

# Production Networks and Firm-level Elasticities of Substitution\*

Brian Cevallos Fujiy<sup>†</sup>

Devaki Ghose<sup>‡</sup>

Gaurav Khanna<sup>§</sup>

**This version: September 2022**

## Abstract

We provide one of the first estimates of elasticities of substitution across inputs supplied by suppliers within the same industry. This elasticity is particularly relevant for the transmission and amplification of supply shocks across the production network. We obtain new real-time administrative tax data on product-level prices and quantities with firm-to-firm transactions. We leverage geographic and temporal variation from the Covid-19 lockdowns in India to estimate these firm-level elasticities of substitution and quantify the fall in trade. If suppliers are complements rather than substitutes in production, this shock can amplify by further transmitting downstream and upstream through the supply chain. We find that even at this very granular supplier level, suppliers are highly complementary, with an estimated elasticity of 0.55. Causally estimating these micro-level elasticities of substitution at the firm level allows us to understand how shocks propagate through supply chains, affecting aggregate GDP. We use our elasticities and simulate the impact of the Covid-19 lockdowns to find that under our estimated elasticities, the overall fall in output is substantial and widespread.

*Keywords:* shock propagation, resilience, Covid-19, networks

*JEL Codes:* F41, F44

---

\*We thank participants at the Danish International Economics Workshop, World Bank, CEPR's Globalisation, Strategic Autonomy and Economic Resilience in Time of Crisis, Annual Mid-Atlantic International Trade Workshop, Midwest International Economic Development Conference, STEG's Annual Conference, IMF's Jacques Polak Annual Research Conference, Bank of Italy's Global Value Chains, and Lund University for their comments and suggestions. We thank David Baqaee, Ana Cecilia Fieler, Banu Demir, Pablo Garriga, Andrei Levchenko, Sebnem Kalemli-Ozcan, Scott Orr, Ludovic Panon, Carolina Sanchez and Sebastian Sotelo for insightful discussions. We thank Ahmed Magdy and Eli Mogel for excellent research assistance and Sushil Khanna for guidance on the data and details on the analysis. The views expressed in this paper do not represent the views of the World Bank or its partner organizations and solely represent the authors' personal views.

<sup>†</sup>University of Michigan, [bcfujiy@umich.edu](mailto:bcfujiy@umich.edu).

<sup>‡</sup>The World Bank, [dghose@worldbank.org](mailto:dghose@worldbank.org).

<sup>§</sup>University of California, San Diego, [gakhanna@ucsd.edu](mailto:gakhanna@ucsd.edu).

# 1 INTRODUCTION

The ability of firms to substitute inputs across suppliers is critical for the resilience of supply chains and the transmission of supply shocks. If it is difficult for firms to substitute across suppliers after an adverse supply shock, the shock will amplify by transmitting further downstream through the supply chain. The importance of this mechanism was reflected during the Covid-19 pandemic, where supply chain disruptions drove dramatic reductions in GDP worldwide. For instance, India reported a -7.3% growth rate for the 2020/21 financial year, one of the most significant contractions worldwide and the largest decline in GDP since India's independence.<sup>1</sup> In this paper, we quantify the importance of firm-level elasticities of substitution across suppliers of intermediate inputs to explain large fluctuations in GDP. We provide new estimation strategies and estimates for these elasticities by leveraging regional variation in supply-side shocks induced by the Indian government's massive lockdown policy. We show that this elasticity is key to partly explaining the dramatic decline of the Indian economy during the Covid-19 pandemic. Using new big data computational techniques, we quantify this decline *directly* using information on the economy-wide firm-to-firm network.

We pose two main research questions. First, are suppliers of intermediate inputs within an industry complements or substitutes? The answer to this question determines how shocks propagate throughout supply chains. We expect shocks to propagate less across firm networks if input-suppliers are substitutable. However, if input-suppliers are complements, the effects of adverse shocks can easily propagate through buyer-supplier networks. Second, we ask, how does this newly estimated elasticity affect firm-level sales, and ultimately GDP, by propagating and amplifying shocks through firm-level input-output linkages?

Two unique features of our setting allow us to answer these questions credibly. First, India had a distinct mosaic of lockdown policies, whereby the roughly 600 districts were classified into three different zones with varying degrees of restrictions. This allows us to isolate variation in the ability to trade and transport goods over this period. Second, we obtain new granular and high-frequency administrative data on the universe of establishment-to-establishment transactions for a region in India, with unique information on unit values and HS-product classifications. These data, while not used before, allow us to estimate new elasticities at the firm (rather than industry) level, and across different suppliers for a firm.

We find that inputs within the same HS-4 industry but across different suppliers are highly complementary. Our estimated elasticity of substitution across suppliers is 0.55. In various specification tests employing different combinations of fixed effects and different sources of variation, we find that the estimated elasticities lie within a range of 0.49 to 0.65. Our new

---

<sup>1</sup><https://www.economicsobservatory.com/how-has-Covid-19-affected-indias-economy>. More broadly, GDP fell by -3.3% and -2.2% during the 2020/21 financial year for emerging market and developing countries, respectively.

elasticities show that even within the same HS-4 industry, inputs across firms are highly complementary. As such, even at the very micro level, firm-specific negative shocks contribute to GDP fluctuations. In contrast, [Atalay \(2017\)](#) estimate elasticities at the industry (rather than firm) level. We also estimate the more aggregate firm-level elasticity of substitution across different industries, and find complementarity across industries, in line with [Atalay \(2017\)](#).

As discussed by [Taschereau-Dumouchel \(2020\)](#) and [Baqae and Farhi \(2019\)](#), the literature so far provides little guidance about estimates of the firm-level elasticity of substitution between suppliers within industries, even though it is a crucial parameter driving the propagation of shocks. While other work estimates elasticities of substitution across industries ([Atalay, 2017](#)), between products from different countries ([Boehm et al., 2019](#)), or across intermediate goods ([Carvalho et al., 2021](#)), such estimates do not yet exist for substitution elasticities across suppliers within the same industry. Estimating elasticities of substitution across different suppliers has been especially challenging for two reasons. First, it is difficult to find detailed information on firm-to-firm transactions with product-specific unit values, reported by each firm. Second, it is challenging to find exogenous sources of variation in firm-level prices that allows one to credibly estimate these elasticities.

We provide estimates of firm-level elasticities of substitution across suppliers by leveraging the nationwide, sudden and unprecedented lockdown imposed by the Indian government in March 2020. Importantly, these lockdowns were not homogeneous: districts were categorized into *Green* (mild lockdown), *Orange* (medium lockdown) and *Red* (severe lockdown). Since the lockdowns were sudden and unexpected, they were likely implemented independent of economic fundamentals, and induced strong variation in transactions between firms across India.<sup>2</sup> We use this variation to estimate the firm-level elasticities of substitution across suppliers.

Yet, Covid-19 was not just a supply shock. [Baqae and Farhi \(2020\)](#) point out the pandemic outbreak was a combination of exogenous shocks to the quantities of factors supplied, the productivity of producers, and the composition of final demand by consumers across industries. To estimate the elasticity of substitution across suppliers of inputs for a particular product produced by a firm, we leverage variation in input prices driven by the sudden restrictions in economic activity due to lockdowns in districts where these suppliers were located. In addition, we leverage variation in trade costs arising from restrictions in economic activity in districts through which the goods need to pass through, from the seller to the buyer. While our instruments help derive the necessary variation, to further isolate supply shocks from other shocks, we control for an entire array of high-dimensional fixed effects, such as product-by-month and buyer-by-month fixed effects, to account for demand-side shocks. Given the richness of our product data, we can also include buyer-by-product and seller-by-product fixed effects. We further control for various other factors, such as firms' exposure to foreign shocks transmitted

---

<sup>2</sup><https://www.bbc.com/news/world-asia-india-56561095>, <https://thewire.in/government/india-Covid-19-lockdown-failure>

through trade [Hummels et al. \(2014\)](#), and the caseload and severity of Covid-19 cases.

This paper has three main sections. First, we present reduced-form evidence on the impact of adverse supply shocks on key firm-level variables such as unit values (prices) and the number of transactions (quantities). We leverage the Indian government’s sudden lockdown measure that affected firm-to-firm trade across districts, depending on whether firms fall in the *Red* zone (strict lockdown), *Orange* zone (moderate lockdown), or *Green* zone (mostly no lockdown). We find that the prices of intermediate inputs rose during the lockdown, especially if either buyers or sellers were located in *Orange* or *Red* zones. In districts where the seller is in a strict lockdown zone (orange or red), transactions fell dramatically, compared to either the case where the buyer is in a lockdown zone or both are in green zones.

Second, we modify a standard multi-sector firm-level model of input-output linkages by augmenting the production function with substitution across suppliers within the same industry. We derive analytical expressions that relate the relative values of quantities purchased of the same HS-4 product from different suppliers, to the equilibrium relative prices. That is, within each HS-4 product category, we quantify how substitutable inputs are between the different suppliers. We find that this elasticity of substitution is close to 0.55. Thus, following [Baqae and Farhi \(2020\)](#), after considering second-order effects, adverse firm-level shocks get amplified in the aggregate by propagating through firm-to-firm linkages while positive shocks get dampened. We further explore whether these elasticities differ across industries, and find that in a handful of industries, suppliers within the same industries are actually substitutes, whereas in others, they are highly complementary. This shows that we should be mindful of heterogeneity across industries in understanding how shocks propagate through supply chains.

Finally, we use the estimated elasticities to analyze how input complementarities at the firm level affect aggregate economic outcomes, and so, how important these complementarities are in explaining GDP declines during the Covid-19 pandemic. We find that a 25% productivity shock relegated only to firms in the red zone reduces overall GDP by 10.96%. This fall would be 2.02 pp less in a model where firms in the same industry are substitutes ( $\epsilon = 1.75$ ), and 0.75 pp more when firms in the same HS-4 industry are almost Leontief ( $\epsilon = 0.001$ ). Given that the quarterly GDP of this state was close to 32.5 billion USD in 2020, the additional losses due to firm-level complementarities translate into 655 million USD (about 19 USD per capita per quarter), compared to the case when firms are substitutes. Next, we investigate whether aggregate GDP losses in the face of large productivity shocks are less if policy-makers allow large firms (high final sales) or more connected firms (more direct and indirect linkages) to operate. We show that as the level of complementarity and the magnitude of the adverse shock increases, it pays more to save the more connected firms. Much importance, both in policy and academic circles, has been paid to large firms, as [Hulten \(1978\)](#) emphasized the importance of firm sizes in the propagation of shocks through production networks. We show that in the face of large adverse shocks and high levels of complementarity across suppliers, the more

connected firms are more important than large firms in shock propagation through the network.

**Related Work.** Our paper connects with two strands of literature. First, we speak to the literature on shock propagation and amplification through supply chains and production networks (Barrot and Sauvagnat, 2018; Carvalho et al., 2021; Peter and Ruane, 2022; Boehm et al., 2019; Korovkin and Makarin, 2020; Ferrari, 2022). There are at least three challenges in this literature. First, most firm-to-firm data either do not contain product-level (unit) prices from each supplying firm, or lack the required variation in such prices to estimate elasticities of substitution across suppliers within an industry.<sup>3</sup> Second, and relatedly, limited identifying variation in prices at the buyer-supplier level allows existing work to estimate substitution elasticities across industries or across domestic and foreign industries, but not across suppliers within an industry. In contrast, we provide one of the first estimates of the elasticity of substitution across suppliers within an industry: a parameter that is crucial in determining how shocks propagate. Third, the lack of firm-level elasticities across suppliers has so far constrained our assessment of the importance of nodal firms, such as the largest or the most connected firms, in the propagation of shocks through production networks.

We contribute to the literature in each of these dimensions. First, we measure unit prices and quantities at the seller-buyer-product-transaction level. We derive price changes from supply and transportation disruptions in lockdown-affected districts and estimate the firm-level elasticity of substitution between suppliers within an industry. We then quantify this elasticity's importance for amplifying firm-specific supply shocks through a roundabout production network (Baqee and Farhi, 2019). We address previously unanswered questions on the importance of nodal or large firms in shock amplification. We exploit computational innovations in big data to compute the second-order effects of productivity shocks using the entire matrix of production linkages. This innovation helps quantify the non-linear effects of productivity shocks *directly* using the network, without relying on approximations using final sales.<sup>4</sup>

Our paper is also related to research on trade collapses during adverse shocks (Behrens et al., 2013; Giovanni and Levchenko, 2009; Bricongne et al., 2012), and shock transmission through GVCs during Covid, via disruptions to imports/exports or aggregate production (Bonaldo et al., 2021; Baqee and Farhi, 2020; Cakmakli et al., 2021; Demir and Javorcik, 2020; Gerschel et al., 2020; Heise et al., 2020; Lafrogne-Roussier et al., 2021; Bas et al., 2022; Chakrabati et al., 2021). In contrast, we analyze how domestic transactions were affected during Covid

---

<sup>3</sup>Carvalho et al. (2021) observe a binary measure of whether firms were connected via buyer-supplier relationships rather than quantities and unit values associated with such transactions. They use a proportionality assumption which precludes estimating the elasticity of substitution across suppliers within an industry, as a buyer sourcing from two suppliers in the same industry will source the same amount given the assumption. Although lacking firm-to-firm price data, Dhyne et al. (2022) structurally estimate a similar elasticity in the context of imperfect competition models where they restrict the elasticity to be larger than 1 for mark-ups to make sense.

<sup>4</sup>As firm-to-firm data become common (Panigrahi, 2021; Demir et al., 2021; Dhyne et al., 2021; Alfaro-Urena et al., 2020), our methods can be used to quantify shock propagation through large/complex networks.

lockdowns in a large developing country. Our key policy motivation stems from the observation that policymakers worldwide are interested in quantifying the trade-off between strict lockdowns that prevent the spread of the virus but affect GDP through complex buyer-seller networks and more lenient measures that increase production and trade but potentially spread the virus. More importantly, even beyond the immediate Covid crisis, our estimates of how substitutable suppliers are within an industry will help policymakers quantify the economy-wide effects of any disruptive events (e.g., natural disasters or sanctions) on trade and production.

## 2 DATA AND CONTEXT

**Firm-to-firm trade.** Our primary data source is daily establishment-level transactions with distinct information on establishment locations.<sup>5</sup> This data is provided by the tax authority of a large Indian state with a diversified production structure, roughly 50% urbanization rates, and high levels of population density. To compare its size in terms of standard firm-to-firm transaction datasets, the population of this Indian state is roughly three times the population of Belgium, seven times the population of Costa Rica, and two times the population of Chile.

The data contains daily transactions between all registered establishments in this state and all registered establishments in India and abroad, from April 2018 to October 2020. This data is collected by the tax authority due to the creation of the *E-way Bill* system in April 2018, which was created to increase compliance for tax purposes. This is a major advantage in comparison to standard VAT firm-to-firm datasets with severe under-reporting, especially in developing countries. By law, any person dealing with the supply of goods and services whose transaction value exceeds 50,000 Rs (700 USD) must generate E-way bills. Transactions with values lower than 700 USD can also be registered, but it is not mandatory. The E-way bill is generated before transportation (usually via truck, rail, air, or ship), and the driver of the vehicle must carry the bill with them, or the entire extent of goods can be confiscated. Our data is generated from these E-way bills. This implies that our network is likely representative of relatively larger firms, but this threshold is sufficiently low such that we are confident we are capturing small firms as well.

Each transaction reports a unique tax code identifier for both the selling and buying establishments, all the items contained within the transaction, the value of the whole transaction, the value of the items being traded up to 8-digit HS codes,<sup>6</sup> quantity of each item, units, and

---

<sup>5</sup>While we use the term ‘firm’ in most parts of the paper, these data are actually at the more granular establishment level, and we can identify the parent firms for each establishment as well.

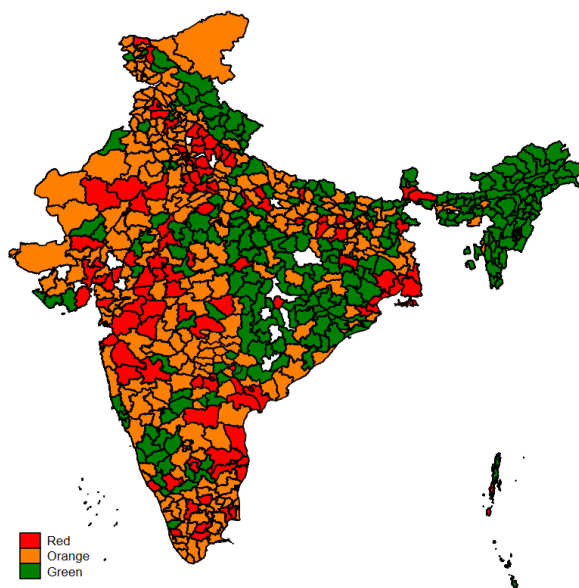
<sup>6</sup>The data partially reports items up to 8-digit HS codes. Until April 2021, in India it was only mandatory to report 4-digit HS codes of goods traded. See <https://economictimes.indiatimes.com/small-biz/gst/six-digit-hsn-code-in-gst-made-mandatory-from-april-1/articleshow/81780235.cms?from=mdr>. 97% of transactions report 4-digit HS codes, 40% report 8-digit HS codes. Given this, our main specifications are based on 4-digit HS codes.

mode of transportation. Each transaction also reports the ZIP code of the selling and buying firms, which we use to merge with other district-level data.

Since the data report both value and quantity of traded items, we construct unit values for each transaction. We also calculate average unit values at the 4-digit HS/month/seller/buyer level, the number of transactions and total value of the goods transacted. This is the foundation of our firm-to-firm dataset that we use in the analysis.

**Lockdowns.** On March 25th 2020, India unexpectedly imposed strict lockdown policies nationwide. The designated severity of the lockdown varied by districts, and was implemented nationwide at the district level, where each district was classified between *Red*, *Orange*, and *Green* zones according to the severity of Covid cases in each district. Yet, at that time, there were barely any Covid cases in India, as the entire country averaged about 50 cases a day (as opposed to about 400,000 cases a day the following year).

FIGURE 1: India's lockdown zones in March, 2020



**Notes:** Map shows the lockdown zones across Indian districts announced on March 25, 2020.

In Figure 1 we map the distribution of lockdowns across India. Districts in the red zone saw the strictest lockdown measures, with rickshaws, taxis and cabs, public transport, barber shops, spas, and salons remaining shut. E-commerce was allowed for essential services. Orange and green zone districts saw fewer restrictions. In addition to the activities allowed in red zones, orange zones allowed the operation of taxis and cab aggregators, as well as the inter-district movement of individuals and vehicles for permitted activities. In addition to the activities allowed in orange zones, buses were allowed to operate with up to 50% seating capacity and



bus depots with 50% capacity in green zones.<sup>7</sup>

Throughout the paper, we use this color scheme as the treatment across Indian districts. In particular, each firm is located within a district, so treated firms are located within a *Red*, *Orange*, or *Green* district between March and May 2020.

**Physical and cultural distance.** We use different measures of *distance* which we include as controls in our empirical results. The measures of geographic distance between districts calculate the length of the shortest distance between district centers. The measure of linguistic distance between Indian districts is from Kone et al. (2018) who using the commonly used ethno-linguistic fractionalization (EFL) index (Mira, 1964). This index measures the probability of two randomly chosen individuals from different districts speaking the same language.

**Other controls.** We control for different firm and district level time varying variables such as data on monthly number of cases, deaths, and recoveries from Covid-19 for all India at the district level from [www.Covidindia.org](http://www.Covidindia.org). For each firm, we construct two variables that measure the firm's exposure to global demand and supply shocks that vary at the HS-4 product and country level, following Hummels et al. (2014). The construction of these exposure variables are described in detail in online data Appendix C.

**Summary statistics.** We present some key summary statistics from the administrative trade data in Table A1. Panels A and B report the unique numbers of sellers, buyers, total sales (in million rupees), and total number of transactions separately in months January-March, April-June, and July-September, for years 2019 and 2020. The most noticeable pattern from the data is the large drop in all variables in 2020 in comparison to 2019, particularly during the April-June period, which coincided with the lockdown policies.

The total value of sales and the number of transactions both fell by almost 60% during April-June of 2020 compared to 2019. For reference, the fall in the value of sales was only 25% after the strict centralized lockdown was over (July-September) and only 15.6% before the lockdown (January-March) compared to the corresponding months in 2019.

To further understand the composition of economic activity of the Indian state of our analysis, in Table A2 we show what types of goods firms within the state sell and buy, and to which destinations. In our state, firms are mostly in the business of selling vegetables, plastics, and minerals; and of buying machinery, metals, and vegetables. In terms of the type of trade, firms in our state mostly sell to firms in other Indian states. This contrasts with how firms in our state buy intermediates, where the share of purchases that come from within the state is almost

---

<sup>7</sup><https://economictimes.indiatimes.com/news/politics-and-nation/lockdown-3-0-guidelines-for-red-zone/activities-prohibited/slideshow/75503925.cms>



the same as from other Indian states. Finally, exports and imports represent a non-negligible but rather small share of both sells and purchases.

