Investigating the Topological Origins of Market Power in Ecuador's Production Network

Master's Thesis
Department of Economics
University of Zurich

Miguel Vázquez Vázquez

Supervisor: Prof. Dr. Alessandro Ferrari

Course of Studies	Economics
Student ID	16-712-598
E-Mail	miguel.vazquezvazquez@uzh.ch
Date	21st May 2024

Abstract

[*] This paper explores the uncharted relationship between market power and topological firm characteristics within Ecuador's production network. Grounded in the premise that neither a firm's position in the network nor the markups imposed on its customers are random, I investigate whether the former can help explain the latter. Leveraging a rich dataset on Ecuadorian firm-to-firm transactions, I construct several measures of network strength at the seller level. Simultaneously, I exploit tax filings to estimate firm-level markups with a common approach from the industrial organisation literature. My analysis reveals suggestive evidence relating higher firm-level markups to a few measures of strong network positions. Even though the statistical significance is high, it is susceptible to the markup specification and the inclusion of controls for unobserved characteristics of firms. Moreover, magnitudes are small, with firms required to move between the first and third quartiles of a network metric distribution to gain, in the best case, only a couple of percentage points in markups.

1 Introduction

Economic production networks are intricate and asymmetrical, as any spontaneous network revolving around human activities (Jackson, 2019). The 2008 financial crisis highlighted how crucial some firms can be in safeguarding or undermining general economic stability, with governments needing to bail out several financial institutions. The goal was to avoid catastrophic ripple effects that the bankruptcy of these institutions holding key economic intermediary roles could cause. Recently, light was again shed upon the importance of intra-firm links as governments worldwide imposed stringent lockdowns. The goal was to contain the spread of the COVID-19 pandemic; the unintended side effects were significant supply chain disruptions with consequences still felt today (Carvalho et al., 2020b).

At the same time, a growing body of literature focuses on market power and its evolution over the past decades. In this context, a common way to measure it is through the notion of the markup — i.e., the margin between price and marginal cost. For example, De Loecker et al. (2020) find that aggregate markups have substantially increased since 1980. The effect traces back to a reallocation of market shares from firms at the lower end of the markup distribution to firms at the upper end, in addition to a sharp increase in firm-level markups for those at the top.

Could there be a connection between market power and topological characteristics? As the availability of high-quality firm-to-firm transaction data increases, it opens a window for a deeper look into the implications of production networks' structures. Identifying key features of the networks might help better understand economic mechanisms such as market power. Over the years, the literature has proposed different explanations for market power. From market structure changes such as mergers (Stiebale and Szücs, 2022) or lower antitrust enforcement (Gallardo and Philippon, 2018) that reduce competition, up to technological improvements that increase productivity hence reducing marginal costs (De Loecker et al., 2021). A combination of both is also possible, with Chiavari and Goraya (2024) suggesting that technological change in production favouring intangible capital can change market structures.

My research aims to investigate and potentially unveil a relationship between being in strong positions within the production network and exerting market power. In other words, I want to examine whether there is evidence of firms

internalising the strength of their network position when making pricing decisions. Ex-ante, I expect firms exhibiting high degrees of network strength along different dimensions to have relatively higher markups. In this context, accounting for the intrinsic nature of each different industry is particularly important, considering that I anticipate network importance to be heterogeneous at the industry level.

My work leverages rich administrative data from Ecuador, including detailed tax filings and intra-firm transactional data. The paper's core relies on a restricted sample of around 50,000 unique firms observed from 2008-2011. The roughly 125,000 firm-year observations account for over half of the network's intra-firm transaction value. More details on the data sources and preparation work can be found in Section 2.

After cleaning and preparing the data, I begin my analysis in Section 3 by estimating firm-level markups on the tax filings data using the methodology brought forward by De Loecker and Warzynski (2012). While initially recovering markups using different variable input definitions and production function forms, I eventually settled for a gross output Translog production function and a combined variable input to recover firm-level markups for my baseline results.[1] I obtain an estimated median firm-level markup of 1.08 across all years and firms. There is sizeable markup heterogeneity across industries, which I define based on ISIC Revision 3.1 divisions (2-digit level). For instance, firms in the three largest industries display a median markup of 1.01. In contrast, the median lies at 1.33 for all the other industries combined.[2] Additionally, despite much movement in the within industry markup ranking of firms over time and a considerable decrease in firm-level markup dispersion, the aggregate markup remains remarkably constant during the four years. Despite the differences in levels, these findings are consistent across the other markup specifications.

I move on to the network analysis part in Section 4. There, I begin by constructing several measures to capture the strength of sellers' position in the network based on intra-firm transactions involving over a million firms. The thirteen measures can be categorised into five areas: importance for the buyer,

---

[1]By contrast, De Loecker and Warzynski (2012) use a value-added Translog production function and exploit labour inputs only in their baseline specification.

[2]It is worth highlighting that the three most prominent industries in my sample are, perhaps coincidentally, the three divisions of ISIC Revision 3.1 section "Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods" (G) and together constitute more than 50% of my sample.

relationship quality, competition, market targeting, and diversification. Most metrics exhibit extremely skewed distributions and do not display high correlations among each other, especially in linear terms. The latter is a favourable outcome, indicating that the metrics capture various network structure dimensions. Although there is firm-level variation across the years, medians and standard deviations are reasonably stable over time. I find that most of the metrics display remarkable heterogeneity between industries. Within an industry group blending primary sector, manufacturing and construction, the median firm is its industry's sole supplier for around 6% of its customers. For a median firm in the remaining industries, the value is virtually 0%. Mostly, these differences are sensible and can be understood through the nature of the industries' operations.

My analysis culminates in Section 5 as I match the estimated markups with the network metrics to investigate the possibility of market power originating in the network. I find suggestive evidence relating higher markup to most network strength measures. The statistical significance of the effect, and sometimes the direction, is susceptible to the markup specification and the inclusion of controls for unobserved characteristics of firms. In addition, magnitudes are slim as they require firms to move between the first and third quartiles of a network metric distribution to gain, in the best case, only one or two percentage points of price-cost margin. Lastly, the correlational nature of my analysis does not allow me to say whether there is evidence of firms adjusting their prices due to their topological characteristics or whether the inverse is true, i.e., that firms may gain network strength thanks to their market power. This is only one of the various limitations I mention when discussing my results.

Finally, I conclude my work in Section 6 by suggesting several possibilities for future research. These include but are not limited to, defining competition markets differently or investigating the other side of the market power coin, i.e., markdowns, through the lens of buyer-based network metrics.

Related Literature

One of the most widely used ways to measure market power in industrial organisation and trade literature is through markups. Nevertheless, markups are usually not observed. There are two possible routes to obtain estimates: the *production approach* and the *demand approach*. While neither is trivial, as sim-

ultaneity bias plagues both, the former is arguably simpler.[3] As such, I focus on the literature concerning the estimation of production functions, which has long been a critical area of economic research.

As first documented by Marschak and Andrews in 1944, the main problem of estimating production functions arises from the fact that the econometrician does not methodologically select the set of inputs used for output production but is instead a result of optimised firm behaviour. Usually, the input decisions and the resulting output are jointly influenced by factors that the econometrician does not observe, e.g., productivity. Not accounting for them in the estimation procedure inevitably leads to biased results. The formalisation of this issue has pushed the literature to focus on ways to identify production functions and productivity correctly.

Over time, two classes of methods established themselves as standard practices for production function estimation: *dynamic panel methods* (Arellano and Bond, 1991; Blundell and Bond, 1998, 2000) and *proxy variable methods* (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Wooldridge, 2009; De Loecker and Warzynski, 2012; Ackerberg et al., 2015; Gandhi et al., 2020).[4] The critical difference between the two lies in the assumptions made about the productivity process to overcome the bias resulting from its nonobservance.

The history of incorporating network analysis tools into economic contexts, a discipline sometimes referred to as *economics of networks*, is relatively recent. However, its popularity has been growing in the last couple of decades and has potential for further growth, primarily due to three reasons: (i) increasingly greater theoretical model complexity that accommodates network-like interactions; (ii) rapid growth in the availability of detailed network data in various contexts; and (iii) exponential increase in computational capacities that allow handling such big data (Jackson, 2016).

Concretely, several economics subfields have embraced network analysis tools in recent times. Some examples are papers examining the effect of social networks on employment (Calvó-Armengol and Jackson, 2004, 2007; Mayer, 2011), studies investigating how to improve the design of financial networks to make them more robust to shocks (Acemoglu et al., 2015; Cabrales et al.,

---

[3]In the second approach, demand is estimated directly to obtain the price elasticities required to compute markups with this alternative. Besides suffering from the well-known price endogeneity bias, estimating demand directly is notoriously tricky as it also requires rarely available data on prices and quantities, in addition to precise definitions of markets. Berry and Haile (2021) provide a recent overview of the issues surrounding demand estimation.

[4]In the literature, *proxy variable methods* are sometimes referred to as *control function methods* due to using a vaguely-specified function to control for productivity.

2017), and articles proving that idiosyncratic microeconomic shocks can be amplified while navigating through production networks to originate aggregate fluctuations (Acemoglu et al., 2012; Grassi, 2017).

Nevertheless, there have been few attempts to investigate a potential direct relationship between market power and network-related metrics, even less so at the firm level. Grassi (2017) shows that network centrality, among other characteristics, acts as a mechanism to transform microeconomic productivity shocks affecting markups into aggregate fluctuations. However, this investigation only considers network measures at the industry level. Zavala (2022) exploits the same data I use to show that agricultural goods exporters exert monopsonistic power on Ecuadorian farmers, who face markdowns on their produce. Yet, the author does not use network-related metrics in his analysis besides the classic market concentration measures. Finally, attempts to examine the relationship between markups and upstreamness — an index proposed by Antràs et al. (2012) to capture a firm's distance to the final consumer — found divergent results (Colonescu, 2021; Gradzewicz and Mućk, 2023). Regardless, both studies examine this relationship at the industry level. The most similar project to mine is a work in progress by Carvalho et al. (2020a), where they also try to relate a firm's position in the production network to market power. Nevertheless, they take rather simplistic market power metrics such as profit margins while also working on a new market power measure recovered directly on the network through a *bottleneck firm* framework.

Thus, my contribution to the literature is twofold. Firstly, I expand the economics of networks literature by filling the gap of an unexplored potential relationship between market power and network position strength at the firm level. Initially, I do so by proposing distinct measures to capture the strength of a selling position in a production network and providing a short overview of the complex production network through them. Consecutively, I present suggestive statistical evidence that the two concepts might somehow be related while providing possible underlying mechanisms.

Secondly, I deliver estimates for Ecuadorian firms' markups over multiple years in a literature otherwise heavily focused on the United States or, more generally, developed economies. Work from De Loecker and Eeckhout (2018) and Diez et al. (2018) represent exceptions, as both papers focus on long-term trends over the last four decades across different continents. Interestingly, they both find discrepancies between markup trends in advanced compared to emerging economies. While the former group displays positive growth,

the trend for the latter, especially in South America, is relatively constant or even slightly decreasing. When it comes to Ecuador specifically, Rodríguez-Moreno and Rochina-Barrachina (2019) provide the only piece of evidence on the country's markups. Yet, they only focus on manufacturing firms using 2010 census data.

2 Data

The primary data source for this project is establishment-level administrative data obtained from the Ecuadorian Internal Revenue Services (*Servicio de Rentas Internas*, SRI) for research purposes.[5] This data has already been used in multiple papers (Adão et al., 2022; Carrillo et al., 2017, 2023a,b; Zavala, 2022). The available data is extremely comprehensive: it includes social security payments, customs forms, firm ownership registries, and more. Nevertheless, due to remarkable discrepancies across the different datasets, I only use income tax form submissions, purchase annexes recording intra-firm transactions, and the firm registry containing information such as the industry and region of a firm's operations. Concretely, I use the tax filing data to estimate markups while constructing the network metrics based on the transaction data.

Despite the richness of the administrative data, my project requires additional external data. In particular, I retrieve Producer Price Indexes (PPI) from Ecuador's statistical institute (*Instituto Nacional de Estadística y Censos*, INEC), which I use as revenue and cost deflators for the industries where such PPIs are available.[6] For other industries, data from Ecuador's Central Bank (*Banco Central del Ecuador*, BCE) is used to construct a general deflator, starting from production-side GDP series at constant and current prices (Banco Central del Ecuador, 2020). Additionally, I download Penn World Table's series on price levels of capital stock in Ecuador to construct a deflator specific to assets (Feenstra et al., 2015). Lastly, I obtain crosswalk tables from United Nations (2002) to convert between different International Standard Industrial Classification (ISIC) Revisions. Specifically, INEC uses Revision 3 for the industry-specific PPIs. In contrast, SRI uses Revision 3.1 in the firm registry, which I will refer to henceforth when no precise revision is mentioned.

---

[5]Beware that throughout my work, I use the terms "establishment" and "firm" interchangeably.

[6]During the years considered in this project, INEC constructs PPIs for ISIC Revision 3.1's sections "Agriculture, hunting and forestry" (A), "Fishing" (B), "Mining and quarrying" (C), and "Manufacturing" (D).

In principle, the temporal coverage for the two primary datasets I use spans from 2008 to 2015. Lamentably, the relative earliness of the firm registry snapshot date — about halfway through 2012 — forces me to discard the latter four years of data altogether. This exclusion is necessary as I cannot accurately assign industry information critical for my work to new firms entering the space after the snapshot. After preparing the different datasets, I merge the tax filing and transaction data separately with the firm registry to immediately discard all firms without industry information.[7] Additionally, I also merge the PPIs to both datasets to get deflated amounts before continuing.

Furthermore, my analysis is restricted to a small subsample of all firms because the markup estimation procedure requires information on fixed assets. In practice, only the income tax filings of incorporated firms and unincorporated firms subject to mandatory bookkeeping, which combine for a fifth of all income tax filers, include such information. This restriction acts as a bottleneck for my analysis sample, as almost all these observations can be found in the much broader transaction data.[8] In the end, I work with around 50,000 establishments over four years, summing up to around 125,000 observations. With this, I conclude the brief data introduction and refer to Appendix A for further details on data sources and preparation processes.

3 Market Power

3.1 Formulating Market Power: The Markup

The first part of my analysis revolves around estimating market power across firms in Ecuador's formal economy over the 2008-2011 period. In the industrial organisation literature, the most widely used measure to capture market power is the markup, generally denoted as μ and defined as the difference between the output price and the marginal production cost. From a theoretical point of view, firms in a perfectly competitive market of homogeneous goods compete for market shares by underbidding each other until no profit remains — i.e., until the price is identical to the marginal cost and the markup is 1 — since

---

[7]It is noteworthy that I also deliberately exclude the state-owned oil producer *Petroecuador* given its disproportionate size compared to its little relevance to my work.

[8]The transaction data contains well above 1 million firms, even after excluding the ones with missing industry information. The observations I can match constitute only 3.5% of all valid seller-year pairs in the network. Yet, they accrue a share of all intra-firm transaction value and count just above 50%.

the price is the only discriminant between the firms' different goods. If, on the other hand, there is imperfect competition due to, for example, heterogeneous goods causing monopolies or oligopolies, firms acquire market power and can charge prices above marginal costs and still be in the market. Since I do not observe either prices or marginal costs, I have to resort to an estimation procedure to recover markups.

For this work, I decided to employ the widely used ratio estimator of De Loecker and Warzynski (2012) to compute markups for every firm and year μ_{it}.[9] This methodology allows me to avoid measuring the cost of capital and making assumptions on the returns to scale.[10] Additionally, it does not require specifying output market competition type nor modelling consumer demand. Instead, it exploits differences in timing between the firm's choice of dynamic inputs (i.e., capital), the realisation of productivity, and the selection of variable inputs (i.e., materials, labour). Concretely, firms are assumed to optimise each period's variable input mix through a cost function minimisation problem conditional on capital (set one period ahead) and productivity (observed just before the optimisation process).

Following De Loecker and Warzynski (2012) I model an economy with N firms (indexed by i) across S industries (indexed by s) over several periods T (indexed by t). Through cost minimisation — a minimal assumption on firm behaviour —, firms optimise their production goal Q_{it} every period t, subject to an output function $F_s(\cdot)$ that is shared across firms in the same industry s through common input technology parameters β_s.[11] Theoretically, $F_s(\cdot)$ and thus β_s may be allowed to change across time, but I decide to keep them constant.[12]

Given that I do not observe value-added and to avoid assuming a fixed

---

[9]Hall (1988) initially proposed the ratio estimator to compute industry-level markups. De Loecker and Warzynski (2012) expanded the methodology to estimate markups at a microeconomic level.

[10]Particularly, measuring the cost of capital would not be possible as I do not observe the typically considered paid dividends in my data. Instead, I would need to rely on estimates requiring additional assumptions.

[11]It is important to remark that a trade-off arises when deciding the narrowness of the industry categorisation here. Picking wider industries might lead to oversimplifying the problem by assuming identical coefficients for the production function to potentially different industries. On the other hand, using narrower industry definitions might increase the risk of discarding industries altogether due to insufficiently sized estimation samples.

[12]The same trade-off involved in the decision of the industry narrowness applies analogously here. Thus, I pool all years to maintain respectable sample sizes for more industries. This decision is grounded in the fact that I do not expect technology to have significantly changed over the small four-year period I analyse.

output-proportional usage of intermediate inputs, I use a gross output production function instead of the authors' default choice of a value-added production function. This decision allows me to identify markups through different variable inputs than solely labour and has the added advantage of maintaining a larger sample size.[13] Thus, gross output Q_{it} is defined as:

$$Q_{it} = F_s(\Omega_{it}, M_{it}, L_{it}, K_{it}) = \Omega_{it} F_s(M_{it}, L_{it}, K_{it}), \tag{1}$$

where Ω_{it} is the productivity parameter which I allowed to be heterogeneous across firms and time and assumed to be Hicks-neutral, while M_{it}, L_{it}, K_{it} are the material, labour, and capital inputs, respectively. Across my work, I experiment with both Cobb-Douglas and Translog production functions, the most commonly used production forms in this literature. For simplicity of notation, I will proceed with the description of the methodology using the Cobb-Douglas specification:[14]

$$Q_{it} = \Omega_{it} M_{it}^{\beta_s^M} L_{it}^{\beta_s^L} K_{it}^{\beta_s^K}, \tag{2}$$

where β_s^X represent the time-invariant technological parameters of any input X_{it} (be it M_{it}, L_{it}, or K_{it}) specific to firm i's industry s.

The firm's optimisation problem can be represented with a Lagrangian objective function $\mathcal{L}(\cdot)$:

$$\mathcal{L}(M_{it}, L_{it}, K_{it}, \Lambda_{it}) = P_{it}^M M_{it} + P_{it}^L L_{it} + P_{it}^K K_{it} - \Lambda_{it}(F_s(\cdot) - Q_{it}), \tag{3}$$

where P_{it}^X is the respective input price of any input X_{it} and Λ_{it}, the Lagrangian multiplier, is to be interpreted as the marginal cost of production.[15] Writing out the first order conditions for any input X_{it} gives the following expression:

$$\frac{\partial \mathcal{L}}{\partial X_{it}} = P_{it}^X - \Lambda_{it} \frac{\partial F_s(\cdot)}{\partial X_{it}} \overset{!}{=} 0. \tag{4}$$

---

[13]In fact, De Loecker and Warzynski (2012) generate value-added by subtracting intermediate input from operative revenue. This results in unusable negative value-added amounts when the former is greater than the latter.

[14]Note that, technically, Cobb-Douglas is nothing else than a restrictive simplification of Translog where all the parameters capturing input interactions are forced to zero.

[15]To see this, take the derivative of the objective function, i.e., the total cost to be optimised, with respect to the output constraint: $\frac{\partial \mathcal{L}(\cdot)}{\partial Q_{it}} = \Lambda_{it}$. By definition, the marginal cost is the increase in total cost given by the production of a further single unit, thus Λ_{it}.

By further rearranging terms and multiplying by X_{it}/Q_{it} on both sides I get:

$$\frac{P_{it}^X}{\Lambda_{it}}\frac{X_{it}}{Q_{it}} = \frac{\partial F_s(\cdot)}{\partial X_{it}}\frac{X_{it}}{Q_{it}} \equiv \theta_{it}^X, \tag{5}$$

where θ_{it}^X is the output elasticity of input X.[16] Finally, I multiply and divide by the output price P_{it} on the left-hand side to get:

$$\frac{P_{it}^X X_{it}}{P_{it}Q_{it}}\frac{P_{it}}{\Lambda_{it}} = \theta_{it}^X. \tag{6}$$

Recalling the definition of markup — output price over marginal cost — it is now easy to get an expression for it by moving the input share to the right side of the equation:

$$\mu_{it} \equiv \frac{P_{it}}{\Lambda_{it}} = \theta_{it}^X \left(\frac{P_{it}^X X_{it}}{P_{it}Q_{it}}\right)^{-1}. \tag{7}$$

From Equation (7), it is clear that the choice of the input X is a crucial aspect of the recovery of the markup using this methodology. Specifically, the selection lies among the available variable inputs[17] and the selected input should be as frictionless as possible such that firms can perfectly adjust their demand for it at the beginning of every period according to their expected productivity and conditional on the dynamic inputs available. Failing to fulfil this assumption will result in biased estimates for the output elasticities. De Loecker and Warzynski (2012) argue that if adjustment costs exist, they will enter the output elasticity estimates and confound the markup measurement, thereby inflating it. In Appendix B, I investigate further how to motivate this decision by looking at the within-firm variation of input demand over time, whereas, in the following subsection, I show how I recovered the partial output elasticity parameters θ^X by estimating the production functions.

