firm spends when allocating its endowment f_E in redeployable (Re) and non-redeployable (NRe) assets, then nominal tangible assets can be simply written as $TA(\tau, q) = C(\tau)q(c, \tau)$. Considering that the amount of collateral is fixed by banks for firms of the same size, we have that the balance sheet value (nominal value) of the tangible assets used for collateral, namely TA, should be decreasing in τ once controlling for firm size. The intuition here is that all firms are required to collect the same *amount* of TA per unit of output, but firms with higher τ will obtain that required amount at a lower cost (lower *nominal* value). Also, for the cut-off firm $\tilde{\tau}$ we have $TA(\tilde{\tau}, q) = C(\tilde{\tau})q(c, \tau)$. It then follows

[15] In order to take into account the oversampling and to retrieve the sample representativeness of the firms' population, a weighting scheme (where weights are inversely proportional to the variance of an observation) is set up according to firm's industry and class size. All our regression results are thus computed by taking into account this weighting scheme, except where otherwise specified. Detailed information on the distribution of firms by country/size class and industry can be retrieved on the Efige website

Figure 1: Distribution of τ



that the nominal value $TA(\tilde{\tau}, q)$ represents the highest value of tangible assets owned by a (cutoff) firm of size $q(c, \tau)$.[16]

From here we can estimate τ non-parametrically in three stages. First, we create size ranges of firms, by industry (deciles, and then quintiles and twentiles as robustness). Secondly, for each size range within each industry, we identify the upper bound level of (nominal) tangible assets recorded by firms (average TA of top 5% of firms, robustness with 1%): this value of TA represents the TA level of the cut-off firm(s). Finally, we compute the firm-specific τ by dividing the TA level of the cut-off firm for each firm-level value of nominal TA, for every size/industry partition. We retrieve an index ≥ 1 which we then bound between 0 (cutoff firms) and 1 (maximum financial capability).[17] The estimated distribution of τ is left-skewed, as we can see in Figure 1.[18]

Considering all combinations of size ranges (quintiles, deciles, twentiles) and of cutoff levels of tangible assets (1%, 5%), we obtain several version of our firm-specific financial capability measures, which we will use as sensitivity checks in testing our Propositions.

[16]Recall that $C(\tilde{\tau})$ is the upper bound of the cost function each firm has to face while investing in a given amount of tangible assets.

[17]Our identification strategy crucially relies on the fact that financial capability and productivity / marginal costs are independent, as a correlation would result in biased measures. While we will explicitly purge total factor productivity estimates from the effects of financial capability, the firm-level measure of τ retrieved here is in any case uncorrelated (.008) with the firm-level labor productivity (value added per employee) measured in the sample.

[18]Note that in our theoretical model τ is assumed to be distributed as a uniform, however this choice simply allows to ease computations, assuming a pareto distribution would not alterate any of our findings.

3.3 Estimation of productivity and markups

In order to estimate markups and productivity at the the firm level and over time, we start from De Loecker and Warzynski (DLW-2012), which introduced a method to estimate markups by employing expenditure on inputs and elasticity of output to the use of inputs in production, all information retrievable from nominal balance sheet data available to us.