Before using the lockdown variation to understand how firm to firm transactions are affected, we verify the stringency of these lockdowns in Figure A3 using google mobility data. The data shows how the number of visitors to (or the time spent in) categorized places change compared to baseline days. The baseline day is the median value from the 5-week period Jan 3 – Feb 6, 2020.<sup>8</sup> As is clear from the graph, until March 2020, there were essentially no differences in mobility trends across red, orange, or green zones. But starting in April 2020, we see that there is a substantial reduction in different types of activities (time spent in retail and recreation, grocery and pharmacy, parks, commuting, and workplaces) in red zones compared to green zones; with orange zones in between. People in red zones also spend more time at home compared to people in either orange or green zones. We notice that starting August 2020, a few months after the centralized lockdown was over, these differences start to reduce, and by December 2020 these differences, especially in workplace mobility, becomes small.

### 3 REDUCED-FORM EVIDENCE

In this section we describe our empirical strategy, and provide evidence showing the role of lockdown policies on key outcome variables for firm-to-firm trade. We show that the sudden Covid-19 lockdown policies between March and May 2020 led to a rise in unit values, and a fall in the monthly number of transactions between firms.<sup>9</sup> In subsequent sections, we exploit this variation to estimate firm-level elasticities of substitution across intermediate inputs.

#### 3.1 Empirical specifications

Our main reduced form specifications employ difference-in-differences specifications where we compare the unit values and the number of transactions both at seller level and seller-buyer level across *Red*, *Orange* and *Green* districts, before and after the lockdown. In our analysis at the seller level, the omitted (control) group are sellers located in *Green* districts and the base month is February 2020, two months before the enforcement of lockdown policies. At the seller-buyer level, the omitted group are sellers and buyers located in *Green* districts and the base month is February 2020.

---

<sup>8</sup>Source: [https://support.google.com/covid19-mobility/answer/9824897?hl=en&ref\\_topic=9822927](https://support.google.com/covid19-mobility/answer/9824897?hl=en&ref_topic=9822927)

<sup>9</sup>To see a similar application of this empirical strategy for domestic violence and economic activity in India, see Ravindran and Shah (2020) and Beyer et al. (2021).

**Seller-level regressions.** We estimate the following specification:

$$Y_{si,t} = \iota_{i,o(s)} + \iota_{i,t} + \sum_{t \neq 1} \beta_t Red_{o(s)} + \sum_{t \neq 1} \gamma_t Orange_{o(s)} + X\delta + \epsilon_{si,t}, \quad (1)$$

where  $Y_{si,t}$  are either unit values or the log number of transactions for seller  $s$  in HS-4 industry  $i$  in month  $t$ ,  $\iota_{i,t}$  are 4-digit HS-by-month fixed effects,  $\iota_{i,o(s)}$  are industry-by-district fixed effects (i.e. fixed effects based on the district  $o$  where seller  $s$  resides),  $X$  are controls that include number of Covid cases, deaths, and recoveries, and exposure to international demand and supply shocks as discussed in Appendix C. We control for the Covid cases and deaths since these are the variables on which the government based its lockdown decisions (Ravindran and Shah, 2020). The covariates of interest are  $Red_{o(s)}$  and  $Orange_{o(s)}$ . The first one is an indicator variable that equals 1 if seller  $s$  located in district  $o(s)$  experienced a severe lockdown, 0 otherwise. The second one equals 1 if seller  $s$  located in district  $o(s)$  experienced a mid-level lockdown, 0 otherwise. The excluded category are  $Green_o$  districts, where mild lockdown was imposed. The estimates of interest are  $\beta_t$  and  $\gamma_t$ . Our base time category is February 2020 which is just before lockdowns began. Standard errors are clustered at the seller's origin district level.

**Seller-buyer level regressions.** At the seller-buyer level we estimate the specification:

$$Y_{si,b,t} = \sum_{(x,z) \in \Omega} \sum_{t \neq 1} \beta_t^{xz} \left( \gamma_{o(s)}^x \times \gamma_{d(b)}^z \right) + \delta_{o(s)} + \delta_{d(b)} + \delta_{i,t} + \beta_1 \log dist_{od} + X\delta + \epsilon_{si,b,t} \quad (2)$$

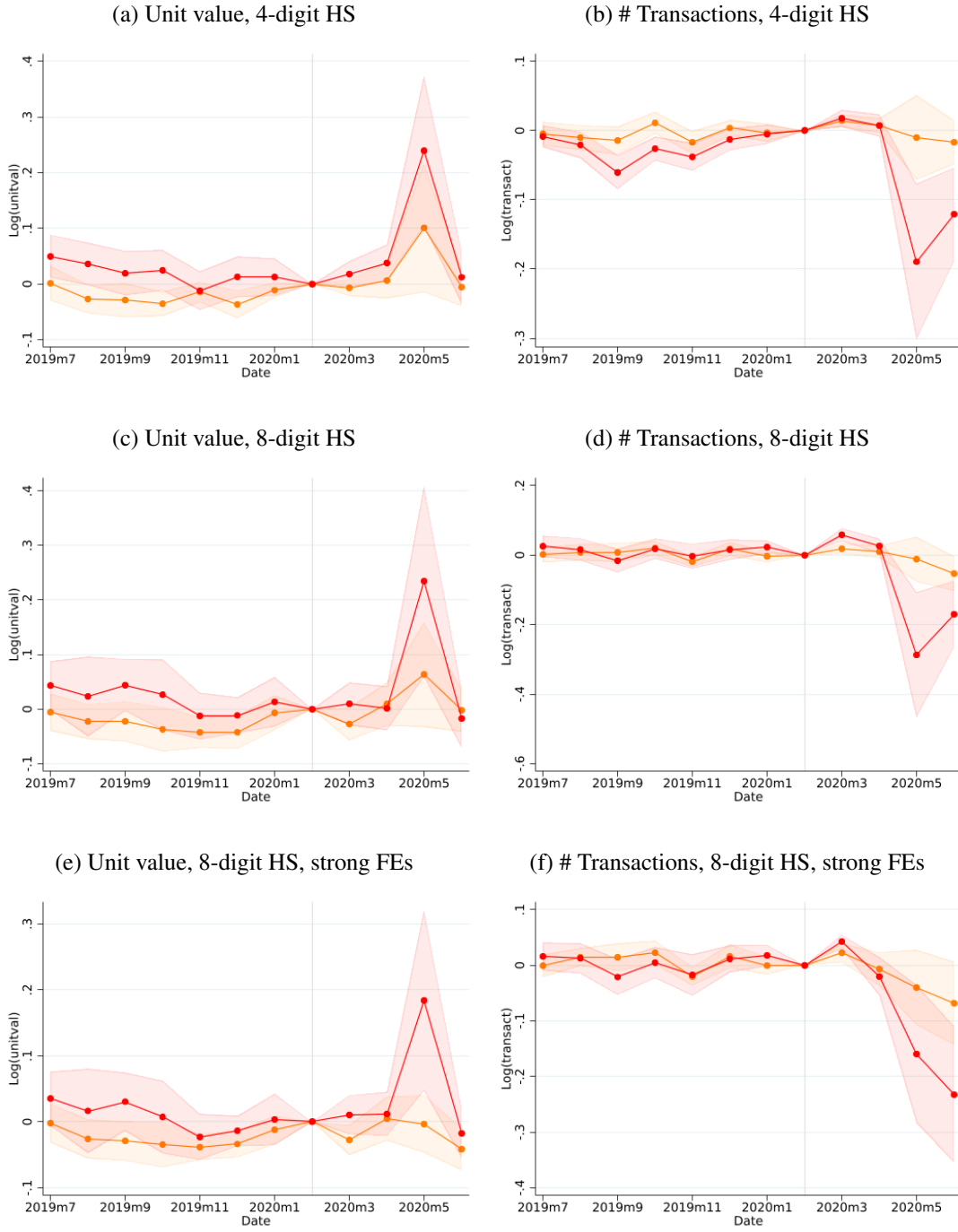
where  $Y_{si,b,t}$  are unit values or number of transactions in logs between seller  $s$  in HS-4 industry  $i$  and a buyer  $b$  in month  $t$ .  $\delta_{o(s)}$ ,  $\delta_{d(b)}$ , and  $\delta_{i,t}$  are origin, destination, industry-by-month fixed effects.  $dist_{od}$  is a vector of cultural and geographic distance variables, and  $X$  are controls that include number of Covid-19 cases, deaths, recoveries and exposures to international demand and supply shocks. The first term of the right-hand side contains our estimates of interest.  $(x,z) \in \Omega$  is a duple that contains the color  $x$  of seller's district, and the color  $z$  of buyer's district.  $\Omega$  is the set that includes all pairs except  $(Green, Green)$ , such that this is the excluded category when estimating Equation (2).  $\gamma_{o(s)}^x$  and  $\gamma_{d(b)}^z$  are thus dummy variables that equal 1 when seller  $s$  is located in district  $o$  located in lockdown zone  $x$ , and when buyer  $b$  is located in district  $d$  located in lockdown zone  $z$ , respectively. The estimates of interest are  $\beta_t^{xz}$ . Our base time category is February 2020 which is just before lockdowns began. Standard errors are two-way clustered at the origin and destination district level.

### 3.2 Reduced-Form Facts

In this section we present two facts from the specifications we laid out in the previous section.

**Fact 1: Sellers' unit values disproportionately rose and trade fell in more severe lockdown zones.** In the first two panels of Figure 2 we plot the coefficients  $\beta_t$  and  $\gamma_t$  from Equation (1), which represent changes in log unit values and log number of transactions with respect to cases in *Green* districts in February 2020 (i.e. the base category). In May 2020, sellers' unit values in *Red* districts rose by around 25pp with respect to the base category, and in *Orange* districts rose by around 10pp with respect to the base category.

FIGURE 2: Seller-level reduced-form event studies



**Notes:** This figure is comprised of 6 plots. Each plot shows estimates for  $\beta_t$  and  $\gamma_t$  from Equation (1). The values of the estimates are all in comparison to sellers in *Green* districts in February 2020. The dependent variable on the left side is in log unit values; on the right side, in log number of transactions. Each row varies by the definition of an industry, and the fixed effects included in the regression. In the first row, an industry is 4-digit HS codes and fixed effects HS/month and district. In the second row, an industry is 8-digit HS codes and fixed effects HS/month and district. In the third row, an industry is 8-digit HS codes and fixed effects HS/month and district/HS. Standard errors are clustered at the district level. All controls mentioned in the paper are included. The shaded area are confidence intervals.

At the same time, sellers' number of transactions in *Red* districts declined by around 20pp, and in *Orange* districts declined by around 3pp with respect to the base category. Additionally, as expected by the severity of the lockdown policies by color, the rise in unit values, and fall in number of transactions was larger for sellers in *Red* districts than for *Orange* ones. In both figures, we find no evidence of pre-trends, implying that there were likely no differences in the trends of unit values or number of transactions between red, orange, and green districts before the lockdown.

The middle two panels of Figure 2 repeats the same exercise with a finer industry definition, using 8-digit HS codes. Results remain virtually the same. In the last row of Figure 2 we include a stronger set of fixed effects (e.g., district-by-industry), and results remain the same.

FIGURE 3: Unit Value, Seller-Buyer Level Regressions

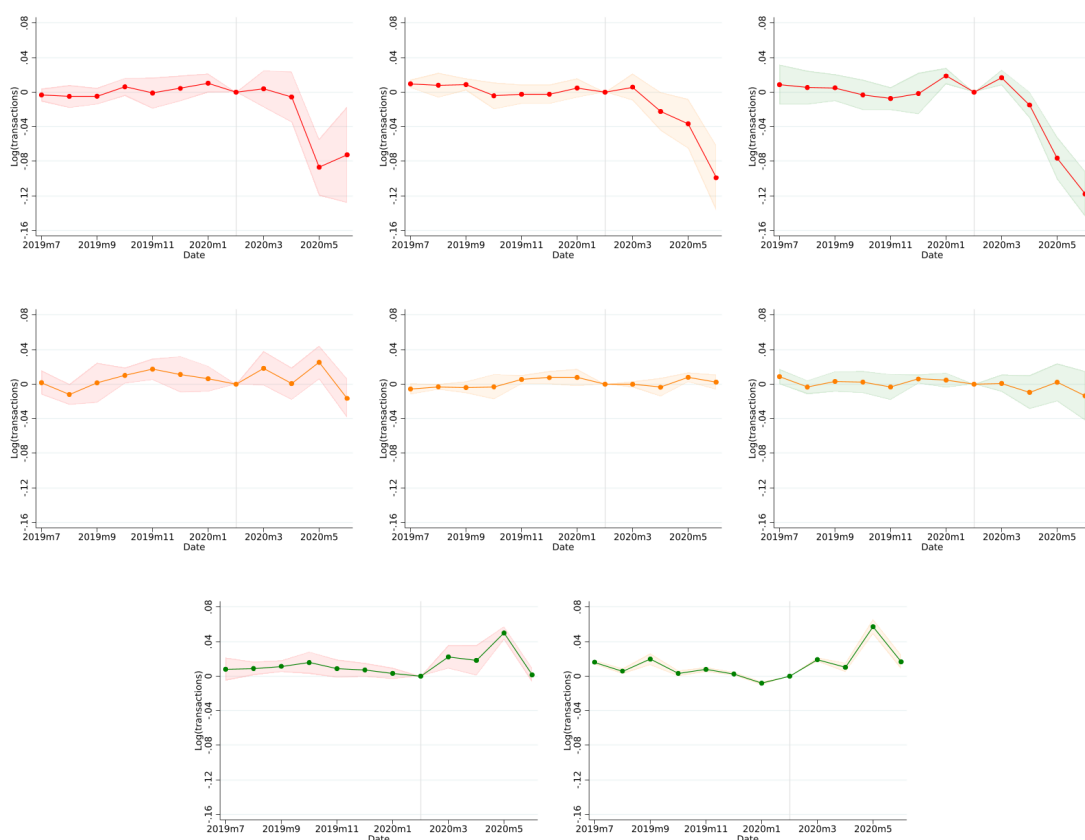


**Notes:** In each plot, the horizontal axis is the month, and the vertical one is the estimate of interest associated with log unit values as in Equation (2) for each month. Regressions include industry/month, origin district, and destination district fixed effects. Standard errors are two-way clustered at the origin and destination state level. An industry is 4-digit HS codes. All controls mentioned in the paper are included. The vertical line in January 2020 splits pre and post-lockdown periods. The baseline category are sellers and buyers located in *Green* districts on January 2020. The color of the line denotes the color of the district the seller is located, while the color of the shaded confidence interval denotes the color of the district the buyer is located.

**Fact 2: Equilibrium unit values rose and number of transactions fell in more severe lockdown zones.** We now report the results from our seller/buyer-level specification. In Figures 3 and 4 we report the estimates for  $\beta_t^{xz}$  in Equation (2), where the estimates are in comparison to cases when both sellers and buyers were located in *Green* districts in February 2020.

In the first row of Figure 3 we plot the coefficients from regression (2) where the seller is in the red zone, and the buyer is in red, orange, and green zones respectively. Similarly, in the second row of Figure 3, we plot the coefficients from regression (2) where the seller is in the orange zone, and in the third row, we plot the coefficients from regression (2) where the seller is in the green zone (and the buyer is in red and orange zones respectively).

FIGURE 4: Number of Transactions, Seller-Buyer Level Regressions



**Notes:** In each plot, the horizontal axis is the month, and the vertical one is the estimate of interest associated to log number of transactions as in Equation (2) for each month. Regressions include industry/month, origin district, and destination district fixed effects. Standard errors are two-way clustered at the origin and destination state level. An industry is 4-digit HS codes. All controls mentioned in the paper are included. The vertical line in January 2020 splits pre and post-lockdown periods. The baseline category are sellers and buyers located in *Green* districts on January 2020. The color of the line denotes the color of the district the seller is located, while the color of the shaded confidence interval denotes the color of the district the buyer is located.

There are two main lessons from these figures. First, even after controlling for bilateral resistance terms, trade costs, and additional covariates, unit values rose and number of transactions fell with respect to the base category (both buyer and seller in green zones). The rise in unit values was as much as 45pp, and the fall in transactions as high as 12pp. Second, these

changes seem to be proportional to the severity of the lockdowns for both sellers and buyers. Once again, there is no evidence of differential pre-trends across zones leading up to the shock.

Our two facts jointly imply that prices where either seller or buyers were located in red districts were higher during the lockdown in comparison to districts where the lockdowns were mild (green zones). This suggests that the lockdown indeed induced variation in prices that we will later leverage to estimate elasticities of substitution across intermediates.

## 4 MODEL

We build a quantitative general equilibrium model of firm-to-firm trade based on [Baqaee and Farhi \(2019\)](#), where the productive sector is perfectly competitive.<sup>10</sup> We adapt the general nested CES structure to reflect the possibility that suppliers within the same industry could be substitutes or complements, derive estimating equations, and use the model to simulate the effects of negative productivity shocks on GDP. Firms combine inputs in a CES fashion under three tiers. In the first tier, firms combine labor and aggregated intermediates. In the second tier, aggregated intermediates are a combination of intermediates by industry composites. In the third tier, industry composites are constructed by suppliers of intermediates.

There are  $N$  firms producing  $N$  goods using the production function

$$y_{nj} = A_n \left( w_{nl} (l_n)^{\frac{\alpha-1}{\alpha}} + (1 - w_{nl}) (x_{nj})^{\frac{\alpha-1}{\alpha}} \right)^{\frac{\alpha}{\alpha-1}}, \quad (3)$$

where  $n$  is the firm and  $j$  is the firms' industry.  $l_n$  is the labor used by firm  $n$ ,  $x_{nj}$  is the composite intermediate input used by firm  $n$  in industry  $j$ ,  $\alpha$  is the elasticity of substitution between labor and the composite material input and  $w_{nl}$  is the intensity of labor in production. The composite material input in turn consists of inputs from the  $I$  different industries in the economy, and is:

$$x_{nj} = \left( \sum_{i=1}^I w_{i,nj}^{\frac{1}{\zeta}} (x_{i,nj})^{\frac{\zeta-1}{\zeta}} \right)^{\frac{\zeta}{\zeta-1}}, \quad (4)$$

where  $\zeta$  is the elasticity of substitution between inputs from different industries, and  $w_{i,nj}$  is the importance of input from industry  $i$  for buyer  $b$  in industry  $j$ .  $x_{i,nj}$  are intermediate inputs from industry  $i$  going to firm  $n$  in industry  $j$ ,<sup>11</sup> which are in turns constructed as

$$x_{i,nj} = \left( \sum_{m=1}^{N_i} \mu_{mi,nj}^{\frac{1}{\epsilon}} x_{mi,nj}^{\frac{\epsilon-1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}}, \quad (5)$$

<sup>10</sup>We do not rely on models featuring market power ([Edmond et al., 2018](#); [Alvarez et al., 2021](#)) since the evidence from the data suggests that the market structure in this Indian state is closer to perfect competition. The median HHI across 4-digit HS industries is 0.1041, which implies that industries are unconcentrated.

<sup>11</sup>We exclude foreign intermediate goods since they are not exposed to Indian Covid-19 lockdown shocks.



where  $x_{mi,nj}$  are intermediate inputs from firm  $m$  in industry  $i$  sold to firm  $n$  in industry  $j$ , and  $\mu_{mi,nj}$  is the importance of input from supplier  $i$  in industry  $m$  in the production of buyer  $j$  in industry  $n$ . We consider a fixed set of firms  $F$  and industries  $I$ , where  $N = |F|$  is the total number of firms in the economy, and  $N_i$  is the number of firms in industry  $i$ .  $\epsilon$  is the elasticity of substitution across suppliers within the same industry. The above production functions work for reproducible factors. For non-reproducible factors, in our case labor, the production function is an endowment:  $Y_f = 1$ .

Industry 0 represents the final consumption of the household and is given by

$$C = \left( \sum_i^N w_{0i} (c_i)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \quad (6)$$

where  $\sum_i w_{0i} = 1$  and  $\sigma$  is the elasticity of substitution in consumption.

**Model in standard-form.** To write the economy in standard form as in [Baqaee and Farhi \(2020\)](#), we define a new input output matrix  $\hat{\Omega}$  which has dimension  $2+N+I$ , where the first dimension represents the household's consumption aggregator, the next dimension corresponds to factors, here only labor, the next  $N$  dimensions are the  $N$  firms that supply inputs to the CES aggregates and the next  $I$  dimensions are the CES aggregates of intermediate inputs of these firms that directly go into the firm's production function. Let us denote the vector of elasticities by  $\hat{\theta}$ , where  $\hat{\theta} = (\sigma, \alpha, \zeta, \epsilon)$ .

Formally, a nested-CES economy in standard form is defined by  $(\hat{\Omega}, \hat{\theta})$ . What distinguishes factors from goods is that factors cannot be produced. The  $(2+N+I) \times (2+N+I)$  input-output matrix  $\hat{\Omega}$  is the matrix whose  $(i, j)$  element is equal to the steady-state value of  $\Omega_{ij} = \frac{p_i x_{ij}}{p_i y_i}$ , which is the expenditure share of the  $i$ th firm on inputs from the  $j$ th supplier as share of the total revenue of firm  $i$ , where, note that, every supplier is a CES aggregate. The Leontief inverse is  $\psi = (1 - \Omega)^{-1}$ . Intuitively, the  $(i, j)$ th element of  $\psi$  (the Leontief inverse) is a measure of  $i$ 's total reliance on  $j$  as a supplier. It captures both the direct and indirect ways through which  $i$  uses  $j$  in its production. Let us also denote the sales of producer  $i$  as a fraction of GDP by  $\lambda_i$ , where  $\lambda_i = \frac{p_i y_i}{\sum_j p_j c_j}$ .