---

[16]Beware that under the Cobb-Douglas specification the output elasticities θ_{it}^X will be reduced to the technology parameters β_s^X themselves, while in the Translog case they will be a function of multiple technology parameters. For example, $\theta_{it}^M = \beta_s^M + 2\beta_s^{MM}M_{it} + \beta_s^{ML}L_{it} + \beta_s^{MK}K_{it} + \beta_s^{MLK}L_{it}K_{it}$. This equation highlights the differences in the estimated output elasticities between Cobb-Douglas and Translog despite utilising the same method. Due to Translog allowing input usage to enter the equation, output elasticities will be heterogeneous across firms and time. On the other hand, Cobb-Douglas elasticities will be identical for all firms in the same industry and constant over time.

[17]Given my data, the reasonable possibilities would be either material costs, labour costs, or the sum of both. Other variables used in the literature are electricity and fuel costs. Unfortunately, I do not observe the former. At the same time, data on the latter is highly sparse, which would leave me with very few observations.

3.2 Recovering Markups: Output Elasticity Estimation

As I move to the data, a notorious issue plaguing the industrial organisation literature emerges: input and output quantities are unobserved. In their stead are expenditures on inputs (e.g., labour costs) and operative revenue, respectively. To address this issue, I will use industry-wide Producer Price Indexes (PPIs) as deflators, as is common in the literature. Nevertheless, not observing prices still risks downward biasing my output elasticity estimates and markups, according to Klette and Griliches (1996). My baseline specification is particularly susceptible because of the Translog production form. Adopting this functional form implies that firm-level price differences not captured by deflators also enter the output elasticity equations through input use intensity.[18] Fortunately, De Ridder et al. (2022) show that, despite the incorrectly rendered levels, the correlation of firm-level markups with other firm characteristics (e.g. the network metrics I constructed) is only marginally affected. The bias-ridden estimated markups should still highly correlate with the true ones.

Concretely, the proxy variable approach involves estimating production functions, partly using lagged inputs as instrumental variables, to get to the output elasticities θ_{it}^{X} of any input X. Markups can then be easily constructed by reading the share of expenditures on input X in total sales from the data as per Equation 7. First, let me take logs of Equation (2):

$$q_{it} = \omega_{it} + \beta_s^M m_{it} + \beta_s^L l_{it} + \beta_s^K k_{it}, \tag{8}$$

where lowercase variables indicate logs instead of levels. Then, let me note that in the data, I only observe $y_{it} = q_{it} + \epsilon_{it}$ rather than the isolated q_{it}, with the ϵ_{it} term capturing both measurement error as well as unexpected variations to final output compared to the production goal Q_{it} set through optimisation. Thus, I rewrite as:

$$y_{it} = \beta_s^M m_{it} + \beta_s^L l_{it} + \beta_s^K k_{it} + \omega_{it} + \epsilon_{it}. \tag{9}$$

---

[18]Using Cobb-Douglas would give me the advantage of implicitly assuming identical output elasticities across all firms within the same industry, simply equal to the respective technology parameters of the production function, thus partly avoiding the issue. Concretely, all markups would be correct up to a constant. However, this would come at a cost: any natural variation in output elasticities across firms in the same industry would break this assumption, causing the bias to pass onto the markup estimates. Practically, firms with underestimated output elasticities will have underestimated markups and vice versa, implying distortions in the distribution of markups, regardless of the chosen production function form.

Now, because the productivity term ω_{it} is unobserved, if I were to estimate the equation above with a simple OLS model, my results would suffer from the ubiquitous simultaneity or transmission bias. Firms with higher productivity will tend to produce higher output levels and, at the same time, decide to employ higher levels of inputs. Concretely, the more flexible the input, the more the technological parameters β^X will be overestimated as firms are better able to adjust their demand based on the productivity they expect to have. Coefficients for labour and capital might sometimes be even underestimated when these inputs face significant adjustment frictions (Gandhi et al., 2020).

Following the work of Levinsohn and Petrin (2003), which builds on the original ideas of Olley and Pakes (1996), I proxy for ω_{it} by first assuming that a firm's demand for materials depends on the (scaled) observed productivity conditional on the dynamic inputs:

$$m_{it} = m_t(k_{it}, \omega_{it}), \tag{10}$$

and then inverting it to get:

$$\omega_{it} = h_t(m_{it}, k_{it}), \tag{11}$$

whereby I make the implicit assumption that $m_t(\cdot)$ is strictly monotone such that invertibility is guaranteed, i.e., $h_t(\cdot) \equiv m_t^{-1}(\cdot)$. Now that I have an expression for ω_{it} only based on observables, I can back out the error term ϵ_{it} out of the observed y_{it}, which is precisely the unique goal of the first stage:

$$y_{it} = \beta_s^M m_{it} + \beta_s^L l_{it} + \beta_s^K k_{it} + h_t(m_{it}, k_{it}) + \epsilon_{it}. \tag{12}$$

Empirically, I run the first stage by non-parametrically modelling y_{it} as a third-degree polynomial expansion of all inputs, which I denote as $\phi_t(\cdot)$:

$$y_{it} = \phi_t(m_{it}, l_{it}, k_{it}) + \epsilon_{it}. \tag{13}$$

The first stage gives me a consistent estimate for the error term $\hat{\epsilon}_{it}$ as well as indirectly an estimate for the expected output:

$$\hat{q}_{it} = y_{it} - \hat{\epsilon}_{it} = \hat{\phi}_t(m_{it}, l_{it}, k_{it}). \tag{14}$$

Acknowledging the critique moved by Ackerberg et al. (2015), I refrain from

identifying any of the β_s^X coefficients in the first stage. Instead, now enters another key assumption of control function methods: firm productivity's law of motion follows a first-order Markov process:

$$\omega_{it} = g(\omega_{it-1}) + \xi_{it}, \tag{15}$$

where ξ_{it} captures unexpected innovation. In the second stage, this assumption allows me to explicitly identify ω_{it} and thus get consistent estimates for the technological parameters based on a few timing assumptions. Let me describe precisely how it works. First, I compute ω_{it} (and ω_{it-1}, analogously) from:

$$\omega_{it} = \hat{\phi}_t(m_{it}, l_{it}, k_{it}) - \beta_s^M m_{it} + \beta_s^L l_{it} + \beta_s^K k_{it}, \tag{16}$$

$$\Leftrightarrow \omega_{it} = \hat{q}_{it} - \beta_s^M m_{it} + \beta_s^L l_{it} + \beta_s^K k_{it}, \tag{17}$$

$$\Leftrightarrow \omega_{it} = \hat{q}_{it} - \vec{\beta}_s \vec{x}_{it}, \tag{18}$$

where the first substitution follows Equation (14) and the second is simply a transformation into vector notation. Thus, a firm is required to have a valid filing in two consecutive years to enter the second stage.[19] Then, I regress ω_{it} on ω_{it-1} to get an estimate for $g(\cdot)$,[20] whereby I indirectly get an estimate for the innovation term: $\hat{\xi}_{it}$.

The trick of the methodology lies in the fact that the innovation term should not be correlated with current period dynamic inputs as k_{it} since, timing-wise, they are assumed to be decided before observing current period productivity ω_{it}, where ξ_{it} materialises. At the same time, the term cannot possibly correlate with previous period variable inputs such as m_{it-1} or l_{it-1} since they take place before ξ_{it}. Still, it is allowed and expected to correlate with their respective current period amounts. Thus, the previous period variable inputs act as instruments for their current period counterparts in Equation (18) — whereby the implicit assumption that makes them suitable instruments is that

---

[19]In line with the extensive work of Olley and Pakes (1996) where they demonstrate the selection bias arising from enforcing a balanced sample, I decide to utilise an unbalanced panel instead of artificially balancing it by removing firms that are not present across all four years. Thus, appearing in two consecutive years is not a trivial matter.

[20]Note that potentially one could again use polynomial expansions of the lagged productivity term to represent $g(\cdot)$ here. Concretely, I encountered singularity issues during the matrix inversion processes performed in the Generalised Method of Moments (GMM) when I employed polynomials of a higher degree. Thus, in my case, the current period's productivity term depends on a first-degree function of the previous period's productivity term plus the unexpected innovation.

input usage over time is auto-correlated —, while current period dynamic inputs instrument for themselves. Concretely, the process just described happens by running a Generalised Method of Moments (GMM) model that iteratively optimises $\vec{\beta}_s$ by minimising the following moments:

$$\mathbb{E}\left(\xi_{it}(\vec{\beta}_s)\begin{bmatrix} m_{it-1} \\ l_{it-1} \\ k_{it} \end{bmatrix}\right) = \vec{0}, \forall s. \tag{19}$$

At this point, I would like to highlight that the entire estimation procedure is analogous when using a combined variable input — I assume that a firm's demand for variable inputs, instead of materials, depends on productivity before inverting and performing the same steps with just one variable input — as well as when using a Translog production form — although there will be more parameters to be estimated and moments to be minimised.

Since the procedure is done separately for each industry, defining what constitutes an industry is a pivotal decision. According to the most common approach in the literature, I decided to define industries based on ISIC divisions (i.e., 2-digit level). Consequently, all firms within the same ISIC division will share technology parameters or, in the Cobb-Douglas specifications, even output elasticities. This definition is a good compromise that allows me to go deeper than ISIC sections — which are too broad and might combine industries with very different technologies (e.g. manufacturing of food versus wood products) — without losing too many observations due to the exclusion of excessively small-sized industries. Concerning that, I set the exclusion threshold at 100 first-stage observations.[21] Overall, I retain 42 industries out of the 60 existing ISIC divisions.[22]

---

[21]Roughly, this threshold corresponds to about 50 observations in the second stage. For more detailed information on the exact number of observations and unique firms available in the 1st and 2nd estimation stages for each of the industries, including the excluded ones, I refer to Table C.1 in the Appendix.

[22]As a consequence of the exclusions, some of the ISIC sections get dropped altogether: "Electricity, gas, and water supply" (E), "Financial intermediation" (J), "Activities of private households as employers and undifferentiated production activities of private households" (P), and "Extraterritorial organisations and bodies" (Q). Nevertheless, their total exclusion is welcome as the first two are traditionally strongly regulated industries where prices are likely dictated or heavily influenced by the government. At the same time, the latter two are very peculiar industries which most likely do not abide by the fundamental assumption of cost-minimising behaviour.

3.3 Estimation Results: Selecting Baseline Specification

In Table 1, I present the median estimated output elasticities and corresponding markups based on the production function form and the variable input exploited to recover the markup.

Table 1: Estimation Results – Output Elasticities and Markups

Production Function Form	Variable Input	Input Cost's Revenue Share	Input's Partial Output Elasticity	Markup
Cobb-Douglas	V	0.77 (0.13)	0.95 (0.08)	1.19 (0.31)
Translog	V		0.85 (0.16)	1.08 (0.30)
Cobb-Douglas	M	0.66 (0.23)	0.66 (0.11)	0.92 (1.06)
Translog	M		0.68 (0.29)	1.10 (1.32)
Cobb-Douglas	L	0.09 (0.12)	0.34 (0.11)	3.87 (21.48)
Translog	L		0.12 (0.40)	0.61 (33.87)

Notes: By specification, this table reports firm-level medians (across all firms and years) for the input cost as share of revenue, the estimated output elasticities and the markup. The different specifications are a combination of assuming different production functions (Cobb-Douglas or Translog) and using different variable inputs to recover markups (M being material inputs, L being labour, and V being the sum of both). Standard deviations reported in parentheses as per De Loecker and Warzynski (2012).

For reference, the table also includes the respective input cost's median revenue share, which, together with the output elasticity, gives life to the markup as presented in Equation 7. Overall, the shares somewhat expectedly correlate with the output elasticities. In a setting with perfect competition in factor and product markets, any input's output elasticity is precisely mirrored by its cost revenue share. Also, according to expectations, revenue shares and output elasticities of the combined input V are roughly given by summing the respective values of M and L.

The dispersion of the markups estimated via labour inputs L is strikingly high. However, it might not be the only problem as their median levels also seem pretty unusual, ranging from the table's lowest 0.61 to the highest 3.87 median values. For the Cobb-Douglas case, markups are so high due to the unreasonably low labour cost revenue shares. In contrast, Translog markups are so low due to obtaining negative output elasticities for more than a third of the sample. For these reasons, I instantly give up on utilising markups recovered via only labour costs any further, neither herein nor in the Appendix.

Even though markups estimated using the other single input, i.e., materials M, seem more stable and reasonable, the Cobb-Douglas median level still sits

suspiciously low at 0.92.[23] Due to relatively higher variance in revenue shares and output elasticities, dispersion for materials-based markups is still higher than when using the combined input V, whose markups have a standard deviation of about 0.3 regardless of the production form.

In general, Translog production estimates are more dispersed. Multiple factors could contribute to this result. First, Translog-based partial output elasticities can differ for each firm and period, thus adding a potential source of variation.[24] Second, the GMM problem with the Translog form involves the parallel optimisation of at least twice the number of parameters than its Cobb-Douglas equivalent. In turn, more iterations are required for the estimation procedure to deliver precise results while making it less stable regarding the initial set of parameters. I will address the latter issue in more detail when I discuss my final results later.

The median markups originally reported by De Loecker and Warzynski (2012), which range from 1.17 to 1.28 across all specifications, are considerably larger than mine. From my estimates, only the Cobb-Douglas one based on the combined inputs fits in this range. At a contextual level, the study by Rodríguez-Moreno and Rochina-Barrachina (2019) is closest to mine as it is the only evidence about markups in Ecuador. When using their specification (Translog production, materials inputs only), my median markup for manufacturing firms in 2010 lies at 1.74, 0.31 upwards of theirs. At the same time, it is essential to reiterate that the *levels* of markups are not necessarily the object of interest, as De Loecker and Warzynski themselves acknowledge the possibility of them being biased, but rather their distribution, evolution, and correlation with other firm characteristics such as the network-related metrics that I construct later on.

As the ultimate goal of my work is not to compare markup estimation specifications, I pick one baseline specification with which I present a few facts about Ecuadorian markups before relating it to the network metrics later in the paper. Following the recommendations of Traina (2018), which suggests including *all* variable inputs to get a correct picture of markups, I decided to

---

[23]The median here is dragged down by the many under-unity markups in the ISIC section G industries, as one can observe in Appendix Figure D.1b.

[24]Recall that, in contrast, they will be identical across time and firms in the same industry in the Cobb-Douglas case. Even though I pool all years when estimating the production function parameters, I can still get time-varying output elasticities in the Translog case thanks to input usage intensities entering the output elasticity equations, as long as the intensities vary over time.

focus on the markups recovered via V instead of M.[25] In terms of production function, Translog assumptions better suit my objective of studying firm-level correlations with network metrics, as they allow for more heterogeneity at the observation level.[26] Furthermore, F-tests constructed to compare the Cobb-Douglas and Translog OLS models consistently picked Translog as the better-fitting model for most industries, with input interaction coefficients often being statistically significant.[27]

To gain further insight, I inspected the relationship between firm size and markups, which is not yet entirely consolidated in the literature.[28] I observed that the relationship between log firm size (measured through sales or variable costs) and markups changes wholly based on the assumed production function form. I get a positively-skewed U-shaped relationship using Translog, whereas, with Cobb-Douglas, I get a negative L-shaped relationship. There is an explanation for this latter shape. Variable cost growth goes hand-in-hand with output growth (i.e., firm size), whereas more fixed costs do not adjust quickly. Thus, it is expected, and confirmed by my data upon investigation, that smaller firms have relatively lower variable cost shares while their fixed cost make up a more significant share of their total cost. Imperfect substitutability between variable and fixed inputs creates a threshold for this trade-off beyond which one does not see variable cost shares increase. Hence, the curve takes an L shape. Thus, markups will mechanically be lower for larger firms in the Cobb-Douglas case due to them reflecting, up to an industry-specific constant, inverse variable input expenditure sale-shares. For this additional reason, I preferred to focus on Translog-derived markups to avoid the fallacy of essentially relating network-based metrics to a firm-size proxy.

---

[25]I suggest to the reader to consult Appendix B for a more in-depth analysis on the decision of the input variable.

[26]De Ridder et al. (2022) warn against using Cobb-Douglas production function as, in their own words, it "reduce[s] the informativeness of markup estimates". De Loecker and Warzynski (2012) also use Translog for the same reasons, although they recover markups through labour inputs only as they operate with a value-added production function.
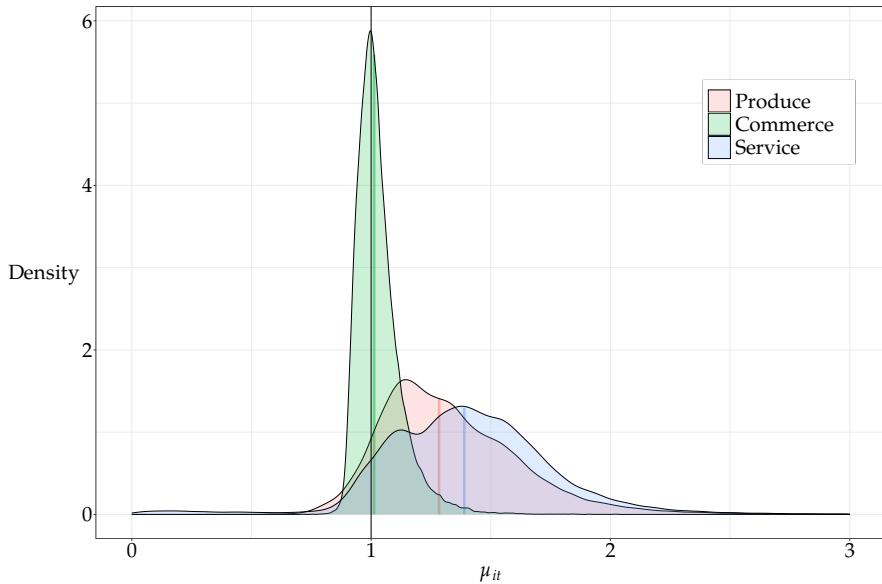
[27]Interestingly, the only six industries for which the test cannot be rejected at the 95% confidence level are among the seven industries with the lowest number of observations. Thus, the cause is likely to be the overly penalising reduction in degrees of freedom that Translog's additional coefficients cause in these small samples.

[28]For example, despite the consensus that larger firms charge higher markups (Autor et al., 2020; De Loecker et al., 2020), a recent report by Di Mauro et al. (2023) points in the opposite direction.

3.4 Some Facts About Markups in Ecuador

An obvious filter through which to look at the markups is that of industrial heterogeneity. The high number of industries (42) for which I estimate markups would make any attempt at concise and clear visualisation fail. Thus, I split them instead into three major industry groups based on the nature of their operations. The fact that ISIC section G, i.e., activities focusing on *commerce*, alone makes up slightly more than half of my observations in the markup estimation sample gives me sufficient reasons to isolate its results. It also splits the economy well. Earlier sections, from A to F, capture all industries that *produce* in the strict sense. Later ones, H to Q — or, in my concrete case, only to O —, are those involved in providing some *service*. The share of observations for these two groups is about one-quarter and one-fifth, respectively.

Figure 1: Firm-Level Markup Distribution by Industry Group

Notes: The red, green, and blue areas correspond to the density of firm-level markups μ_{it} for the three industry groups comprising the ISIC Revision 3.1 industries A01 to F45, G50 to G52, and H55 to Q99, respectively. Colour-coded line segments represent group medians. A black vertical line denotes equality between prices and marginal costs, i.e., $\mu_{it} = 1$. All years are pooled together. Though considered for the density computation, markups below 0 and above 3 were trimmed from the plot.

Figure 1 plots the density distributions of my baseline firm-level markup estimates by industry group. While the first and third industry groups' variances seem similar, the distribution for *service* firms is slightly upward shifted, exemplified by the higher median at 1.39. On the other hand, *produce* firms' distribution has a lower but still respectable median of 1.28 and seems to be a

little right-skewed. The picture looks quite different for the *commerce* industry group. The median sits very low at 1.01 (narrowly below 1 for two of the three industries within the group), and firms concentrate much more on unity. One might naively think its low spread is only due to the smaller group size and within-group industry heterogeneity, but this is untrue. Replicating this figure by ISIC sections still has section G as the group with the lowest variance by some margin, despite several ISIC sections consisting of a single industry (A, B, F, H, L, M, N).