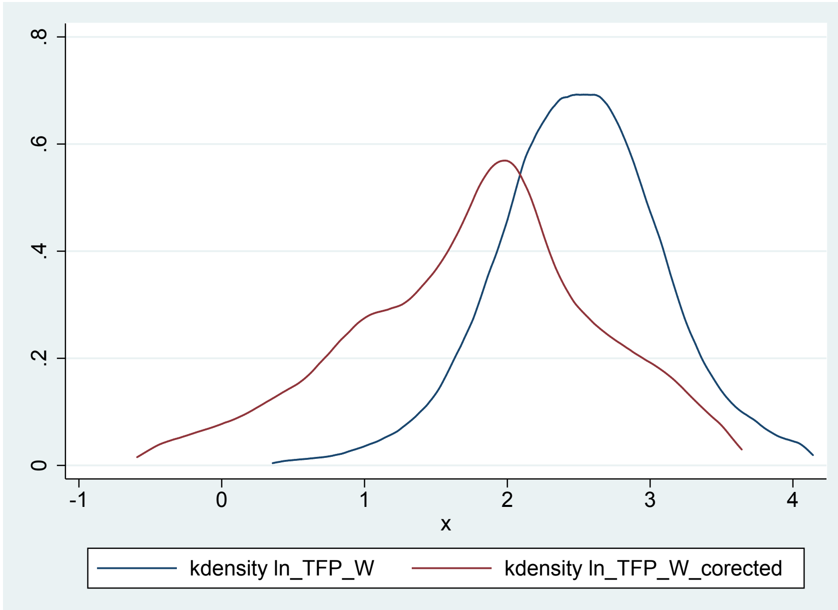
Technically, we have estimated our production function coefficients relying on Wooldridge (2009), which proposes to improve on the Ackerberg, Caves and Frazer (ACF-2009) algorithm originally employed in De Loecker and Warzynski (2012) through the use of a GMM framework, in order to obtain efficient estimates for the marginal productivities of capital and labor. These efficiently estimated coefficients have then been employed to compute firm level markups, as well as total factor productivity estimates (TFP, henceforth).

Still, a key issue in our production function estimation is due to the fact that, on top of the standard simultaneity bias, TFP estimation could suffer from the potential endogeneity between financial capability and the amount of total fixed assets (capital) used by the firm in production, since more financially capable firms can more easily obtain fixed assets. If financial capability is not controlled for in the estimation, the latter ends up in the error term, leading to a positive correlation between the same error term and capital. This in turn leads to potentially biased production function coefficients affecting (downward) the estimated TFP and markup. For these reasons we have modified the Woolridge (2009) algorithm including the financial capability variable as an additional control (as in De Loecker, 2007 for the export status).

Figure 2 plots the retrieved (corrected) TFP vs. the uncorrected one obtained through the standard Woolridge (2009) algorithm, and confirms the downward bias of productivity estimation when not controlling for the firm specific financial capability.

Table 2 below reports instead the median values and standard deviations of four different firm-level markups computed by using our TFP measures. The first two employ the Woolridge (2009) algorithm for computing TFP, both in the standard and corrected version discussed above. As a robustness check, we have also estimated production function coefficients through the ACF-2009 routine as in De Loecker and Warzynski (2012), and then used the retrieved coefficients to construct an alternative measure of markups. Also in this case we have corrected the standard ACF algorithm for the unobserved differential access of firms to tangible assets. Markups computed under the ACF algorithm could then be regressed on TFP computed la Woolridge (2009) to address any claim of potential spurious correlation between the two measures. All our results qualitatively hold in any case with

Figure 2: Productivity kernel graphs



different combinations of markups and TFP.

Table 2: Markup estimates: median values and standard deviations

Estimation method	Median	Standard deviation
Wooldridge (no correction)	1.2063	1.2465
Wooldridge (correction)	1.2152	1.1446
ACF (no correction)	1.0668	0.6317
ACF (correction)	1.0886	0.9905

4 Empirical analysis

4.1 Test of Proposition 1

Through our theoretical framework we can structurally test our firm-level markup equation (16), expecting a positive correlation between markups and firm-specific productivity (inverse of marginal costs), as well as between markups and financial capability.

Specifically, recalling our markup equation

$$\mu(c,\tau) = \frac{1}{2}\left[\omega c_D - \phi - c + \theta(\tau)\right]$$

we estimate the latter at the firm-year level, with the dependent variable $\mu(c,\tau)$ being the markup estimated through DeLoecker-Warzynski (2012), as previously discussed. In terms of covariates, ωc_D and ϕ are fixed effects or controls (depending on specification), c is (the inverse of) our TFP measure, corrected for τ and estimated in section 3.3, while $\theta(\tau)$ is the (normalized) non-parametric estimation of the level of financial capability retrieved as described in section 3.2.