The input output covariance operator is given by

$$\text{Cov}_{\Omega_k}(\psi_{(i)}, \psi_{(j)}) = \sum_{l=1}^{2+N+I} \Omega_{kl} \psi_{li} \psi_{lj} - \left( \sum_{l=1}^{2+N+I} \Omega_{kl} \psi_{li} \right) \left( \sum_{l=1}^{2+N+I} \Omega_{kl} \psi_{lj} \right). \quad (7)$$

This operator measures the covariance between the  $i$ th and the  $j$ th columns of the Leontief inverse using the  $k$ th row of the input output matrix as distribution. The second-order macroeconomic impact of microeconomic shocks in this economy is given by:

$$\frac{d^2 \log Y}{d \log A_j d \log A_i} = \frac{d \lambda_i}{d \log A_j} = \sum_k (\theta_k - 1) \lambda_k \text{Cov}_{\Omega(k)}(\Psi_{(i)}, \Psi_{(j)}). \quad (8)$$

For detailed derivation of this, see the Appendix of [Baqee and Farhi \(2019\)](#). To get an intuition of how firm-level shocks can propagate through supply chains, consider a specific example: firm  $j$ , located in the red zone, suffers a negative productivity shock, given by  $d \log A_j < 0$ .

The second order term captures the reallocation effect: In response to a negative shock to industry  $j$ , all industries  $k$  that are downstream of  $j$  may readjust their demand for all other inputs. Crucially, the impact of such readjustments by any given  $k$  on the output of industry  $i$  depends on the size of industry  $k$  as captured by its *Domar* weight  $\lambda_k$ , the elasticity of substitution  $\theta_k$  in  $k$ 's production function, and the extent to which the supply chains that connect  $i$  and  $j$  to  $k$  coincide with one another, as given by the covariance term.

#### 4.1 Equations to estimate firm-level elasticity of substitution across suppliers

Using the model outlined above, in this section we derive the firm-level elasticity of substitution across suppliers within an industry. We introduce a notation change to facilitate the exposition: a firm  $n$  can be either a buyer  $b \in F$  or a seller  $s \in F$ . A firm  $b$  in industry  $j \in I$  maximizes profits subject to its technology and to a CES bundle of intermediate inputs:

$$\max_{\{l_{bj}, x_{si,bj}\}} p_{bj} y_{bj} - w_{bj} l_{bj} - \sum_i \sum_s p_{si,bj} x_{si,bj}$$

subject to (3), (4), and (5).  $\epsilon$  from Equation (5) is the elasticity of substitution across different suppliers within the same industry. This is the key elasticity we want to estimate. Note that the results of this estimation procedure holds with any CES production function with an arbitrary number of nests, as long as the lowest nest consists of suppliers within the same HS-4 industry. Details about the optimization problem are in Appendix D.1. The maximization problem yields the following expression:

$$\log \left( \frac{PM_{si,bj}}{PM_{i,bj}} \right) = (1 - \epsilon) \log \left( \frac{p_{si,bj}}{p_{i,bj}} \right) + \log (\mu_{si,bj}), \quad (9)$$

where  $p_{i,bj} = \left( \sum_{s'} (p_{s'i,bj}^{1-\epsilon} \mu_{s'i,bj}) \right)^{\frac{1}{1-\epsilon}}$  is a CES price index,  $PM_{si,bj} \equiv p_{si,bj} x_{si,bj}$ , and  $PM_{i,bj} \equiv \sum_s PM_{si,bj}$ , and  $\log (\mu_{si,bj})$  is the error term. This is our main estimating equation for the firm-level elasticity of substitution parameter  $\epsilon$  which we take to the data, as will be described in detail in Section 5.

## 4.2 Equations to estimate firm-level elasticity of substitution across industries

In this section, we derive conditions from the model to estimate the firm-level elasticity of substitution across industries. We rewrite the maximization problem of the firm such that it maximizes

$$\max_{\{l_{bj}, x_{i,bj}\}} p_{bj} y_{bj} - w_{bj} l_{bj} - \sum_i p_{i,bj} x_{i,bj}$$

subject to (3), (4), and  $p_{i,bj} = \left( \sum_s \mu_{si,bj} p_{si,bj}^{1-\epsilon} \right)^{\frac{1}{1-\epsilon}}$ .  $\zeta$  from Equation (4) is the firm-level elasticity of substitution across industries  $i$  we estimate. Notice that in this case, we need values for  $\epsilon$  and  $\mu_{si,bj}$  to calculate prices. We consider  $\epsilon = \hat{\epsilon}$ , where  $\hat{\epsilon}$  is our estimate, and we recover  $\mu_{si,bj}$ . Details on the optimization problem are in Appendix D.2.1. The maximization problem yields the following expression:

$$\log \left( \frac{PM_{i,bj}}{PM_{bj}} \right) = (1 - \zeta) \log \left( \frac{p_{i,bj}}{p_{bj}} \right) + \log(w_{i,bj}), \quad (10)$$

where  $p_{bj} = \left( \sum_{i'} \left( p_{i',bj}^{1-\zeta} w_{i',bj} \right) \right)^{\frac{1}{1-\zeta}}$  is a CES price index,  $PM_{i,bj} \equiv p_{i,bj} x_{i,bj}$ , and  $PM_{bj} \equiv \sum_i PM_{i,bj}$ , and  $\log(w_{i,bj})$  is the error term. This is our estimating equation for the firm-level elasticity of substitution  $\zeta$  which we take to the data, as described in Section 5.

## 5 ESTIMATION

In this section, we discuss how we estimate the primary elasticities in our model. The vector of parameters is  $\hat{\theta} = (\sigma, \alpha, \zeta, \epsilon)$ . We set the elasticity of substitution between different consumption varieties  $\sigma = 4$  (Broda and Weinstein, 2006), and the elasticity of substitution between labor and the composite intermediate input  $\alpha = 0.5$  (Baqae and Farhi, 2019). We now estimate the firm-level elasticity of substitution across suppliers ( $\epsilon$ ) and the firm-level elasticity of substitution across industries ( $\zeta$ ) leveraging variation in the lockdown zones.

### 5.1 Estimating equations for $\epsilon$ and $\zeta$

In order to estimate  $\epsilon$  from Equation (9), the first major challenge we face is that the price index  $p_{i,bj}$  includes the unobserved quantity  $\mu_{si,bj}$  which denotes the importance of input from supplier  $s$  in industry  $i$  in the production of buyer  $b$  in industry  $j$ . This unobserved quantity could depend on a number of factors such as unobserved input demand shocks or the buyer's preference for certain inputs. In order to construct changes in price indices that are observable, we follow Redding and Weinstein (2020) in assuming that the overall importance of an indus-

try in a buyer's input use does not change between two consecutive months, even though the importance of inputs from suppliers within an industry can change.<sup>12</sup> We arrive at the equation below that links the overall expenditure share on a certain supplier's input in an industry to the corresponding relative price:

$$\log \left( \frac{\widehat{PM}_{si,bj,t}}{\widehat{PM}_{i,bj,t}} \right) = \omega_{b,t} + \omega_{i,t} + \omega_{b,i} + \omega_{s,i} + (1 - \epsilon) \log \left( \frac{\widehat{p}_{si,bj,t}}{\widehat{p}_{i,bj,t}} \right) + \log \left( \widehat{\lambda}_{i,bj,t} \widehat{s}_{i,bj,t}^* \right) + X\beta + \xi_{si,bj,t}, \quad (11)$$

where  $\widehat{x}_t = \frac{x_t}{x_{t-1}}$  are variables in changes with respect to the previous month.  $[\omega_{b,t}, \omega_{i,t}, \omega_{b,i}, \omega_{s,i}]$  is a set of fixed effects, including buyer-by-month, product-by-month, buyer-by-product, and seller-by-product fixed effects.  $\widehat{p}_{i,bj,t} = \prod_{s \in \Omega_{i,bj,t}^*} p_{si,bj,t}^{\frac{1}{N_{i,bj,t}^*}}$  is a geometric mean of unit values across common suppliers, where  $\Omega_{i,bj,t}^* \equiv \Omega_{i,bj,t} \cap \Omega_{i,bj,t-1}$  is the set of common suppliers for buyer  $b$  that appear in both the current and previous month, and  $N_{i,bj,t}^* \equiv |\Omega_{i,bj,t}^*|$  is the number of common suppliers for buyer  $b$  in month  $t$ .  $X$  are controls, including exposure to foreign demand and supply shocks, the number and severity of Covid cases, and geographic and cultural distance.

Our setup has the advantage that we can decompose the change in price buyer  $b$  pays for inputs from seller  $s$  between  $\widehat{p}_{i,bj,t}$ , the change in expenditure share  $\widehat{s}_{i,bj,t}^*$  and a [Feenstra \(1994\)](#) correction term  $\widehat{\lambda}_{i,bj,t}$  that takes into account the fact that sellers enter and exit in the data. More details are in Appendix D.1.3. Standard errors are two-way clustered at the origin and destination state level.

Now, to estimate  $\zeta$  from Equation (10), there are two issues to address. First, notice that the price index  $p_{i,bj}$  is a function of (unobservable) demand shocks  $\mu_{si,bj,t}$ , and  $\epsilon$ . Second, the price index  $p_{bj,t}$  is also a function of unobservable industry-level demand shocks  $w_{i,bj,t}$ , which makes their computation challenging.

First, we construct price indices as  $p_{i,bj,t} \equiv \left( \sum_s \mu_{si,bj,t} p_{si,bj,t}^{1-\widehat{\epsilon}} \right)^{\frac{1}{1-\widehat{\epsilon}}}$ , where  $\widehat{\epsilon}$  are estimated previously,  $p_{si,bj,t}$  come directly from the data, and demand shocks  $\mu_{si,bj,t}$  are constructed recursively. This recursive construction of demand shocks come from predicting residuals from Equation (11) and setting an initial value for shocks  $\mu_{si,bj,0}$  (Appendix D.2.2).

Second, we construct buyer-level price indices  $p_{bj,t}$  following [Redding and Weinstein \(2020\)](#). We assume that the overall importance of the composite intermediates at HS-4 level in the production function does not change between consecutive months. As such, we can construct this price independent of industry-level demand shocks  $w_{i,bj,t}$  after controlling for buyers' expenditure shares by industry. More details about this are in Appendix D.2.1.

<sup>12</sup>This assumption simply requires that, for instance, a shoemaker's overall preference for leather in shoe-manufacturing does not change, although its preference for leather from certain suppliers can change. That is, demand-shocks may change  $\mu_{si,bj,t}$  (e.g., the demand for leather from certain suppliers), but the geometric mean of  $\mu_{si,bj,t}$  across suppliers within an industry is stable between  $t$  and  $t-1$ . This enables us to construct changes in price indices that are not dependent on  $\mu_{si,bj,t}$ , but are directly observed in the data (details in Appendix D.1.2).

We then derive the following expression we take directly to the data:

$$\log \left( \frac{\widehat{PM}_{i,bj,t}}{\widehat{PM}_{bj,t}} \right) = \omega_{b,t} + \omega_{i,t} + \omega_{b,i} + (1 - \zeta) \log \left( \frac{\widehat{p}_{i,bj,t}}{\widehat{p}_{bj,t}} \right) + \log(\tilde{s}_{bj,t}) + \xi_{i,bj,t}, \quad (12)$$

where  $[\omega_{b,t}, \omega_{i,t}, \omega_{b,i}]$  are a set of buyer-by-month, product-by-month, and buyer-by-product fixed effects.  $\tilde{p}_{bj,t} \equiv \prod_{i=1}^{N_{bj,t}} \tilde{p}_{i,bj,t}^{\frac{1}{N_{bj,t}}}$  is the geometric mean of unit values across industries that buyer  $b$  sources from, and  $\tilde{s}_{bj,t} \equiv \prod_{i=1}^{N_{bj,t}} \tilde{s}_{i,bj,t}^{\frac{1}{N_{bj,t}}}$  is the geometric mean of expenditure shares across industries. Detailed derivations are in Appendix D.2.

## 5.2 Addressing endogeneity concerns

OLS estimates of  $\epsilon$  are biased if unobserved demand-side shocks (changing  $\mu_{si,bj,t}$ ) drive changes in prices and expenditure shares. The firm-level elasticity of substitution is a function of the slope of the buyer's input demand curve, and hence simultaneous shifts in the demand and supply curves induced by the Covid-19 shock can also bias our estimates. For example, if Covid-19 induced demand shocks led to contractions in buyers' income and at the same time supply-shocks lead to contractions in the sellers supply, the demand curves will look flatter (estimated  $\epsilon$  higher) compared to the unbiased value of  $\epsilon$ . Additionally, measurement error in input prices, proxied by unit values, may induce attenuation biases.

Our estimation strategy therefore involves using the sudden demarcations of lockdown zones that restrict economic activity in certain Indian districts as an instrumental variable when estimating this equation in two-stage least squares (2SLS). We use the disruptions in prices caused by sudden lockdowns that made it costlier for sellers in *Red* and *Orange* zones to produce and send their intermediate goods. The idea is that, after controlling for the lockdown zones the buyer is located in, exposure to international demand and supply shocks, the number and severity of regional Covid-19 cases, the variation in prices facing a buyer are driven by supply shocks induced by policy mandated sudden changes in the seller's lockdown zones. In addition, since the goods from the seller to the buyer have to transit through several districts located in different lockdown zones facing different severity in the movements of trucks and border controls, changes in the costs of transportation induced by these lockdowns provide another source of exogenous variation to estimate the firm-level elasticity of substitution.

To formalize the intuition behind our identification strategy, following the standard practice in the trade literature, we assume that prices can be separated between prices at the origin and a trade cost. In logs and in changes, this is

$$\log(\widehat{p}_{si,bj,t}) = \log(\widehat{\tau}_{s,b,t}) + \log(\widehat{p}_{si,t}).$$

Here we can see the type of variation driving the two types of instruments we use. First, exogenous shifters to prices at the seller level  $p_{si,t}$ , such as economic restrictions induced by the lockdown zone the seller is located in, help us obtain unbiased estimates of the elasticity  $\epsilon$ . Second, exogenous shifters at the seller-buyer level, for example, changes in transportation costs  $\tau_{s,b,t}$  driven by the lockdown zones of the districts the goods pass through, also induce the needed variation. We now describe each of these instruments and then implement them within our estimation strategy.

**Seller-level instruments.** We need supply-side shifters to obtain unbiased elasticities of substitution. In that sense, shocks induced by the Covid lockdown policies that only impact sellers would provide that variation. In Equation (13) below we formalize this intuition, so

$$\log(\widehat{p}_{si,bj,t}) = \beta^R Red_{o(s)} Lock_t + \beta^O Orange_{o(s)} Lock_t + \epsilon_{si,bj,t}^\nu, \quad (13)$$

where  $Lock_t$  is a dummy variable that equals 1 for the months from March to May of 2020, which are the months when the lockdown policies were implemented, 0 otherwise, and  $Red_{o(s)}$  and  $Orange_{o(s)}$  are indicator variables that equal 1 whenever seller  $s$  was located in *Red* or *Orange* districts, respectively.

**Seller/Buyer-level instruments.** The transportation of supplies from the location of the supplier to the buyer implies going through different districts, each of which are affected by lockdown policies in different ways. Intuitively, a route that contains more *Red* districts should increase the cost of transportation in contrast with a route with no *Red* districts. We construct instruments that capture that idea. We allow trade cost to change over time such that we can leverage the Covid lockdown policy. In particular, we assume

$$\tau_{sb,t} = traveltime_{sb,t}^\sigma.$$

After considering this functional form for trade costs into the expression of prices and log-differencing, we obtain

$$\log(\widehat{p}_{si,bj,t}) = \sigma \log(\widehat{traveltime}_{sb,t}).$$

We leverage the Covid-19 lockdown as an exogenous shifter that only influences travel time between locations of seller  $s$  and buyer  $b$ , as reflected in Equation 14 below.

$$\log(\widehat{p}_{si,bj,t}) = \beta^R Red_{o(s)d(b)} Lock_t + \beta^O Orange_{o(s)d(b)} Lock_t + \epsilon_{si,bj,t}^\nu. \quad (14)$$

Detailed derivations are in Appendix D.1.4.  $Red_{o(s)d(b)}$  and  $Orange_{o(s)d(b)}$  are the share of districts designated as *Red* and *Orange*, respectively, along the route between seller  $s$  and buyer

b. We constructed these variables using Dijkstra algorithm for least-cost routes. Details about the implementation of this algorithm are in Appendix C.

Finally, we also instrument the changes in relative prices in Equation 12 to estimate  $\zeta$ . We do this because of potential unobservable industry-level demand shocks that also induce an upward bias to estimates of  $\zeta$ . To construct our instruments, we leverage the seller-level and seller/buyer-level instruments we used to estimate  $\epsilon$  and calculate weighted averages across suppliers to instrument on the change of relative prices for buyers. The intuition is that buyers that purchased inputs either from a larger share of sellers in *Red* zones, or from sellers located in districts where the route is comprised or a larger share of *Red* zones were more exposed to Covid-19 lockdowns. More details are in Appendix D.2.3.

**Discussion of instruments.** The instruments induce buyers of certain types to be more affected than others based on their production networks. The Local Average Treatment Effect (LATE) may not represent the Average Treatment Effect (ATE) if buyers in *Red*, *Orange*, and *Green* zones already traded intensively with sellers in certain lockdown zones, and there is heterogeneity in responses. For instance, if buyers in *Red* traded mostly with sellers in *Red*, then our instrument may estimate effects on firms induced by having more *Red* sellers, and so it would upweight effects on buyers in *Red*. In Figure A2 we run two sets of balance check to investigate these patterns. These checks show that, in general, sellers from *Red*, *Orange*, and *Green* zones had similar interactions with buyers from *Red*, *Orange*, and *Green* zones.

We also consider whether certain industries source intensively from firms located in certain zones. For instance, if all the rubber supply of firms in this production network comes from suppliers in *Red* zones, then buyers of rubber would find it increasingly difficult to find suppliers. Once again, if there is heterogeneity in responses by industry, our estimated LATE elasticity would weigh the rubber industry higher than non-rubber industries. While not a source of bias, it does affect the interpretation of the estimated parameter. In Figures A2g and A2h, we plot the shares of total purchases of each industry that are sourced from firms in *Red*, *Orange*, and *Green* zones. With the exception of the small HS industry 19 (arms and ammunitions), there is no noticeable degree of concentration of suppliers from any particular zone.

### 5.3 Elasticity estimation results

In this section we show results of the estimation of both firm-level elasticities of substitution across suppliers within an industry, and then across industries.



### 5.3.1. Firm-level elasticities of substitution across suppliers

First, we report OLS estimates in Table 1. The implied elasticities exhibit a robust value of 0.78 across all the different specifications. In column (1), we include both buyer/month and HS/month fixed effects. In column (2) we also include buyer/HS and seller/HS fixed effects. We obtain a similar elasticity of 0.77. To test whether our estimates vary by industry aggregation, in columns (3) and (4) the estimations are based on 6-digit and 8-digit HS codes. The elasticities are around 0.75, so the estimates do not significantly change. Since these elasticities are below 1, these estimates suggest that, at the firm level, suppliers act as complements rather than substitutes for buyers. This is important for aggregate incomes since, from Equation (8) we can see that, once we take into account second order effects, an elasticity of substitution less than 1 implies that the aggregate impacts of negative shocks are amplified.

TABLE 1: OLS, firm-level elasticity of substitution across suppliers

	(1)	(2)	(3)	(4)
$\log\left(\frac{\hat{p}}{\bar{p}}\right)$	0.2171	0.2222	0.2506	0.2441
	(0.0133)	(0.0147)	(0.0324)	(0.0352)
$\epsilon$	0.7828	0.7777	0.7493	0.7558
$R^2$	0.4177	0.4601	0.4838	0.4958
Obs	2028039	1966591	851483	993583
HSN digits	4	4	6	8
Buyer/month FE	Y	Y	Y	Y
HSN/month FE	Y	Y	Y	Y
Buyer/HSN FE		Y	Y	Y
Seller/HSN FE		Y	Y	Y

**Notes:** OLS estimates from Equation (11). The first row reports the estimates associated with changes in relative unit values in logs. Standard errors are two-way clustered at the origin and destination state level, and are reported in parentheses below each estimate. The fifth row reports the implied value for  $\epsilon$ , which is 1 minus the estimate on the first row. The table contains four columns. Each column correspond to different specifications on how we define an industry (4-digit, 6-digit, or 8-digit HS codes) and of fixed effects, as pointed out by the last five rows of the table. All specifications include the controls mentioned in the paper.