Overall, this figure highlights how homogeneous *commerce* firms are in terms of markups while also showing how firms in these industries have, on average, lower markups than firms in other industries. Both make sense for a couple of reasons. Firstly, the three industries that constitute ISIC section G are the three largest industries in my data. One would indeed expect that increased competition prevents firms from quickly raising markups. Thus, I expect them to concentrate. Secondly, these *commerce* industries are involved in buying-and-selling activities that supposedly generate small intrinsic value-added. Thus, I expect them to focus on moving large quantities to achieve higher revenues rather than charging high markups, which should be closer to unity.[29] Appendix Figure D.1 shows the robustness of these results across different specifications.
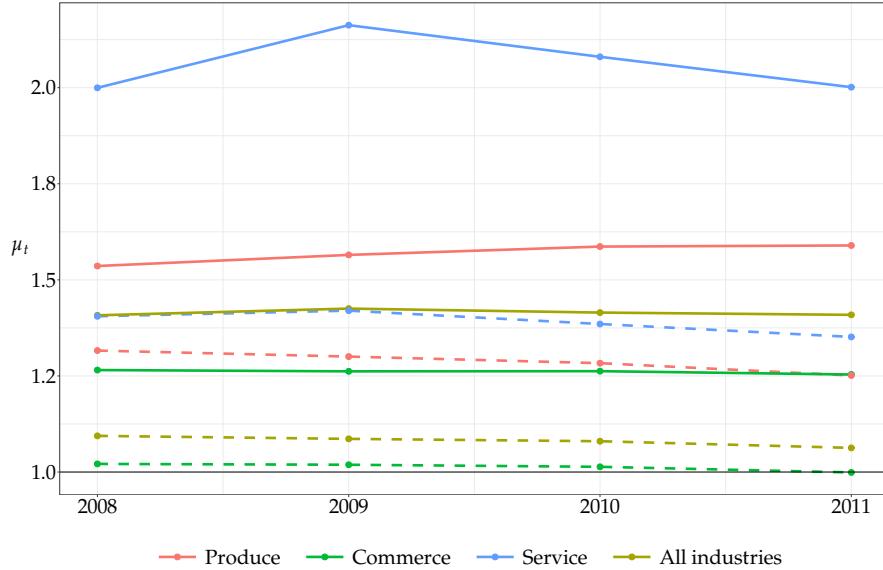
An additional lens through which to look at markups is their evolution over time. Unfortunately, the limited available time window makes it hard to discuss proper trends. Nevertheless, it is still something worth inspecting, especially considering the timing of the Global Financial Crisis that made Ecuador's GDP growth fluctuate from 6.4% in 2008 to 0.6% in the following year, and back up to 7.9% by 2011 (The World Bank, 2024).

Figure 2 depicts the evolution of markups along three dimensions: aggregate ($\bar{\mu}_t$), median ($\ddot{\mu}_t$), and dispersion measured by the standard deviation of firm-level markups ($\sigma_{\mu_{it}}$). Aggregate markups in Subfigure 2a seem to have stayed relatively stable across the four years, which is consistent with a pattern that Diez et al. (2018) observed for emerging economies over a much wider time frame. The only exception seems to be the aggregate markup for

---
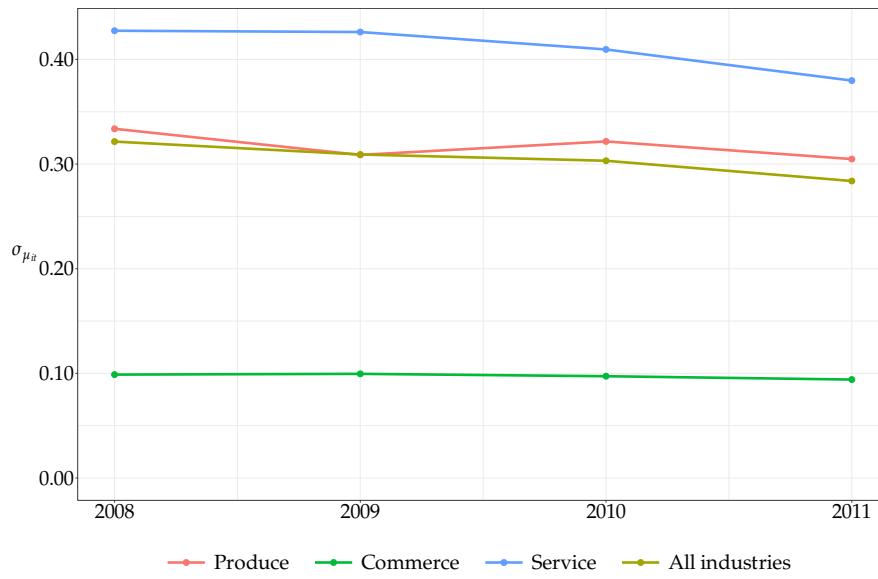
[29]This suspicion is confirmed by the data. The top three selling industries by transaction count are the three industries in ISIC section G by quite some margin: the third (G52) has double the number of transactions as the fourth. The same holds for the top three buying industries, though the distance to the fourth here is smaller. I considered the entire cleaned transactional data (barring missing industry firms) and not only the firms in my markup estimation sample. Doing so would have made this result rather mechanical since these industries comprise more than half of my sample.

Figure 2: Evolution of Markups by Industry Group

(a) Aggregate Markup (—) and Median Firm-Level Markup (- -)

(b) Dispersion of Firm-Level Markups

Notes: The red, green, and blue lines picture the specific markup trends for the three industry groups comprising the ISIC Revision 3.1 industries A01 to F45, G50 to G52, and H55 to Q99, respectively. The yellow lines combine all industries. In the above Subfigure, aggregate markups $\bar{\mu}_t$ (solid line) are constructed by weighting individual markups with the yearly variable input cost share over the whole sample, i.e., $\bar{\mu}_t = \sum_i \frac{V_{it}}{\sum_j V_{jt}} \mu_{it}^V$, instead of the more common sales shares, following the critique raised by Edmond et al. (2023). A black horizontal line denotes equality between prices and marginal costs, i.e., $\mu_t = 1$.

the *produce* industry group, for which the aggregate markup grows over time. Additionally, *service* industries registered a somewhat puzzling and abrupt in-

crease in 2009 (spurred by industries I61 and I64) before gradually reverting to the 2008 level.

On the other hand, median markups — also depicted in Subfigure 2a) — follow a clear downward evolution across all industry groups. The *commerce* industry group has perhaps the most moderate slope among them, while the other two industry groups' medians run relatively parallel. The *service* group displays the widest gap between aggregate and median markup by a large margin, implying that some high-markup firms with prominent input cost shares pull the aggregate markup way up. It is interesting to see that the median and aggregate markup trends are opposite for the *produce* industries. In general, the fact that median markups go down while aggregate markups stay constant or even increase points towards offsets by above-median-markup firms. They could be raising markups or gaining weight due to, for example, low-markup firms scaling down on costs following the crisis.

Like median markups, the dispersion of firm-level markups declines over time, both for each industry group and across the whole economy, as Subfigure 2b markedly shows. Beware that this is not a mechanical result. It could be possible for each industry group's dispersion to decline and the overall dispersion to increase. The *commerce* industry group displays the slowest decline, but this is reasonable given its already shallow initial level.

One can get a different perspective on markups' evolution over time by shifting the focus towards intra-industry firm ranking changes. One way to think about it is to divide each industry's markup distribution into deciles and look at how firms change deciles over time. By doing this, one discovers an outstanding dynamism. Overall, from one year to the next, there are ranking exchanges from firms in any decile of their industry's markup distribution to any other decile. I will report some essential numbers as trying to visualise this proved cumbersome. On average, across all years and industries, roughly 19% of the firms move downwards in their own industry's markup distribution by one decile, 6% by more than one. Simultaneously, 15% of the firms move up by one decile and 9% by more than one. The remaining 51% of the firms stay in the same decile. These shares are substantially similar over time and across different industry groups.

The general results presented here are somewhat robust to other markup specifications despite the levels of the curves being substantially different, as one can see in Figures D.2 and D.3. To summarise, there seems to be a majority convergence toward lower markups, given that the dispersion and the median

are simultaneously going down. Despite lots of intra-industrial mobility in all directions, a few firms are either gaining weight or raising markups (or both) so that the aggregate markup does not decline.

4   Network Strength

4.1   Constructing Network-Related Metrics

For the second part of my analysis, I focus on conceptualising and constructing different network-based metrics. These measures aim to capture various facets of the broad idea of network strength that could elicit market power in the context of production networks.

Preface

Admittedly, when I started brainstorming for this project, I leaned towards using already existing network metrics, some of which have been greatly overlooked in the economics literature. The network science field offers a wide range of centrality metrics besides the trivial in- and out-degree and centralities. One such example is betweenness centrality, according to which the most central firms are those through which the highest number of shortest paths between any two firms goes through them. Intuitively, this concept could be exciting when investigating, for example, the importance of a firm within value chains. Moreover, the network science field also proposes other metrics that characterise the network besides centrality, such as assortativity, clustering or reciprocity. Lamentably, the computation of many of these measures is prohibitive for a network that covers the entire formal economy of Ecuador, despite the low average density of around 1.13×10^{-5}. Thus, I construct my metrics instead, drawing inspiration from existing ones whenever possible. These can be split into five categories: importance for buyer, relationship quality, competition, market targeting, and diversification.

Before moving on to the description of the network metrics, I deem it necessary to stress that only firms' self-reported domestic *purchase* transactions, and not sales, are available to me when constructing the production network. Thus, I reconstruct sales by simply inverting the direction of the reported purchase transactions. This procedure conceals a couple of caveats. First, I do not observe sales to final domestic consumers, or at least not those where the final

consumer is an individual rather than a firm, nor any sales, intermediate or final, to foreign buyers. This can be problematic since the estimated markups rely on sales that cannot be distinguished into intermediate and final. Even if I could, as with domestic versus foreign revenues, I still could not estimate separate markups unless I make assumptions on the proportions of costs used for intermediate versus final sales or domestic versus foreign sales. I discuss the implications of this issue after presenting the regression results.

The second caveat concerns the mere fact that I am inverting the direction of purchases to get sales. One might naively think that this would lead to having significantly less accurate information on sales than if I observed self-reported sales directly. The opposite is likely true, considering that firms must file these purchase annexes to deduct the Value-Added Tax (VAT) paid on purchases from the one owed to the tax authorities due to sales.[30] Pomeranz (2015) highlights how this self-enforcement mechanism reduces transaction misreporting due to incentivising firms to create paper trails for their purchases.[31] Additionally, I would not be observing the hundreds of thousands of sellers that are too tiny to self-report sales but still generate competition. In fact, except for firms without industry information, I consider all the firms in the transaction data, regardless of whether they enter the markup estimation sample. The reasoning behind this is wanting to capture the production network with as much fidelity as possible to reality.

Importance for buyers

Coming to the metrics, one way to define a strong network selling position is by examining the importance of the seller's products for its buyers. In particular, significance can then be assessed within a set of comparable products or across the whole range of a firm's inputs. Let me give a concrete example of the former. If buyer j purchases 100 USD worth of goods from industry $s(i)$, seller i's industry, and 30 USD of those are purchased precisely from seller i, then i's *within industry importance* for j is 0.3. I can do this exercise for every combination of buyer and seller in the network. The implicit assumption here is thus that each seller competes directly with all the other sellers in the

---

[30]Given that VAT is levied on the margin (i.e., value-added) generated at each stage of production (from raw material to final consumer sales), the amount of VAT that each firm eventually has to pay to the Government will be a difference between the VAT amounts charged on its sales minus the VAT amounts they spend on their purchases.

[31]In contrast, if firms self-reported their sales, they would have a natural incentive to underreport to pay less VAT or not even charge it to their customers.

economy that operate in the same industry. Formally, I define this pairwise measure as:

$$\textit{Within Industry Importance}_{ij,t} = \frac{X_{ij,t}}{\sum_{k \in s(i)} X_{kj,t}},\qquad(20)$$

where $X_{ij,t}$ is the sum of all sales from firm i to firm j in a given year t, and I divide by the total purchases made by firm j from all firms k in firm i's industry $s(i)$ (including firm i itself) that are active in year t. In the future, I will leave out the time subscript t for simplicity of notation despite all the measures being defined on a year-to-year basis.

Applying a simple adjustment to the measure I just defined allows me to measure importance across *all* buyer's inputs. Concretely, I scale down the *within industry importance* of firm i for j by how vital i's industry is for j's overall operations. Here, again, I use transaction value to quantify the significance of each industry. Continuing with the previous example, if buyer j purchases 200 USD worth of goods from only two industries, 100 USD from $s(i)$ and 100 from another, each industry will weigh 50% for buyer j. Then, the *across industry importance* of i for j is scaled down to 0.15 (0.3×0.5). Formally, I define this scaled measure as:

$$\textit{Across Industry Importance}_{ij} = \frac{X_{ij}}{\sum_{k \in s(i)} X_{kj}} \times \frac{\sum_{k \in s(i)} X_{kj}}{\sum_{k} X_{kj}} = \frac{X_{ij}}{\sum_{k} X_{kj}},\qquad(21)$$

where $\sum_{k} X_{kj}$ captures all purchases of firm j across all its suppliers k. It is easy to see that this measure is very similar to the a_{ij} coefficients of what in the literature is called the input-output matrix. These coefficients also capture the importance of inputs from i as a share of j's costs, though they subtly differ. Usually $\sum_{k} a_{kj} = 1 - (\beta^{X_1} + \cdots + \beta^{X_N})$, where constant returns to scale are assumed and $\beta^{X_1}, \ldots, \beta^{X_N}$ are the output elasticities of other potential inputs besides intermediates (e.g. labour, capital). In this case, I have $\sum_{k} a_{kj} = 1$ such that a_{ij} represents the importance of firm i's inputs as a share of firm j total purchases, rather than total costs.[32]

Relationship quality

Another interesting way to think of a strong network position relates to the quality of a firm's relationships with its partners. In this context, one applic-

---

[32]Please beware that total purchases in the network should not be interpreted as intermediate input purchases. Their nature could be assets, intermediate inputs, or supplies. I am not able to distinguish them.

able concept is that of reciprocity. Intuitively, if a firm i generally sells to firms j from which it does not buy back anything, firm i could charge ceteris paribus higher markups compared to other sellers with more bilateral interests. In network science, the reciprocity of a node is computed as the number of edges that go in both directions — in my context, this would mean that firm i has a relationship with firm j both as supplier and customer — as a share of the total number of edges attached to that node. High values then have a clear interpretation, but low values are ambiguous. Technically, low reciprocity could be due to either selling to many firms while buying from a minority or buying from many firms but selling to few of them. Formally, I define *seller reciprocity* as:

$$\text{Seller Reciprocity}_{ij} = \begin{cases} 1 & \text{if } X_{ij} \neq 0 \ \wedge \ X_{ji} \neq 0, \\ 0 & \text{if } X_{ij} \neq 0 \ \wedge \ X_{ji} = 0, \\ NA & \text{if } X_{ij} = 0, \end{cases} \tag{22}$$

where X_{ji} represents purchases done by firm i from firm j, i.e., sales of firm j to firm i. Further, given that I am more concerned with the seller side of things I condition reciprocity on selling, such that the value will be not assigned, NA, in the cases where firm i does not sell to firm j.

The metric just created measures sellers' reciprocity on the margins, i.e., whether a seller has a bilateral relationship with its buyer. However, looking at the effect on a more continuous range of reciprocity is potentially attractive. For this reason, I construct another measure that accounts for how much the reciprocity favours the seller, conditional on reciprocity existing. Formally, I define it as:

$$\text{Reciprocity Degree}_{ij} = \begin{cases} \frac{X_{ij}}{X_{ij}+X_{ji}} & \text{if } X_{ij} \neq 0 \ \wedge \ X_{ji} \neq 0, \\ NA & \text{if } X_{ij} = 0 \ \vee \ X_{ji} = 0. \end{cases} \tag{23}$$

Note that, by construction, this measure can never be exactly 0 or 1.

A business relationship's quality can also be measured through the lens of *selling frequency*. On the one hand, firms that engage in frequent interactions with the same partners might slowly increase their prices, assuming that partners are reluctant to switch suppliers, perhaps due to the comfort of routine processes or the costs of fostering a new business relationship.[33] On the other

---

[33]This is the same logic for which, for example, one might keep paying an overpriced mobile plan from established providers despite there being better offers from other competitors, as long as the price difference is not excessively large.

hand, (new) sellers with average low-frequency relationships might be trying to attract new customers via lower prices, possibly even below marginal cost. It is thus hard for me to predict how such a measure would relate to markups. Formally, I define:

$$\textit{Selling Frequency}_{ij} = \#x_{ij}, \tag{24}$$

where x_{ij} represents a single sale from firm i to firm j such that $X_{ij} = \sum x_{ij}$, for each year t. Here, I use the $\#$ symbol in the set cardinality sense, i.e., I am counting the number of single operations x_{ij} occurring between seller i and buyer j in a given year t.

Competition

Yet another way to think about network strength is to look at the level of competition. The ubiquitous industry-concentration Herfindahl-Hirschman Index (HHI) is probably the most common way to measure competition. Unfortunately, despite it sometimes being a good measure to compare industries, it is not helpful when wanting to compare the sellers within them, as my goal is. One superficial yet potentially effective way to formalise competition, or lack thereof, is to check how often a seller i is the unique supplier in its industry for any buyer j. Formally, I define the dummy *unique seller* with the help of a previous metric as:

$$\textit{Unique Seller}_{ij} = \begin{cases} 1 & \text{if } \textit{Within Industry Importance}_{ij,t} = 1, \\ 0 & \text{if } \textit{Within Industry Importance}_{ij,t} \neq 1. \end{cases} \tag{25}$$

In addition, one might also be interested in a more continuous definition of competition. For this reason, I compute a measure that tells me, for each industry, how many sellers k buyer j interacts with as a share of all potential sellers (i.e., all firms in the industry $N_{k \in s}$). Formally, I define the *competition intensity* at the buyer–seller's industry level as:

$$\textit{Competition Intensity}_{sj} = \frac{\sum_{k \in s} \#X_{ij}}{N_{k \in s}}, \tag{26}$$

Market targeting

Another network property I am interested in is a seller's market targeting. Concretely, I begin by examining two dimensions: geographical targeting and

industrial targeting. I create a dummy for the former to distinguish buyers in the same province as the seller from those in other provinces. A priori, I would expect that having a *local focus* should correlate with higher markups, possibly due to reduced distribution costs and potentially stronger local brand loyalty. Formally:

$$Local\ Focus_{ij} = \begin{cases} 1 & \text{if } p(i) = p(j), \\ 0 & \text{if } p(i) \neq p(j), \end{cases} \tag{27}$$

where $p(\cdot)$ is a function returning the province of the firm.

Conversely, I create a dummy that captures whether the buyer is in a different industry than the seller. The rationale is that having a *vertical focus*, i.e., selling outside one's industry, might confer more market power to the seller due to, for example, asymmetrical knowledge about the product. Formally:

$$Vertical\ Focus_{ij} = \begin{cases} 1 & \text{if } s(i) \neq s(j), \\ 0 & \text{if } s(i) = s(j), \end{cases} \tag{28}$$

where $s(\cdot)$ returns the industry, as seen earlier.

Related to the concept of monopsony, I would expect firms that do not need to rely too much on a handful of buyers to have higher market power. In extreme cases, if a firm sells its entire output to a single buyer, it will have little to no market power. A metric trying to capture this phenomenon might also indirectly measure potential cases of vertical integration, i.e., when a firm owns its supplier and tailors its production to its needs. I construct a metric conceptually similar to the first one presented but mirrored. Concretely, I want to measure how each buyer is essential to a seller based on the share of the seller's transaction value with the buyer's industry it purchases. Formally:

$$Buyer\ Specialisation_{ij} = \frac{X_{ij}}{\sum_{k \in s(j)} X_{ik}}. \tag{29}$$

Diversification

Finally, I am also interested in examining plain diversification. Generally, I expect firms that diversify more to hold greater market power. To this goal, I create three further metrics which do not warrant a formal definition due to self-explanatory name: *Unique Buyer Count*, *Unique Industry Count*, and *Unique Province Count*. In contrast to all previous metrics, these three are constructed

already at the seller level and thus need no averaging across buyers, which I discuss now.