We test equation (16) for the years 2002-2013 under various specifications. In addition, as heterogeneity in financial capability is relevant only for liquidity constrained firms (as in this case the markup is influenced by the lower cost of the required startup capital), we condition our estimates on whether firms have requested in the considered period a loan from a bank, an information available in our dataset.[19]

Table 3 presents the results. In column (1) we estimate firm-specific financial capability by deciles of sales, and we assume the cutoff level of tangible assets to be the top 5% for each decile. We employ a full set of firm fixed effects to wipe out any unobserved heterogeneity at the firm level that can drive the results, as well as year fixed effects. Results confirm that markups are positively correlated with productivity and that even controlling for productivity financially capable firms display significantly higher markups, as predicted by the theoretical framework. In column (2) we control for the possibility that some financial shock happening over time at the country level (and thus not picked up by our firm FE) might drive the results, introducing a measure of the change in collateral requirement as retrieved from the ECB Bank Lending Survey for the euro area. [20]

Insofar we have identified the effects of productivity and financial capability through the within variation in the data, thus implying that firms can readjust their allocation of capital and, consequently, markups over time. If our theory is valid, however, our results should also hold when we identify through the between variation in the data: controlling for productivity, firms with higher financial capability should also have higher markups. In columns (3) and (4) we thus replicate our analysis without firm fixed effects. We include a set of country*industry fixed effects to capture all possible spurious compositional effects beyond variation at the firm level.[21] We also control for some firms' fixed characteristics that might be correlated with both productivity and financial capability, notably the (logarithm of) firm's age as well as a firm's size.[22] In column (3) we keep the panel dimension through

[19] As a matter of fact, this condition is verified for 14,139 firms in our data (i.e. 96% of the sample).

[20] The ECB Bank Lending Survey shows how in the years 2008 and 2009 collateral requirements by banks have tightened threefold on average in the euro area

[21] Industry fixed-effects are retrieved from Manova (2013) as measures of financial vulnerability (i.e. the extent to which a firm relies on outside capital for its investment).

[22] The effects of firms'size on TFP and financial constraints have been widely discussed in the literature (see

Table 3: Test of Proposition 1

	(1) Within estimator	(2) Within estimator	(3) Between estimator	(4) OLS
	decile of sales, top 5% TA cutoff	decile of sales, top 5% TA cutoff	decile of sales, top 5% TA cutoff	decile of sales, top 5% TA cutoff
	all years	all years	all years	only 2008
Dependent variable	$\ln(\mu)_i$	$\ln(\mu)_i$	$\ln(\mu)_i$	$\ln(\mu)_i$
$\ln(TFP)_i$	1.547***	1.594***	1.363***	1.462***
	(0.0109)	(0.0139)	(0.0123)	(0.0191)
Financial capability$_i$	0.437***	0.484***	0.205***	0.280***
	(0.0189)	(0.0231)	(0.0237)	(0.0375)
Change in collateral requirement		-0.0152*	-0.173*	
		(0.00778)	(0.101)	
Obs.	53,698	35,525	32,149	4,548
R2	0.807	0.836	0.726	0.769
Number of marks	7,873	7,249	6,544	
Firm size and age controls	NO	NO	YES	YES
Firm FE	YES	YES	NO	NO
Country-Industry FE	NO	NO	YES	YES
Year FE	YES	YES	YES	NO
Robust SE	YES	YES	NO	YES

Notes. Dependent variable: (log of) markups estimated as in DeLoecker and Warzynski (2012), using production function coefficients estimated as in Wooldridge (2009). The financial capability variable is computed across deciles of sales, assuming firms having top 5% of TA to be the cutoff firms. TFP is (in log) computed through a modified version of Wooldridge (2009), cleaning production function estimates for firm-level financial capability. Change in collateral requirements indicates the percentage increase/decrease in the collateral requirements by banks. All specifications are estimated with robust standard errors. ***,**,* = indicate significance at the 1, 5, and 10% level, respectively.

a between estimator, while in column (4) we focus on the cross-section for the year 2008 (the year for which most of the firm characteristics observed in our survey were measured). In both these cases the coefficient of financial capability decreases by around a third with respect to the within-estimation, but remains positive and highly significant.