Nevertheless, as we describe in the previous section, it is likely that OLS estimates are contaminated by simultaneous demand shocks that happened during Covid-19. In Table 2 we report 2SLS estimates based on our proposed instruments. We find evidence that inputs across different suppliers of a firm within the same 4-digit HS industry are highly complementary, ranging from 0.49–0.65, depending on the set of fixed effects and instruments we use. Our preferred specification is column (3) with an elasticity of 0.55, where we use both the seller and the seller-buyer level instrument, essentially deriving variation from both sell-

ers' production costs and transportation costs. We include buyer/month and HS/month fixed effects that account for time-varying demand shocks, and also account for entry/exit with the [Feenstra \(1994\)](#) term. Each specification reports a high Kleibergen-Paap F-statistic, indicating that our instruments are statistically relevant. In columns (1) and (2) we use the seller-level and seller/buyer-level instruments separately. The elasticities are 0.49 and 0.6 respectively, which also reflect complementarity. Finally, in column (4) we also include buyer/HS and seller/HS fixed effects, and the elasticity rises to 0.66.

TABLE 2: 2SLS, firm-level elasticity of substitution across suppliers

	(1)	(2)	(3)	(4)
$\log\left(\frac{\hat{p}}{\bar{p}}\right)$	0.5042	0.3945	0.4538	0.3409
	(0.2129)	(0.0933)	(0.1389)	(0.1068)
$\epsilon$	0.4957	0.6054	0.5461	0.6590
Obs	2854292	2028039	2028039	1966591
K-PF	48.232	133.688	143.413	248.977
Seller IV	Y		Y	Y
Bilateral IV		Y	Y	Y
Buyer/month FE	Y	Y	Y	Y
HSN/month FE	Y	Y	Y	Y
Buyer/HSN FE				Y
Seller/HSN FE				Y

**Notes:** IV-2SLS estimates from Equation (11). The set of common suppliers of buyer  $b$  is  $\Omega_{i,bj,t}^* = \Omega_{i,bj,t} \cap \Omega_{i,bj,t-1}$ . That is, a supplier  $s$  of buyer  $b$  is considered *common* if they also traded during the previous month. The first stage uses either bilateral or seller-level instruments, as pointed out by rows six and seven. Bilateral instruments correspond to Equation (14), while seller-level instruments correspond to Equation (13). The first row reports estimates associated with changes in relative unit values in logs. Standard errors are two-way clustered at the origin and destination state level, and are reported in parentheses below each estimate. The fourth row reports the Kleibergen-Paap F statistic from the first stage. The fifth row reports the implied value for  $\epsilon$ , which is 1 minus the estimate on the first row. An industry is 4-digit HS codes and the treatment period is March-May 2020. The table contains four columns. Each column corresponds to different combinations of instruments and of fixed effects, as pointed out by the last six rows. All specifications include the controls mentioned in the paper.

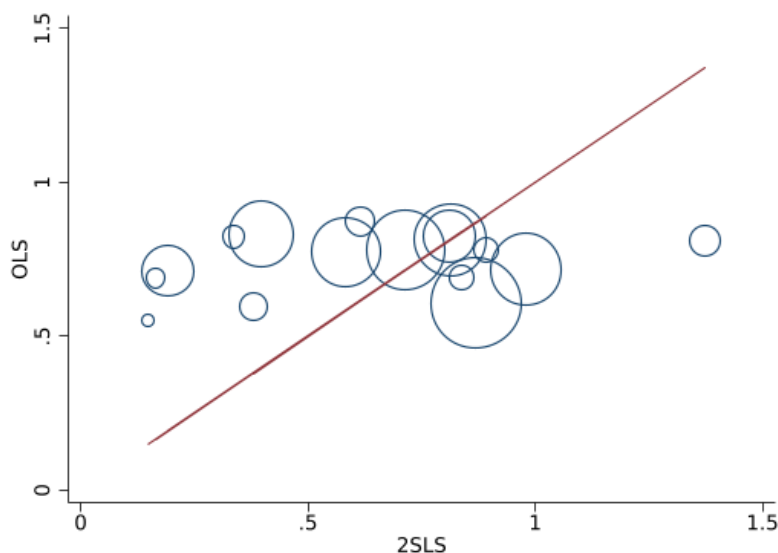
The IV estimates for  $\epsilon$  are smaller than the OLS estimates. As discussed in Section 5.2, the bias is in the expected direction if we expect the Covid-19 shock to also induce negative demand shocks, thereby biasing up OLS estimates of  $\epsilon$ . We may expect that our estimated elasticity be lower for the sub-sample of buyers who did not have more than one supplier to source inputs from. In Table A3, we restrict our sample to cases when a buyer traded with at-least two sellers in two consecutive periods. Column (3), our preferred specification, yields an elasticity of substitution of 0.58, very close to the estimate from our main specification.

To examine differences by the level of aggregation of the industry, we rerun our main

specification in Table A4 using HS-6 and HS-8 as industry definitions. Finer industry classifications (e.g., HS-8) may imply that there are fewer suppliers one may be able to source from, and so we may expect a lower elasticity of substitution between suppliers. In columns (1) and (3) we replicate our main specifications, with elasticities of 0.43 (for HS-6) and 0.06 (for HS-8) respectively. These numbers reflect even higher degrees of complementarity when we consider a more granular notion of industry. Overall, these patterns suggest that inputs are highly specific for buying firms.

**Elasticity Heterogeneity by Industry.** We now analyze whether the degree of substitution across suppliers varies by industry. The idea is that firms that source from highly specific intermediate inputs (i.e. processed foods) should report a lower elasticity of substitution across suppliers than firms that source from more general inputs (e.g. textiles). In Table A5 and Figure 5 we show the estimates of this elasticity of substitution across twenty one broad industries (HS section). We find that the OLS elasticity of substitution across industries lies in the range of 0.7–0.9. Once we instrument for the unit values with the Covid-19-induced lockdown variation, we find that there is wider heterogeneity across industries in the estimate of this elasticity of substitution. Indeed, we find that that *Processed foods* yield an elasticity of 0.19, while *Textiles* yield an elasticity of 0.81. Also, while for the majority of the industries we find evidence for complementarity, there are some industries such as Plastics, Vegetables, and Handicrafts where suppliers within an HS-4 industry are likely substitutes.

FIGURE 5: Elasticities  $\epsilon$  by seller's industry



**Notes:** The vertical axis is the firm-level elasticity of substitution by the industry of the seller, estimated by OLS. The horizontal axis is estimated by IV-2SLS. An industry is an HS section. The size of each bubble is determined by total sales in the corresponding industry. See Table A5 for industry-specific numbers.

### 5.3.2. Firm-level elasticities of substitution across industries

In Table 3, we report our estimates for the firm-level elasticity of substitution across industries. In column (1) we show the OLS estimate of  $\zeta = 0.91$ , which reflects complementarity between industries. Columns (2) and (3) show cases when we define industries more granularly. In this case, the elasticities are around 0.8, which also reflects complementarity between industries.

TABLE 3: Firm-level elasticity of substitution across industries

	(1)	(2)	(3)	(4)	(5)	(6)
$\log\left(\frac{\hat{p}}{\bar{p}}\right)$	0.0842 (0.0039)	0.2014 (0.0045)	0.1996 (0.0048)	0.3136 (0.1060)	0.1712 (0.0040)	0.1996 (0.0048)
$\zeta$	0.9157	0.7985	0.8003	0.6863	0.4368	0.4721
Obs	1292329	794376	766804	1292329	794376	766804
K-PF	.	.	.	27.284	17.950	15.868
Estimator	OLS	OLS	OLS	2SLS	2SLS	2SLS
HSN digits	4	8	8	4	8	8
HSN/month FE	Y	Y	Y	Y	Y	Y
Buyer/month FE	Y	Y	Y	Y	Y	Y
Buyer/HSN FE			Y			Y

**Notes:** IV-2SLS estimates from Equation (10). Price indices are constructed by recovering the residuals used in the corresponding specification when estimating  $\epsilon$  and the corresponding estimate of  $\epsilon$ . The first three columns are OLS estimates of  $\zeta$ ; the last three, 2SLS of estimates of  $\zeta$  using both weighted averages of both bilateral or seller-level instruments across sellers. Bilateral instruments correspond to Equation (14), while seller-level instruments correspond to Equation (13). Each column corresponds to a different combination of fixed effects and definition of industry. Columns (1)-(2) and (4)-(5) correspond to our preferred specification when estimating  $\epsilon$  and 4-digit and 8-digit HS codes. In columns (3) and (6) we also include buyer/HS fixed effects. The first row reports the estimates associated with changes in relative unit values in logs. Standard errors are clustered at the buyer's district level, and are reported in parentheses below each estimate. The fourth row reports the Kleibergen-Paap F statistic from the first stage. The fifth row reports the implied value for  $\epsilon$ , which is 1 minus the estimate on the first row. The sixth row denotes whether estimators are OLS or 2SLS. The sixth row mentions the definition of industry. The last three rows indicate the combination of fixed effects.

In columns (4)-(6) we report our estimates of  $\zeta$  under 2SLS estimation after using a weighted average of instruments across buyers' sellers as discussed in Section 5.2. Our specification in column (4) reports a value of 0.68, reflecting that simultaneous negative demand and supply shocks during Covid led to an underestimation of  $\zeta$  under OLS. This elasticity is higher than the 2SLS elasticity of substitution across suppliers ( $\epsilon = 0.55$ ), reflecting a lower degree of complementarity across industries compared to suppliers.<sup>13</sup> In columns (5) and (6), similar values for this elasticity hold when we define an industry as 8-digit HS codes, and after

<sup>13</sup>This finding is consistent with the literature in macroeconomics (Houthakker, 1955; Bachmann et al., 2022; Lagos, 2006).

the inclusion of buyer/HS fixed effects. Finally, F-stats are high, which reflects the statistical relevance of our weighted averaged instruments.

Unlike the elasticity of substitution across suppliers within an industry, there have been previous attempts in the literature to estimate the elasticity of substitution across industries. In particular, other work has estimated a wide range of values for parameters akin to  $\zeta$  depending on the aggregation of the industry and on the research question. Our elasticity is close to [Boehm et al. \(2019\)](#) who estimate an elasticity between 0.40–0.62 for different inputs at HS-10 for non-Japanese firms and 0.2 for Japanese firms. [Atalay \(2017\)](#) finds an estimate of around 0.1 for 30 aggregated industries using US data.

## 6 QUANTIFICATION

In this section, we use both data from our production network and our newly estimated elasticities to quantify the role of these elasticities in the propagation of shocks. To do this, we need to write down the Leontief matrix in standard form. Given the production structure of our economy, we need four submatrices: (i) firm purchases from 4-digit HS industries, (ii) firm sales to 4-digit HS industries, (iii) labor employed by each firm, and (iv) final sales by each firm. The first two submatrices are directly constructed from the firm-to-firm trade data from the pre-Covid period of March 2019 to February 2020. Labor employed and final sales by firms are obtained by merging in firm-level data from Indiamart, which contains information on firm-level employment and final sales.<sup>14</sup> For more details for this, see Appendix C.

There are 1293 industries. The average firm is connected to 10 industries as a buyer and 5 industries as a seller. The most connected buyer and seller buys from and sells to over 500 industries. We use this 94,555 by 94,555 input output matrix consisting of firm-level sales and purchases from these 1293 industries at the HS-4 level to understand how complementarities at the firm-level affect the propagation of shocks through the firm production networks.

For more details on the derivation of the shock propagation equation and its numerical implementation, see details in Appendix E. While previous work also quantify the effect of firm-level shocks on aggregate GDP, they mostly rely on changes in firm-level final sales rather than the direct production network. Using the production network directly, exponentially increases computational complexity from the order of  $N$  to  $(N+I) \times (N+I)$  (where  $N$  is the no of firms, and  $I$  is the no of industries.) As such, we use computational innovations in big data to implement this procedure.

Note that our quantification exercises in this section are conditional on the industries that firms interact with being given at the extensive margin (even though a firm can change its set of buyers/suppliers). We therefore need to empirically assess whether the set of HS-4 industries

---

<sup>14</sup><https://www.indiamart.com/>

a buyer buys from and the set of HS-4 industries that a seller sells to, changes between the pre and the post Covid period. We do this by inspecting whether both sellers and buyers in each industry continued to trade in their corresponding industries after Covid-19 lockdowns. In Figure A4 we show the industry-level distribution of share of sellers that sold and buyers that purchased goods from that industry during both time periods  $t$  and  $t - 1$ , where  $t$  is a 6-month window before and after the lockdowns.<sup>15</sup> In the figure we see that, for both sellers and buyers, these two distributions are very similar to each other. The overall stability in Figure A4 shows that the assumption that the industries that firms interact with does not change is tenable when analyzing the impact of negative productivity shocks.

### 6.1 *How much does the firm-level elasticity of substitution matter?*

In this section, we assess the importance of the estimated firm-level elasticity of substitution across suppliers, by studying how this elasticity determines the impacts of adverse firm-productivity shocks on aggregate GDP. 29% of all firms in our data lie in the red zone. In this counterfactual, we conservatively shock only the productivity of firms located in the red zone by 25% (with no direct impacts on other firms).

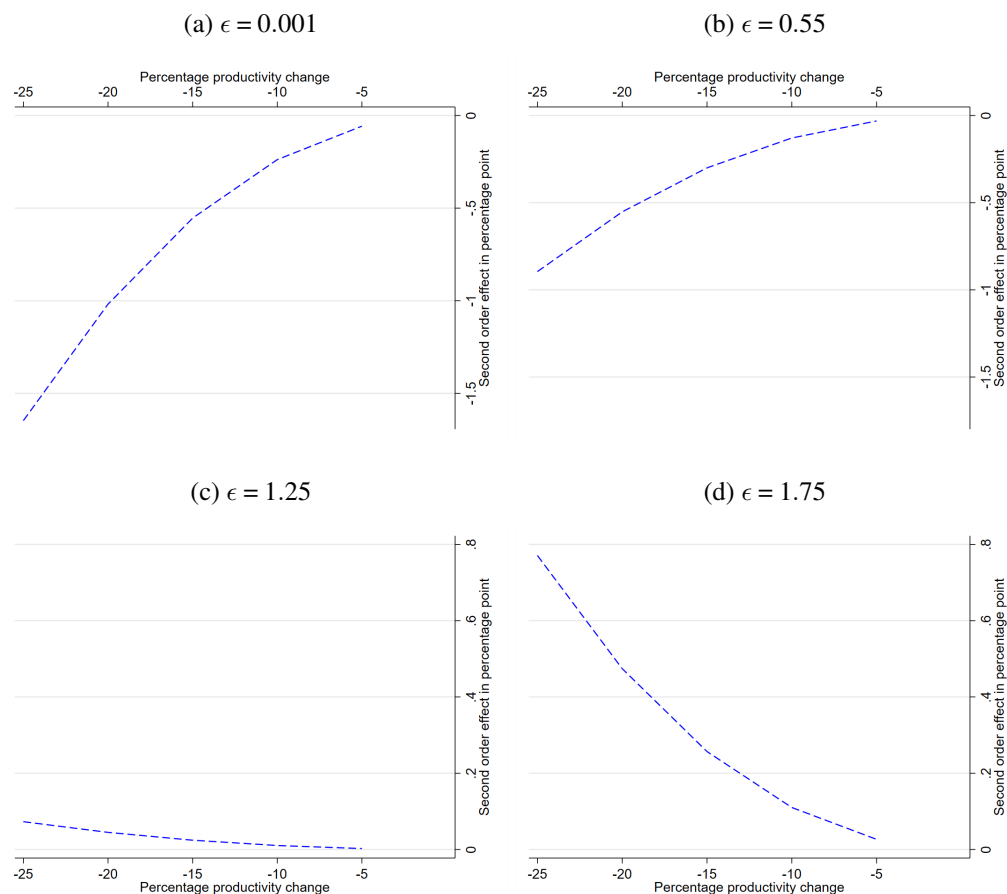
We find that a 25% productivity shock to firms in the red zone reduces GDP by 10.96% (as an empirical benchmark, the state’s annual GDP fell by 11.3% in 2020/21). This fall would be 2.017 pp less in a model where firms in the same HS-4 industry are considered substitutes ( $\epsilon = 2$ ), and 0.75 pp more when firms in the same HS-4 industry are considered almost Leontief ( $\epsilon = 0.001$ ). In terms of GDP losses, given that the quarterly GDP of this state was close to 32.5 billion USD in 2020-2021, the additional losses due to firm-level complementarities translate into 655 million USD (about 19 USD per capita per quarter), compared to the case when firms are substitutes. To put these numbers into perspective, [Baqae and Farhi \(2019\)](#) showed that complementarities only at the industry level (with an elasticity of substitution 0.001) amplify the effect of a negative 13% shock in the oil-industry on GDP by about 0.61%. Note that, the differences in GDP that arise from changing values of firm-level elasticities of substitution across suppliers, only changes the second order effects on GDP.

How important are these second-order effects that we have estimated? To assess the importance of these second-order effects, we simulate different levels of negative productivity shocks for 4 different values of elasticities of substitution and plot the second-order percentage point change in GDP due to these shocks in Figure 6. The top two plots show these differences for a very high level of complementarity 0.001 and our estimated elasticity 0.55, respectively. The bottom two pictures show the additional change in GDP due to the second order when firms are substitutes (1.25 and 1.75). Two things are clear from these pictures: First, for the

<sup>15</sup>For the pre-Covid period,  $t$  is June 2019-October 2019, and  $t - 1$  is June 2018-October 2018. For the post-Covid period,  $t$  is June 2020-October 2020, and  $t - 1$  is June 2019-October 2019.

same negative productivity shock, the second-order effects are much larger when the firms are more complementary. Second, given the same value of the elasticity of substitution, as the magnitudes of the productivity shocks increase, the second order effects become more and more important, that is, there are non-linear effects.

FIGURE 6: How important are second order effects?



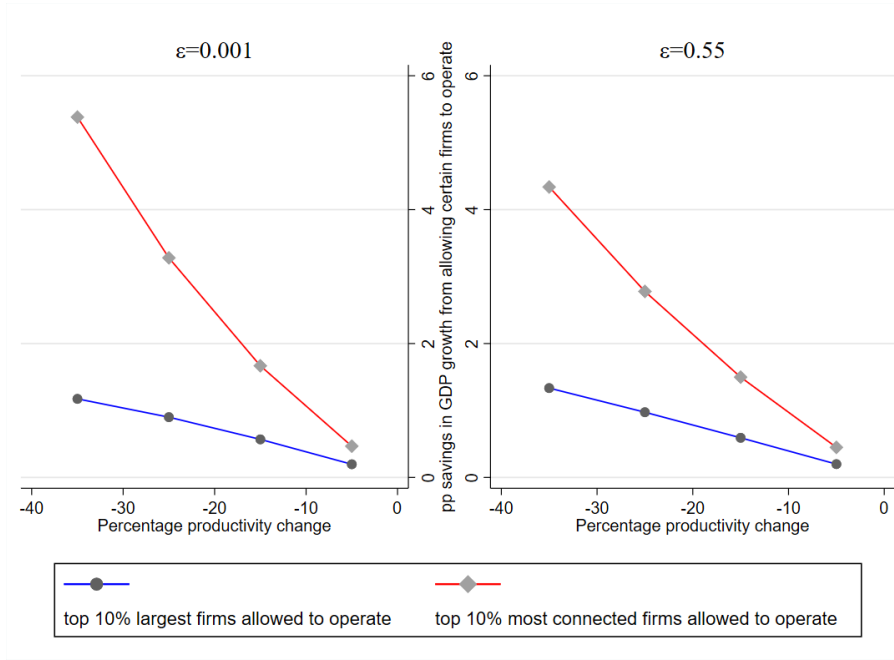
**Notes:** These figures plot the percentage change in productivity on the x-axis. On the y-axis we plot the second order change in GDP in percentage points, for the corresponding change in productivity. Sub-figures (a) and (b) plot these effects when the elasticity of substitution across suppliers within the same industry  $\epsilon = 0.001$ , and  $\epsilon = 0.55$ , respectively. Sub-figures (c) and (d) plot these effects for  $\epsilon = 1.25$  and  $1.75$  respectively.

Observing the bottom two pictures, we can clearly see that as firms become more and more substitutes, the second-order effects actually dampen the negative first-order effects, and more so, for higher values of productivity shocks. That is, unlike the first-order effects which only depend on firm sizes, complementarities at the firm level non-linearly amplify the effects of negative productivity shocks. This reflects similar amplification patterns that (Baqaee and Farhi, 2019) documented, but at the industry level.

These graphs illustrate to us the importance of second-order effects that are largely driven



FIGURE 7: Should we Protect Large or Connected Firms?



**Notes:** These figures plot the percentage change in productivity on the x-axis. On the y-axis we plot the percentage point savings in GDP growth from allowing the top 10% largest or top 10% most connected firms to operate. The left panel plots this change when the elasticity of substitution across suppliers within the same industry  $\epsilon = 0.001$ , and the right panel shows these changes when  $\epsilon = 0.55$

by complementarities at the firm level, especially for large short-lived negative productivity shocks such as Covid-19. For a very long time, since (Hulten, 1978), policy-makers and researchers have emphasized the importance of firm sizes in the propagation of shocks. In the next counterfactual we are going to investigate how important are large firms versus connected firms in the propagation of shocks.

### 6.1.1. Counterfactual GDP changes for large and connected firms

In this counterfactual, we look at the fall in GDP if instead of shocking the 27,320 firms in red-zones, we instead shock the largest and the most connected 27,320 firms in the state. Largest firms are measured by *Domar* weights (final sales share). The most connected firms are measured by the Leontief inverse, which measures the direct and indirect connections of suppliers.