Averaging at the seller level

Until now, most of the metrics I formalised, besides the last defined unique counts, are specified at the seller-buyer level. However, I require network characteristics at the seller level to relate them to markups. Consequently, I aggregate by taking averages across buyers. Here, there are a few caveats to consider. The first relates to the weighting. Naturally, having a *within industry importance* of 1 with a buyer to which a seller only supplies 10 USD worth of goods might not be the same as the same value while supplying 1,000 times as much. For this reason, I create weighted and weighted (by selling value) versions of all the measures. The second caveat concerns that the network only captures connections between sellers and buyers that concretely transact with one another. Hence, all my measures are conditional on actual selling. Despite the intriguing potential and relevancy of considering unrealised transactions, its concrete implementation is far from easy due to the difficulty of knowing which firms could be potential buyers. My attempts at doing so produced unrealistic and uninterpretable results, so I refrained from pursuing this avenue and focused instead on metrics conditional on selling.
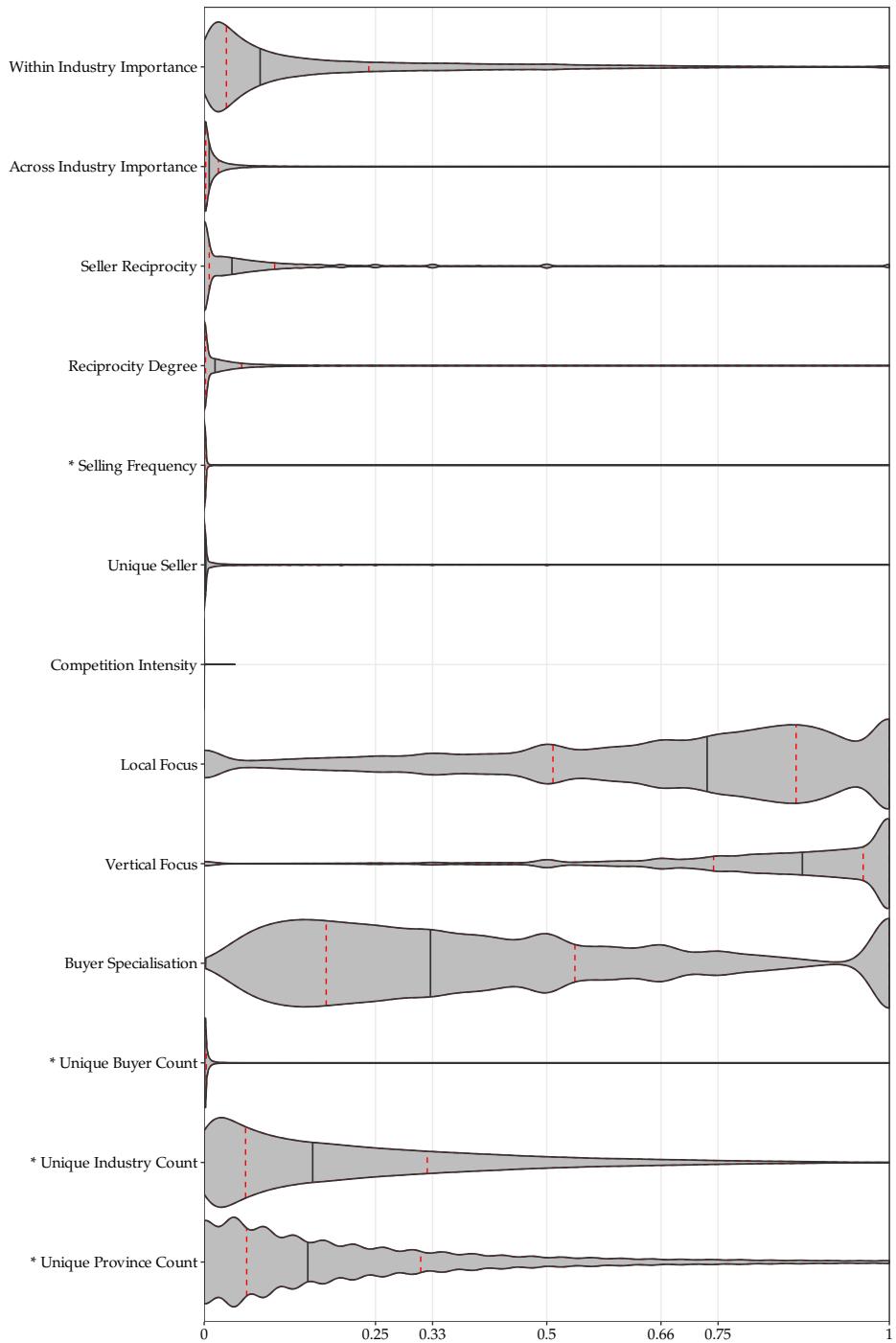
4.2 Some Facts About the Ecuadorian Production Network

I believe presenting facts based on the whole sample of sellers in the network for which I built metrics would bear little interest, as I cannot consider most of those firms when relating the network strength metrics to markups. Thus, I focus on the above 97% of the estimation sample firm-year pairs that match in the network data as sellers.[34]

To begin with, I plot the distributions of the variables to get an idea of what I am working with. For reasons of space, Figure 3 plots only the distributions of my unweighted network metrics, while the distributions of the weighted ones can be found in Appendix Figure E.1. Except for *buyer specialisation*, whose weighting makes higher values much more common, the overall picture

---

[34]This matched subsample constitutes only 3.5% of all seller-year pairs in the network. Yet, they accrue a share of all intra-firm transaction value and count just above 50%. To check the matching shares of individual industries with the network refer to Appendix Table C.1. On the other hand, to see a comparison between the matched sample and the remaining sellers in the network refer to Appendix E.

Figure 3: Distribution of Unweighted Network Metrics

Notes: Violin plots display the mirrored distribution of each network metric. For each distribution, a solid black line marks the median, while dashed red lines represent the 25th and 75th percentiles, respectively. For visualisation purposes, metrics marked with an asterisk (*) have been minimum-maximum-scaled to force them into the remaining metrics' natural 0-1 range.

is the same between weighted and unweighted variables. For some metrics, weighting makes the distributions slightly less skewed; the opposite is true for others.
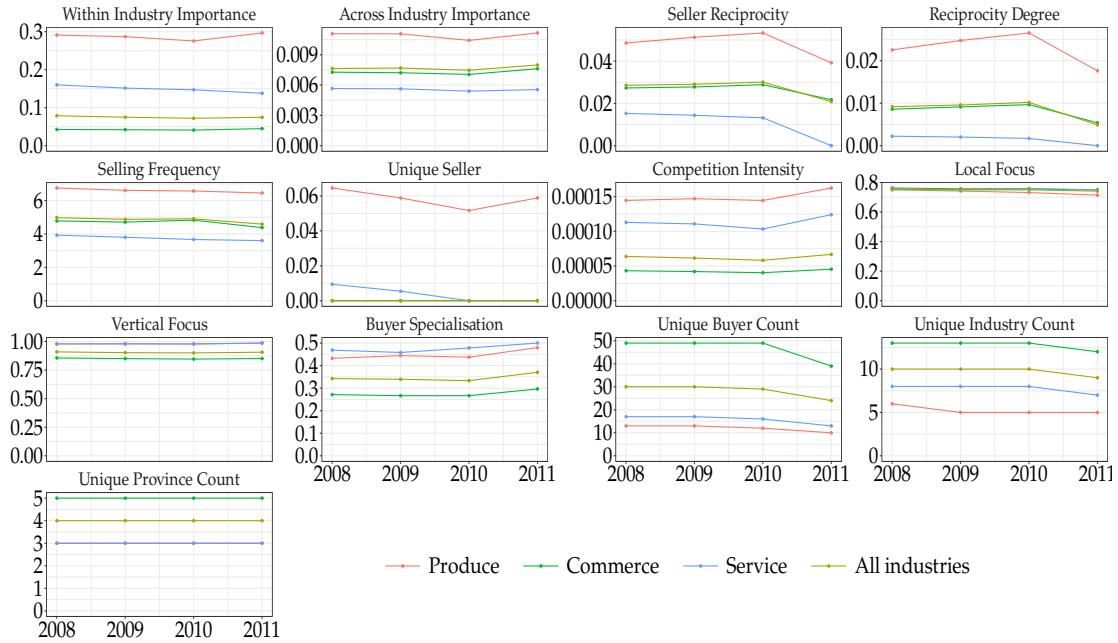
The metrics in the first three categories —- importance for buyers, relationship quality, and competition — display (very) skewed distributions concentrating around 0. Conversely, most sellers seem to have a *local* and *vertical focus* regarding market targeting, thus showing left-skewed distributions. Instead, the *buyer specialisation* metric is, relative to other metrics, more uniformly distributed along the whole 0 to 1 range, though it appears bimodal. Indeed, density decreases as the value increases until around 0.9 before suddenly growing again around 1. This bunching effect around the value of 1 could be due to firms generally serving few customers or vertical integration. Regarding diversification metrics, the unique counts show the distribution that one would expect: right-skewed. Particularly in the distributions of binary variables, one can easily spot bunching around basic fractions such as one-third or two-thirds, likely due to the firms with a small customer base.

Now, looking into industry heterogeneity and evolution over time is also necessary. Considering that appropriately aggregating the network metrics I constructed might be more challenging than aggregating markups, as the network metrics themselves are aggregations across all the buyers of a seller, I opt instead to display medians. Figure 4 illustrates the evolution of the median values of each metric by industry group.[35] Overall, there seems to be almost no variation over time. A few metrics display trend changes for 2011, but I reckon this is due to the relatively high number of purchase annexes filed that year compared to the previous years. Thus, I focus on industry heterogeneity in what follows.

Generally, firms involved in more strictly interpreted producing activities display the highest medians across several network metrics expected to imply higher market power. However, when it comes to diversification metrics, they display the lowest means in all three unique counts, which might, in part, explain the previous finding. They also show a somewhat high median *buyer specialisation* coefficient. Thus, firms in the *produce* industry group seem somewhat more specialised by providing goods to a relatively restricted sample of customers, for which they are relatively more valuable. Nevertheless, they

---

[35]In contrast to what was done for the markups in Figure 2, I save some space by refraining from illustrating the evolution of dispersion, given that, essentially, there are no changes over time.

Figure 4: Aggregate Network Metrics

Notes: The red, green, and blue lines picture the specific markup trends for the three industry groups comprising the ISIC Revision 3.1 industries A01 to F45, G50 to G52, and H55 to Q99, respectively. The yellow lines combine all industries. Note that for *vertical focus* and *unique province count*, the blue and red lines perfectly overlap. The same holds for *unique seller* between the yellow and green lines.

also display relatively higher levels in the reciprocity metrics and *competition intensity*, which I do not hypothesise to confer higher market power.
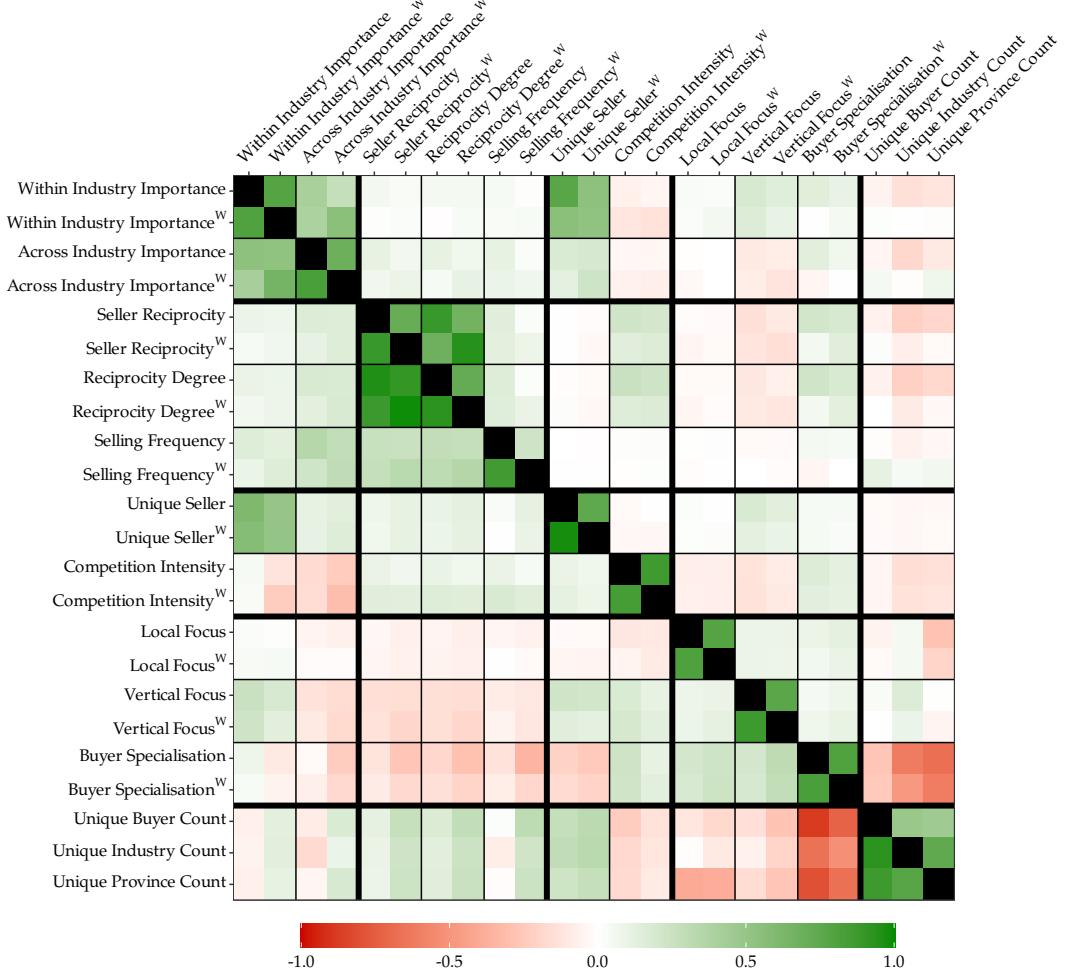
On the other side of the spectrum are *commerce* firms, which, by the nature of their operations, are expected to interact with many customers rather than focus on a few selected. This belief is reflected by their low *within industry importance*, *unique seller*, *buyer specialisation* medians, and their high diversification numbers given by the unique counts. *Service* firms lie between the other two groups in most metrics. Notable exceptions are in their lower reciprocity values, which I think results from the nature of their operations, which do not require as many inputs as other industries.

The figure conveys a clear message. Similarly to what was witnessed with markups and according to my expectations, there is indeed industry heterogeneity across industries, or at least across industry groups. Perhaps the only exceptions to this fact are the market targeting variables, where industry groups barely differ. Therefore, it is crucial to consider it when investigating the relationship between market power and network strength. On a side note, it is worth highlighting that a firm comparison between industries along the

competition intensity measure is unjustified since groups mechanically cluster. This is due to the metric's common denominator for all sellers in a given industry.

Finally, given the high number of variables created, a natural question arises: how do these metrics relate? Figure 5 answers the question by provid-

Figure 5: Correlation Matrix of Constructed Network Metrics

Notes: For any combination of two variables, the respective lower-triangle cell depicts the Spearman rank-correlation coefficient ρ, whereas the upper-triangle cells depict the Pearson linear correlation coefficient r. Vertical and horizontal thin lines group different specifications of the same variable, while thick lines separate the different categories of variables. Note that unique count variables do not have a corresponding weighted version as they are defined already at the seller level and are thus not averages. Green and red hues imply positive and negative correlations, respectively, with colour intensity representing strength. The trivial diagonal cells are coloured black to improve readability.

ing a correlation matrix between all variables, including both weighted and unweighted versions. Because a standard correlation matrix incorporates re-

dundant information due to mirrored cells along the diagonal, I show two different correlation measures on each side of the diagonals instead. The upper triangle reports the Pearson linear correlation coefficient r, while the lower triangle co Spearman rank-correlation coefficient ρ. The upper triangle seems mildly lighter overall than the lower one, indicating a higher correlation in terms of ranking than what linearity would imply. In general, I believe it is better to focus on the lower triangle since assuming the relationship among the metrics to be linear might be debatable.

Correlations are positive and decisive in the cells just around the diagonal. The implication is that unweighted and weighted metrics are reasonably similar, which is sensible to expect. Coefficients are also high between pairs of variables constructed very similarly or even on top of each other, such as the two importance for buyer metrics, *within industry importance* and *unique seller*, and the two reciprocity metrics.

Despite the figure being rather green, there are also several negative correlations. Some are expected, such as *competition intensity* with the importance metrics. Indeed, when facing more competition, it should be more challenging to achieve high-importance scores. Others are perhaps more puzzling, such as the market targeting category with the reciprocity metrics. For example, I would have imagined that being highly focused on a few buyers would make it likelier to engage in reciprocal transactions. Instead, the direction seems the opposite, possibly pointing to a unilateral vertical integration effect again. It is worth pointing out that correlations are indeed lightly positive when considering the Pearson coefficient. *Vertical focus* correlating positively with *within industry importance* but negatively with the following (in order) four metrics can be attributed to the fact that each industry's highest share of trade is with itself. Overall, the strongest negative correlations occur between the *buyer specialisation* metric and the unique counts, which is intuitive.

5   Network Origins of Market Power

5.1   Results

My work's leading innovation and goal is to investigate the relationship between market power and network strength. I perform this analysis by combining the

markups estimated in the first part of the paper[36] with firms' topological characteristics constructed in the second part, which are grounded in their network positioning as sellers. Due to the originality of my work, I do not set out to uncover any causal relationship. Instead, I rely on simple ordinary least squares (OLS) regressions to find out whether there is any significant correlation suggesting a link between the two.

Generally, the network metrics I constructed do not necessarily allow for a straightforward understanding of the results. The standard one-unit-increase interpretation is unrealistic because most are bounded between 0 and 1. Thus, it would imply moving between the two extremes, which, in reality, rarely happens. For this reason, I interpret most of my results using the interquartile range (IQR) as the basis for the relationships' magnitudes. Exceptionally, I take logs for variables not naturally bounded by 0 and 1 and interpret them with percentage increases. As always with linear regressions, the underlying assumption is that the effect is linear and thus identical along the entire range of the variables' distributions.

To begin with, I would like to see how the industry- and time-fixed effects, but also the input controls recommended by De Loecker and Warzynski (2012) and firm-fixed effects, shape the relationship at hand. In Table 2, I deliberately

Table 2: Regression Results Decomposition

	IQR	(1)	(2)	(3)	(4)	(5)
Within Industry Importance	0.20	0.2832*** (0.0082)	0.2825*** (0.0082)	0.0616*** (0.0072)	0.0440*** (0.0074)	-0.0051 (0.0056)
Year FE		X	4	4	4	4
Industry FE		X	X	42	42	42
Input usage controls		X	X	X	✓	✓
Firm FE		X	X	X	X	49,726
Observations		126,902	126,902	126,902	126,902	126,902
Adjusted R-squared		0.04	0.04	0.63	0.65	0.94

Notes: This table reports regression results of the unweighted within industry importance on markups. The relationship is decomposed by incrementally increasing controls. For fixed effects, I report the number of groups. Standard errors clustered at the firm level reported in parentheses. Significance levels given by: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

pick the first constructed metric, the unweighted *within industry importance,*

---

[36]Note that there are 240 observations with negative markups in my baseline markup specification. Since markups cannot be negative, I assume they are wrongly estimated and do not consider these observations for the results in this section.

and report the OLS results of regressing it on markups while incrementally adding controls to illustrate this idea.[37]

The first column shows the raw relationship across all firms in all industries and years. I find that moving from the first to the third quartile of the *within industry importance* metric, represented by an increase of 0.2 in the measure, is associated with an increase in the markup of about (0.2 × 0.2832 × 100 =) 5.7 percentage points. Recall that my median markup of 1.08 implies an 8% price margin over marginal cost. Thus, the effect's magnitude is considerable, representing a 70% increase in the margin around the median. Furthermore, the effect is precisely estimated, although it is worth mentioning that the uncertainty regarding the markup estimation itself is not considered here.

Adding year-fixed effects in column (2) excludes time trends and has virtually no impact on the coefficient. This is somewhat expected, given the small time variation detected in markups and network-based metrics. Considering that this finding is consistent across all network metrics, I will henceforth always include time controls without further discussing the implications of the decision.

On the contrary, additionally controlling for a firm's industry of operation, as I do in column (3), dramatically changes the relationship. Though still precisely estimated, the effect is now around one-fifth in magnitude. Hence, a big part of the effect reported before stems from industry heterogeneity.[38] Specifically, industries with higher average *within industry importance* values also charge higher markups on average and vice versa, resulting in an overestimated coefficient. This fact is mainly driven by the three *commerce* industries, which, besides being the largest three industries, have among the lowest average *within industry importance* scores and markups. The coefficient in the third column implies that a firm in the first quartile of the importance metric has, on average and compared to an industry peer in the third quartile, an expected markup smaller by around one percentage point.

Adding the input usage controls in column (4) as suggested by De Loecker and Warzynski (2012) when using Translog markups further reduces the coef-

---

[37]Appendix Tables F.1 and F.2 visualise this exercise for all the other unweighted and weighted metrics, respectively.

[38]Including industry-fixed effects has diverse effects across the different metrics. In some cases, adding industry fixed-effect flips the coefficient's sign in both directions.

ficient's magnitude, if ever slightly.[39] Finally, controlling for firm fixed effects in the last column nullifies the result completely, which becomes extremely small and imprecisely estimated. Thus, unobserved idiosyncrasies at the firm level seem influential in explaining the markups. After controlling for such effects, there is not enough variation left in the markups for the *within industry importance* metric to explain.

Next, I present the regression results for all the constructed variables. Given what I just learned, I consistently report results that include time- and industry-fixed effects, as well as the suggested input usage controls. I instead distinguish between specifications with and without firm-fixed effects. Table 3 reports these results subdivided by network strength category. Here, I considered the unweighted metrics. Appendix Table F.3 reports qualitatively similar results for the weighted versions.