We then proceed with some sensitivity and robustness checks. In Table (4) we change the estimation procedure of τ as well as employ different measures of firm-level markups (see Table 2), while always controlling for firm fixed effects. Namely, in column (1) we estimate financial capability by shrinking the size ranges of firms' sales to quintiles, and widening the cutoff level threshold of tangible assets in each industry to the top 10% of firms. In column (2) we do the opposite, broadening the size ranges to twentiles, and narrowing the cutoff

for example Hadlock and Pierce, 2010). We introduce this control in the form of a categorical variable, varying from 1 to 4 based on the firm having between 10-19, 20-49, 50-249 or more than 250 employees, respectively. The choice of a categorical variable is driven by the willingness of reducing the possible endogeneity with TFP and other firm-specific controls. All our results are confirmed if we substitute the natural log of the number of employee to the size categories.

Table 4: Test of Proposition 1 - Sensitivity

	(1)	(2)	(3)	(4)
	Within estimator	Within estimator	Within estimator	Within estimator
	quintile of sales, top 10% TA cutoff	twentiles of sales, top 1% TA cutoff	decile of sales, top 5% TA cutoff	decile of sales, top 5% TA cutoff
	all years	all years	all years	only 2008
Dependent variable	ln(μ)i Wooldridge (correction)	ln(μ)i Wooldridge (correction)	ln(μ)i ACF (no correction)	ln(μ)i ACF (correction)
ln(TFP)i	1.587***	1.588***	0.703***	1.584***
	(0.0137)	(0.0138)	(0.00874)	(0.0110)
Financial capabilityi	0.390***	0.466***	0.467***	0.682***
	(0.0212)	(0.0256)	(0.0177)	(0.0192)
Change in collateral requirement	-0.0142*	-0.0135*	-0.0181***	-0.0257***
	(0.00782)	(0.00784)	(0.00642)	(0.00732)
Obs.	35,525	35,393	40,034	39,777
R2	0.835	0.834	0.646	0.838
Number of marks	7,249	7,222	7,519	7,536
Firm size and age controls	NO	NO	NO	NO
Firm FE	YES	YES	YES	YES
Country-Industry FE	NO	NO	NO	NO
Year FE	YES	YES	YES	YES
Robust SE	YES	YES	YES	YES

Notes. Dependent variable: (log of) markups estimated as in DeLoecker and Warzynski (2012), using production function coefficients estimated, respectively, as in Wooldridge (2009) with a correction for financial capability (Columns 1 and 2), as in the standard ACF(2015) algorithm (Column 3), and with the correction for financial capability (Column 4). The financial capability variable is now computed across quintiles of sales, with cutoffs firm considered those with top 10% largest TA (Column 1), or twentiles of sales, with cutoff firms considered those with top 1% largest TA (Column 2). All specifications are estimated with robust standard errors. ***,**,* = indicate significance at the 1, 5, and 10% level, respectively.

level of tangible assets to the top 1% of the distribution. In columns (3) and (4) we go back to our original measure of financial capability, but we split the methods through which TFP and markups are estimated, in order to avoid picking up some spurious correlation. Namely, in column (3) we employ our corrected measure of TFP estimated through the Woolridge (2009) algorithm, but we use as dependent variable markups estimated through the standard ACF (2015) method. In column (4) we repeat the exercise using firm level markups estimated with the ACF (2015) algorithm corrected for the financial capability of firms. All our results remain unchanged.

In terms of robustness checks, we need to assess whether our measure of financial capability keeps its significance also when we control for firm-specific variables to which it could be potentially endogenous. To that extent, we use two questions available in the Efige survey for the year 2008. A first question inquires on the number of banks used by the firm. The question is answered by almost the entire sample and shows an average of three banks per firm (two for the median firm). The intuition is that a management better connected to a relatively high number of banks might obtain an advantage in the cost of its collateral. The second question relates to the R&D investments incurred by the firm. The idea is that a higher financial capability might lead to higher R&D investments, and from here higher productivity and thus a higher markup. In this case, the relation between financial capability and markups might be spuriously driven by this omitted variable. Column (1) of Table (5) replicates column (5) of Table (3), reporting our benchmark specification for the cross-sectional sample previously discussed.