In Figure 7, we plot the results from conducting two different counterfactuals for two different values of elasticities of substitution: 0.001 and 0.55, where in the first counterfactual the largest 10% firms in the red zones are allowed to operate and in the second counterfactual, the most connected 10% firms in the red zones are allowed to operate. In the x-axis, we plot different values of negative productivity shock, starting from -5% to -35%. In the y-axis, the blue line represents the percentage point difference in GDP when the largest 10% firms are

allowed to operate compared to the baseline case. The red line represents the same, but when the most connected firms are allowed to operate. Two things are notable from these graphs: First, as the level of complementarity increases, it pays more to save the more connected firms. Second, for the same level of complementarity, as the magnitude of the adverse productivity shock increases, it pays even more to save the more connected firms.

This counterfactual illustrates that when evaluating which firms to value more when it comes to the effects on aggregate GDP, policy-makers should be looking at not just large firms but also the connected firms. In fact, the connected firms become more important for a large, yet short-lived, shock and when suppliers are highly complementary.

## 7 CONCLUSION

In this paper, we use highly disaggregated firm-to-firm transaction data from a large Indian state and provide one of the first estimates of elasticities of substitution across suppliers within the same industry at the firm level. We provide new estimation strategies and estimates for these elasticities by leveraging regional variation in supply-side shocks induced by the Indian government's massive lockdown policy. We find that inputs are highly complementary even at this very granular level. This elasticity crucially determines aggregate impacts and the transmission of shocks across the network, but has previously eluded the literature ([Baqae and Farhi, 2019](#)). The combined advantage of having product-level unit values and quasi-experimental variation in supply-side shocks allows us to overcome previous challenges in the literature, and credibly estimate this elasticity across suppliers within an industry.

Since inputs are complementary, adverse shocks to even a small subset of firms that are highly linked in the supply chain can negatively affect the aggregate economy by propagating through firm networks. When we conservatively shock only the productivity of firms located in the red zone by 25%, we find that if suppliers within the same industry were substitutes instead of complements, the fall in aggregate quarterly GDP in the state under study would be about 655 million USD lower, or about 19 USD per capita lower per quarter. Using new computational techniques in the field of big data, we can quantify this decline *directly* using information on the economy-wide firm-to-firm network without relying on any first-order approximations. Our methods thus provide new techniques to quantify shocks through large and complex production networks. Using data on the entire production network in the state, we identify the nodal and the largest firms in the network and show that as the level of complementarity and the magnitude of the negative productivity shock increase, it pays more to save the more connected firms. Our findings have implications for policymakers worldwide, who often face difficult trade-offs in crisis regarding which firms to bail out.

## REFERENCES

- Alfaro-Urena, A., I. Manelici, and J. P. Vasquez (2020). The Effects of joining Multinational Supply Chains: New Evidence from Firm-to-Firm Linkages. *Available at SSRN 3376129*.
- Alviarez, V., M. Fioretti, K. Kikkawa, and M. Morlacco (2021). Two-Sided Market Power in Firm-to-Firm Trade.
- Atalay, E. (2017). How Important are Sectoral Shocks? *American Economic Journal: Macroeconomics* 9(4), 254–80.
- Bachmann, R., D. Baqaee, C. Bayer, M. Kuhn, A. Löschel, B. Moll, A. Peichl, K. Pittel, M. Schularick, et al. (2022). What if? the economic effects for germany of a stop of energy imports from russia. *ECONtribute Policy Brief* 28, 2022.
- Baqaee, D. and E. Farhi (2020). Nonlinear Production Networks with an Application to the Covid-19 Crisis. Technical report, National Bureau of Economic Research.
- Baqaee, D. R. and E. Farhi (2019). The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten’s Theorem. *Econometrica* 87(4), 1155–1203.
- Barrot, J.-N. and J. Sauvagnat (2018). Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks. *The Quarterly Journal of Economics* 131(3), 1543–1592.
- Bas, M., A. M. Fernandes, and C. Paunov (2022). How Resilient Was Trade to Covid-19.
- Behrens, K., G. Corcos, and G. Mion (2013). Trade Crisis? What Trade Crisis? *Review of Economics and Statistics* 95(2), 702–709.
- Beyer, R., T. Jain, and S. Sinha (2021). Lights out? Covid-19 Containment Policies and Economic Activity. *Covid-19 Containment Policies and Economic Activity (May 29, 2021)*.
- Boehm, C. E., A. Flaaen, and N. Pandalai-Nayar (2019). Input Linkages and the Transmission of Shocks: Firm-level Evidence from the 2011 Tōhoku Earthquake. *Review of Economics and Statistics* 101(1), 60–75.
- Bonadio, B., Z. Huo, A. A. Levchenko, and N. Pandalai-Nayar (2021). Global Supply Chains in the Pandemic. *Journal of International Economics*, 103534.
- Bricongne, J.-C., L. Fontagné, G. Gaulier, D. Taglioni, and V. Vicard (2012). Firms and the Global Crisis: French Exports in the Turmoil. *Journal of International Economics* 87(1).
- Broda, C. and D. E. Weinstein (2006). Globalization and the Gains from Variety. *The Quarterly journal of economics* 121(2), 541–585.
- Cakmakli, C., S. Demiralp, S. Kalemli-Ozcan, S. Yesiltas, and M. Yildirim (2021). The Economic Case for Global Vaccinations: An Epidemiological Model with International Production Networks. Technical report, National Bureau of Economic Research.
- Carvalho, V. M., M. Nirei, Y. U. Saito, and A. Tahbaz-Salehi (2021). Supply Chain Disruptions: Evidence from the Great East Japan Earthquake. *The Quarterly Journal of Economics* 136(2).

- Chakrabati, A., K. Mahajan, and S. Tomar (2021). Aggregate Shocks, Domestic Trade Collapse and Regional Realignment.
- Demir, B., A. C. Fieler, D. Xu, and K. K. Yang (2021). O-ring Production Networks. Technical report, National Bureau of Economic Research.
- Demir, B. and B. Javorcik (2020). Trade Finance Matters: Evidence from the Covid-19 Crisis. *Oxford Review of Economic Policy*.
- Dhyne, E., A. K. Kikkawa, and G. Magerman (2022). Imperfect competition in firm-to-firm trade. *Journal of the European Economic Association* 20(5), 1933–1970.
- Dhyne, E., A. K. Kikkawa, M. Mogstad, and F. Tintelnot (2021). Trade and Domestic Production Networks. *The Review of Economic Studies* 88(2), 643–668.
- Edmond, C., V. Midrigan, and D. Y. Xu (2018). How Costly are Markups? Technical report, National Bureau of Economic Research.
- Fajgelbaum, P. D. and E. Schaal (2020). Optimal Transport Networks in Spatial Equilibrium. *Econometrica* 88(4), 1411–1452.
- Feenstra, R. C. (1994). New Product Varieties and the Measurement of International Prices. *The American Economic Review*, 157–177.
- Ferrari, A. (2022). Inventories, demand shocks propagation and amplification in supply chains. *arXiv preprint arXiv:2205.03862*.
- Gerschel, E., A. Martinez, and I. Mejean (2020). Propagation of Shocks in Global Value Chains: The Coronavirus Case. *Notes IPP* (53).
- Giovanni, J. d. and A. A. Levchenko (2009). Trade Openness and Volatility. *The Review of Economics and Statistics* 91(3), 558–585.
- Heise, S. et al. (2020). How did China’s Covid-19 Shutdown affect US Supply Chains? Technical report, Federal Reserve Bank of New York.
- Houthakker, H. S. (1955). The pareto distribution and the cobb-douglas production function in activity analysis. *The Review of Economic Studies* 23(1), 27–31.
- Hulten, C. R. (1978). Growth Accounting with Intermediate Inputs. *The Review of Economic Studies* 45(3), 511–518.
- Hummels, D., R. Jørgensen, J. Munch, and C. Xiang (2014). The Wage Effects of Offshoring: Evidence from Danish Matched Worker-Firm data. *American Economic Review* 104(6).
- Kone, Z. L., M. Y. Liu, A. Mattoo, C. Ozden, and S. Sharma (2018). Internal Borders and Migration in India. *Journal of Economic Geography* 18(4), 729–759.
- Korovkin, V. and A. Makarin (2020). Production networks and war. *arXiv preprint arXiv:2011.14756*.
- Lafragne-Roussier, R., J. Martin, and I. Méjean (2021). Supply Shocks in Supply Chains: Evidence from the Early Lockdown in China. Technical report, Mimeo, CREST.
- Lagos, R. (2006). A model of tfp. *The Review of Economic Studies* 73(4), 983–1007.
- Mira, A. N. (1964). Moscow: Miklukho-Maklai Ethnological Institute. *Department of Geodesy and Cartography, State Geological Committee of the Soviet Union*.

- Panigrahi, P. P. (2021). *Endogenous Spatial Production Networks: Theory, Estimation, and Evidence from Indian Firm-to-Firm Linkages*. Ph. D. thesis, University of California, Berkeley.
- Peter, A. and C. Ruane (2022). The Aggregate Importance of Intermediate Input Substitutability. *Princeton University, memo*.
- Ravindran, S. and M. Shah (2020). Unintended Consequences of Lockdowns: Covid-19 and the Shadow Pandemic. Technical report, National Bureau of Economic Research.
- Redding, S. J. and D. E. Weinstein (2020). Measuring Aggregate Price Indices with Taste Shocks: Theory and Evidence for CES preferences. *The Quarterly Journal of Economics* 135(1).
- Taschereau-Dumouchel, M. (2020). Cascades and Fluctuations in an Economy with an Endogenous Production Network. *Available at SSRN 3115854*.

# Appendix for online publication only

## A APPENDIX TABLES

TABLE A1: Summary statistics

	<i>Panel A: 2019</i>		
	Jan-March	April-June	July-September
Number of sellers	135,849	131,996	133,897
Number of buyers	193,660	188,708	189,219
Total sales (mln. rupees)	962,688	908,361	1,036,831
Number of transactions	7,772,883	7,808,325	7,934,706
	<i>Panel B: 2020</i>		
	Jan-March	April-June	July-September
Number of sellers	113,121	69,171	86,696
Number of buyers	164,153	114,353	135,056
Total sales (mln. rupees)	811,755	369,645	775,478
Number of transactions	7,362,508	3,201,081	4,782,336

**Notes:** This table is comprised of two panels. Panel A contains information about the number of sellers, buyers, transactions, and total sales for periods January-March, April-June, July-September for year 2019. Panel B is the same as Panel A, but for 2020.

TABLE A2: Distribution of economic activity by industry and type of transaction

<b>HS section</b>	<b>Sales share</b>	<b>Purchase share</b>
Animales	1.5034	.7723
Vegetables	15.2982	11.2945
Fats	2.2934	2.6251
Processed foods	4.2172	5.5548
Minerals	13.1241	10.2353
Chemicals	9.8288	9.0791
Plastics	13.1516	9.1410
Leather	.1618	.1677
Wood	2.5110	1.2130
Wood derivatives	1.0783	1.3598
Textiles	3.6342	6.4576
Clothing	1.3428	.9107
Handcrafts	1.0190	1.9337
Jewelry	1.7005	1.4980
Metal	10.4473	12.1969
Machinery	10.9909	13.5771
Transport equipment	4.7124	8.4147
Surgical instrum.	1.4478	1.6478
Arms and ammo	.0057	.0095
Miscellaneous	1.2263	1.4936
Art	.3043	.4166
<b>Type of transaction</b>		
Within-state	72.6822	52.2224
Inter-state	23.2183	44.5151
Foreign	4.0994	3.2623

**Notes:** The table is comprised of an upper panel and a lower panel. In the upper panel we show the share of sales and purchases from/to our Indian state of analysis by HS section. In the lower panel we show the share of sales to and purchases from our Indian state, by whether the buyer or seller is within the state, in another state of India, or abroad. Statistics were calculated using data for all 2019.



TABLE A3: 2SLS, firm-level elasticity of substitution across (at least two) suppliers

	(1)	(2)	(3)	(4)
$\log\left(\frac{\hat{p}}{\bar{p}}\right)$	0.2383 (0.1206)	0.3381 (0.0627)	0.4121 (0.1236)	0.3688 (0.1146)
$\epsilon$	0.7616	0.6618	0.5878	0.6311
Obs	851120	599918	599918	544819
K-PF	58.989	97.958	233.084	527.534
Seller IV	Y		Y	Y
Bilateral IV		Y	Y	Y
Buyer/month FE	Y	Y	Y	Y
HSN/month FE	Y	Y	Y	Y
Buyer/HSN FE				Y
Seller/HSN FE				Y

**Notes:** IV-2SLS estimates from Equation (11). The set of common suppliers of buyer  $b$  is  $\Omega_{i,bj,t}^* = \Omega_{i,bj,t} \cap \Omega_{i,bj,t-1}$ . That is, a supplier  $s$  of buyer  $b$  is considered *common* if they also traded during the previous month. We only consider the cases when a buyer traded with at least two common suppliers in a given period. The first stage uses either bilateral or seller-level instruments, as pointed out by rows six and seven. Bilateral instruments correspond to Equation (14), while seller-level instruments correspond to Equation (13). The first row reports the estimates associated with changes in relative unit values in logs. Standard errors are two-way clustered at the origin and destination state level, and are reported in parentheses below each estimate. The fourth row reports the Kleibergen-Paap F statistic from the first stage. The fifth row reports the implied value for  $\epsilon$ , which is 1 minus the estimate on the first row. An industry is 4-digit HS codes and the treatment period is March-May 2020. The table contains four columns. Each column corresponds to different combinations of instruments and of fixed effects, as pointed out by the last six rows. All specifications include the controls mentioned in the paper.

TABLE A4: Alternative specifications: 2SLS, firm-level elasticity of substitution across suppliers

	(1)	(2)	(3)	(4)
$\log\left(\frac{\hat{p}}{\bar{p}}\right)$	0.5687 (0.2086)	0.5476 (0.1818)	0.9371 (0.3856)	0.8063 (0.3305)
$\epsilon$	0.4312	0.4523	0.0628	0.1936
Obs	879997	851483	1026381	993583
K-PF	37.629	121.309	42.335	87.990
HSN digits	6	6	8	8
Seller IV	Y	Y	Y	Y
Bilateral IV	Y	Y	Y	Y
Buyer/month FE	Y	Y	Y	Y
HSN/month FE	Y	Y	Y	Y
Buyer/HSN FE		Y		Y
Seller/HSN FE		Y		Y

**Notes:** IV-2SLS estimates from Equation (11). The set of common suppliers of buyer  $b$  is  $\Omega_{i,bj,t}^* = \Omega_{i,bj,t} \cap \Omega_{i,bj,t-1}$ . That is, a supplier  $s$  of buyer  $b$  is considered *common* if they also traded during the previous month. In all specifications, the first stage uses both bilateral and seller-level instruments as pointed in rows seven and eight. Bilateral instruments correspond to Equation (14), while seller-level instruments correspond to Equation (13). The first row reports the estimates associated with changes in relative unit values in logs. Standard errors are two-way clustered at the origin and destination state level, and are reported in parentheses below each estimate. The fourth row reports the Kleibergen-Paap F statistic from the first stage. The fifth row reports the implied value for  $\epsilon$ , which is 1 minus the estimate on the first row. An industry is either 6-digit or 8-digit HS codes as pointed out by the sixth row, and the treatment period is March-May 2020. The table contains four columns. Each column corresponds to different combinations of HS codes and of fixed effects, as pointed out by the last five rows. All specifications include the controls mentioned in the paper.

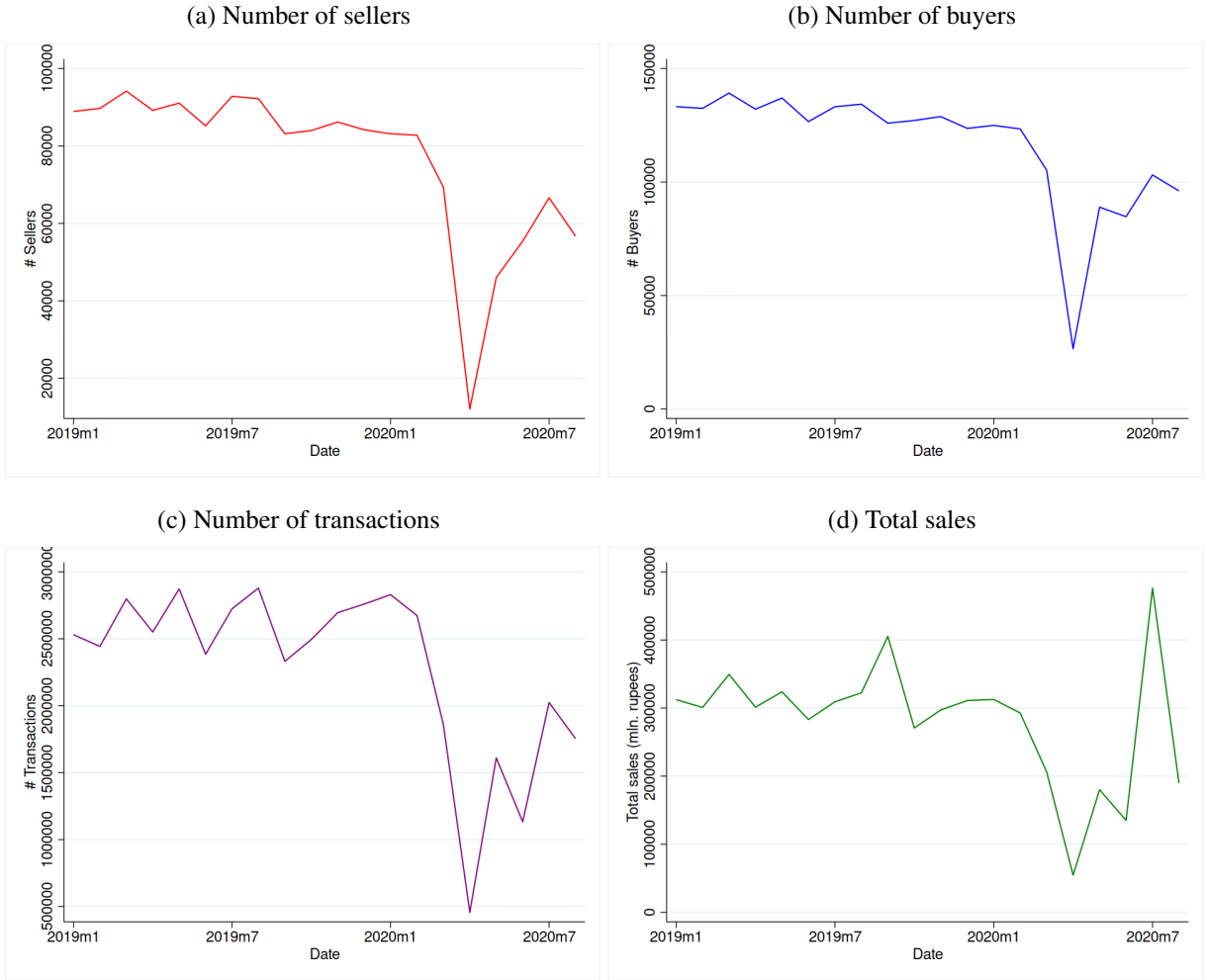
TABLE A5: Firm-level elasticities of substitution across suppliers, by HS section

Section	Name	OLS elast.	2SLS elast.
1	Animals	0.6892	0.1648
2	Vegetables	0.7799	0.7149
3	Fats	.	.
4	Processed foods	0.7125	0.1917
5	Minerals	0.8326	0.3974
6	Chemicals	0.7735	0.5828
7	Plastics	0.7179	0.9796
8	Leather	.	.
9	Wood	0.8728	0.6154
10	Wood derivatives	0.7812	0.8915
11	Textiles	0.8249	0.8103
12	Clothing	0.8232	0.3360
13	Handcrafts	0.6737	.
14	Jewelry	0.8104	1.3721
15	Metal	0.8145	0.8142
16	Machinery	0.6072	0.8691
17	Transport equipment	.	.
18	Surgical instruments	0.5954	0.3799
19	Arms and ammo	0.4140	.
20	Miscellaneous	0.6903	0.8383
21	Art	0.5514	0.1486

**Notes:** Each row corresponds to an industry, which is defined as a HS section. The second column contains the name of the industry. The third and fourth columns report the estimated elasticities by OLS and 2SLS as in Equation (11). Both OLS and 2SLS estimators include HS/month, buyer/month, buyer/HS, and seller/HS fixed effects. Standard errors are two-way clustered at both origin and destination states. All specifications include the controls mentioned in the paper. Missing elasticities were not able to be estimated due to a lack of statistical power.