Besides *competition intensity*, all coefficients in column (1), i.e., without firm-fixed effects, are statistically highly significant. In terms of magnitudes, the effects are not so substantial. Among the highest are the ones in the market targeting category. Sellers who are orientated towards their province and those who sell relatively high percentages of their product to a unique buyer display lower markups. Moving from the first to the third quartile in any of the two variables implies an average markup decrease by more than two percentage points. The most prominent effect among the variables regressed in logs belongs to the *unique province count*. Concretely, a 100% increase in the variable, i.e., doubling the number of provinces served, corresponds to a 2.83 percentage points higher markup, on average.

While some of the signs of the coefficients are as I had predicted, e.g. the ones in the importance for the buyer and diversification categories or the *buyer specialisation* metric, some others are pretty unexpected. For example, the two reciprocity measures displaying positive coefficients go against my hypothesis that firms would charge higher markups when they have less to lose due to a unilateral relationship. Furthermore, it is crucial to consider that I only observe purchases for the firms that file the annexes, which is not a random sample. Hence, having higher reciprocity values captures, indirectly, selling to relatively larger firms, as tiny ones often do not file such annexes. This

---

[39]Concretely, I include variable input and capital levels to account for firm-level price differences not captured by the deflators, which are introduced due to input intensities entering the output elasticity equations. By doing so, I am also controlling for firm size, as measured by the cost side, which is likely to correlate with markups and a good portion of the network metrics.

Table 3: Regression Results for All Network Metrics

		IQR	(1)	(2)
Importance for buyer	Within Industry Importance	0.20	0.0440 *** (0.0074)	-0.0051 (0.0056)
	Across Industry Importance	0.02	0.2025 *** (0.0290)	-0.0035 (0.0137)
Relationship quality	Seller Reciprocity	0.09	0.0692 *** (0.0067)	-0.0049 (0.0039)
	Reciprocity Degree	0.04	0.0563 *** (0.0065)	-0.0054 (0.0046)
	log(Selling Frequency)	6.93	0.0189 *** (0.0015)	-0.0064 *** (0.0010)
Competition	Unique Seller	0.03	-0.0371 *** (0.0097)	-0.0006 (0.0072)
	Competition Intensity	0.00	-1.4258 (0.7345)	0.1616 (0.7715)
Market targeting	Local Focus	0.39	-0.0606 *** (0.0038)	0.0020 (0.0033)
	Vertical Focus	0.24	-0.0141 *** (0.0038)	-0.0073 (0.0039)
	Buyer Specialisation	0.42	-0.0610 *** (0.0041)	0.0060 (0.0033)
Diversification	log(Unique Buyer Count)	84	0.0151 *** (0.0010)	-0.0012 (0.0011)
	log(Unique Industry Count)	16	0.0147 *** (0.0012)	-0.0035 ** (0.0013)
	log(Unique Province Count)	6	0.0283 *** (0.0017)	-0.0011 (0.0012)
Year FE			✓	✓
Industry FE			✓	✓
Input usage controls			✓	✓
Firm FE				✓
Observations			126,902	126,902

Notes: This table reports results of regressing different network-related metrics on markups. Each row captures an independent regression, i.e. the respective variable is regressed alone on markups, with the inclusion of the respective fixed effects and controls reported at the bottom. Concretely, Column (2) differs from Column (1) only through the inclusion of firm fixed effects. Interquartile ranges (IQR) are reported in levels even for the variables regressed in logs. Standard errors clustered at the firm level reported in parentheses. Significance levels given by: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

caveat only makes the positive sign of the coefficients even more bizarre, as one would expect less market power when trading with larger firms. Another odd result is the negative coefficient for the *unique seller* metric. I had expected to find a positive relationship between being more often the unique seller of one's industry. Yet, some vertical integration mechanism could be at play here, which would go hand in hand with lower markups on average. Finally, I also believe that targeting firms outside one's industry and focusing on local markets would correlate with higher markups, but this does not seem to be the case after all.

Regardless of the hypothesised mechanisms and the realised small magnitudes, once I control for firm-fixed effects in column (2), practically all my effects become statistically insignificant as their magnitude drops faster than their standard errors. The only exceptions to this phenomenon lie in the log of the *selling frequency* and the log of the *unique industry count*. However, despite retaining their statistical significance, the coefficients' magnitude becomes abysmal and, more importantly, changes sign. According to this set of results, having more frequent transactions and diversifying through a higher number of industries served correlates negatively with markups after controlling for firm-specific characteristics.

Appendix F includes regression results based on different markup specifications. Figures F.4 and F.5 use Cobb-Douglas production with the same variable input V, whereas Figures F.6 and F.7 keep Translog production while using only materials M as variable input to recover markups, respectively for unweighted and weighted metrics. With both alternative specifications, the results are qualitatively quite different from those of my baseline. Despite the small magnitudes, significant effects are found for almost all metrics, even when including firm-fixed effects.

5.2 Discussion

Validity limitations

To begin this discussion, I believe it is paramount to emphasise the correlational nature of my research. My analysis focuses on suggestive evidence; no causality claim can or should be made based on the results. Several distinct threats prevent me from doing so by undermining the internal validity of the estimates.

As a starter, there might be omitted variable bias through, e.g. regulatory

frameworks, which can directly affect both my dependent and independent variables. Although some of the most regulation-prone industries are already excluded from my analysis (see Appendix Table C.1 for details on excluded industries), stringent frameworks regulating competition and prices might affect my results for the remaining firms in a way that I cannot control for.[40] Another variable of concern could be firm size, although I somewhat account for it by including variable inputs and capital controls.[41]

Measurement errors in my variables are other conceivable threats, as everything is constructed based on self-reported amounts.[42] For example, I highlighted earlier how unexpectedly high discrepancies between different data sources prevented me from combining more granular data to estimate markups. For my dependent variable, measurement error does not pose a problem as long as it is random, meaning it does not correlate with my independent variables. Arguably, one could expect larger firms to make fewer reporting mistakes. Given that I expect firm size to correlate with markups and network metrics, the randomness of the measurement errors is not guaranteed. On the other hand, measurement errors in the transaction data will pass through to the independent variables built upon it. Consequently, my results could suffer from attenuation bias such that, ignoring other issues, one can take them as a lower bound of the actual effect.

A further threat to proper identification lies in the direction of the cause-effect relationship. Can firms internalise the market power originating in the network in their economic decisions to achieve higher markups? Or, vice versa, do firms strengthen their network position thanks to the market power they already possess, captured by the markups? Both could be true at the same time, resulting in simultaneity bias. Observing exogenous shocks to my independent variables, e.g. through an instrumental variable approach, would be the most common way to disentangle this duality. Unfortunately, this lies beyond the scope of my work.

Finally, two separate sources of sampling bias might affect the external validity of the estimates as well, even within the Ecuadorian context. The first

---

[40]Concretely, one possible concern is that I do not treat state-owned firms differently although they might face less strict requirements in terms of reporting.

[41]I already discussed how the relationship between markups and firm size looks like in my data towards the end of Section 3.

[42]Although there is data about tax filing amendments, which has the potential to reduce measurement error, I was not able to use it. The amendment forms' structure is more complicated than the original forms, and I could not access the necessary documentation to process it.

originates in the non-observation of informal firms, which constitute a considerable component of the Ecuadorian economy, employing 37% of the working population in 2011 (The World Bank, 2013). The second relates to the forced exclusion of most formal firms from the initially available sample. Thanks to their more detailed tax filings, only incorporated firms and unincorporated firms subject to mandatory bookkeeping could be considered for the markup estimation procedure. For these reasons, any result should be evaluated in the context of Ecuador's most prominent firms, roughly 50,000 strong. It should not be extrapolated to the entire universe of firms in the country, as the dynamics governing market power might differ.

Technical limitations

Despite De Loecker and Warzynski (2012) admitting that the levels of their ratio estimator are biased[43] they argue that the estimation using any variable input should yield the same markups. Nevertheless, the firm-level markups I obtain using material costs, labour costs, or the sum of both differ considerably not only in levels but also in correlations.[44] My contrasting findings are consistent with a recent paper by Raval (2023). The author shows overwhelming evidence for different markup distributions, even negatively correlated with each other, using labour versus material costs as variable input.[45] Consequently, my regression results with the network metrics are not robust across markup specifications. My findings are either barely affected or entirely nullified by including firm-fixed effects, depending on which markup specification I use.

Another technical issue with the markup estimation procedure concerns the initial parameters passed to the GMM problem. As I have described in Section 3, GMM iteratively optimises the technological parameters of the production function by minimising a specified set of moments. Given an initial

---

[43]A critique by Bond et al. (2021) shows that this happens when prices are unobserved and (deflated) revenues are used instead of physical output, as is the case in this work and most of the literature.

[44]However, at a more aggregate or distributional level, the general trends align across specifications, as presented in Section 3.

[45]My markup estimates based on labour costs display very poor or even negative correlation with the other more reasonable estimates based on materials only or a combination of materials and labour. These latter estimates at least correlate positively with each other. However, the magnitudes are modest, ranging from 0.18 to 0.64 depending on the exact pair and type of correlation coefficient, Pearson or Spearman. In general, specifications that share the same definition of variable cost display higher correlations than specifications sharing the same production function form.

set of parameter values, the procedure is deterministic. Notwithstanding, I have found out that even slight changes to the initial values can make GMM return significantly different results. This incident occurred even for some of the biggest industries and with the Cobb-Douglas specification. Thus, it is neither a question of sample size nor model complexity. For full disclosure, I initialised my values with the OLS estimates in the Cobb-Douglas setting and as zeroes in the Translog one. My approach imitates what I believe De Loecker and Warzynski (2012) do in their replication package. However, I could not find any discussion on this issue in their materials or, more generally, in the markup estimation literature.

Contextual limitations

A different type of limitation that is worth mentioning concerns the definition of what I call a firm. With ownership data readily available, I could have grouped tax IDs with joint ownership, as Adão et al. (2022) do. Nevertheless, I refrained from doing so as I feared the estimation of production functions on aggregated firms of potentially two completely different industries would have led to misspecified markups. As a starter, to which industry would one assign the consolidated firm? This decision alone will have significant implications for the firm's estimated markup. Thus, I highlight here that my analysis is grounded on a definition of a firm that mimics the economic establishment level rather than the perhaps ordinary ownership level. Potential troubles concern the fact that firms with shared ownership might transact with each other outside of market rules, potentially affecting both the estimated markups and the network metrics.

As mentioned in Section 2, I was forced to discard half of the data in terms of the timeline to avoid biasing the sample since I could not observe industry information for new firms. This is a grand limitation regarding how much my work can inform about trends, as I can only analyse four years of data. Ideally, a more recent snapshot from the firm registry or a methodology for accurately predicting the industry of a firm's operations would greatly expand the contextual relevance of my work.

On a different note, I also do not feel confident discussing entry and exit effects realistically, as a firm could potentially be part of the sample in 2008 and 2010 while not being included in 2009 and 2011, even though it is still active in reality. This happens following my cleaning steps for the tax filing

data, where the concept of entry and exit gets blurred by firms submitting invalid forms for my markup estimation procedure that get discarded. The rationale for this decision was to preserve as many observations as possible for the markup estimation while at the same time not imputing missing years.

Finally, one cannot forget the issue related to the fact that markups and network metrics are estimated based on two different datasets. As mentioned in Section 4, the fact that I do not observe final sales transactions is particularly problematic since there is a discrepancy between the measure of revenue considered in the markup estimation (which includes final sales) and that considered for the network metrics (which does not). Concretely, one could believe markups to be higher on final sales due to, e.g. individuals having less leverage than a firm purchasing the same products as intermediate inputs. If true, I should estimate higher markups for firms focused on final consumption. However, these same firms would show poor network metrics, again due to their focus on final consumption. Hence, I could get biased results due to omitting a relevant variable.[46] Thus, I close the cycle and the discussion by reconnecting with the validity limitations mentioned at the beginning.

6 Conclusion

To conclude, I recapitulate that the primary goal of this paper was to investigate a previously untested hypothesis, namely that market power may originate from concrete topological advantages that firms acknowledge and then incorporate into their price decision-making process. My analysis reveals that statistically significant firm-level relationships between markups and different indicators of network strength can be found in the data. Nevertheless, based on the exact specification of markups, these effects may vanish or change direction when including firm-fixed effects. At any rate, their implied magnitudes are modest, and their implied direction is sometimes puzzling. Given that the findings cannot be interpreted causally, I choose not to extend my discoveries to propose any policy implications.

Notwithstanding, I encourage future research to follow up on this novel idea as my results provide at least some suggestive evidence that there might be some connection between market power and network strength, even though I might not have grasped the full extent or mechanism behind it yet. As with

---

[46]Note that a similar argument can be made for exports, which are considered in the markup estimation but not observed in the transactional network.

any research project, time and space constraints prevent me from digging deeper into various dimensions. Now, I want to take a chance to list some of the possible avenues for future research that I feel are exciting.

Before embarking on the project, I firmly believed there was potential for further incorporating and adapting network science methods into economics. My work strengthened this belief further. Lamentably, I could not adopt, as I had hoped, readily available measures such as betweenness centrality due to computational constraints. Hopefully, future research can apply this and other potentially attractive network science tools to answer economics questions, perhaps with smaller networks or approximations. Although I briefly considered aggregating my network at a higher level to use such tools, I eventually refrained from doing so, preferring to remain at the firm level as I had set out to. Nevertheless, many of the network science concepts I wanted to use initially can still provide insightful results even at the industry level, and I hope future research will attempt this.

On the markup estimation side, I focused on the production function approach by employing the methodology developed by De Loecker and Warzynski (2012). Alternatively, it could be interesting to see how the general results and the markup facts change based on different approaches to markup estimation. Other realistic possibilities would include dynamic panel methods in the spirit of Arellano and Bond (1991) and Blundell and Bond (2000) or a new approach developed by Kirov et al. (2021), which directly utilises revenue data without the need to convert it to physical output. In the context of markup estimation, I wish researchers would provide more information on how their different markup specifications correlate besides simply reporting medians and standard deviations. Additionally, I hope that in the future, researchers explicitly disclose their initial values when using GMM and how changes to them affect their markups, given the disproportional role they might play.

Recent research from Di Mauro et al. (2023) suggests that contrary to generally held beliefs, larger firms might charge lower markups than their smaller competitors. The findings naturally provoke questions about where the supposed market power manifests, if not through higher markups in the product market. The same researchers provide a possible explanation: markdowns on the factor market. A recent paper from Zavala (2022) that uses the same data as I do here further supports this idea. The author shows how powerful exporters exert their monopsonistic market power by charging markdowns on Ecuadorian farmers' produce. Thus, an interesting spin-off of my work would

be to investigate the other side of market power, namely markdowns instead of markups. One could create similar but inverse buyer-side network-related metrics and study how they relate to the firm-level markdowns that firms impose on their suppliers.

Yet another potential avenue for future research is investigating the role of vertical integration in detail. Thanks to the availability of ownership data, examining the network effects of vertical integration, i.e., firms taking control of upstream or downstream firms, on markups was a real possibility. Unfortunately, time constraints did not allow me to do so.

Lastly, experimenting with other definitions of competition markets in the network could generate different results. Particularly appealing is the idea of either going narrower on the industry classification, including a spatial dimension — for example, provinces or regions based on Ecuador's diverse geomorphology —, or combining both. In parallel with this, it would be interesting to investigate the relationship between market power and networks when considering potential buyers. I tried to do it but it did not work out as intended, perhaps narrower competitions markets would help in this sense.

References

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., and Tahbaz-Salehi, A. (2012). The network origins of aggregate fluctuations. *Econometrica*, 80(5):1977–2016.

Acemoglu, D., Ozdaglar, A., and Tahbaz-Salehi, A. (2015). Systemic risk and stability in financial networks. *American Economic Review*, 105(2):564–608.

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83(6):2411–2451.

Adão, R., Carrillo, P., Costinot, A., Donaldson, D., and Pomeranz, D. (2022). Imports, exports, and earnings inequality: Measures of exposure and estimates of incidence. *The Quarterly Journal of Economics*, 137(3):1553–1614.

Antràs, P., Chor, D., Fally, T., and Hillberry, R. (2012). Measuring the upstreamness of production and trade flows. *American Economic Review*, 102(3):412–16.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(2):277–297.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2020). The Fall of the Labor Share and the Rise of Superstar Firms*. *The Quarterly Journal of Economics*, 135(2):645–709.

Banco Central del Ecuador (2020). Estadísticas Sector Real - Cuentas Nacionales Anuales. `https://contenido.bce.fin.ec/documentos/Administracion/CuentasNacionalesAnuales.html`. Accessed: 2024-02-22.

Berry, S. T. and Haile, P. A. (2021). Foundations of demand estimation. In Ho, K., Hortaçsu, A., and Lizzeri, A., editors, *Handbook of Industrial Organization*, volume 4, chapter 1, pages 1–62. Elsevier.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1):115–143.

Blundell, R. and Bond, S. (2000). Gmm estimation with persistent panel data: an application to production functions. *Econometric Reviews*, 19(3):321–340.

Bond, S., Hashemi, A., Kaplan, G., and Zoch, P. (2021). Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data. *Journal of Monetary Economics*, 121:1–14.

Cabrales, A., Gottardi, P., and Vega-Redondo, F. (2017). Risk sharing and contagion in networks. *The Review of Financial Studies*, 30(9):3086–3127.

Calvó-Armengol, A. and Jackson, M. O. (2004). The effects of social networks on employment and inequality. *The American Economic Review*, 94(3):426–454.

Calvó-Armengol, A. and Jackson, M. O. (2007). Networks in labor markets: Wage and employment dynamics and inequality. *Journal of Economic Theory*, 132(1):27–46.

Carrillo, P., Donaldson, D., Pomeranz, D., and Singhal, M. (2023a). Ghosting the tax authority: Fake firms and tax fraud in ecuador. *American Economic Review: Insights*, 5(4):427–44.

Carrillo, P., Donaldson, D., Pomeranz, D., and Singhal, M. (2023b). Misallocation in firm production: A nonparametric analysis using procurement lotteries. Working Paper 31311, National Bureau of Economic Research.

Carrillo, P., Pomeranz, D., and Singhal, M. (2017). Dodging the taxman: Firm misreporting and limits to tax enforcement. *American Economic Journal: Applied Economics*, 9(2):144–64.

Carvalho, V., Elliott, M., and Spray, J. (2020a). Network Bottlenecks and Market Power.

Carvalho, V., Elliott, M., and Spray, J. (2020b). Supply Chain Bottlenecks in a Pandemic. COVID-19 Economic Research, Cambridge Institute for New Economic Thinking.

Chiavari, A. and Goraya, S. S. (2024). The Rise of Intangible Capital and the Macroeconomic Implications. Working Paper, .

Colonescu, C. (2021). Price markups and upstreamness in world input-output data. *Acta Universitatis Sapientiae, Economics and Business*, 9:71–85.

De Loecker, J. and Eeckhout, J. (2018). Global market power. Working Paper 24768, National Bureau of Economic Research.

De Loecker, J., Eeckhout, J., and Mongey, S. (2021). Quantifying market power and business dynamism in the macroeconomy. NBER Working Papers 28761, National Bureau of Economic Research, Inc.

De Loecker, J., Eeckhout, J., and Unger, G. (2020). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics*, 135(2):561–644.

De Loecker, J. and Warzynski, F. (2012). Markups and firm-level export status. *American Economic Review*, 102(6):2437–71.

De Ridder, M., Grassi, B., and Morzenti, G. (2022). The hitchhiker's guide to markup estimation. Poid working papers, Centre for Economic Performance, LSE.

Di Mauro, F., Mertens, M., and Mottironi, B. (2023). Sources of large firms' market power and why it matters. Voxeu column, Center for Economic Policy Research.

Diez, F. J., Leigh, D., and Tambunlertchai, S. (2018). Global Market Power and its Macroeconomic Implications. IMF Working Papers 2018/137, International Monetary Fund.

Edmond, C., Midrigan, V., and Xu, D. Y. (2023). How costly are markups? *Journal of Political Economy*, 131(7):1619–1675.

Feenstra, R. C., Inklaar, R., and Timmer, M. P. (2015). The next generation of the penn world table. *American Economic Review*, 105(10):3150–82.

Gallardo, G. G. and Philippon, T. (2018). How eu markets became more competitive than us markets: A study of institutional drift. *Political Economy - Development: Political Institutions eJournal*.

Gandhi, A., Navarro, S., and Rivers, D. A. (2020). On the identification of gross output production functions. *Journal of Political Economy*, 128(8):2973–3016.

Gradzewicz, M. and Mućk, J. (2023). Globalisation and the fall of markups. *The World Economy*.

Grassi, B. (2017). IO in I-O: Size, Industrial Organization, and the Input-Output Network Make a Firm Structurally Important. Working Papers 619, IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University.

Hall, R. E. (1988). The relation between price and marginal cost in u.s. industry. *Journal of Political Economy*, 96(5):921–947.