Table 5: Proposition 1 - Robustness

	(1) OLS	(2) OLS	(3) OLS	(4) OLS
	decile of sales, top 5% TA cutoff	decile of sales, top 5% TA cutoff	decile of sales, top 5% TA cutoff	decile of sales, top 5% TA cutoff
	only 2008	only 2008	only 2008	only 2008
Dependent variable	$\ln(\mu)_i$	$\ln(\mu)_i$	$\ln(\mu)_i$	$\ln(\mu)_i$
$\ln(\text{TFP})_i$	1.462***	1.459***	1.461***	1.458***
	(0.0191)	(0.0188)	(0.0191)	(0.0188)
Financial capability$_i$	0.280***	0.296***	0.281***	0.297***
	(0.0375)	(0.0367)	(0.0375)	(0.0367)
$\ln(\text{Banks})_i$		0.111***		0.110***
		(0.0117)		(0.0118)
Investments in R&D$_i$			0.0203*	0.00983
			(0.0111)	(0.0109)
Obs.	4,548	4,500	4,548	4,500
R2	0.769	0.777	0.770	0.777
Firm size and age controls	YES	YES	YES	YES
Firm FE	NO	NO	NO	NO
Country-Industry FE	YES	YES	YES	YES
Year FE	NO	NO	NO	NO
Robust SE	YES	YES	YES	YES

Notes. Dependent variable: (log of) markups estimated as in DeLoecker and Warzynski (2012), using production function coefficients estimated as in Wooldridge (2009). Financial capability is computed by decile of sales, assuming firms having the top 5% of TA to be the cutoff firms. TFP is (in log) computed through a modified version of Wooldridge (2009), cleaning production function estimates for firm-level financial capability. The (log of) number of banks indicates the number of banks used by the firm. Investment in R&D assumes value one if the firm invested in innovation in the period 2005-2008. All specifications are estimated with robust standard errors. ***,**,* = indicate significance at the 1, 5, and 10% level, respectively.

Column (2) and (3) show that controlling for the (log of) number of banks a firm has relationships with, or the fact that a firm has undertaken an investment in R&D, does not affect the sign and significance of our measure of financial capability. In column (4) we control for both variables jointly, with no changes in significance.

4.2 Test of Proposition 2

As a last exercise we also provide indirect evidence of our second theoretical result, and namely the fact that a trade shock leads to pro-competitive effects of trade mediated by credit constraints, with larger collateral requirements leading to lower reductions in average markups.

To test for this, we consider an augmented version of our markup equation. Specifically, we introduce a time-specific 'trade shock' dummy variable that measures the sudden, ample and symmetric negative trade shock incurred by European countries during the credit crisis of 2008/09 (Baldwin, 2009); and a country-specific 'collateral' dummy variable taking into account the fact that, based on data from the European Lending Survey and national sources, Italy and Spain report on average a higher collateral requirement by banks vs. other countries in the sample. If our Proposition 2 is correct, we should observe a positive sign of the crisis dummy, as a negative trade shock leads to higher markups; and a negative sign of the interaction between the trade shock and the collateral dummy, as the effect of the trade shock will be smaller the higher is the collateral requirement.

Table 6 reports the results of our estimation. In column (1) and (2) we employ a specification of the markup equation where we do not control for financial capability since, as discussed in the theoretical section, the latter should have the same pro-competitive effect of the trade shock on markups. In columns (3) and (4) we re-introduce financial capability, as this should lead to a better specified markup equation. Column (1) and (3) test the effect of the trade shock on markups in the narrow time-window of 2007 and 2008, while column (2) and (4) enlarge the time-window estimation to cover the period 2006-2009.

Results are consistent with our theory across specifications. The effect of the trade shock is negatively signed and significant, as lower trade leads to higher markups; while the interaction of the trade shock dummy with high collateral requirement is negatively signed and significant: higher collateral requirements dampen the pro-competitive effects of trade. Interestingly, we also observe that high collateral requirements per se are on average negatively and significantly correlated to markups.[23] However, given the nature of our proxy

[23]The intuition for the latter result is that when banks pledge for more collateral, at the firm level this increases costs, and thus leads to a reduction of markups. At the industry level, some firms would not be able