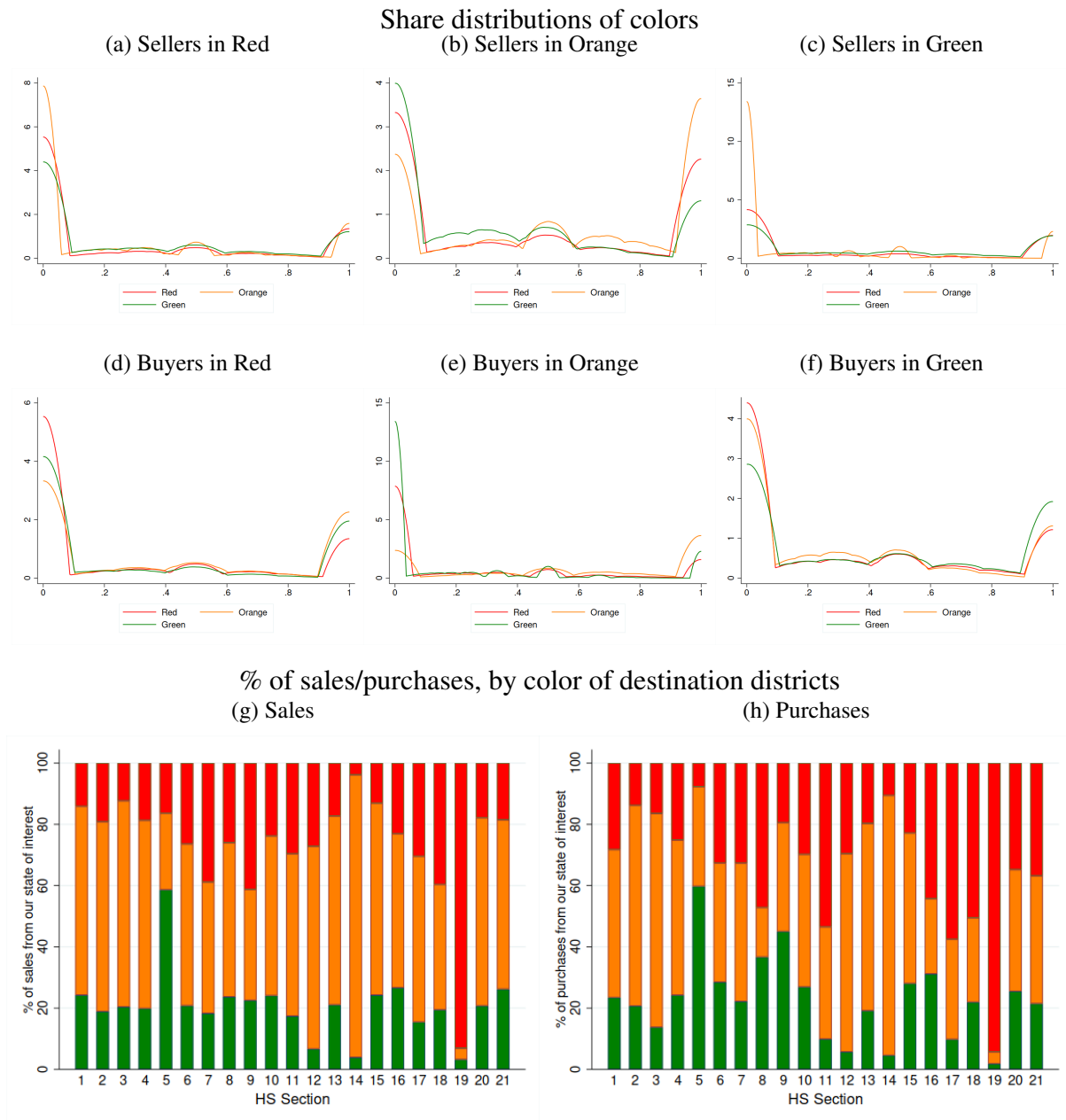
## B APPENDIX FIGURES

FIGURE A1: Variation over time in aggregate outcomes



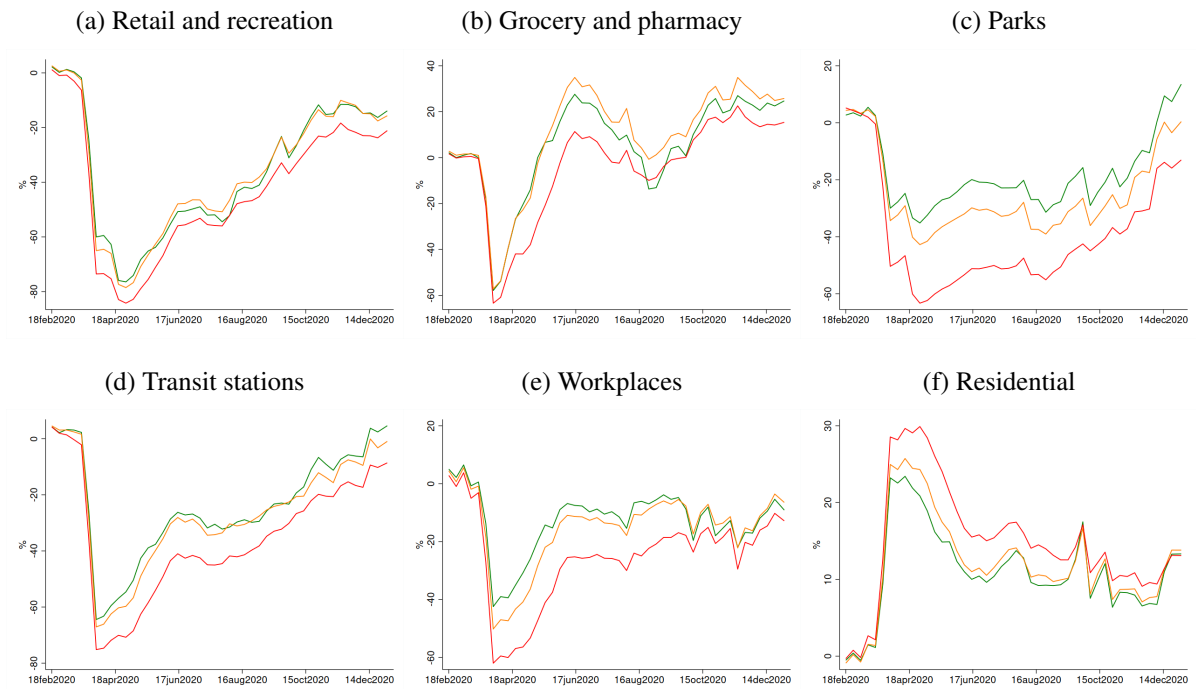
**Notes:** This figure is comprised by 4 panels. In each panel, the horizontal axis is a month, and the vertical axis is a different aggregate outcome. In the first panel, we show the number of sellers that reported a transaction by month. In the second panel, we show the number of buyers that reported a transaction by month. In the third panel, we show the number of transactions that were reported in a given month. In the fourth panel, we show total sales for a given month.

FIGURE A2: Distribution of links and sales across lockdown zones



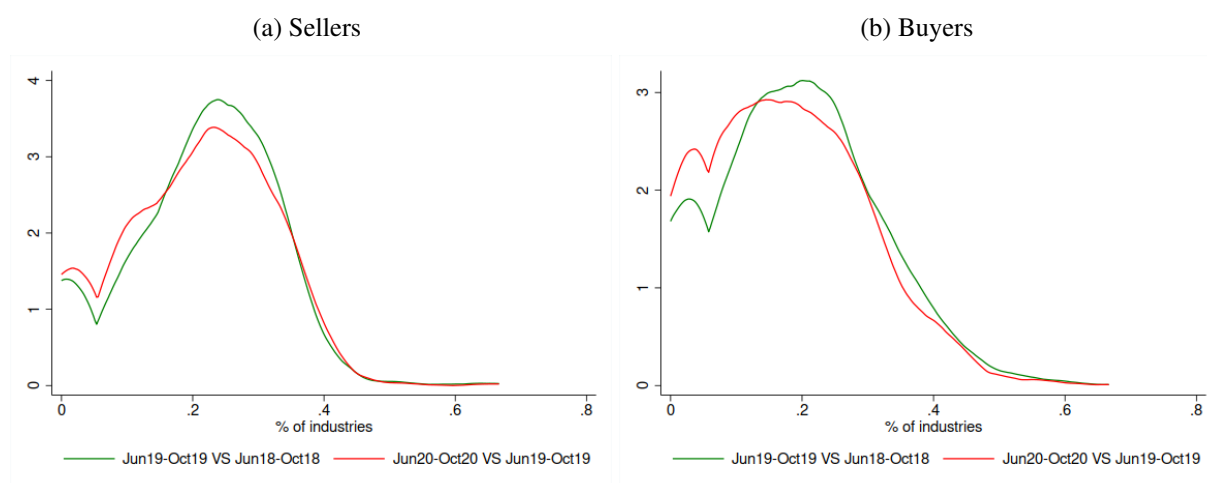
**Notes:** This figure is comprised by two set of panels. The first six figures are the first panel, and the last two figures are the second panel. First we explain the first panel. In the three upper figures, each panel plots the distribution of the share of buyers located in *Red*, *Orange*, or *Green* districts. Each figure corresponds to sellers located in their corresponding color district. In the middle three figures, each figure plots the distribution of the share of sellers located in *Red*, *Orange*, or *Green* districts. Each figure corresponds to buyers located in their corresponding color district. The time period is April 2018 - February 2020. Now, about the second panel, in the left panel, for each HS section (horizontal axis), we plot the share of total sales of firms located in our large Indian state by color of selling districts. In the right panel, for each HS section (horizontal axis), we plot the share of total purchases of firms located in our large Indian state by color of buying districts. The time period for this data is the full 2019 year.

FIGURE A3: Google mobility trends by lockdown zone



**Notes:** These plots are based on *Google Mobility Trends* data, which shows how visits and length of stay at different places change compared to a baseline. The baseline is the median value, for the corresponding day of the week, during January 3rd - February 6th 2020. The raw data is at the daily frequency for each district in India. We collapse this data at the weekly frequency, and at the zone level. Each panel corresponds to mobility in different places.

FIGURE A4: Change in Industry Links Before/After Lockdowns



**Notes:** The figure is comprised of two density plots. On the left we study sellers; on the right, buyers. In that figure we plot the distribution of the share of sellers that sold goods from a given industry in both period  $t$  and  $t - 1$ , where these periods are one year apart, and an industry are 4-digit HS codes. The green density are periods before Covid-19 lockdowns, where  $t$  is between June 2019 and October 2019, and  $t - 1$  is between June 2018 and October 2018. The red density are periods after Covid-10 lockdowns, where  $t$  is between June 2020 and October 2020, and  $t - 1$  is between June 2019 and October 2019.

## C DATA

**Exposure variables.** We have two exposure variables:  $ED_{si,t}$  and  $IM_{si,t}$ . The first one denotes the exposure of firm  $s$  selling product  $i$  to global demand shocks in month  $t$ . The second one denotes the exposure of firm  $s$  selling product  $i$  to global supply shocks in month  $t$ . First, we construct these exposures by country, such that

$$ED_{si,x,t} = \left( \frac{Y_{si,x,0}}{\sum_{x'} Y_{si,x',0}} \right) X_{i,x,t}$$

$$IM_{si,m,t} = \left( \frac{Y_{si,m,0}}{\sum_{m'} Y_{si,m',0}} \right) M_{i,m,t},$$

where  $Y_{si,x,0}$  is the value of goods of seller  $s$  of product  $i$  shipped to country  $x$  in the beginning of the sample,  $Y_{si,m,0}$  is the value of goods of seller  $s$  of product  $i$  shipped from country  $m$  in the beginning of the sample,  $X_{i,x,t}$  is the value of export demand from country  $x$  for product  $i$  in month  $t$ , excluding demand for Indian products, and  $M_{i,m,t}$  is the value of import demand to country  $x$  for product  $i$  in month  $t$ , excluding demand for Indian products. We then do a weighted sum of these measures across countries, such that

$$ED_{si,t} = \sum_x \left( \frac{Y_{s,x,0}}{\sum_{x'} Y_{s,x',0}} \right) ED_{si,x,t}$$

$$IM_{si,t} = \sum_m \left( \frac{Y_{s,m,0}}{\sum_{m'} Y_{s,m',0}} \right) IM_{si,m,t}$$

**Labor and sales.** Our firm-to-firm dataset lacks data on number of employees and final sales. Then, the objective is to predict values for number of employees and final sales for all buyers and sellers of the dataset. We do this by obtaining data on number of employees and total sales from an external dataset for a subset of our firms, run an OLS regression of both labor and final sales on observable variables in our firm-to-firm dataset, store the OLS estimates, and use them to predict labor and final sales for all firms.

We scraped data on number of employees and total sales from the website *IndiaMART*,<sup>16</sup> India's largest B2B digital platform. We scraped around 300,000-400,000 firm profiles, and then sent them to the tax authority to be matched with our firm-to-firm trade dataset. The matching procedure yielded 50,720 unique firms.

Each firm reports its number of employees and annual turnover (sales), both reported in brackets. The reported brackets for sales are: up to 50 Lakh, 50 Lakh-1 Crore, 1-2 Crore, 2-5 Crore, 5-10 Crore, 10-25 Crore, 25-50 Crore, 50-100 Crore, 100-500 Crore, 500-1,000 Crore, 1,000-5,000 Crore, 5,000-10,000 Crore, more than 10,000 Crore. First, we convert each reported number into rupees, since sales in the trade dataset is reported in rupees.<sup>17</sup> Then, for each firm we assign the median value of its corresponding sales bracket. For the last bracket, we consider the upper bound to be 100,000 Crore. The reported brackets for labor are: up to 10

<sup>16</sup><https://www.indiamart.com/>

<sup>17</sup> 100,000 rupees = 1 Lakh; and 10,000,000 rupees = 1 Crore.



employees, 11-25, 26-50, 51-100, 101-500, 501-1000, 1001-2000, 2001-5000, more than 5000 employees. For each firm we assign the median value of its corresponding labor bracket. For the last bracket, we consider the upper bound to be 50,000 employees.

We then run the following OLS regressions:

$$\begin{aligned}\log(labor_n) &= \alpha_0 + \alpha_1 \log(sales_n) + \alpha_2 \log(distance_n) + \epsilon_i^l \\ \log(final_n) &= \beta_0 + \beta_1 \log(sales_n) + \beta_2 \log(distance_n) + \epsilon_i^f,\end{aligned}$$

where  $sales_n$  are total sales of intermediates of firm  $n$  and  $distance_n$  is the average distance in kilometers of all firms' registered transactions,  $labor_n$  is the number of employees constructed as previously explained, and  $final_n$  is final sales. We constructed final sales by subtracting total intermediate sales from total sales, where we construct the former directly from our firm-to-firm dataset. In the vast majority of cases, this difference was positive, which reassures that IndiaMART indeed reports total sales. Whenever the differences were negative, we input a value of 0, which implies that all firm's sales are of intermediates.

We obtain the following estimated elasticities:  $(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2) = (-2.1138, 0.2502, 0.2853)$ , and  $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = (9.8848, 0.3665, 0.4227)$ . They are estimated under robust standard errors, and are all significant at the 1% confidence level. We then use these estimates to predict labor and final sales to all firms in our dataset.

**Dijkstra algorithm** We now list the steps of a Dijkstra algorithm we used to construct our the seller/buyer-level instruments. We obtained a set of *shapefiles* of district administrative boundaries for India according to India's 2011 census. We reprojected the shapefiles into an *Asian/South Equidistance Conic* projection, which is the projection that best preserves the distance measurements. Once shapefiles are reprojected, the objective is to construct a transportation network between Indian districts.

First, we obtain the centroid of each district in India. Then, we construct a network structure according to the set of centroids. There are many ways to construct a network, so we need to take a stance on how to form the connections between centroids. For each centroid, we generate connections to the  $k$  closest centroids according to Euclidean distances.<sup>18</sup> We follow [Fajgelbaum and Schaal \(2020\)](#) and consider  $k = 8$  such that we consider the main cardinal directions (i.e. north, south, east, west, north-east, south-east, north-west, south-west).

We now run the Dijkstra algorithm. For all district pairs, the algorithm provides us with the list of all districts that comprise the route between the district pair, and the distance of each leg that comprise the route. Using the name of the districts, we use the lockdown data to assign a lockdown color to each district along the route, and obtain our seller/buyer-level instruments. Our first instrument is the share of districts in a route that are *Red*, *Orange*, or *Green*. When calculating these shares, we rule out the zone where the buyer resides so we don't consider demand-side shocks in our instrument. Using the distance of each leg, our second instrument is the share of meters of the route that are *Red*, *Orange*, or *Green*. We consider a leg to be of

<sup>18</sup>Consider the set of nodes  $\Phi$ , where  $K \equiv \Phi$  is the number of nodes. The number of connections per node  $k$  could range from 0 up to  $K$ , where each represent extreme cases of network formation.  $k = 0$  is a network without connections, so it is not possible to run a Dijkstra algorithm since it is not possible to go from one node to another.  $k = K$  is a fully-connected network, where all nodes are connected with each other. Running a Dijkstra algorithm on this scenario is trivial since the shortest distance between any pair of nodes is their connection itself. Therefore, a feasible number of connections per node must be  $k \in (0, K)$ .

color  $x = \text{Red}, \text{Orange}, \text{Green}$  whenever the origin district was of color  $x$ . In this case we also ignore the color of the district where the buyer resides.

## D DERIVATIONS

### *D.1 Estimation of firm-level elasticities of substitution across suppliers*

In this section we describe the steps to derive the firm-level elasticity of substitution across suppliers. First, we describe the model and the equations we take to the data. Second, explain how we construct price indices we need to estimate this elasticity. Third, we describe how we deal with the entry/exit of suppliers for the estimation. Finally, we explain how we construct the seller-level and seller/buyer-level instruments we use to causally estimate our elasticity.

#### **D.1.1. Expression to estimate firm-level elasticities of substitution across suppliers**

A firm  $b$  in industry  $j \in F$  maximizes profits subject to its technology and to a CES bundle of intermediate inputs:

$$\begin{aligned} \max \quad & p_{bj}y_{bj} - w_{bj}l_{bj} - \sum_i \sum_s p_{si,bj}x_{si,bj} \\ \text{s.t.} \quad & y_{bj} = A_b \left( w_{bl} (l_{bj})^{\frac{\alpha-1}{\alpha}} + (1-w_{bl}) (x_{bj})^{\frac{\alpha-1}{\alpha}} \right)^{\frac{\alpha}{\alpha-1}}, \\ & x_{bj} = \left( \sum_i w_{i,bj}^{\frac{1}{\zeta}} x_{i,bj}^{\frac{\zeta-1}{\zeta}} \right)^{\frac{\zeta}{\zeta-1}}, \\ & x_{i,bj} = \left( \sum_s \mu_{si,bj}^{\frac{1}{\epsilon}} x_{si,bj}^{\frac{\epsilon-1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}} \end{aligned}$$

The first order condition with respect to  $x_{si,bj}$  is

$$\begin{aligned} [x_{si,bj}] : & p_{bj} \left( \frac{\alpha}{\alpha-1} \right) y_{bj} (\dots_{bj})^{-1} (1-w_{bl}) \left( \frac{\alpha-1}{\alpha} \right) x_{bj}^{\frac{\alpha-1}{\alpha}-1} \\ & \left( \frac{\zeta}{\zeta-1} \right) x_{bj} (\dots_{bj})^{-1} w_{i,j} \left( \frac{\zeta}{\zeta-1} \right) x_{i,bj}^{\frac{\zeta-1}{\zeta}-1} \\ & \left( \frac{\epsilon}{\epsilon-1} \right) x_{i,bj} (\dots_{i,bj})^{-1} \mu_{si,bj}^{\frac{1}{\epsilon}} \left( \frac{\epsilon-1}{\epsilon} \right) x_{si,bj}^{\frac{\epsilon-1}{\epsilon}-1} = p_{si,bj}, \\ & = p_{bj} y_{bj} (\dots_{bj})^{-1} (1-w_{bl}) x_{bj}^{\frac{\alpha-1}{\alpha}} \\ & (\dots_{bj})^{-1} w_{i,j} x_{i,bj}^{\frac{\zeta-1}{\zeta}} \\ & (\dots_{i,bj})^{-1} \mu_{si,bj}^{\frac{1}{\epsilon}} x_{si,bj}^{\frac{\epsilon-1}{\epsilon}} = p_{si,bj}, \end{aligned}$$

where  $(\dots)$  are components that we do not write in detail since they cancel out eventually. Now,

consider the first order conditions with respect to  $x_{si,bj}$  and  $x_{s'i,bj}$  and divide them, such that

$$\begin{aligned}
\frac{\mu_{si,bj}^{\frac{1}{\epsilon}} x_{si,bj}^{-\frac{1}{\epsilon}}}{\mu_{s'i,bj}^{\frac{1}{\epsilon}} x_{s'i,bj}^{-\frac{1}{\epsilon}}} &= \frac{p_{si,bj}}{p_{s'i,bj}}, \\
\frac{x_{si,bj}^{-\frac{1}{\epsilon}} p_{si,bj}^{\frac{1}{\epsilon}}}{x_{s'i,bj}^{-\frac{1}{\epsilon}} p_{s'i,bj}^{\frac{1}{\epsilon}}} &= \frac{p_{si,bj}^{1-\frac{1}{\epsilon}} \mu_{si,bj}^{-\frac{1}{\epsilon}}}{p_{s'i,bj}^{1-\frac{1}{\epsilon}} \mu_{s'i,bj}^{-\frac{1}{\epsilon}}}, \\
(x_{si,bj} p_{si,bj})^{-\frac{1}{\epsilon}} \left( p_{s'i,bj}^{\frac{\epsilon-1}{\epsilon}} \mu_{s'i,bj}^{-\frac{1}{\epsilon}} \right) &= p_{si,bj}^{\frac{\epsilon-1}{\epsilon}} \mu_{si,bj}^{-\frac{1}{\epsilon}} (x_{s'i,bj} p_{s'i,bj})^{-\frac{1}{\epsilon}}, \\
(x_{si,bj} p_{si,bj}) (p_{s'i,bj}^{1-\epsilon} \mu_{s'i,bj}) &= p_{si,bj}^{1-\epsilon} \mu_{si,bj} (x_{s'i,bj} p_{s'i,bj}), \\
(PM_{si,bj}) (p_{s'i,bj}^{1-\epsilon} \mu_{s'i,bj}) &= p_{si,bj}^{1-\epsilon} \mu_{si,bj} (PM_{s'i,bj}), \\
(PM_{si,bj}) \sum_{s'} (p_{s'i,bj}^{1-\epsilon} \mu_{s'i,bj}) &= p_{si,bj}^{1-\epsilon} \mu_{si,bj} \sum_{s'} (PM_{s'i,bj}), \\
(PM_{si,bj}) p_{i,bj}^{1-\epsilon} &= p_{si,bj}^{1-\epsilon} \mu_{si,bj} PM_{i,bj}, \\
\frac{PM_{si,bj}}{PM_{i,bj}} &= \left( \frac{p_{si,bj}}{p_{i,bj}} \mu_{si,bj}^{\frac{1}{1-\epsilon}} \right)^{1-\epsilon}, \\
\log \left( \frac{PM_{si,bj}}{PM_{i,bj}} \right) &= (1-\epsilon) \log \left( \frac{p_{si,bj}}{p_{i,bj}} \right) + \log (\mu_{si,bj}).
\end{aligned}$$

where  $PM_{si,bj} \equiv p_{si,bj} x_{si,bj}$ ,  $p_{i,bj}^{1-\epsilon} \equiv \sum_{s'} p_{s'i,bj}^{1-\epsilon} \mu_{s'i,bj}$ , and  $PM_{i,bj} \equiv \sum_{s'} PM_{s'i,bj}$ .