Jackson, M. O. (2016). The Past and Future of Network Analysis in Economics. In *The Oxford Handbook of the Economics of Networks*. Oxford University Press.

Jackson, M. O. (2019). *The Human Network : How Your Social Position Determines Your Power, Beliefs, and Behaviors.* Knopf Doubleday Publishing Group.

Kirov, I., Mengano, P., and Traina, J. (2021). Measuring markups with revenue data. *Available at SSRN 3912966.*

Klette, T. J. and Griliches, Z. (1996). The inconsistency of common scale estimators when output prices are unobserved and endogenous. *Journal of Applied Econometrics*, 11(4):343–361.

Levinsohn, J. and Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies*, 70(2):317–341.

Marschak, J. and Andrews, W. H. (1944). Random simultaneous equations and the theory of production. *Econometrica*, 12(3/4):143–205.

Mayer, A. (2011). Quantifying the effects of job matching through social networks. *Journal of Applied Economics*, 14(1):35–59.

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297.

Pomeranz, D. (2015). No taxation without information: Deterrence and self-enforcement in the value added tax. *American Economic Review*, 105(8):2539–69.

Raval, D. (2023). Testing the Production Approach to Markup Estimation. *The Review of Economic Studies*, 90(5):2592–2611.

Rodríguez-Moreno, J. and Rochina-Barrachina, M. (2019). ICT use, investments in R&D and workers' training, firms' productivity and markups: The case of ecuadorian manufacturing. *The European Journal of Development Research*, 31.

Stiebale, J. and Szücs, F. W. (2022). Mergers and market power: evidence from rivals' responses in european markets. *The RAND Journal of Economics*.

The World Bank (2013). Ecuador - The faces of informality : Ecuador - Las caras de la informalidad (Spanish). `http://documents.worldbank.org/curated/en/942561468246331465/Ecuador-Las-caras-de-la-informalidad`. Accessed: 2024-05-15.

The World Bank (2024). GDP growth (annual %). `https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?end=2011&locations=EC&start=2008`. Accessed: 2024-05-06.

Traina, J. (2018). Is aggregate market power increasing? production trends using financial statements. *Production Trends Using Financial Statements (February 8, 2018)*.

United Nations (2002). International Standard Industrial Classification of All Economic Activities (ISIC) - Revision 3.1. Statistical Papers Series M No. 4, Rev.3.1, Department Of Economic And Social Affairs.

Wooldridge, J. M. (2009). On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters*, 104(3):112–114.

Zavala, L. (2022). Unfair trade? monopsony power in agricultural value chains. *The World Bank*.

A   Data Preparation

Preface

The first hurdle I encountered during data preparation lies in the seemingly innocuous snapshot date of the firm registry. Naturally, firms that commenced operations after this snapshot date, somewhere around half-2012, are absent from the registry. This results in systematically missing information about their respective industries, an essential aspect of this study.

Attempts to deduce a firm's industry based on its observed selling and purchasing patterns in transaction data produced results significantly better than random assignment. However, the accuracy of this approach was insufficient for me to rely on. Therefore, to avoid biasing the sample and potentially skewing my estimates by selectively excluding specific firms (i.e., the new entrants), I was compelled to discard the latter four years of available data (2012-2015), retaining only the first half (2008-2011).

An additional challenge arose during the initial stages of constructing the panel structure required for the markup estimation procedure. While integrating various datasets, I encountered discrepancies that complicated the process. I initially intended to use data from social security payments, domestic purchase annexes, and customs forms because of their purportedly higher quality and granularity. However, the magnitudes of wages and other benefits reported in the social security datasets generally do not align well with the labour costs documented in tax filings. Similarly, discrepancies are evident in the export and import amounts derived from customs forms. Further, poor matching is also noted in the domestic purchase annexes, compounded by an inability to determine whether the purchases are designated for production (i.e., materials), supplies, or assets. Given these issues, I hesitate to use these transactions as a proxy for intermediate inputs in production.

Consequently, I rely instead exclusively on the amounts reported in the tax filings, recognising that these variables should, in theory, exhibit greater internal consistency within each firm-year observation than combining data from different sources.

Tax filing data

The income tax data consists of two distinct types of submission forms. The 101 form is for juridical persons, such as corporations, while the 102 form

is for natural persons, primarily self-employed individuals. The 102 form filers are further divided into two subcategories: those subject to mandatory book-keeping because they either started their business with at least $60,000 of capital or have yearly revenues or costs exceeding, respectively, $100,000 or $80,000. These firms must file a longer 102 form, which resembles the 101 form. The other subcategory consists of natural persons with annual revenue exceeding $10,000 but not subject to mandatory book-keeping; they file a shorter 102-form that covers personal costs and revenues. Due to the inability to observe assets in the short 102 form, I must disregard all short 102 form filers, which represent above 80% of all tax filers.

Cleaning raw tax form submissions starts with creating accounting items, such as labour costs, by summing the appropriate individual tax cells for each firm and year. To maintain consistency, I disregard self-reported totals and subtotals, relying instead on the calculated sums across individual tax cells.[1] I then round all values to the nearest dollar and drop all filings where any key variables (tangible fixed assets, operative revenue, material costs, labour costs) equal zero.[2]

Further, I deal with duplicate submissions by keeping the latest ones and then deliberately selecting the observations with the highest amounts across all variables for the remaining few cases.[3] Finally, I manually exclude the state-owned oil producer and exporter *Petroecuador*[4] from my dataset before

---

[1]An exception to this rule occurs when the calculated total tangible fixed assets – which I use as a measure of capital – are negative. In such cases, I assume the reported depreciation to be overstated and use the self-reported total tangible fixed assets, which already account for depreciation. Most of these instances involve zero or fully depreciated reported fixed assets, suggesting these observations would be dropped later regardless.

[2]Tangible fixed assets encompass land, finished and unfinished buildings, facilities, ships, aircraft, barges, furniture, furnishings, machinery and equipment, computer equipment and software, vehicles, and transportation equipment, minus accumulated depreciation for all these items. Operative revenue includes net domestic sales and net exports. Material costs are calculated by summing differences in initial and final inventories of raw materials, goods not produced by the firm, products in process, and finished products, adding net purchases of domestic and imported raw materials, and goods not produced by the firm. Labour costs consist of wages, salaries, other remunerations and compensations, social benefits and contributions to social security, and professional fees for domestic and foreign contractors.

[3]A few submissions without a timestamp are excluded beforehand.

[4]I exclude *Petroecuador* for two main reasons: (i) Ecuador was part of the Organization of the Petroleum Exporting Countries (OPEC) during 2008-2011, indicating that *Petroecuador*'s prices were not set by the market (OPEC, 2024); (ii) *Petroecuador* is significantly larger than any other firm in Ecuador, which could distort my analysis despite its limited relevance to my research questions.

trimming outliers based on material and labour revenue shares.[5] The final dataset covers over 50,000 firms, comprising over 130,000 valid filings across the four years.

Purchase annex data

To formalise the Ecuadorian production network, I describe the data cleaning steps applied to the purchase annexes, which track intra-firm transactions. In its raw form, it is worth mentioning that the dataset contains disaggregated individual transactions among firms (i.e., there could be multiple transactions on the same day between the same two firms). First, I removed entries with missing purchase dates and those with identical buyer and seller IDs. Next, I adjust the transaction values to ensure they conform to the statutory 12% VAT rate when the reported VAT amount suggests otherwise.[6] Subsequently, I discard all negative transactions and those with suspiciously high purchase values (above $100,000,000), assuming they resulted from errors, such as filing the buyer or seller identification number in the wrong field. This threshold was set because, realistically, only *Petroecuador* could be involved in transactions of such magnitude.[7] Finally, I correct instances of double-reporting across different fields on the form before summing values and counting frequencies at the buyer-seller-year level. Consolidated amounts are subsequently rounded to the nearest dollar. Over the four years, the resulting cleaned dataset encompasses approximately 35 million observations aggregated from 250 million single transactions between 2.7 million unique sellers and 130 thousand unique buyers. The total value exchanged across the cleaned dataset is over 280 billion US dollars from 2007.

Firm registry data

Cleaning the firm registry uniquely involved removing firm entries with no information on the industry of their operations and those where the reported

---

[5]Specifically, I remove all observations where either the material or labour revenue shares are in the top or bottom 2.5% on a year-by-year basis. The average cutoffs for material share are 4.6% at the bottom and 105.9% at the top, while for labour share, they are 0.2% at the bottom and 70.4% at the top.

[6]When the VAT amount is missing, the transaction value remains unchanged.

[7]The state-owned oil exporter *Petroecuador* does indeed have five transactions exceeding $100,000,000 across the four years. These are the only transactions above this threshold that I retain.

industry fell under Ecuador's somewhat vague special sections.[8]

Price deflator data

Regarding the deflators, the cleaning process was relatively minimal. Although INEC's industry-specific PPIs are more granular than ISIC classes, as they use Ecuador's narrower classification, I had to employ a crosswalk to align ISIC revisions. I was, therefore, limited to the ISIC class level. Concretely, to compute deflators for ISIC Revision 3.1 classes, I calculated the average of the PPIs provided for all matching ISIC Revision 3 classes. The same approach was applied to higher-level deflators but with weighted averages based on the number of matched class levels. Ultimately, all deflators, including the one for capital and the general deflator, were adjusted to have 2007 as the base year, just before my observation period.

Merging datasets
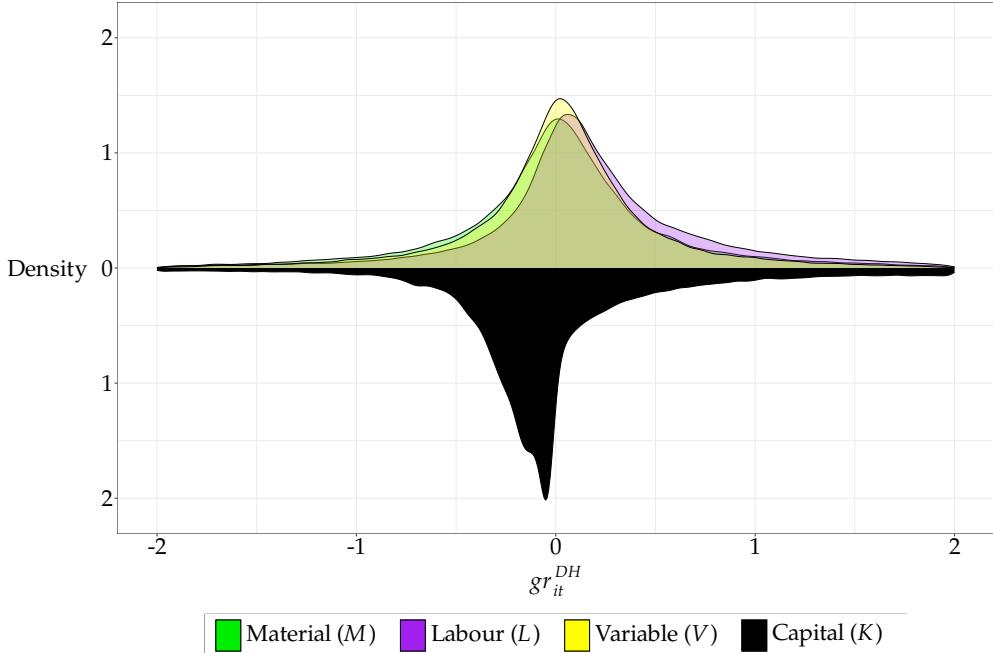
It is crucial to reiterate here that knowing a firm's industry is non-negotiable for my work. Indeed, the output elasticities I need to derive markups are estimated by pooling firms according to their industry, and many of the network metrics I construct rely on the assumption that sellers compete for buyers' purchases with other sellers within the same industry. Thus, when matching the firm registry to the tax filings data and the transactional data, which I use for the markup estimation procedure and network metrics construction, respectively, I immediately discard all firms that are unmatched and thus for which I lack industry information. In the case of tax filing data, this affects only a handful of observations, as the set of firms I start with comprises relatively larger firms that are unlikely not to appear in the registry. The same holds for the number of buyers in the transactional data. Regarding sellers, I lost around 1.3 million unique units, almost half the initial number. These are tiny sellers, though: they account for around 3% of the single transactions and 4% of the total transaction value.

---

[8]These sections include "Private sector salaried work" (R), "Public sector salaried work" (S), "Without economic activity" (T).

B Variable Input Decision

I decided to test my options for variable inputs by checking their *variability*, in terms of how much firms adjust their demand from year to year. Figure B.1

Figure B.1: Inputs Variability

Notes: This figure illustrates the Davis and Haltiwanger (1992) growth rate gr_{it}^{DH} for different inputs. Capital, the only dynamic input, is filled in black and flipped along the vertical axis to facilitate the comparison among the variable inputs: materials (green) labour (purple), and the combination of both (yellow).

plots the densities of such adjustments measured using the Davis and Halti-wanger (1992) growth rate, which I define for any input X as:[9]

$$gr_{it}^{DH} = \frac{(X_{it} - X_{it-1})}{\frac{1}{2}(X_{it} + X_{it-1})}. \tag{30}$$

The proposed growth rate is a monotonic transformation of the more typical $gr_{it} = \frac{(X_{it} - X_{it-1})}{X_{it-1}}$ with the advantages of being a good approximation for the standard one for small rates and being bounded between -2 and 2, with 0 meaning no growth.[10] It is clear to see by the high peak and narrowness of its density curve that capital is the least variable input, as expected. The peak

---

[9]Given that I do not observe quantities I used deflated expenses instead.

[10]Note that by construction I will never reach the gr_{it} values of -2 or 2 as they would require firm i's input X to move towards $X_{it} = 0$ or to move from $X_{it-1} = 0$, respectively, which is impossible due to the a priori exclusion of any observation with input values of zero.

and most of the area are to the left of zero, most likely due to depreciation. Capital was not under my choices and is included in the figure only as a proof of concept. Material and labour inputs might appear similar in variability at first, though labour is arguably more variable considering the skewness of the x-axis. The peak for labour is also shifted entirely to the right than zero, implying that it is common for firms to increase their labour inputs. A variable combining both materials and labour, henceforth $V = M + L$, has a density which, unsurprisingly, is a combination of the two single densities, with a higher peak just between the ones of the separate inputs.

Another way to examine this issue is to check the inputs' correlation with revenue. The idea is that the more the input is variable and free to adjust according to production targets, the more it will correlate with revenue. Table B.1 illustrate the results of such correlation exercises. Here, materials in-

Table B.1: Correlations Between Different Inputs and Revenue

Input	Pearson's r		Spearman's ρ	
	Linear	Log	Linear	Log
Materials (M)	0.901	0.946	0.939	
Labour (L)	0.785	0.740	0.698	
Variable (V)	0.916	0.978	0.975	
Capital (K)	0.034	0.530	0.490	

Notes: The variable input V is given by summing M and L. Capital is included just as proof of concept and does not belong to the possible variable inputs. Spearman's ρ is identical for linear and log variables due to the log being a monotonic transformation.

puts show much more variability than labour inputs with respect to revenues across all the correlation measures. This is indeed expected. The combined variable input shows even higher correlations with revenue, though I fear this might be somewhat mechanical. As someone who reads the main text knows, I eventually settled on the combined variable inputs for my baseline specifications. This is not a novelty as Traina (2018), Diez et al. (2019), and De Loecker et al. (2020) are some of the papers that have used a combination of materials and labour as variable input, under the name of *cost of goods sold*, while using the same markup estimation procedure. Curiously, Diez et al. (2019) highlight that using labour costs — a choice made by De Loecker and Warzynski (2012) and many other papers that use value-added as a measure for

output — can be problematic when dealing with highly regulated labour markets due to, e.g., the implied adjustment costs of hiring and firing workers. According to MacIsaac Donna (1999), Ecuador is subject to such a strict regulatory framework — at least in the Latin American context — which leaned me into avoiding using this input alone for my estimation. Along the same lines, Petrin and Sivadasan (2013) identify Ecuador as having the highest expected discounted cost of firing a worker among Latin American and OECD countries in 1999, based on data from a study made by Heckman and Pages (2003). On the other hand, Traina (2018) highlights in his work the importance of considering *all* variable costs to get the correct picture on markups, despite De Loecker and Warzynski (2012) arguing that the methodology should yield the same markups regardless of choice.

C Information on Samples and Matching

Table C.1: Industry Distribution of Markup Estimation Sample and Match with Network

Code	Industry Description	1st Stage		2nd Stage		Share Matched in Network	
		Observations	Unique firms	Observations	Unique firms	Observations	Unique firms
A01	Agriculture, hunting and related service activities	8,175	3,372	4,544	2,184	96.2%	96.9%
A02	*Forestry, logging and related service activities*	*75*	*35*	*39*	*19*	*-*	*-*
B05	Fishing, aquaculture and service activities incidental to fishing	2,641	1,034	1,534	710	97.8%	97.1%
C10	*Mining of coal and lignite; extraction of peat*	*1*	*1*	*0*	*0*	*-*	*-*
C11	Extraction of crude petroleum and natural gas; service activities incidental to oil and gas extraction, excluding surveying	161	65	92	42	98.8%	96.9%
C12	*Mining of uranium and thorium ores*	*1*	*1*	*0*	*0*	*-*	*-*
C13	Mining of metal ores	123	63	56	32	80.5%	84.1%
C14	Other mining and quarrying	178	79	93	49	95.5%	93.7%
D15	Manufacture of food products and beverages	3,525	1,209	2,268	923	98.3%	98.1%
D16	*Manufacture of tobacco products*	*4*	*1*	*3*	*1*	*-*	*-*
D17	Manufacture of textiles	935	310	617	246	97.1%	96.8%
D18	Manufacture of wearing apparel; dressing and dyeing of fur	1,650	629	986	440	97.3%	96.0%
D19	Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear	305	110	191	82	99.0%	98.2%
D20	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials	513	192	306	131	97.9%	97.4%
D21	Manufacture of paper and paper products	280	89	189	72	99.6%	98.9%
D22	Publishing, printing and reproduction of recorded media	2,097	742	1,302	552	99.9%	99.6%
D23	*Manufacture of coke, refined petroleum products and nuclear fuel*	*48*	*14*	*34*	*13*	*-*	*-*
D24	Manufacture of chemicals and chemical products	1,720	526	1,177	452	99.4%	98.7%
D25	Manufacture of rubber and plastics products	884	280	599	238	99.2%	97.5%
D26	Manufacture of other non-metallic mineral products	779	281	471	200	99.0%	98.6%
D27	Manufacture of basic metals	317	105	209	82	99.1%	100.0%
D28	Manufacture of fabricated metal products, except machinery and equipment	1,235	453	761	333	99.4%	98.7%
D29	Manufacture of machinery and equipment n.e.c.	1,245	474	739	329	99.3%	98.3%
D30	*Manufacture of office, accounting and computing machinery*	*15*	*6*	*9*	*4*	*-*	*-*
D31	Manufacture of electrical machinery and apparatus n.e.c.	442	168	266	108	98.0%	94.6%
D32	*Manufacture of radio, television and communication equipment and apparatus*	*59*	*21*	*38*	*16*	*-*	*-*
D33	Manufacture of medical, precision and optical instruments, watches and clocks	501	186	304	132	97.8%	96.2%
D34	Manufacture of motor vehicles, trailers and semi-trailers	264	84	176	71	99.6%	100.0%
D35	Manufacture of other transport equipment	120	52	65	33	97.5%	96.2%
D36	Manufacture of furniture; manufacturing n.e.c.	1,172	430	714	310	96.8%	96.5%
D37	*Recycling*	*59*	*24*	*31*	*15*	*-*	*-*
E40	*Electricity, gas, steam and hot water supply*	*91*	*37*	*53*	*23*	*-*	*-*
E41	*Collection, purification and distribution of water*	*10*	*5*	*3*	*1*	*-*	*-*
F45	Construction	4,462	2,097	2,214	1,124	92.3%	91.6%
G50	Sale, maintenance and repair of motor vehicles and motorcycles; retail sale of automotive fuel	10,237	3,618	6,426	2,723	99.1%	98.7%
G51	Wholesale trade and commission trade, except of motor vehicles and motor-cycles	39,517	14,701	24,161	10,561	97.7%	96.8%
G52	Retail trade, except of motor vehicles and motorcycles; repair of personal and household goods	20,734	8,047	12,154	5,447	98.2%	98.0%
H55	Hotels and restaurants	4,898	1,900	2,913	1,314	98.3%	98.3%
I60	Land transport; transport via pipelines	1,732	916	765	412	95.6%	95.4%
I61	Water transport	101	42	57	27	96.0%	95.2%
I62	*Air transport*	*34*	*16*	*15*	*10*	*-*	*-*
I63	Supporting and auxiliary transport activities; activities of travel agencies	1,004	544	417	241	96.0%	96.3%
I64	Post and telecommunications	439	202	221	108	98.2%	97.5%
J65	*Financial intermediation, except insurance and pension funding*	*96*	*44*	*50*	*23*	*-*	*-*
J66	*Insurance and pension funding, except compulsory social security*	*14*	*8*	*5*	*3*	*-*	*-*
J67	*Activities auxiliary to financial intermediation*	*61*	*32*	*26*	*13*	*-*	*-*
K70	Real estate activities	1,931	939	950	513	88.6%	88.9%
K71	Renting of machinery and equipment without operator and of personal and household goods	395	196	185	101	97.0%	95.4%
K72	Computer and related activities	767	336	402	191	97.8%	96.1%
K73	*Research and development*	*87*	*40*	*42*	*22*	*-*	*-*
K74	Other business activities	9,486	4,480	4,618	2,398	97.5%	97.1%
L75	Public administration and defence; compulsory social security	619	313	293	148	95.2%	94.9%
M80	Education	596	315	264	144	85.9%	86.0%
N85	Health and social work	2,236	935	1,236	587	96.2%	95.4%
O90	*Sewage and refuse disposal, sanitation and similar activities*	*42*	*18*	*21*	*10*	*-*	*-*
O91	Activities of membership organisations n.e.c.	1,007	489	472	250	86.9%	84.9%
O92	Recreational, cultural and sporting activities	514	240	255	130	94.2%	93.3%
O93	Other service activities	648	307	324	171	96.0%	94.8%
P95	*Activities of private households as employers of domestic staff*	*7*	*4*	*3*	*2*	*-*	*-*
P96	*Undifferentiated goods-producing activities of private households for own use*	*0*	*0*	*0*	*0*	*-*	*-*
P97	*Undifferentiated service-producing activities of private households for own use*	*0*	*0*	*0*	*0*	*-*	*-*
Q99	*Extraterritorial organisations and bodies*	*8*	*3*	*4*	*2*	*-*	*-*

Notes: Industries are defined according to ISIC Revision 3.1 divisions (2-digits level, where the prefix letter indicates the corresponding ISIC Revision 3.1 section). For the 1st stage I consider all valid tax filings as described in Appendix Section A. For the 2nd stage all valid observations are considered for which a valid observation is also available in the previous year. Finally, for the markup estimation procedure I exclude all industries (formatted in *red cursive*) where the number of observations available for the 1st stage is below 100. The last two columns report the shares of observations and unique firms that find a match in the transaction network out of the 1st stage values. Obviously, excluded industries are not matched. Note that further observations may be excluded later on, for example in the case of negative markups in a given specification.