Table 6: Test of Proposition 2

	(1) OLS	(2) OLS	(3) OLS	(4) OLS
			decile of sales, top 5% TA cutoff	decile of sales, top 5% TA cutoff
	2007-2008	2006-2009	2007-2008	2006-2009
Dependent variable	$\ln(\mu)_i$	$\ln(\mu)_i$	$\ln(\mu)_i$	$\ln(\mu)_i$
$\ln(TFP)_i$	1.426***	1.421***	1.451***	1.443***
	(0.0527)	(0.0362)	(0.0506)	(0.0349)
Financial capability$_i$			0.337***	0.304***
			(0.0597)	(0.0458)
High collateral requirement (HCR)	-0.435***	-0.558***	-0.466***	-0.510***
	(0.0712)	(0.0617)	(0.0649)	(0.0561)
Negative trade shock (NTS)	0.0491***	0.0652***	0.0508***	0.0707***
	(0.0102)	(0.0182)	(0.00990)	(0.0185)
HCR*NTS	-0.0520**	-0.0470**	-0.0423**	-0.0362*
	(0.0172)	(0.0177)	(0.0184)	(0.0189)
Obs.	9,371	17,662	9,371	17,662
R2	0.756	0.753	0.763	0.759
Firm size and age controls	YES	YES	YES	YES
Country-Industry FE	YES	YES	YES	YES
Year FE	NO	YES	NO	YES
Robust SE	YES	YES	YES	YES

Notes. Dependent variable: (log of) markups estimated as in DeLoecker and Warzynski (2012). Financial capability is computed by decile of sales, assuming firms having the top 5% of TA to be the cutoff firms. TFP is (in log) computed through a modified version of Wooldridge (2009), cleaning production function estimates for firm-level financial capability. Negative trade shock is a dummy=1 if year 2008 or 2009. High collateral requirement is a dummy = 1 if Italy or Spain. All specifications are estimated with standard errors clustered at the country-year level. ***,**,* = indicate significance at the 1, 5, and 10% level, respectively.

we cannot be sure of the actual channel here, as it might simply be the case that markups are lower in Italy and Spain for market reasons.

5 Conclusions

In this paper we have introduced financial frictions in a framework of monopolistically competitive firms with endogenous markups and heterogeneous productivity, as in Melitz and Ottaviano (2008). Before producing, firms need to invest part of their fixed costs in start-up capital to be used as collateral in order to obtain a loan necessary to finance production costs. In addition to productivity, firms are also heterogeneous in their financial capability: some firms obtain start-up capital at better conditions, thus decreasing their cost of collateral.

The theoretical model predicts that, conditional on productivity, a higher financial capability is associated to higher markups at the firm level, while higher collateral requirements mediate the pro-competitive effects of trade shocks on average across firms.

These theoretical results are structurally tested on a representative sample of manufacturing firms covering a subset of European countries during the financial crisis. Through the theory we also obtain a structural non-parametric measure of firm-level financial capability from balance sheet information, through which we can derive TFP measures unbiased for the (unobserved) capital formation process.

In terms of future lines of research, the impact of collateral requirements on the pro-competitive effects of trade should be more thoroughly explored with more detailed firm-level data. To this extent, starting from the markup equation, it should be also possible to derive a structural measure of financial constraints at the firm level, that could be successfully used in firm-level analyses of the impact of financial frictions on the industry equilibrium. But we leave this topic to another paper.

to satisfy the liquidity constraint as the repayment function $R(c)$ becomes larger. Hence, the least efficient firms in the market would not obtain the loan from banks and exit, generating a fall in the production cost cutoff c_D, and thus a reduction in average markups.

Reference

[To be added]

Appendix

A Existence and uniqueness of cost cutoff

Equation (19) sets the expected profits of a firm facing the choice of entering the market as:

$$\pi^e = \frac{Lk}{4\gamma c_M^k(1-a)} \int_0^{c_D} \int_a^1 \left[\omega c_D - c - \phi + \theta(\tau)\right]^2 c^{k-1} \, d\tau \, dc = f_E$$