### D.1.2. Constructing price indices

In this section we derive the expressions that allows us to construct price indexes based on observable data. First, go back to the derivation in Appendix D.1, where

$$(PM_{si,bj}) p_{i,bj}^{1-\epsilon} = p_{si,bj}^{1-\epsilon} \mu_{si,bj} PM_{i,bj}.$$

In the data we observe the production network over time, so we introduce a time dimension such that

$$(PM_{si,bj,t}) p_{i,bj,t}^{1-\epsilon} = p_{si,bj,t}^{1-\epsilon} \mu_{si,bj,t} PM_{i,bj,t},$$

where  $t$  is a month. We can now express this equation in changes, such that

$$\left( \widehat{PM}_{si,bj,t} \right) \widehat{p}_{i,bj,t}^{1-\epsilon} = \widehat{p}_{si,bj,t}^{1-\epsilon} \widehat{\mu}_{si,bj,t} \widehat{PM}_{i,bj,t},$$

where  $\widehat{x}_t \equiv \frac{x_t}{x_{t-1}}$ . Our objective is for  $\widehat{p}_{i,bj,t}$  not to depend on  $\widehat{\mu}_{si,bj,t}$ , which are not observable. To do this, we rely on Redding and Weinstein (2020). The key assumption is that the overall importance of an industry in a buyer's input use is time-invariant. Concretely, the geometric mean of  $\mu_{si,bj,t}$  across common sellers is constant. From the maximization problem of the firm,

we obtain the following expression for the CES price index at the buyer level:

$$p_{i,bj,t} = \left( \sum_{s \in \Omega_{i,bj,t}} \mu_{si,bj,t} p_{si,bj,t}^{1-\epsilon} \right)^{\frac{1}{1-\epsilon}},$$

where  $\Omega_{i,bj,t}$  is the set of all sellers that provided to buyer  $b$  in time  $t$ . We apply Shephard's Lemma to this CES price function, which in turn yields an expression for expenditure share:

$$s_{si,bj,t} = \frac{\mu_{si,bj,t} p_{si,bj,t}^{1-\epsilon}}{p_{i,bj,t}^{1-\epsilon}},$$

where  $s_{si,bj,t} \equiv \frac{PM_{si,bj,t}}{\sum_{s \in \Omega_{i,bj,t}} PM_{si,bj,t}}$ . We can then rewrite this expression such that

$$p_{i,bj,t} = p_{si,bj,t} \left( \frac{\mu_{si,bj,t}}{s_{si,bj,t}} \right)^{\frac{1}{1-\epsilon}}, \forall s \in \Omega_{i,bj,t}.$$

This expression in changes is

$$\hat{p}_{i,bj,t} = \hat{p}_{si,bj,t} \left( \frac{\hat{\mu}_{si,bj,t}}{\hat{s}_{si,bj,t}} \right)^{\frac{1}{1-\epsilon}}.$$

Now, common suppliers for a buyer  $b$  in time  $t$  is the set of suppliers  $\Omega_{i,bj,t}^*$  that sold to buyer  $b$  in the current and previous period (i.e.  $\Omega_{i,bj,t}^* \equiv \Omega_{i,bj,t} \cap \Omega_{i,bj,t-1}$ ), where  $N_{i,bj,t}^* \equiv |\Omega_{i,bj,t}^*|$  is the number of common sellers for buyer  $b$  in time  $t$ . We now apply a geometric mean to this expression, such that

$$\begin{aligned} \hat{p}_{i,bj,t}^{N_{i,bj,t}^*} &= \prod_{s=1}^{N_{i,bj,t}^*} \left\{ \hat{p}_{si,bj,t} \left( \frac{\hat{\mu}_{si,bj,t}}{\hat{s}_{si,bj,t}} \right)^{\frac{1}{1-\epsilon}} \right\}, \\ \hat{p}_{i,bj,t}^{N_{i,bj,t}^*} &= \prod_{s=1}^{N_{i,bj,t}^*} \hat{p}_{si,bj,t} \prod_{s=1}^{N_{i,bj,t}^*} \hat{\mu}_{si,bj,t}^{\frac{1}{1-\epsilon}} \prod_{s=1}^{N_{i,bj,t}^*} \hat{s}_{si,bj,t}^{\frac{1}{\epsilon-1}}, \\ \hat{p}_{i,bj,t} &= \prod_{s=1}^{N_{i,bj,t}^*} \hat{p}_{si,bj,t}^{\frac{1}{N_{i,bj,t}^*}} \left( \prod_{s=1}^{N_{i,bj,t}^*} \hat{\mu}_{si,bj,t}^{\frac{1}{N_{i,bj,t}^*}} \right)^{\frac{1}{1-\epsilon}} \prod_{s=1}^{N_{i,bj,t}^*} \left( \hat{s}_{si,bj,t}^{\frac{1}{N_{i,bj,t}^*}} \right)^{\frac{1}{\epsilon-1}}, \\ \hat{p}_{i,bj,t} &= \hat{p}_{i,bj,t}^{\frac{1}{N_{i,bj,t}^*}} \hat{s}_{i,bj,t}^{\frac{1}{\epsilon-1}} \left( \prod_{s=1}^{N_{i,bj,t}^*} \hat{\mu}_{si,bj,t}^{\frac{1}{N_{i,bj,t}^*}} \right)^{\frac{1}{1-\epsilon}}. \end{aligned}$$

We now formally state the assumption we require to move forward, which is

$$\tilde{\mu}_{i,bj,t} = \prod_{s=1}^{N_{i,bj,t}^*} \mu_{si,bj,t}^{\frac{1}{N_{i,bj,t}^*}} = \prod_{s=1}^{N_{i,bj,t}^*} \mu_{si,bj,t-1}^{\frac{1}{N_{i,bj,t}^*}} = \tilde{\mu}_{i,bj,t-1}.$$

Then, the last term of our expression is

$$\begin{aligned}
\prod_{s=1}^{N_{i,bj,t}^*} \widehat{\mu}_{si,bj,t}^{\frac{1}{N_{i,bj,t}^*}} &= \prod_{s=1}^{N_{i,bj,t}^*} \left( \frac{\mu_{si,bj,t}}{\mu_{si,bj,t-1}} \right)^{\frac{1}{N_{i,bj,t}^*}}, \\
&= \frac{\prod_{s=1}^{N_{i,bj,t}^*} \mu_{si,bj,t}^{\frac{1}{N_{i,bj,t}^*}}}{\prod_{s=1}^{N_{i,bj,t}^*} \mu_{si,bj,t-1}^{\frac{1}{N_{i,bj,t}^*}}}, \\
&= \frac{\widetilde{\mu}_{i,bj,t}}{\widetilde{\mu}_{i,bj,t-1}}, \\
&= 1.
\end{aligned}$$

So our final expression boils down to

$$\widehat{p}_{i,bj,t}^{1-\epsilon} = \frac{\widehat{p}_{i,bj,t}^{1-\epsilon}}{\widetilde{s}_{i,bj,t}},$$

where  $\widetilde{p}_{i,bj,t} \equiv \prod_s p_{si,bj,t}^{\frac{1}{N_{i,bj,t}^*}}$  is a geometric mean of unit values across common suppliers, and  $\widetilde{s}_{i,bj,t} \equiv \prod_s s_{si,bj,t}^{\frac{1}{N_{i,bj,t}^*}}$  is a geometric mean of expenditure shares across common suppliers. Notice that we have reached to our objective, since now  $\widehat{p}_{i,bj,t}$  is independent of  $\mu_{si,bj,t}$ . Finally, the expression we take to the data is

$$\begin{aligned}
\left( \widehat{PM}_{si,bj,t} \right) \widehat{p}_{i,bj,t}^{1-\epsilon} &= \widehat{p}_{si,bj,t}^{1-\epsilon} \widehat{\mu}_{si,bj,t} \widehat{PM}_{i,bj,t}, \\
\left( \widehat{PM}_{si,bj,t} \right) \widehat{p}_{i,bj,t}^{1-\epsilon} \widetilde{s}_{i,bj,t}^1 &= \widehat{p}_{si,bj,t}^{1-\epsilon} \widehat{\mu}_{si,bj,t} \widehat{PM}_{i,bj,t}, \\
\frac{\widehat{PM}_{si,bj,t}}{\widehat{PM}_{i,bj,t}} &= \left( \frac{\widehat{p}_{si,bj,t}}{\widehat{p}_{i,bj,t}} \right)^{1-\epsilon} \left( \widetilde{s}_{i,bj,t} \widehat{\mu}_{si,bj,t} \right), \\
\log \left( \frac{\widehat{PM}_{si,bj,t}}{\widehat{PM}_{i,bj,t}} \right) &= (1-\epsilon) \log \left( \frac{\widehat{p}_{si,bj,t}}{\widehat{p}_{i,bj,t}} \right) + \log \left( \widetilde{s}_{i,bj,t} \widehat{\mu}_{si,bj,t} \right), \\
\log \left( \frac{\widehat{PM}_{si,bj,t}}{\widehat{PM}_{i,bj,t}} \right) &= (1-\epsilon) \log \left( \frac{\widehat{p}_{si,bj,t}}{\widehat{p}_{i,bj,t}} \right) + \log \left( \widetilde{s}_{i,bj,t} \right) + \log \left( \widehat{\mu}_{si,bj,t} \right).
\end{aligned}$$

### D.1.3. Addressing entry/exit of suppliers

In this section we explain how we address the fact that seller and buyer matches do not happen in every period (i.e. entry and exit of sellers). The concern is that not taking into account the fact that sellers and buyers do not trade in every period could induce a bias in the estimation of  $\epsilon$ . We address this by including a correction term by [Feenstra \(1994\)](#) in our regressions. First, notice we can write down the expenditure share as

$$s_{si,bj,t} \equiv \lambda_{i,bj,t} s_{si,bj,t}^*,$$

where  $\lambda_{i,bj,t}$  is the *Feenstra* correction term, and  $s_{si,bj,t}^*$  is the expenditure share with respect to total expenditure on common suppliers. Notice that these terms are constructed as

$$\begin{aligned} s_{si,bj,t} &\equiv \frac{PM_{si,bj,t}}{\sum_{s \in \Omega_{i,bj,t}} PM_{si,bj,t}}, \\ \lambda_{i,bj,t} &\equiv \frac{\sum_{s \in \Omega_{i,bj,t}^*} PM_{si,bj,t}}{\sum_{s \in \Omega_{i,bj,t}} PM_{si,bj,t}}, \\ s_{si,bj,t}^* &\equiv \frac{PM_{si,bj,t}}{\sum_{s \in \Omega_{i,bj,t}^*} PM_{si,bj,t}}. \end{aligned}$$

In changes, the expression for expenditure shares is

$$\widehat{s}_{si,bj,t} = \widehat{\lambda}_{i,bj,t} \widehat{s}_{si,bj,t}^*.$$

Then, the geometric mean for expenditure shares is

$$\begin{aligned} \widehat{s}_{si,bj,t} &= \prod_{s=1}^{N_{i,bj,t}^*} \widehat{s}_{si,bj,t}^{\frac{1}{N_{i,bj,t}^*}}, \\ &= \prod_{s=1}^{N_{i,bj,t}^*} \left( \widehat{\lambda}_{i,bj,t} \widehat{s}_{si,bj,t}^* \right)^{\frac{1}{N_{i,bj,t}^*}}, \\ &= \widehat{\lambda}_{i,bj,t} \prod_{s=1}^{N_{i,bj,t}^*} \left( \widehat{s}_{si,bj,t}^* \right)^{\frac{1}{N_{i,bj,t}^*}}, \\ &= \widehat{\lambda}_{i,bj,t} \widehat{s}_{si,bj,t}^*. \end{aligned}$$

So the final expression we take to the data is

$$\begin{aligned} \log \left( \frac{\widehat{PM}_{si,bj,t}}{\widehat{PM}_{i,bj,t}} \right) &= (1 - \epsilon) \log \left( \frac{\widehat{p}_{si,bj,t}}{\widehat{p}_{i,bj,t}} \right) + \log \left( \widehat{s}_{si,bj,t} \right) + \log \left( \widehat{\mu}_{si,bj,t} \right), \\ &= (1 - \epsilon) \log \left( \frac{\widehat{p}_{si,bj,t}}{\widehat{p}_{i,bj,t}} \right) + \log \left( \widehat{\lambda}_{i,bj,t} \widehat{s}_{si,bj,t}^* \right) + \log \left( \widehat{\mu}_{si,bj,t} \right), \\ &= (1 - \epsilon) \log \left( \frac{\widehat{p}_{si,bj,t}}{\widehat{p}_{i,bj,t}} \right) + \log \left( \widehat{\lambda}_{i,bj,t} \right) + \log \left( \widehat{s}_{si,bj,t}^* \right) + \log \left( \widehat{\mu}_{si,bj,t} \right). \end{aligned}$$

#### D.1.4. Addressing endogeneity concerns

The equation from the previous section is what we take to the data. Nevertheless, there are further endogeneity issues that would contaminate our estimates for  $\epsilon$ . In particular, Covid lockdowns could have also induced changes in demand, which in turn would bias our estimates. For example, if Covid shocks also induce negative demand shocks, our estimates would then be biased upwards. In this section we derive our instruments. First, we consider non-arbitrage in shipping, so prices at the origin and destination between sellers and suppliers are related as

$$p_{si,bj,t} = p_{si,t} \tau_{sb,t},$$

where  $p_{si,t}$  is the marginal cost (MC) of production of good  $i$  for seller  $s$  in month  $t$ ,  $\tau_{sb,t}$  is the iceberg cost of transporting the good from seller  $s$  to buyer  $b$  in month  $t$ . Now, we can then express this in changes, such that

$$\widehat{p}_{si,bj,t} = \widehat{p}_{si,t} \widehat{\tau}_{sb,t}.$$

In logarithms, we have

$$\log(\widehat{p}_{si,bj,t}) = \log(\widehat{p}_{si,t}) + \log(\widehat{\tau}_{sb,t}).$$

These two components of price imply two instruments. First, our seller-level instrument that uses variation in MC at the seller-product level due to lockdown measures at the seller's district. To isolate variation in marginal costs driven by seller's lockdown zone, we interact the lockdown dummy ( $Lock_t$ ) which takes the value 1 between March and May with dummy variables  $Red_{o_s}$  and  $Orange_{o_s}$  that equal 1 whenever seller  $s$  was located in a district  $o$  that was either *Red* or *Orange* during the lockdown. Then, our excluded instruments are

$$\log(\widehat{p}_{si,t}) = \beta^R Red_{o(s)} Lock_t + \beta^O Orange_{o(s)} Lock_t + \epsilon_{si,bj,t}^\nu.$$

Now we explain how we construct the instrument at the seller/buyer level. We have to take a stance about the functional form of the trade cost  $\tau_{sb,t}$ . We assume that trade costs are proportional to the travel time of the transportation of intermediate inputs, such that

$$\tau_{sb,t} = TravelTime_{sb,t}^\sigma.$$

If we express this in changes, we get

$$\widehat{\tau}_{sb,t} = \widehat{TravelTime}_{sb,t}^\sigma.$$

We exploit variation from the Covid-19 lockdown, which induced exogenous variation in the travel time between location pairs of sellers and buyers. Given this, we assume the following difference-in-differences setup for travel time:

$$\widehat{TravelTime}_{sb,t} = \exp(\gamma^R Red_{o(s)d(b)} Lock_t + \gamma^O Orange_{o(s)d(b)} Lock_t + \nu_{si,bj,t}),$$

where  $Red_{o(s)d(b)}$  and  $Orange_{o(s)d(b)}$  are the share of number of districts or of distance designated as *Red* and *Orange*, respectively, along the route between seller  $s$  and buyer  $b$ . We constructed these variables using Dijkstra algorithms. Further details about this are in Appendix C. Combining the expression for changes in travel time due to the lockdown and trade costs, we get the

following expression for our seller/buyer level excluded instruments

$$\log(\widehat{\tau}_{sb,t}) = \beta^R Red_{o(s)d(b)} Lock_t + \beta^O Orange_{o(s)d(b)} Lock_t + \nu_{si,bj,t}$$

## D.2 Estimation of firm-level elasticities of substitution across industries

In this section we describe the steps to derive the firm-level elasticity of substitution across industries. First, we describe the model and the equations we take to the data. Second, we describe how we construct price indices we need to estimate this elasticity. Finally, we describe the instrument we use to causally estimate our elasticity.