Even though, surprisingly, some firms in the markup estimation sample do not appear in the transaction data, the vast majority do. One can check matching ratios by industry in Table C.1.

The other way around does not hold though. In fact, the vast majority of sellers in the network are pretty small firms, and they only appear in the data through the purchase annexes of larger firms. As such, most of them are not required to file detailed tax forms. Table C.2 pitches the matched sample, which is used for the regression analysis, against the unmatched network observations. Additionally, it also provides summary statistics for the combination of the two, i.e., the full network sample.

Table C.2: Summary Statistics of Sellers in Network

Summary statistic	Matched observations	Unmatched observations	All network observations
Dummy: Files purchase annexes	0.927	0.058	0.093
Number of buyers supplied	110.6	6.2	10.4
Number of industries supplied	13.3	2.8	3.2
Transaction frequency	1,042.1	37.3	77.9
Transaction value	1,167,400	40,875	86,400
Average amount per transaction	5,322	1,543	1,696
Number of unique sellers	49,842	1,408,007	1,429,650
Share of all	3.49%	98.49%	100.00%
Number of total single transactions	132,474,030	112,638,767	245,112,797
Share of all	54.05%	45.95%	100.00%
Sum of total transactions value	148,409,288,369	123,388,050,062	271,797,338,431
Share of all	54.60%	45.40%	100.00%
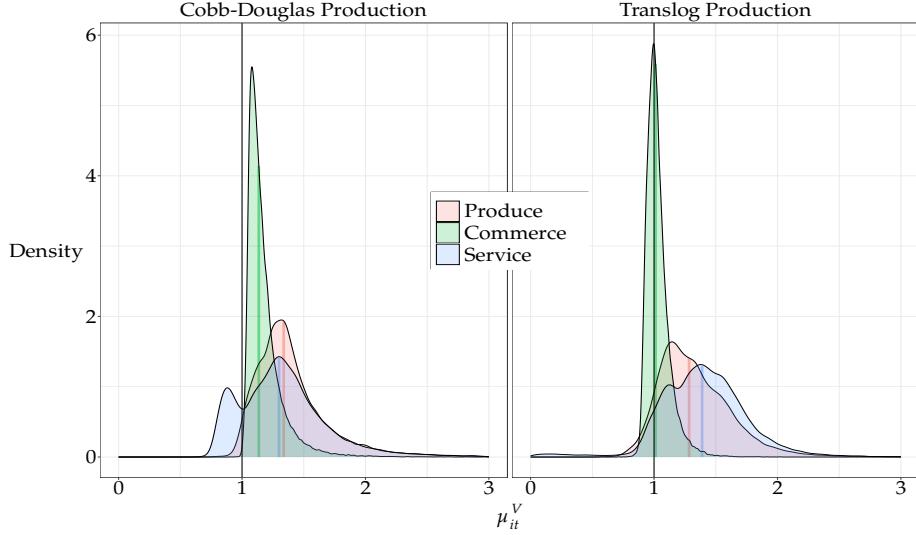
Notes: The reported values are averages across all sellers and all years for each sample The decision of pooling all years together stems from the minimal variance of these summary statistics over time. For the construction of this table I used the nominal values, although the network metrics are computed on deflated values.

The slightly over 125,000 observations that are matched in the network data only constitute 3.5% of all valid seller-year pairs in the network. Yet, they accrue a share of all intra-firm transaction value and count just above 50%. There is no need for formal F-tests to confirm the marked differences between the two sets. Sellers in the intersection set are much more likely to file purchase annexes, thus appearing as buyers as well; roughly 93% of the year-seller observations do so. This number drops to 6% for network-only sellers. The number of unique buyers and industries served by the average seller in the intersection set is much higher than those for the sellers for whom I could not estimate markups. The same holds for transaction frequencies, total values, and average amounts.
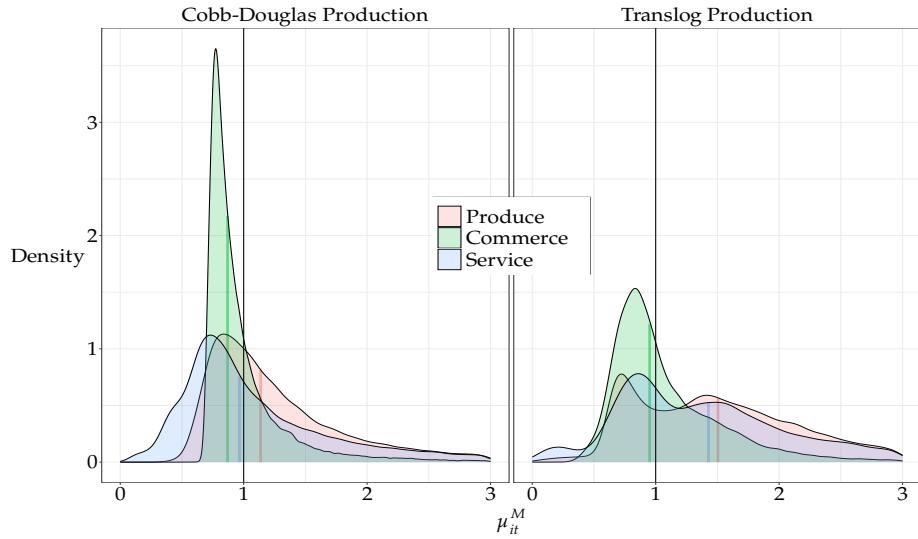
D Market Power: Robustness to Different Markup Specifications

Figure D.1: Firm-level Markup Distribution by Industry Group

(a) Based on Materials and Labour Combined (μ_{it}^V)
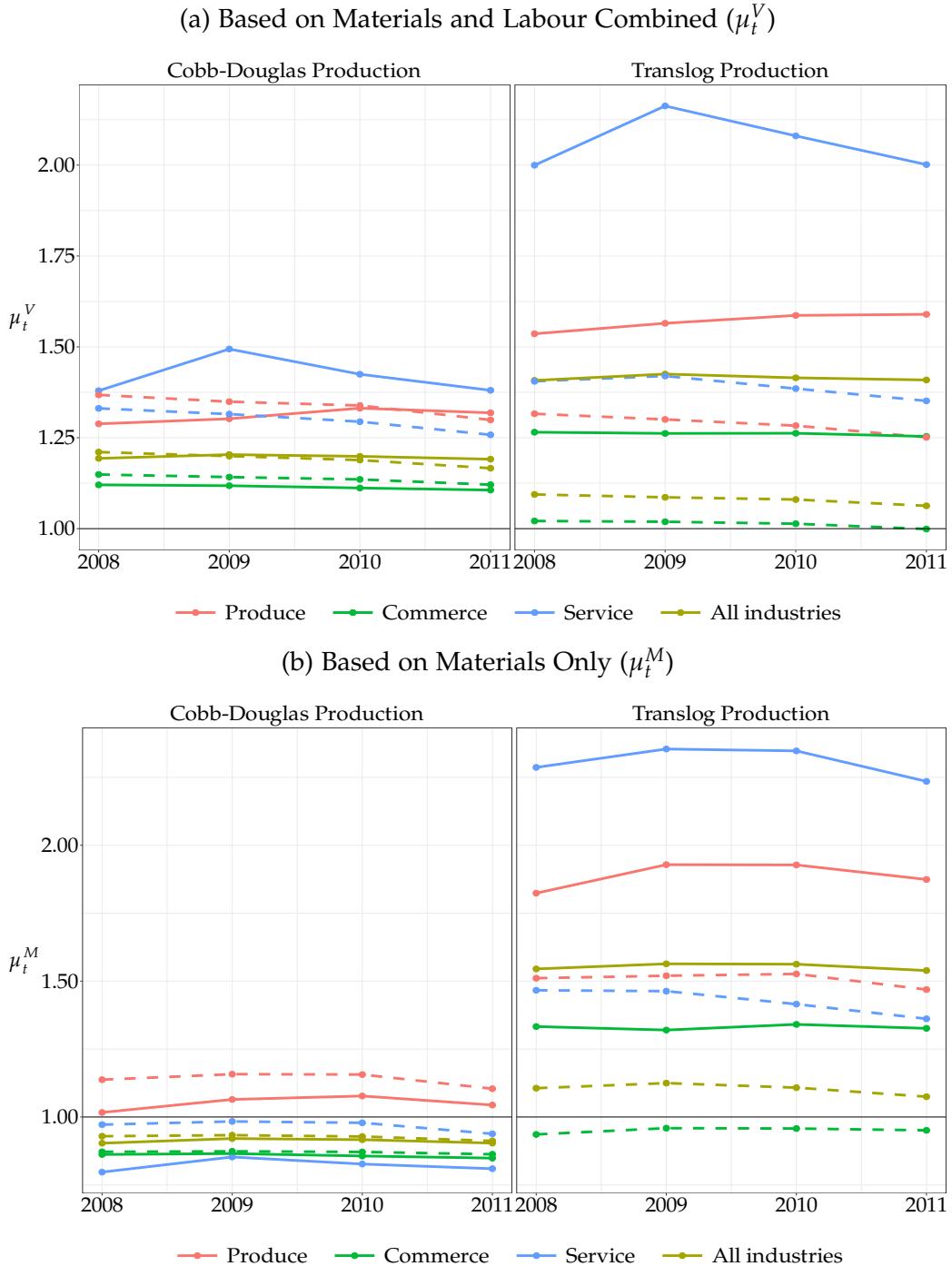
(b) Based on Materials Only (μ_{it}^M)

Notes: The figure illustrates the distribution of firm-level markups μ_{it}^V in Subfigure (a), respectively μ_{it}^M in Subfigure (b), by industry group. The red, green, and blue areas correspond to the density of markups for the three industry groups comprising the ISIC Revision 3.1 industries A01 to F45, G50 to G52, and H55 to Q99, respectively. Colour-coded line segments represent group medians. A black vertical line denotes equality between prices and marginal costs, i.e., $\mu_{it}^X = 1$. All years are pooled together. Though considered for the density computation, markups below 0 and above 3 were trimmed from the plot. The Translog panel of Subfigure D.1a emulates Figure 1 in the main text.
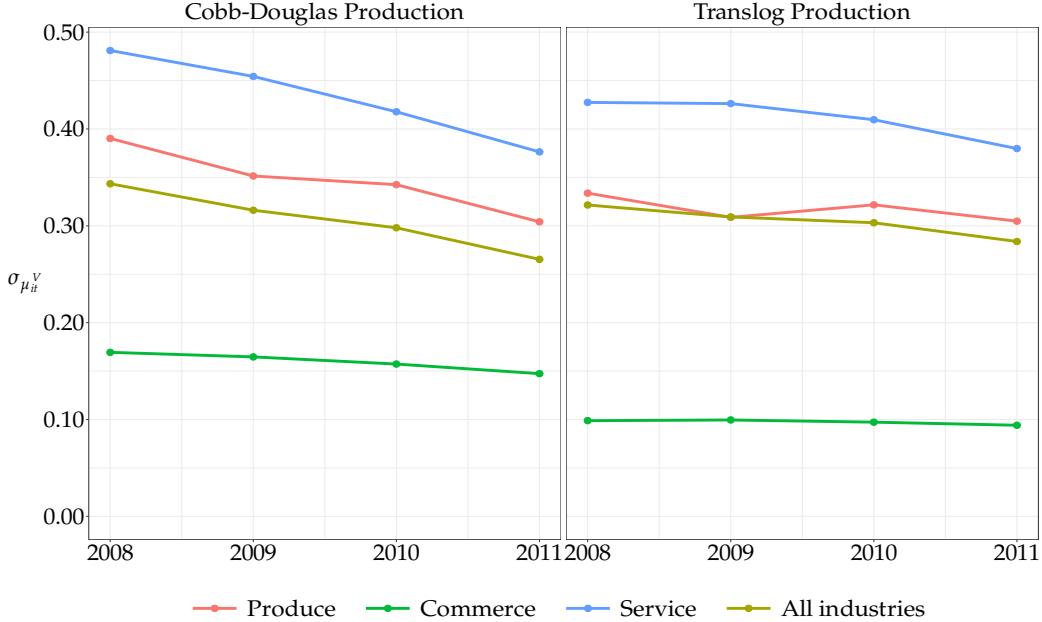
Figure D.2: Evolution of Aggregate Markup (—) and Median Firm-Level Markup (- -) by Industry Group
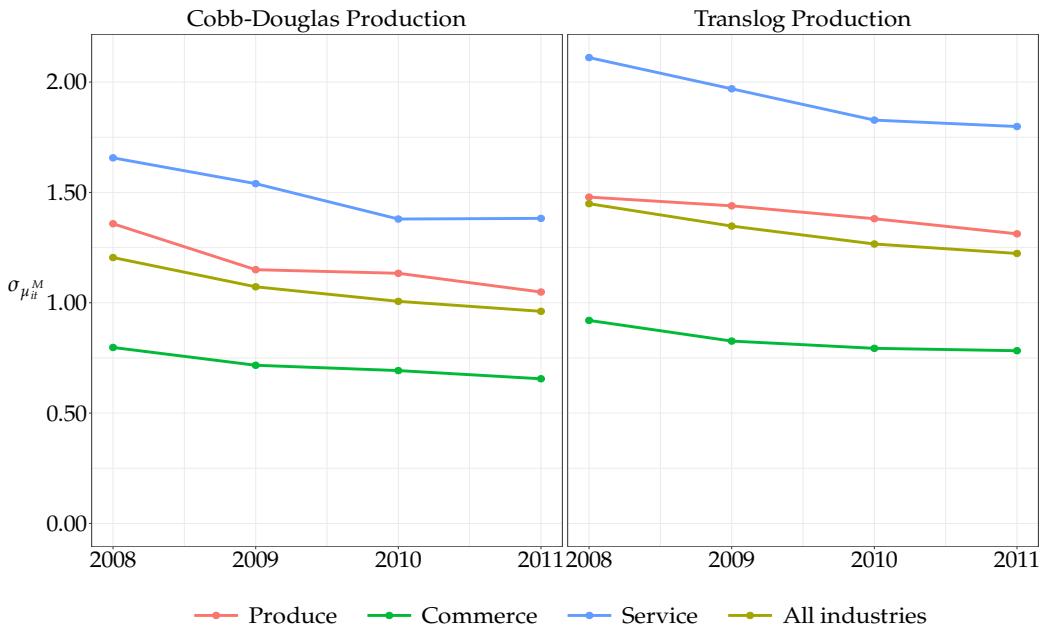
(a) Based on Materials and Labour Combined (μ_t^V)

(b) Based on Materials Only (μ_t^M)

Notes: The red, green, and blue lines picture the specific markup trends for the three industry groups comprising the ISIC Revision 3.1 industries A01 to F45, G50 to G52, and H55 to Q99, respectively. The yellow lines combine all industries. Aggregate markups $\bar{\mu}_t$ (solid line) are constructed by weighting individual markups with the yearly corresponding variable input cost share over the whole sample, i.e., $\bar{\mu}_t^X = \sum_i \frac{X_{it}}{\sum_j X_{jt}} \mu_{it}^X$, instead of the more common sales shares, following the critique raised by Edmond et al. (2023). A black horizontal line denotes equality between prices and marginal costs, i.e., $\mu_t^V = 1$. The Translog panel of Subfigure D.2a emulates Subfigure 2a in the main text.

Figure D.3: Evolution of Firm-level Markup Dispersion

(a) Based on Materials and Labour Combined (μ_{it}^V)
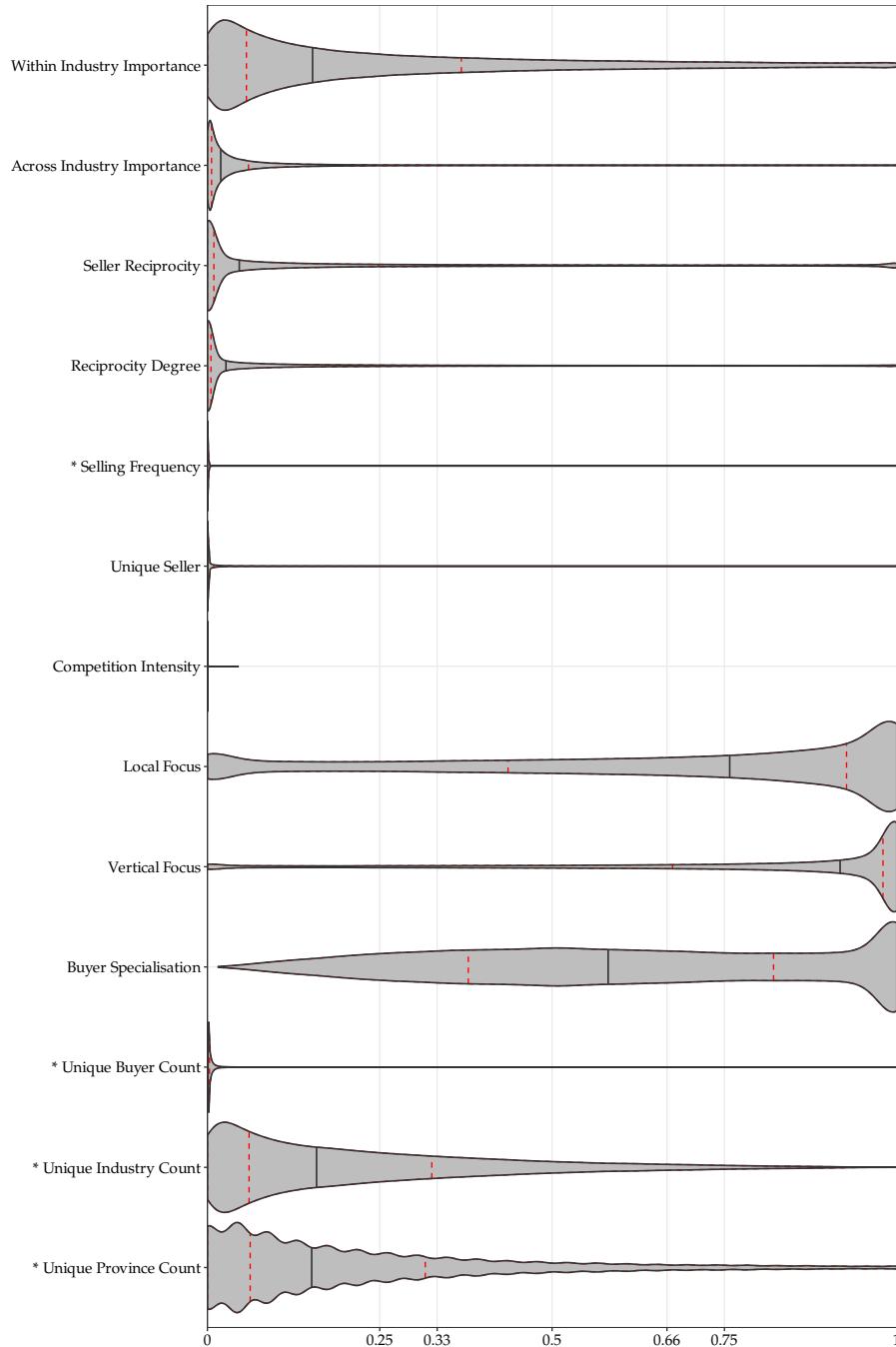
(b) Based on Materials Only (μ_{it}^M)

Notes: The red, green, and blue lines picture the specific markup trends for the three industry groups comprising the ISIC Revision 3.1 industries A01 to F45, G50 to G52, and H55 to Q99, respectively. The yellow lines combine all industries. The Translog panel of Subfigure D.3a emulates Subfigure 2b in the main text.