Solving the integral yields

$$\pi^e = \frac{Lk}{4\gamma c_M^k(1-a)} c_D^k \left[A c_D^2 + B c_D + C\right] = f_E$$

with the terms A, B and C being respectively equal to:

$$A = 4(1-a)\left[\frac{1}{2+k} - \frac{2\omega}{1+k} + \frac{\omega^2}{k}\right]$$

$$B = \frac{4\omega + k\omega - k}{k(1+k)}\left[2a\phi - 2\phi + f_E\beta(ln4 + a - 1) - 2f_E\beta ln(1+a)\right]$$

$$C = -\frac{1}{k(1+a)}\left[4a^2\phi^2 - 4\phi^2 + f_E\beta\phi(ln4 + a - 1) + f_E^2\beta^2(ln16 - 3 + 2a + a^2 + aln16)\right] + \\ + \frac{1}{k(1+a)}\left[4(1+a)f_E\beta(f_E\beta + 2\phi)ln(1+a)\right]$$

Now define $f(c_D)$ as:

$$f(c_D) = \pi^e - f_E = A c_D^{k+2} + B c_D^{k+1} + C c_D^k - \frac{4 f_E \gamma c_M^k}{Lk(1-a)}$$

By Rolle's Theorem, between two solutions of $f(c_D) = 0$ there is always a solution of $f'(c_D)$. Hence, if $f'(c_D) = 0$ at least two positive values of the cost cutoff exist. Moreover, as long as the second positive cost cutoff is $> c_M$, the latter also implies the uniqueness of c_D.

By taking the first derivative of $f(c_D)$ we obtain

$$f'(c_D) = (k+2)A c_D^{k+1} + (k+1)B c_D^k + kC c_D^{k-1}$$

where $A > 0$ and $C > 0$ always, while $B < 0$ for a very broad range of parameters in our model. Hence, by Cartesio's Rule, $f'(c_D) = 0$ has at least two positive solutions, i.e. there is a solution to $f(c_D) = 0$.

B Derivative of cost cut-off with respect to L

By applying Dini's implicit function theorem, we obtain:

$$\frac{\partial c_D}{\partial L} = -\frac{\partial \pi^e(L, c_D(L))/\partial L}{\partial \pi^e(L, c_D(L))/\partial c_D}$$

The derivative of the expected profit function with respect to L is equal to:

$$\frac{\partial \pi^e(L, c_D(L))}{\partial L} = \frac{k c_D^k}{4\gamma c_M}(A c_D^2 + B c_D + C) > 0$$

with A, B and C having been defined in Appendix A. The denominator is instead equal to:

$$\frac{\partial \pi^e(\beta, c_D(\beta))}{\partial c_D} = (k+2) A c_D^{k+1} + (k+1) B c_D^k + k C c_D^{k-1}$$

and is always positive in correspondence of positive values of the expected profits. Hence, we have that:

$$\frac{\partial c_D}{\partial L} = -\frac{\partial \pi^e(L, c_D(L))/\partial L}{\partial \pi^e(L, c_D(L))/\partial c_D} < 0$$

Looking at how collateral requirements β affect the above derivative, by comparing the two partial derivatives of π^e it can be easily shown that a higher β will increase the parameters B and C in both the numerator and the denominator of $\frac{\partial c_D}{\partial L}$, with the effect being however larger on the denominator. As a result a higher β will translate *ceteris paribus* into a lower value of $\frac{\partial c_D}{\partial L}$.

C Descriptive statistics

From the Amadeus dataset linked with Efige we derive information on Tangible Fixed Assets, Sales as a proxy of output and the number of employees. We also use two questions available in the Efige survey as a robustness to our estimations of firm-specific financial capability. A first question inquires on the number of banks used by the firm. The question is answered by almost the entire sample and shows an average of three banks per firm (two for the median firm). A second question inquires whether firms have undertaken investments in R&D in the period 2005-2008.

Table 7 reports descriptive statistics for the year 2008, i.e. the year referred to in the questions related to financial capability and investment in R&D.

Table 7: Descriptive statistics

	Obs.	Mean	Std. Dev.	Min	Max
Tangible Fixed Assets (2008)	12035	1903	4582.88	1,002	50204
Sales (2008)	10554	10986	24694.42	194	250214
Employees (2008)	9583	66	113.94	10	1062
Number of Banks	14571	2.99	2.02	1	14
Investments in R&D	14759	59.90%	0.49	0	1