### D.2.1 Expressions to estimate firm-level elasticities of substitution across industries

We rewrite the initial maximization problem, so

$$\begin{aligned} \max \quad & p_{bj} y_{bj} - w_{bj} l_{bj} - \sum_i p_{i,bj} x_{i,bj} \\ \text{s.t.} \quad & y_{bj} = A_b \left( w_{bl} (l_{bj})^{\frac{\alpha-1}{\alpha}} + (1-w_{bl}) (x_{bj})^{\frac{\alpha-1}{\alpha}} \right)^{\frac{\alpha}{\alpha-1}}, \\ & x_{bj} = \left( \sum_i^I w_{i,bj}^{\frac{1}{\zeta}} x_{i,bj}^{\frac{\zeta-1}{\zeta}} \right)^{\frac{\zeta}{\zeta-1}}, \\ & p_{i,bj} = \left( \sum_s \mu_{si,bj} p_{si,bj}^{1-\epsilon} \right)^{\frac{1}{1-\epsilon}} \end{aligned}$$

The first order condition with respect to  $x_{i,bj}$  is

$$\begin{aligned} [x_{i,bj}] : & p_{bj} \left( \frac{\alpha}{\alpha-1} \right) y_{bj} (\dots_{bj})^{-1} (1-w_{bl}) \left( \frac{\alpha-1}{\alpha} \right) x_{bj}^{\frac{\alpha-1}{\alpha}-1} \\ & \left( \frac{\zeta}{\zeta-1} \right) x_{bj} (\dots_{bj})^{-1} w_{i,bj}^{\frac{1}{\zeta}} \left( \frac{\zeta}{\zeta-1} \right) x_{i,bj}^{\frac{\zeta-1}{\zeta}-1} = p_{i,bj}, p_{i,bj} \\ & = p_{bj} y_{bj} (\dots_{bj})^{-1} (1-w_{bl}) x_{bj}^{\frac{\alpha-1}{\alpha}} \\ & (\dots_{bj})^{-1} w_{i,bj}^{\frac{1}{\zeta}} x_{i,bj}^{\frac{\zeta-1}{\zeta}}, \end{aligned}$$

where  $(\dots)$  are components that we do not write explicitly since they eventually cancel out. Now, consider the same first order conditions with respect to  $x_{i',bj}$  and divide them, such that



$$\begin{aligned}
& \frac{p_{bj} y_{bj} (\dots_{bj})^{-1} (1 - w_{bl}) x_{bj}^{\frac{\alpha-1}{\alpha}} (\dots_{bj})^{-1} w_{i,bj}^{\frac{1}{\zeta}} x_{i,bj}^{\frac{-1}{\zeta}}}{p_{bj} y_{bj} (\dots_{bj})^{-1} (1 - w_{bl}) x_{bj}^{\frac{\alpha-1}{\alpha}} (\dots_{bj})^{-1} w_{i',bj}^{\frac{1}{\zeta}} x_{i',bj}^{\frac{-1}{\zeta}}} = \frac{p_{i,bj}}{p_{i',bj}}, \\
& \frac{w_{i,bj}^{\frac{1}{\zeta}} x_{i,bj}^{\frac{-1}{\zeta}}}{w_{i',bj}^{\frac{1}{\zeta}} x_{i',bj}^{\frac{-1}{\zeta}}} = \frac{p_{i,bj}}{p_{i',bj}}, \\
& \frac{w_{i,bj}^{\frac{1}{\zeta}} x_{i,bj}^{\frac{-1}{\zeta}} p_{i,bj}^{\frac{-1}{\zeta}}}{w_{i',bj}^{\frac{1}{\zeta}} x_{i',bj}^{\frac{-1}{\zeta}} p_{i',bj}^{\frac{-1}{\zeta}}} = \frac{p_{i,bj} p_{i,bj}^{\frac{-1}{\zeta}}}{p_{i',bj} p_{i',bj}^{\frac{-1}{\zeta}}}, \\
& \frac{w_{i,bj}^{\frac{1}{\zeta}} (x_{i,bj} p_{i,bj})^{\frac{-1}{\zeta}}}{w_{i',bj}^{\frac{1}{\zeta}} (x_{i',bj} p_{i',bj})^{\frac{-1}{\zeta}}} = \frac{p_{i,bj}^{\frac{\zeta-1}{\zeta}}}{p_{i',bj}^{\frac{\zeta-1}{\zeta}}}, \\
& \left( \frac{w_{i,bj}^{\frac{1}{\zeta}} (x_{i,bj} p_{i,bj})^{\frac{-1}{\zeta}}}{w_{i',bj}^{\frac{1}{\zeta}} (x_{i',bj} p_{i',bj})^{\frac{-1}{\zeta}}} \right)^{-\zeta} = \left( \frac{p_{i,bj}^{\frac{\zeta-1}{\zeta}}}{p_{i',bj}^{\frac{\zeta-1}{\zeta}}} \right)^{-\zeta}, \\
& \frac{w_{i',bj} (x_{i,bj} p_{i,bj})}{w_{i,bj} (x_{i',bj} p_{i',bj})} = \frac{p_{i,bj}^{1-\zeta}}{p_{i',bj}^{1-\zeta}}, \\
& PM_{i,bj} (w_{i',bj} p_{i',bj}^{1-\zeta}) = PM_{i',bj} (w_{i,bj} p_{i,bj}^{1-\zeta}), \\
& \sum_{i'} PM_{i,bj} (w_{i',bj} p_{i',bj}^{1-\zeta}) = \sum_{i'} PM_{i',bj} (w_{i,bj} p_{i,bj}^{1-\zeta}), \\
& PM_{i,bj} \sum_{i'} w_{i',bj} p_{i',bj}^{1-\zeta} = w_{i,bj} p_{i,bj}^{1-\zeta} \sum_{i'} PM_{i',bj}, \\
& PM_{i,bj} p_{bj}^{1-\zeta} = w_{i,bj} p_{i,bj}^{1-\zeta} PM_{bj}, \\
& \frac{PM_{i,bj}}{PM_{bj}} = \frac{w_{i,bj} p_{i,bj}^{1-\zeta}}{p_{bj}^{1-\zeta}}, \\
& \frac{PM_{i,bj}}{PM_{bj}} = \left( w_{i,bj}^{\frac{1}{1-\zeta}} \frac{p_{i,bj}}{p_{bj}} \right)^{1-\zeta}, \\
& \log \left( \frac{PM_{i,bj}}{PM_{bj}} \right) = (1-\zeta) \log \left( \frac{p_{i,bj}}{p_{bj}} \right) + \log (w_{i,bj}),
\end{aligned}$$

where  $PM_{bj} \equiv \sum_i PM_{i,bj}$ , and  $p_{bj} = \left( \sum_i w_{i,bj} p_{i,bj}^{1-\zeta} \right)^{\frac{1}{1-\zeta}}$ . As we did for the estimation of the elasticity of substitution across suppliers, we introduce a time dimension, apply Shephard's lemma to this CES price function, and also assume that the overall importance of the composite

intermediates is time-invariant, so

$$\begin{aligned}
s_{i,bj,t} &= \frac{w_{i,bj,t} p_{i,bj,t}^{1-\zeta}}{p_{bj,t}^{1-\zeta}}, \\
p_{bj,t} &= p_{i,bj,t} \left( \frac{w_{i,bj,t}}{s_{i,bj,t}} \right)^{\frac{1}{1-\zeta}}, \\
\widehat{p}_{bj,t} &= \widehat{p}_{i,bj,t} \left( \frac{\widehat{w}_{i,bj,t}}{\widehat{s}_{i,bj,t}} \right)^{\frac{1}{1-\zeta}}, \\
\widehat{p}_{bj,t}^{N_{bj,t}} &= \prod_{i=1}^{N_{bj,t}} \widehat{p}_{i,bj,t} \left( \frac{\widehat{w}_{i,bj,t}}{\widehat{s}_{i,bj,t}} \right)^{\frac{1}{1-\zeta}}, \\
\widehat{p}_{bj,t}^{N_{bj,t}} &= \prod_{i=1}^{N_{bj,t}} \widehat{p}_{i,bj,t} \prod_{i=1}^{N_{bj,t}} \widehat{w}_{i,bj,t}^{\frac{1}{1-\zeta}} \prod_{i=1}^{N_{bj,t}} \widehat{s}_{i,bj,t}^{\frac{1}{\zeta-1}}, \\
\widehat{p}_{bj,t} &= \prod_{i=1}^{N_{bj,t}} \widehat{p}_{i,bj,t}^{\frac{1}{N_{bj,t}}} \left( \prod_{i=1}^{N_{bj,t}} \widehat{w}_{i,bj,t}^{\frac{1}{N_{bj,t}}} \right)^{\frac{1}{1-\zeta}} \left( \prod_{i=1}^{N_{bj,t}} \widehat{s}_{i,bj,t}^{\frac{1}{N_{bj,t}}} \right)^{\frac{1}{\zeta-1}}, \\
\widehat{p}_{bj,t} &= \widehat{\widetilde{p}}_{bj,t} \widehat{\widetilde{w}}_{bj,t}^{\frac{1}{1-\zeta}} \widehat{\widetilde{s}}_{bj,t}^{\frac{1}{\zeta-1}}, \\
\widehat{p}_{bj,t} &= \widehat{\widetilde{p}}_{bj,t} \widehat{\widetilde{s}}_{bj,t}^{\frac{1}{\zeta-1}}, \\
\widehat{p}_{bj,t} &= \frac{\widehat{\widetilde{p}}_{bj,t}}{\widehat{\widetilde{s}}_{bj,t}^{\frac{1}{1-\zeta}}},
\end{aligned}$$

where  $\widetilde{p}_{bj,t} \equiv \prod_{i=1}^{N_{bj,t}} \widetilde{p}_{i,bj,t}^{\frac{1}{N_{bj,t}}}$  is the geometric mean of unit values across industries that buyer  $b$  sources from, and  $\widetilde{s}_{bj,t} \equiv \prod_{i=1}^{N_{bj,t}} \widetilde{s}_{i,bj,t}^{\frac{1}{N_{bj,t}}}$  is the geometric mean of expenditure shares across industries. Now, if we also introduce a time dimension into our estimating equation, express it in changes, and consider our expression for unit values, we have

$$\begin{aligned}
PM_{i,bj,t} p_{bj,t}^{1-\zeta} &= w_{i,bj,t} p_{i,bj,t}^{1-\zeta} PM_{bj,t}, \\
\widehat{PM}_{i,bj,t} \widehat{p}_{bj,t}^{1-\zeta} &= \widehat{w}_{i,bj,t} \widehat{p}_{i,bj,t}^{1-\zeta} \widehat{PM}_{bj,t}, \\
\log \left( \frac{\widehat{PM}_{i,bj,t}}{\widehat{PM}_{bj,t}} \right) &= (1-\zeta) \log \left( \frac{\widehat{p}_{i,bj,t}}{\widehat{p}_{bj,t}} \right) + \log \left( \widehat{w}_{i,bj,t} \right), \\
\log \left( \frac{\widehat{PM}_{i,bj,t}}{\widehat{PM}_{bj,t}} \right) &= (1-\zeta) \log \left( \frac{\widehat{p}_{i,bj,t}}{\widehat{\widetilde{p}}_{bj,t}^{\frac{1}{1-\zeta}}} \right) + \log \left( \widehat{w}_{i,bj,t} \right), \\
\log \left( \frac{\widehat{PM}_{i,bj,t}}{\widehat{PM}_{bj,t}} \right) &= (1-\zeta) \log \left( \frac{\widehat{p}_{i,bj,t}}{\widehat{\widetilde{p}}_{bj,t}} \right) + \log \left( \widehat{\widetilde{s}}_{bj,t} \right) + \log \left( \widehat{w}_{i,bj,t} \right).
\end{aligned}$$

### D.2.2. Constructing price index $p_{i,bj,t}$

To estimate  $\zeta$ , we need values for  $p_{i,bj,t}$ , which are not directly observed in the data since  $p_{i,bj,t} \equiv \left( \sum_s \mu_{si,bj,t} p_{si,bj,t}^{1-\epsilon} \right)^{\frac{1}{1-\epsilon}}$ , which is a function of  $\epsilon$  and  $\mu_{si,bj,t}$ . For  $\epsilon$ , we consider  $\epsilon = \hat{\epsilon}$ , where  $\hat{\epsilon}$  is our estimated elasticity. For  $\mu_{si,bj,t}$ , we use the fact that the residuals when estimating  $\epsilon$  are a function of these shocks. Recall that

$$\log \left( \frac{\widehat{PM}_{si,bj,t}}{\widehat{PM}_{i,bj,t}} \right) = (1-\epsilon) \log \left( \frac{\widehat{p}_{si,bj,t}}{\widehat{p}_{i,bj,t}} \right) + X\beta + \phi_{si,bj,t},$$

where  $\phi_{si,bj,t} = \log(\widehat{\mu}_{si,bj,t}) = \log\left(\frac{\mu_{si,bj,t}}{\mu_{si,bj,t-1}}\right) = \log(\mu_{si,bj,t}) - \log(\mu_{si,bj,t-1})$  are the residuals of this estimating equation. By assumption,  $\log(\mu_{si,bj,t})$  are i.i.d and normally distributed shocks with mean  $\mu$  and variance  $\sigma^2$ , so the mean and variance of  $\log(\mu_{si,bj,t}) - \log(\mu_{si,bj,t-1})$  is 0 and  $2\sigma^2$ , respectively. We now construct  $p_{i,bj,t}$  by the following steps:

1. Run the 2SLS regression to obtain the estimate  $\hat{\epsilon}$ ;
2. Recover predicted values for the error term  $\widehat{\phi}_{si,bj,t}$ ;
3. Calculate the empirical mean and variance of  $\widehat{\phi}_{si,bj,t} : \{\widehat{\mu}_\phi, \widehat{\sigma}_\phi^2\}$ ;
4. Recover the values for mean and variance of  $\log(\mu_{si,bj,t})$ , such that: (i)  $\mu = \widehat{\mu}_\phi$  and  $\sigma^2 = \frac{\widehat{\sigma}_\phi^2}{2}$ ;
5. Make a random draw for  $\log(\mu_{si,bj,0})$ , which is drawn from a normal distribution with mean  $\widehat{\mu}_\phi$  and variance  $\frac{\widehat{\sigma}_\phi^2}{2}$ ;
6. For a given  $\mu_{si,bj,0}$ , recover  $\mu_{si,bj,t}$  according to the following law of motion:

$$\begin{aligned} \log \left( \frac{\mu_{si,bj,t}}{\mu_{si,bj,t-1}} \right) &= \widehat{\phi}_{si,bj,t}, \\ \frac{\mu_{si,bj,t}}{\mu_{si,bj,t-1}} &= \exp \left( \widehat{\phi}_{si,bj,t} \right), \\ \mu_{si,bj,t} &= \exp \left( \phi_{si,bj,t} \right) \mu_{si,bj,t-1}; \end{aligned}$$

7. We then construct unit values by

$$p_{i,bj,t} \equiv \left( \sum_s \mu_{si,bj,t} p_{si,bj,t}^{1-\hat{\epsilon}} \right)^{\frac{1}{1-\hat{\epsilon}}}$$

### D.2.3. Constructing instruments

To obtain an exogenous shifter of relative unit values, which we use to obtain an unbiased estimate of  $\zeta$ , we rely on the instruments we use to estimate  $\epsilon$ . Consider the set of instruments  $Z_{si,bj,t}$ . Then, we consider the new set of instruments:

$$W_{i,bj,t} = \bar{Z}_{si,bj,t} = \frac{1}{N_{i,bj,t}} \sum_s Z_{si,bj,t}.$$

For intuition, consider the instrument that varies across both the color zone of the seller and the buyer (i.e. the share of districts of color red in the route between the location of the seller and of the buyer). Then, the new instrument is the simple average of these shares across sellers. Intuitively, the higher the shares of red-colored locations within the routes, the higher the shock on prices

## E SIMULATIONS USING QUANTITATIVE MODEL

### E.1 Deriving expression for shock propagation through GDP

In this section, we discuss details of the simulation using the quantitative model. In order to do that, we first recall the different notations used in the paper.  $N$  is the number of firms,  $I$  is the number of industries.  $\lambda_k$  is the *Domar* weight of firm or sector  $k$ .  $\theta_k$  is the elasticity of substitution corresponding to the  $k^{th}$  reproducible sector.  $\Omega_{li}$  is the  $(l, i)^{th}$  element of the  $(N+I+2)$  input output matrix  $\Omega$ . It therefore measures the direct reliance of  $l$  on  $i$  as a supplier.  $\psi_{li}$  corresponds to the  $(l, i)^{th}$  element of the  $(N+I+2)$  Leontief inverse, and captures the direct and indirect reliance of  $l$  on  $i$  as a supplier. The aggregate change in GDP ( $\Delta \log y$ ) in response to changes in productivity of firm  $j$  ( $\delta \log A_j$ ) up to a second order is given by the following:

$$\Delta \log y = \sum_{j=1}^N \frac{\partial \log y}{\partial \log A_j} (\Delta \log A_j) + \frac{1}{2} \sum_{i=1}^N \sum_{j=1, i \neq j}^N \frac{\partial^2 \log y}{\partial \log A_i \partial \log A_j} (\Delta \log A_i) (\Delta \log A_j) + \frac{1}{2} \sum_{i=1}^N \frac{\partial^2 \log y}{\partial \log A_i^2} (\Delta \log A_i)^2 \quad (15)$$

Following [Baqaee and Farhi \(2019\)](#), after replacing second order terms

$$\begin{aligned}
&= \sum_{j=1}^N \lambda_j (\Delta \log A_j) + \frac{1}{2} \sum_{i=1}^N \sum_{j=1, i \neq j}^N \left( \sum_{k=0}^N (\theta_k - 1) \lambda_k \text{Cov}_{\Omega(k)}(\psi_{(i)}, \psi_{(j)}) \right) (\Delta \log A_i) (\Delta \log A_j) \\
&\quad + \frac{1}{2} \sum_{i=1}^N \left( \sum_{k=0}^N (\theta_k - 1) \lambda_k \text{Var}_{\Omega(k)} \psi_{(i)} \right) (\Delta \log A_i)^2 \\
&= \sum_{j=1}^N \lambda_j (\Delta \log A_j) + \frac{1}{2} \sum_{i=1}^N \sum_{j=1, i \neq j}^N \left( \sum_{k=0}^N (\theta_k - 1) \lambda_k \left( \left( \sum_{l=1}^{N+F} \Omega_{kl} \psi_{li} \psi_{lj} \right) \right. \right. \\
&\quad \left. \left. - \left( \sum_{l=1}^{N+F} \Omega_{kl} \psi_{li} \right) \left( \sum_{l=1}^{N+F} \Omega_{kl} \psi_{lj} \right) \right) (\Delta \log A_i) (\Delta \log A_j) \right. \\
&\quad \left. + \frac{1}{2} \sum_{i=1}^N \left( \sum_{k=0}^N (\theta_k - 1) \lambda_k \left( \left( \sum_{l=1}^{N+F} \Omega_{kl} \psi_{li} \psi_{li} \right) - \left( \sum_{l=1}^{N+F} \Omega_{kl} \psi_{li} \right) \left( \sum_{l=1}^{N+F} \Omega_{kl} \psi_{li} \right) \right) \right) (\Delta \log A_i)^2 \right. \\
&= \sum_{j=1}^N \lambda_j (\Delta \log A_j) + \frac{1}{2} B + \frac{1}{2} C
\end{aligned} \tag{16}$$

We will now write down the expressions for B and C in matrix form in order to evaluate the second order effects.  $\times$  denotes matrix multiplication and  $\cdot$  denotes element by element matrix operations.

### Quantifying B:

To quantify B, the term that mainly captures the second order effects on GDP that operates through changes in firm  $i$ 's *Domar* weight in response to productivity shocks to firm  $j$ , where  $j \in N, j \neq i$ , we introduce the following intermediate matrices which we will define below:  $M$ ,  $N$ ,  $Covar1$ ,  $Covar21$ ,  $Covar22$ , and  $Covar2$ .  $J_{m,n}$  denotes a matrix of ones of size  $m$  by  $n$ .

$$M = \psi \cdot (\Delta \log A)^T \tag{17}$$

$$N = J_{(N+I+2, N+I+2)} \cdot \left( J_{(N+I+2, 1)} \times \left( \psi \cdot (\Delta \log A)^T \right) \right) - \left( \psi \cdot (\Delta \log A)^T \right) \tag{18}$$

$$Covar1 = \Omega \times (M \cdot N) \tag{19}$$

$$Covar21 = \Omega \times M \tag{20}$$

$$Covar22 = \Omega \times N \tag{21}$$

$$Covar2 = Covar21 \cdot Covar22 \tag{22}$$

$$B = \left( (\theta - 1) \cdot \lambda \right) \times (Covar1 - Covar2) \tag{23}$$

**Quantifying C:** The term  $C$ , mainly captures the second order effects on GDP that operates through changes in firm  $i$ 's *Domar* weight in response to productivity shocks to firm  $i$  itself.

$$C = \left( \left( (\theta - 1) \cdot \lambda \right) \times \left( \Omega \times (\psi \cdot \psi) - (\Omega \times \psi) \cdot (\Omega \times \psi) \right) \right) \times \left( \Delta \log A \cdot \Delta \log A \right) \quad (24)$$

In matrix form, we can rewrite equation (15) as:

$$\begin{aligned} \Delta \log y = & \lambda \times \Delta \log A + .5 \left( (\theta - 1) \cdot \lambda \right) \times \left( \text{Covar1} - \text{Covar2} \right) + \\ & .5 \left( \left( (\theta - 1) \cdot \lambda \right) \times \left( \Omega \times (\psi \cdot \psi) - (\Omega \times \psi) \cdot (\Omega \times \psi) \right) \right) \times \left( \Delta \log A \cdot \Delta \log A \right) \end{aligned} \quad (25)$$

## E.2 Numerical implementation in Python

Numerically implementing this exercise is challenging due to the sheer size of the firm-to-firm trade network. We have data on 93260 firms across 1293 industries. This generates a 94,555 by 94,555 input output matrix. The elements inside the input output matrix are very small as the fraction of an industry's output going to a single firm is very small and each industry in turn sources from a large number suppliers. Therefore, and to keep the calculations as precise as possible, we had to use float64 variable types with these matrices, which resulted in matrices larger than most servers' memories. For instance, the Leontief inverse matrix alone took more than 66 GB of storage/memory size. A lot of the calculation's steps required performing matrix multiplication operations on these large matrices. Matrix multiplication is one of the most demanding operations in terms of computing resources in the world of computer science. We break down this computation via a number of state-of-the-art big data computing techniques, thus achieving scalability when applying our techniques to arbitrarily large input output matrices. As detailed firm-to-firm transactions data are becoming more widely available, these techniques will advance the literature quantifying the propagation of shocks through firm networks.

First, we are able to fit datasets larger than RAM using Dask which provides multi-core and distributed and parallel execution on larger-than-memory datasets.<sup>19</sup> We use Dask distributed capabilities to add parallelism to the calculations in computing second order effects which require few matrix multiplication operations on large 94,555 by 94,555 matrices.

Second, we use a computer powered with multiple GPUs. GPUs are essential for the numerous matrix multiplications this process involves. To demonstrate this in numbers, computing 10 columns of Leontief inverse matrix (only 0.000001 %) takes about 4 days on a powerful server with multiple CPUs, 500 GB of RAM and 16 cores. Computing the entire Leontief inverse on a server powered with 4 GPUs took about 1 hour. The part of our work of computing the second order effect, which involves 3 operations of large matrix multiplication would not be practical using CPUs only.

Third, we use the properties of sparse matrices to define matrix multiplications that can ignore large contiguous chunks of zeros, a typical feature of input output matrices.

<sup>19</sup>[https://tutorial.dask.org/00\\_overview.html](https://tutorial.dask.org/00_overview.html)

Fourth, we developed a custom matrix multiplication function to overcome the limitation of the relatively small memory size of GPUs. The custom matrix multiplication function splits the matrix into chunks of full columns (typically in the order of few 1000's of columns), and multiplies the sparse input output matrix by each chunk and then concatenates all result chunks to formulate the final result.