E Network Strength: Weighted Metrics

Figure E.1: Distribution of Weighted Network Metrics

Notes: Violin plots display the mirrored distribution of each network metric. For each distribution, a solid black line marks the median, while dashed red lines represent the 25th and 75th percentiles, respectively. For visualisation purposes, metrics marked with an asterisk (*) have been minimum-maximum-scaled to force them into the remaining metrics' natural 0-1 range. Figure 3 depicts the same plots for the unweighted metrics.

F Network Origins of Market Power: Further Results

Translog, Input V, Unweighted Metrics Decomposition

Table F.1: Regression Results Decomposition – Unweighted Metrics

	IQR	(1)	(2)	(3)	(4)	(5)
Within Industry Importance	0.20	0.2832*** (0.0082)	0.2825*** (0.0082)	0.0616*** (0.0072)	0.0440*** (0.0074)	-0.0051 (0.0056)
Across Industry Importance	0.02	0.4169*** (0.0329)	0.4211*** (0.0330)	0.2730*** (0.0242)	0.2025*** (0.0290)	-0.0035 (0.0137)
Seller Reciprocity	0.09	0.2587*** (0.0082)	0.2573*** (0.0082)	0.0843*** (0.0056)	0.0692*** (0.0067)	-0.0049 (0.0039)
Reciprocity Degree	0.04	0.3448*** (0.0101)	0.3445*** (0.0100)	0.0627*** (0.0063)	0.0563*** (0.0065)	-0.0054 (0.0046)
log(Selling Frequency)	6.93	0.0120*** (0.0015)	0.0116*** (0.0015)	0.0242*** (0.0011)	0.0189*** (0.0015)	-0.0064*** (0.0010)
Unique Seller	0.03	0.2890*** (0.0118)	0.2877*** (0.0118)	-0.0330*** (0.0099)	-0.0371*** (0.0097)	-0.0006 (0.0072)
Competition Intensity	0.00	38.2492*** (1.6036)	38.5207*** (1.6051)	-1.8237* (0.7310)	-1.4258 (0.7345)	0.1616 (0.7715)
Local Focus	0.39	-0.0957*** (0.0055)	-0.0974*** (0.0055)	-0.0705*** (0.0036)	-0.0606*** (0.0038)	0.0020 (0.0033)
Vertical Focus	0.24	0.1088*** (0.0072)	0.1080*** (0.0072)	-0.0156*** (0.0038)	-0.0141*** (0.0038)	-0.0073 (0.0039)
Buyer Specialisation	0.42	0.1129*** (0.0050)	0.1149*** (0.0050)	-0.0757*** (0.0036)	-0.0610*** (0.0041)	0.0060 (0.0033)
log(Unique Buyer Count)	84.00	-0.0161*** (0.0012)	-0.0165*** (0.0012)	0.0192*** (0.0010)	0.0151*** (0.0010)	-0.0012 (0.0011)
log(Unique Industry Count)	16.00	-0.0291*** (0.0016)	-0.0299*** (0.0016)	0.0192*** (0.0011)	0.0147*** (0.0012)	-0.0035** (0.0013)
log(Unique Province Count)	6.00	-0.0317*** (0.0019)	-0.0321*** (0.0019)	0.0347*** (0.0015)	0.0283*** (0.0017)	-0.0011 (0.0012)
Year FE		X	4	4	4	4
Industry FE		X	X	42	42	42
Input usage controls		X	X	X	✓	✓
Firm FE		X	X	X	X	49,726
Observations		126,902	126,902	126,902	126,902	126,902

Notes: This table reports regression results of the unweighted network metrics on markups, decomposing the relationships by incrementally increasing controls. Each row captures an independent regression, i.e. the respective variable is regressed alone on markups, with the inclusion of the respective fixed effects and controls reported at the bottom. For fixed effects, I report the number of groups. Interquartile ranges (IQR) are reported in levels even for the variables regressed in logs. Standard errors clustered at the firm level reported in parentheses. Significance levels given by: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Translog, Input V, Weighted Metrics Decomposition

Table F.2: Regression Results Decomposition – Weighted Metrics

	IQR	(1)	(2)	(3)	(4)	(5)
Within Industry Importance	0.31	0.1411*** (0.0065)	0.1398*** (0.0065)	0.0726*** (0.0050)	0.0532*** (0.0055)	-0.0022 (0.0031)
Across Industry Importance	0.05	0.1310*** (0.0193)	0.1298*** (0.0193)	0.2391*** (0.0161)	0.1800*** (0.0162)	-0.0011 (0.0058)
Seller Reciprocity	0.17	0.1412*** (0.0055)	0.1402*** (0.0055)	0.0772*** (0.0040)	0.0618*** (0.0043)	-0.0034 (0.0021)
Reciprocity Degree	0.11	0.1742*** (0.0064)	0.1734*** (0.0064)	0.0833*** (0.0047)	0.0678*** (0.0048)	-0.0042 (0.0023)
log(Selling Frequency)	32.26	0.0015 (0.0012)	0.0012 (0.0012)	0.0220*** (0.0010)	0.0176*** (0.0011)	-0.0036*** (0.0006)
Unique Seller	0.01	0.1521*** (0.0103)	0.1505*** (0.0103)	-0.0392*** (0.0079)	-0.0378*** (0.0076)	0.0042 (0.0047)
Competition Intensity	0.00	29.0596*** (1.1857)	29.1654*** (1.1851)	0.3135 (0.5669)	0.4213 (0.5581)	0.9830 (0.6444)
Local Focus	0.52	-0.0499*** (0.0042)	-0.0510*** (0.0042)	-0.0410*** (0.0026)	-0.0361*** (0.0028)	0.0017 (0.0021)
Vertical Focus	0.25	0.0829*** (0.0046)	0.0828*** (0.0046)	-0.0176*** (0.0026)	-0.0128*** (0.0029)	-0.0031 (0.0024)
Buyer Specialisation	0.49	0.0915*** (0.0052)	0.0935*** (0.0052)	-0.0730*** (0.0037)	-0.0572*** (0.0043)	-0.0002 (0.0026)
log(Unique Buyer Count)	84.00	-0.0161*** (0.0012)	-0.0165*** (0.0012)	0.0192*** (0.0010)	0.0151*** (0.0010)	-0.0012 (0.0011)
log(Unique Industry Count)	16.00	-0.0291*** (0.0016)	-0.0299*** (0.0016)	0.0192*** (0.0011)	0.0147*** (0.0012)	-0.0035** (0.0013)
log(Unique Province Count)	6.00	-0.0317*** (0.0019)	-0.0321*** (0.0019)	0.0347*** (0.0015)	0.0283*** (0.0017)	-0.0011 (0.0012)
Year FE		X	4	4	4	4
Industry FE		X	X	42	42	42
Input usage controls		X	X	X	✓	✓
Firm FE		X	X	X	X	49,726
Observations		126,902	126,902	126,902	126,902	126,902

Notes: This table reports regression results of the weighted network metrics on markups, decomposing the relationships by incrementally increasing controls. Each row captures an independent regression, i.e. the respective variable is regressed alone on markups, with the inclusion of the respective fixed effects and controls reported at the bottom. For fixed effects, I report the number of groups. Interquartile ranges (IQR) are reported in levels even for the variables regressed in logs. Standard errors clustered at the firm level reported in parentheses. Significance levels given by: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Translog, Input V, Weighted Metrics Results

Table F.3: Regression Results for All Network Metrics

		IQR	(1)	(2)
Importance for buyer	Within Industry Importance	0.31	0.0532 *** (0.0055)	-0.0022 (0.0031)
	Across Industry Importance	0.05	0.1800 *** (0.0162)	-0.0011 (0.0058)
Relationship quality	Seller Reciprocity	0.17	0.0618 *** (0.0043)	-0.0034 (0.0021)
	Reciprocity Degree	0.11	0.0678 *** (0.0048)	-0.0042 (0.0023)
	log(Selling Frequency)	32.26	0.0176 *** (0.0011)	-0.0036 *** (0.0006)
Competition	Unique Seller	0.01	-0.0378 *** (0.0076)	0.0042 (0.0047)
	Competition Intensity	0.00	0.4213 (0.5581)	0.9830 (0.6444)
Market targeting	Local Focus	0.52	-0.0361 *** (0.0028)	0.0017 (0.0021)
	Vertical Focus	0.25	-0.0128 *** (0.0029)	-0.0031 (0.0024)
	Buyer Specialisation	0.49	-0.0572 *** (0.0043)	-0.0002 (0.0026)
Diversification	log(Unique Buyer Count)	84	0.0151 *** (0.0010)	-0.0012 (0.0011)
	log(Unique Industry Count)	16	0.0147 *** (0.0012)	-0.0035 ** (0.0013)
	log(Unique Province Count)	6	0.0283 *** (0.0017)	-0.0011 (0.0012)
Year FE			✓	✓
Industry FE			✓	✓
Input usage controls			✓	✓
Firm FE				✓
Observations			126,902	126,902

Notes: This table reports results of regressing different network-related metrics on markups. Each row captures an independent regression, i.e. the respective variable is regressed alone on markups, with the inclusion of the respective fixed effects and controls reported at the bottom. Concretely, Column (2) differs from Column (1) only through the inclusion of firm fixed effects. Interquartile ranges (IQR) are reported in levels even for the variables regressed in logs. Standard errors clustered at the firm level reported in parentheses. Significance levels given by: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Cobb-Douglas, Input V, Unweighted Metrics Results

Table F.4: Regression Results for All Network Metrics

		IQR	(1)	(2)
Importance for buyer	Within Industry Importance	0.20	-0.0487 *** (0.0087)	-0.0529 *** (0.0106)
	Across Industry Importance	0.02	-0.1545 *** (0.0230)	-0.1339 *** (0.0255)
Relationship quality	Seller Reciprocity	0.09	-0.0485 *** (0.0076)	-0.0303 *** (0.0070)
	Reciprocity Degree	0.04	0.0000 (0.0097)	-0.0330 *** (0.0087)
	log(Selling Frequency)	6.93	-0.0284 *** (0.0011)	-0.0436 *** (0.0021)
Competition	Unique Seller	0.03	-0.0420 ** (0.0133)	-0.0030 (0.0138)
	Competition Intensity	0.00	3.7335 *** (0.9006)	0.4194 (1.1501)
Market targeting	Local Focus	0.39	-0.0136 ** (0.0043)	0.0261 *** (0.0071)
	Vertical Focus	0.24	0.0176 ** (0.0062)	-0.0248 ** (0.0077)
	Buyer Specialisation	0.41	0.1961 *** (0.0040)	0.1324 *** (0.0068)
Diversification	log(Unique Buyer Count)	84	-0.0347 *** (0.0007)	-0.0576 *** (0.0023)
	log(Unique Industry Count)	16	-0.0486 *** (0.0011)	-0.0593 *** (0.0029)
	log(Unique Province Count)	6	-0.0558 *** (0.0013)	-0.0473 *** (0.0025)
Year FE			✓	✓
Industry FE			✓	✓
Input usage controls			✓	✓
Firm FE				✓
Observations			127,128	127,128

Notes: This table reports results of regressing different network-related metrics on markups. Each row captures an independent regression, i.e. the respective variable is regressed alone on markups, with the inclusion of the respective fixed effects and controls reported at the bottom. Concretely, Column (2) differs from Column (1) only through the inclusion of firm fixed effects. Interquartile ranges (IQR) are reported in levels even for the variables regressed in logs. Standard errors clustered at the firm level reported in parentheses. Significance levels given by: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Cobb-Douglas, Input V, Weighted Metrics Results

Table F.5: Regression Results for All Network Metrics

		IQR	(1)	(2)
Importance for buyer	Within Industry Importance	0.31	-0.0867 *** (0.0051)	-0.0541 *** (0.0058)
	Across Industry Importance	0.05	-0.1636 *** (0.0097)	-0.0983 *** (0.0096)
Relationship quality	Seller Reciprocity	0.17	-0.0472 *** (0.0039)	-0.0225 *** (0.0036)
	Reciprocity Degree	0.11	-0.0357 *** (0.0045)	-0.0271 *** (0.0041)
	log(Selling Frequency)	32.25	-0.0276 *** (0.0007)	-0.0299 *** (0.0011)
Competition	Unique Seller	0.01	-0.0199 (0.0104)	0.0113 (0.0078)
	Competition Intensity	0.00	0.1453 (0.6529)	0.2146 (0.8898)
Market targeting	Local Focus	0.52	-0.0075 * (0.0034)	0.0181 *** (0.0042)
	Vertical Focus	0.25	0.0252 *** (0.0040)	-0.0126 ** (0.0045)
	Buyer Specialisation	0.49	0.1477 *** (0.0035)	0.0613 *** (0.0052)
Diversification	log(Unique Buyer Count)	84	-0.0347 *** (0.0007)	-0.0576 *** (0.0023)
	log(Unique Industry Count)	16	-0.0486 *** (0.0011)	-0.0593 *** (0.0029)
	log(Unique Province Count)	6	-0.0558 *** (0.0013)	-0.0473 *** (0.0025)
Year FE			✓	✓
Industry FE			✓	✓
Input usage controls			✓	✓
Firm FE				✓
Observations			127,128	127,128

Notes: This table reports results of regressing different network-related metrics on markups. Each row captures an independent regression, i.e. the respective variable is regressed alone on markups, with the inclusion of the respective fixed effects and controls reported at the bottom. Concretely, Column (2) differs from Column (1) only through the inclusion of firm fixed effects. Interquartile ranges (IQR) are reported in levels even for the variables regressed in logs. Standard errors clustered at the firm level reported in parentheses. Significance levels given by: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Translog, Input M, Unweighted Metrics Results

Table F.6: Regression Results for All Network Metrics

		IQR	(1)	(2)
Importance for buyer	Within Industry Importance	0.20	-0.1203 ** (0.0373)	-0.1029 ** (0.0395)
	Across Industry Importance	0.02	-0.1633 (0.1000)	-0.3413 *** (0.0862)
Relationship quality	Seller Reciprocity	0.09	0.2691 *** (0.0315)	0.0357 (0.0297)
	Reciprocity Degree	0.04	0.3700 *** (0.0400)	0.0538 (0.0365)
	log(Selling Frequency)	6.93	0.0357 *** (0.0049)	-0.0341 *** (0.0074)
Competition	Unique Seller	0.02	-0.3449 *** (0.0618)	-0.0397 (0.0575)
	Competition Intensity	0.00	8.8765 (5.5418)	21.2095 (12.0132)
Market targeting	Local Focus	0.39	0.0078 (0.0170)	0.0407 (0.0236)
	Vertical Focus	0.24	-0.0387 (0.0235)	-0.0033 (0.0265)
	Buyer Specialisation	0.41	0.0487 ** (0.0184)	0.1794 *** (0.0273)
Diversification	log(Unique Buyer Count)	84	0.0050 (0.0035)	-0.0561 *** (0.0075)
	log(Unique Industry Count)	16	0.0043 (0.0055)	-0.0469 *** (0.0094)
	log(Unique Province Count)	6	-0.0012 (0.0064)	-0.0508 *** (0.0077)
Year FE			✓	✓
Industry FE			✓	✓
Input usage controls			✓	✓
Firm FE				✓
Observations			126,334	126,334

Notes: This table reports results of regressing different network-related metrics on markups. Each row captures an independent regression, i.e. the respective variable is regressed alone on markups, with the inclusion of the respective fixed effects and controls reported at the bottom. Concretely, Column (2) differs from Column (1) only through the inclusion of firm fixed effects. Interquartile ranges (IQR) are reported in levels even for the variables regressed in logs. Standard errors clustered at the firm level reported in parentheses. Significance levels given by: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Translog, Input *M*, Weighted Metrics Results

Table F.7: Regression Results for All Network Metrics

		IQR	(1)	(2)
Importance for buyer	Within Industry Importance	0.31	-0.1380 *** (0.0231)	-0.0972 *** (0.0215)
	Across Industry Importance	0.05	-0.0863 (0.0456)	-0.1859 *** (0.0375)
Relationship quality	Seller Reciprocity	0.17	0.2510 *** (0.0196)	0.0306 (0.0164)
	Reciprocity Degree	0.11	0.3166 *** (0.0226)	0.0402 * (0.0189)
	log(Selling Frequency)	32.16	0.0263 *** (0.0033)	-0.0200 *** (0.0045)
Competition	Unique Seller	0.00	-0.3251 *** (0.0490)	-0.0243 (0.0426)
	Competition Intensity	0.00	16.2048 *** (4.5537)	14.5308 (8.5318)
Market targeting	Local Focus	0.52	0.0323 * (0.0136)	0.0053 (0.0167)
	Vertical Focus	0.25	-0.0476 ** (0.0165)	-0.0453 * (0.0186)
	Buyer Specialisation	0.49	0.0362 * (0.0180)	0.0687 ** (0.0213)
Diversification	log(Unique Buyer Count)	84	0.0050 (0.0035)	-0.0561 *** (0.0075)
	log(Unique Industry Count)	16	0.0043 (0.0055)	-0.0469 *** (0.0094)
	log(Unique Province Count)	6	-0.0012 (0.0064)	-0.0508 *** (0.0077)
Year FE			✓	✓
Industry FE			✓	✓
Input usage controls			✓	✓
Firm FE				✓
Observations			126,334	126,334

Notes: This table reports results of regressing different network-related metrics on markups. Each row captures an independent regression, i.e. the respective variable is regressed alone on markups, with the inclusion of the respective fixed effects and controls reported at the bottom. Concretely, Column (2) differs from Column (1) only through the inclusion of firm fixed effects. Interquartile ranges (IQR) are reported in levels even for the variables regressed in logs. Standard errors clustered at the firm level reported in parentheses. Significance levels given by: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Appendix References

Davis, S. J. and Haltiwanger, J. (1992). Gross job creation, gross job destruction, and employment reallocation. *The Quarterly Journal of Economics*, 107(3):819–863.

De Loecker, J., Eeckhout, J., and Unger, G. (2020). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics*, 135(2):561–644.

De Loecker, J. and Warzynski, F. (2012). Markups and firm-level export status. *American Economic Review*, 102(6):2437–71.

Diez, F. J., Fan, J., and Villegas-Sánchez, C. (2019). Global Declining Competition. IMF Working Papers 2019/082, International Monetary Fund.

Edmond, C., Midrigan, V., and Xu, D. Y. (2023). How costly are markups? *Journal of Political Economy*, 131(7):1619–1675.

Heckman, J. and Pages, C. (2003). Law and employment: Lessons from latin america and the caribbean. Working Paper 10129, National Bureau of Economic Research.

MacIsaac Donna, R. M. G. (1999). Do labor market regulations affect labor earnings in ecuador? Policy Research Working Paper WPS1717, World Bank Group.

OPEC (2024). Member Countries. `https://www.opec.org/opec_web/en/about_us/25.htm`. Accessed: 2024-03-04.

Petrin, A. and Sivadasan, J. (2013). Estimating Lost Output from Allocative Inefficiency, with an Application to Chile and Firing Costs. *The Review of Economics and Statistics*, 95(1):286–301.

Traina, J. (2018). Is aggregate market power increasing? production trends using financial statements. *Production Trends Using Financial Statements (February 8, 2018)*.

Statement of Authorship for a written thesis at the Department for Economics at the University of Zurich

I declare that this work titled **"Investigating the Topological Origins of Market Power in Ecuador's Production Network"** has been composed by myself, and describes my own work, unless otherwise acknowledged in the text.

This work has not been and will not be submitted for any other degree or the obtaining of ECTS points at the University of Zurich or any other institution of higher education.

All sentences or passages quoted in this paper from other people's work have been specifically acknowledged by clear cross-referencing to author, work and page(s). Any illustrations which are not the work of the author have been used with the explicit permission of the originator and are specifically acknowledged.

I understand that failure to specifically acknowledge all used work amounts to plagiarism and will be considered grounds for failure and will have judicial and disciplinary consequences according §7ff of the 'Disziplinarordnung der Universität Zürich' as well as §27 of the 'Rahmenverordnung für das Studium in den Bachelor- und Master-Studiengängen der Wirtschaftswissenschaftlichen Fakultät der Universität Zürich'.

With my signature I declare the accuracy of these specifications.

Name: VÁZQUEZ VÁZQUEZ   First Name: MIGUEL

Matriculation Number: 16-712-598   Date: 21/05/2024

Signature: