When Cutting Out the Middleman Backfires: Disintermediation, Wholesale Markups, and Misallocation*

Yiu Hing Barron Tsai†

November 2025

Most Recent Version

Abstract

I study the welfare implications of technology-induced disintermediation using a production network model with endogenous intermediation, wholesaler entry and exit, and markups. Wholesalers economize on the search costs of forming buyer-supplier relationships, but as direct trade technology improves, demand for intermediation falls: marginal wholesalers exit, survivors gain market share, and markups rise. These higher markups distort relative input prices and misallocate resources, partially offsetting the gains from disintermediation. I test the model's predictions using Turkish firm-to-firm transactions, exploiting the staggered rollout of fiber internet. Consistent with theory, provinces with faster fiber growth see less intermediated trade, fewer wholesalers, and higher wholesale markups. Calibrating the model to these responses, endogenous markup increases reduce welfare gains from fiber-induced disintermediation by 30%. The results demonstrate that technologies enabling firms to bypass intermediaries can generate unintended efficiency losses by consolidating wholesale market power, highlighting the potential role for complementary competition policy.

JEL Classification Codes: F12, F14, L13, L81, O33

*I am deeply grateful to my advisors Rafael Dix-Carneiro, Daniel Yi Xu, Laura Castillo-Martinez, Federico Huneeus, and Matthias Kehrig for their guidance and support.

†Department of Economics, Duke University.

1 Introduction

Wholesalers are an integral part of modern production networks, serving as key intermediaries that connect producers and buyers. They help firms avoid the search costs of building many direct buyer-supplier links by pooling these relationships: firms connect once to the intermediary and gain access to a broad network.[1] For instance, in Turkey, wholesalers account for more than half of domestic manufacturing trade in 2012. However, pooling firm relationships requires a large upfront investment by the wholesaler, so profitability hinges on operating at sufficient scale: raising entry barriers, favoring larger intermediaries, and heightening concerns about wholesale market power.[2]

Technological advances—such as Turkey's rapid fiber internet expansion—may enhance the efficiency of direct trade by facilitating tighter coordination between suppliers and customers in manufacturing and product development. By making direct trade more effective, these advances can reduce firms' reliance on wholesalers and help circumvent intermediary market power. Yet the same disintermediation reduces volumes handled by wholesalers, and because intermediation entails sizable fixed investments and scale economies, marginal wholesalers exit. Their exit raises concentration among survivors, who increase markups and distort relative input prices further, potentially exacerbating misallocation even as direct trade proliferates. Consequently, the welfare effects of technology-induced disintermediation are a priori ambiguous. In this paper, I quantify the extent to which increases in wholesale markups—induced by the endogenous exits of wholesalers—offset the gains from technology-induced disintermediation.

I address this question with a model of production network formation featuring endogenous intermediation, wholesaler entry and exit, and markups. My framework builds upon the production network model introduced by Demir et al. (2024) and its spatial extension by Arkolakis, Huneeus and Miyauchi (2023). In these existing models, buyers and suppliers endogenously establish *direct* relationships by posting costly advertisements to participate in matching markets. The key innovation of my model is to explicitly allow firms to trade *indirectly* through wholesalers even when the buyers and suppliers have not established any direct connections. In particular, I assume that sourcing through wholesalers does not require firms to incur search costs; instead, wholesalers impose markups to cover their fixed costs of entry and of searching for upstream suppliers. This assumption captures the notion that by trading indirectly, firms outsource the costly task of forming buyer-supplier networks to wholesalers, thereby economizing on the associated search costs—consistent with my empirical evidence from Turkey that smaller firms,

[1] Recent analyses of firm-to-firm transaction data reveal that production networks are highly sparse, with active buyer-supplier links representing only a small fraction of potential connections. This sparsity has been documented for Belgium (Dhyne et al., 2023) and Chile (Arkolakis, Huneeus and Miyauchi, 2023). Furthermore, larger, more productive firms tend to maintain a greater number of buyers and suppliers. These patterns suggest that establishing buyer–supplier relationships entails sizable fixed costs that only firms with sufficient scale can overcome.

[2] Modern wholesalers also invest heavily in distribution networks and local presence to ensure timely delivery (Ganapati, 2024). These fixed investments further contribute to scale dependence, particularly in global input distribution, where sunk costs and per-shipment fees are substantial (Kasahara and Rodrigue, 2005; Alessandria, Kaboski and Midrigan, 2010).

lacking the necessary scale for direct sourcing, rely more heavily on wholesalers. Consequently, the decision between direct and indirect trade in the model reflects a critical trade-off: firms choosing direct trade incur search and matching costs but benefit from lower per-unit input costs, whereas firms opting for wholesalers avoid search costs at the expense of higher per-unit prices. Lastly, wholesalers compete à la Cournot when reselling input varieties, leading to markups that rise as the number of wholesalers in the market decreases.

I begin the theoretical analysis by studying the social planner's problem to identify the sources of inefficiency in the model. The analysis uncovers two primary inefficiencies. First, double marginalization by wholesalers inflates the prices of indirectly traded inputs relative to directly traded ones, causing a misalignment between relative prices and relative marginal costs of inputs traded directly vs. indirectly. This markup distortion leads to a misallocation of production resources, adversely affecting both the intensive margin—by increasing the volume of directly traded goods per match—and the extensive margin—by generating excessive direct match formation. Second, this excessive direct match formation is further amplified by a congestion externality in matching, as firms do not internalize how their participation lowers the matching rates for others.

While the social planner analysis identifies the wedges in optimality conditions introduced by wholesale markups, it does not directly speak to how they translate into losses in aggregate productivity and welfare.[3] To address this, I next compare the decentralized equilibrium to the first-best benchmark along these two dimensions. This comparison confirms that the divergence in aggregate productivity and welfare strictly increases with the wholesale markup and, in the absence of congestion externalities, vanishes only when the markup equals 1. Building on this result, I use the wholesalers' free entry condition to study how improvements in the efficiency of direct trade influence wholesale market structure. Such shocks reduce firms' reliance on wholesale trade, leading to a contraction in indirect trade share and wholesalers' aggregate profits. Given the presence of substantial fixed entry costs, fewer wholesalers can operate profitably, increasing market concentration. The surviving wholesalers respond by raising their markups, leveraging greater market power. This endogenous increase in wholesale markups exacerbates misallocation and amplifies the wedge between decentralized and efficient levels of aggregate productivity and welfare. Together, these findings reveal a more nuanced view of the welfare implications of technology-induced disintermediation: while improved direct trade productivity enhances efficiency at the firm level, it can also worsen distortions arising from wholesale market power in a general equilibrium setting. This underscores the need for quantitative evaluation to assess the overall welfare impact.

Guided by these theoretical predictions, I next turn to the data and provide causal evidence that

[3]Aggregate productivity is defined as real final output (welfare) per unit of production labor. Aggregate productivity is not a sufficient measure of welfare when network formation and firm entry are endogenous: excessive direct match formation can raise aggregate productivity, despite inducing an inefficient allocation of labor between production and match formation, thereby lowering welfare.

technology-induced disintermediation raises wholesale markups, using the rollout of fiber internet in Turkey as a case study. I assemble a province–pair panel from Turkey's VAT records, which report the value of all domestic formal firm-to-firm transactions above the 5,000 TRY reporting threshold, yielding a near-universe of inter-provincial manufacturing trade. I exploit the staggered rollout of fiber-optic internet as a plausibly exogenous improvement in direct trade technology. Following Demir, Javorcik and Panigrahi (2023), I instrument fiber connectivity with distance to the nearest oil and gas pipeline: optical fiber cables are often laid alongside oil and gas pipelines for monitoring, and a national policy granting internet services providers access to the fiber optic infrastructure built along oil and gas pipelines accelerated expansion. Because the pipeline network predates the internet rollout and was planned for energy logistics, proximity to these pipelines provides plausibly exogenous variation in fiber connectivity. These data and this design allow me to directly test the model's predictions about disintermediation, entry, concentration, and markups.

Using this empirical framework, I establish several key findings. First, province pairs experiencing faster growth in internet connectivity—as measured by the minimum fiber intensity between the two provinces—show a relative decline in the share of indirect trade (Finding 1). This decline is driven by relative increases in both the extensive margin (the number of direct buyer-supplier matches) and the intensive margin (the average trade flow per match) of direct trade (Finding 2). These findings establish that fiber internet expansion facilitates disintermediation by promoting direct trade. Next, I document that provinces with faster fiber internet growth experience a relative decline in the number of wholesalers, with the surviving wholesalers gaining market share (Finding 3). Consistent with the model's predictions, these provinces also experience a relative increase in aggregate wholesale markups (Finding 4). These results strongly support the mechanisms behind my model, confirming the importance of accounting for how technology-induced disintermediation shapes wholesale market structure and its implications for aggregate welfare.

Finally, I conduct a quantitative exercise to evaluate the welfare implications of disintermediation. I do so using a spatial extension of the model to capture heterogeneity across provinces—in both the importance of wholesale trade and the speed of fiber-internet rollout—providing a richer welfare evaluation. The spatial extension also allows me to calibrate shocks that replicate the episode of fiber-internet expansion in Turkey. Specifically, I calibrate shocks to the productivity of directly traded inputs to match (i) the observed decline in indirect trade shares in provinces with relatively faster fiber expansion and (ii) the underlying relative changes of direct and indirect trade flows. The calibrated model successfully reproduces the empirical patterns, including the relative decline in the number of wholesalers and the relative increase in wholesale markups. Quantitatively, fiber-induced disintermediation raised welfare by 4.6 p.p., but higher wholesale markups from increased concentration partly offset these

gains. In particular, rising wholesale markups exacerbated resource misallocation, dampening aggregate welfare improvements by approximately 1.4 p.p. (≈30%). This underscores the value of complementary policies—such as wholesale subsidies—to mitigate markup distortions and fully harness the gains from technology adoption that cut out the middleman.

Related Literature This paper contributes to the growing literature on wholesale intermediaries and connects to a long-standing New Trade Theory insight (Krugman 1979, 1980) that internal scale economies generate imperfect competition. I share the same core economic rationale for the endogenous emergence of wholesalers as recent work: wholesalers economize on fixed transaction costs via aggregation, but performing this aggregation requires sizable fixed investments in local distribution and supplier relationships; the resulting scale-induced limits on entry raise concentration and generate market power. My departure is to examine the implications of this shared mechanism for upstream production: whereas prior studies analyze the downstream distribution game in partial equilibrium, I embed wholesale intermediation in an upstream production network and study how endogenous wholesale markups distort relative input prices and misallocate resources. For example, Ganapati (2024) studies how fixed-cost-intensive technologies reinforce scale, increasing concentration and markups in U.S. wholesale sectors (1992–2012). Similarly, Grant and Startz (2022) analyze how aggregation can endogenously spawn multi-tier intermediation, with markups at each stage to cover fixed entry costs. Both treat upstream production as exogenous.

My paper also contributes to the literature on endogenous production network with wholesale intermediaries.[4] Existing studies in this literature have primarily quantified the aggregate productivity gains from wholesalers reducing matching frictions in production networks (Blum et al., 2023; Manova, Moxnes and Perelló, 2024), but have not examined the inefficiency of network formation when wholesalers are present. By explicitly modeling search costs incurred by final goods producers seeking direct suppliers, I relax the common assumption that directly and indirectly traded inputs are sold at identical prices, allowing me to study the welfare consequences of wholesalers' endogenous market power.[5]

This focus on efficiency implications connects my paper to studies documenting real-world disintermediation driven by technological improvements. For example, Bartkus et al. (2022) evaluate an NGO-led program in the Amazon that provided fishing cooperatives with motorized boats, ice machines,

[4]Relative to the literature examining endogenous production network formation without a specific focus on wholesale intermediaries (e.g., Demir et al. (2024), Dhyne et al. (2022), Dhyne et al. (2023), Eaton, Kortum and Kramarz (2022), Huneeus (2018)), my framework also captures how shocks propagate within production networks, both through the formation of buyer-supplier relationships (Arkolakis, Huneeus and Miyauchi 2023) and via higher-order compositional effects (Baqaee and Farhi 2019). I further highlight how compositional changes can yield first-order welfare impacts through endogenous entry, wholesale markups, and cannibalization.

[5]Perelló (2024) treats intermediation as a passive technology: firms can adopt indirect sourcing to access a wider supplier set and reduce disruption risk, but they pay a fixed brokerage fee that raises variable costs. Because the fixed brokerage fee is exogenous and its nature—resource cost versus pure rent—is left unspecified, the paper does not examine whether wholesale markups distort relative input prices and misallocate resources upstream.

and fuel—technology that enabled fishermen to preserve and transport their catch for direct sale in urban markets. The intervention allowed participating fishermen to bypass traditional middlemen and secure higher sales prices. Similarly, Iacovone and McKenzie (2022) study Agruppa, a start-up platform in Colombia that uses mobile technology to aggregate orders from small fruit and vegetable vendors and deliver produce directly from suppliers. Vendors using the platform reduced their travel and purchase costs, enabling them to pay lower prices for goods compared to buying from wholesale markets.

In both cases, technological innovations that facilitated direct trade delivered clear price improvements to those making the switch—sellers obtaining higher revenues and buyers paying lower input costs—translating into higher welfare. My analysis finds a similar overall welfare improvement from fiber internet expansion, but also reveals a more nuanced effect: disintermediation can raise wholesale market concentration, allowing surviving wholesalers to increase markups. Crucially, because some firms continue to rely on wholesalers, this markup increase widens markup dispersion across intermediate goods and exacerbates the misallocation of production resources. This heightened misallocation partially offsets the aggregate gains from disintermediation, and highlights the importance of complementing technological investments that promote disintermediation with competition policies aimed at limiting markup distortions, so that the full welfare potential of such investments can be realized. This result echoes Grant and Startz (2022), who also caution that general equilibrium considerations can complicate the welfare assessment of technology-induced disintermediation.[6]

Lastly, this paper extends the literature on trade facilitation—in particular, policies that reduce trading frictions through digital infrastructure. While prior research has predominantly focused on the role of digital infrastructure in facilitating international trade (e.g., Fernandes et al. 2019; Malgouyres, Mayer and Mazet-Sonilhac 2021; Akerman, Leuven and Mogstad 2022)[7], my paper shifts the focus to its impact on domestic production networks. It is most closely related to Demir, Javorcik and Panigrahi (2023), who document that fiber internet expansion in Turkey has increased firms' access to input varieties and reduced sourcing concentration. Building on these insights, I provide novel empirical evidence on how fiber internet expansion affects wholesale intermediation, highlighting disintermediation, increased wholesale market concentration, and rising wholesale markups as significant and previously unexplored consequences.

The paper is organized as follows. Section 2 presents a set of motivational facts that guide the setup of the model. Section 3 introduces a model of production network formation with an endogenous composition of indirect trade, wholesale market structure, and markups. Section 4 outlines a set

[6]In their framework, reduction in direct sourcing fixed costs can exacerbate the under-provision of retail varieties, under a generalization of CES preference that allows for a wedge in social and private incentives in creating varieties. I similarly show that technology may worsen efficiency, but through a different mechanism that hinges on endogenous upstream production.

[7]For example, internet rollout has been shown to increase firm-level exports in China (Fernandes et al., 2019), boost firm-level imports in France (Malgouyres, Mayer and Mazet-Sonilhac, 2021), and alter the sensitivity of trade flows to distance in Norway (Akerman, Leuven and Mogstad, 2022).

of theoretical results that provide insights into how wholesale market power affects the efficiency of production network formation and the welfare implications of disintermediation. Section 5 presents empirical evidence from the expansion of fiber internet in Turkey between 2012 and 2019, used as a case study to validate the model's theoretical predictions. Section 6 presents the results of a quantitative exercise simulating shocks that replicate fiber internet expansion to evaluate its welfare impact. Finally, Section 7 concludes.

2 Data and Motivational Facts

2.1 Turkish Firm-Level Data and Firm-to-Firm Transaction Data

The Ministry of Industry and Technology (MoIT) in Turkey integrates administrative records from eight institutions, creating a comprehensive database of all *formal* firms. The primary dataset used in this paper is the VAT record, which reports the value of all domestic firm-to-firm transactions exceeding 5,000 Turkish Liras (≈ USD 840 in 2019). By merging this dataset with the firm registry—containing each firm's province and 4-digit NACE industry code—I construct the near-universe of Turkish inter-provincial trade in manufacturing goods. A key feature of the data is the ability to distinguish between direct trade flows (manufacturers' sales (NACE 10–33) to other manufacturers) and indirect trade flows (manufacturers' sales to wholesale intermediaries (NACE 46)). I also use each firm's sales, wage bill, cost of goods sold, and capital stock, contained in the income statements.

Throughout, I focus exclusively on the direct and indirect trade of manufacturing firms. In 2019, for example, the sample includes 138,503 manufacturing firms and 73,543 wholesalers.[8] The sample period is from 2012 to 2019.

2.2 Motivational Facts about Wholesale Intermediation of Manufacturing Goods in Turkey

In this section, I document a set of facts about wholesale intermediation of manufacturing goods in Turkey to motivate the model I develop in Section 3.

Fact 1: Half of all manufacturing goods trade is intermediated through wholesalers, but the share of indirect trade has been declining

Figure 1 plots the aggregate indirect trade share in the domestic trade of manufacturing goods in Turkey over the sample period (2012-2019). Here, indirect trade refers to total sales of manufacturing goods to wholesalers, direct trade refers to sales to manufacturing firms, and the aggregate indirect trade share is defined as the ratio of indirect trade to the sum of direct and indirect trade. The figure shows

[8]The final sample contains firms that are observed across all data sets.

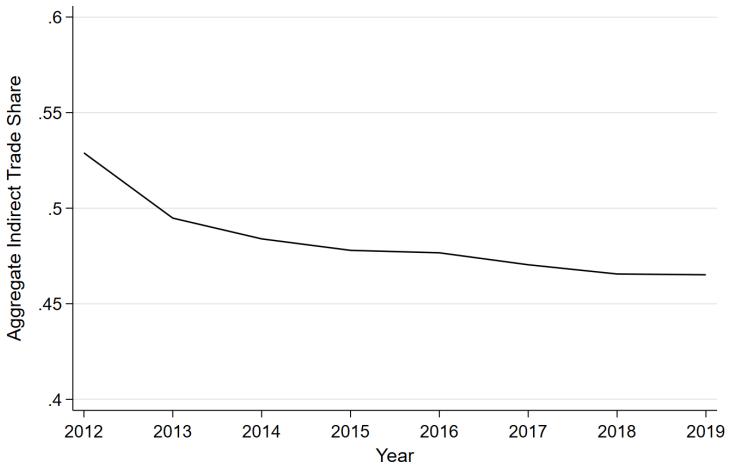

Figure 1: Evolution of Aggregate Indirect Trade Share for Turkey

Note: This chart plots the aggregate indirect trade share in Turkey between 2012 and 2019. Here, indirect trade refers to total sales of manufacturing goods to wholesalers, direct trade refers to sales to manufacturing firms, and the aggregate indirect trade share is defined as the ratio of indirect trade to the sum of direct and indirect trade.

that indirect trade through wholesalers accounted for 53% of the domestic manufacturing goods trade in Turkey in 2012, which declined steadily to 46% by 2019.

Fact 2: Small firms rely more on wholesalers for sourcing

	Indirect Sourcing Sh	**Indirect Sourcing Sh**	**Log Number of Direct Suppliers**	**Log Number of Direct Buyers**
Log Sales	-0.0190***	-0.0146***	0.5512***	0.4367***
	(0.0022)	(0.0012)	(0.0005)	(0.0006)
Industry FE		✓	✓	✓
Province FE		✓	✓	✓
Year FE		✓	✓	✓
Observations	843,543	843,543	795,172	750,942
R-squared	0.008	0.130	0.647	0.499

Table 1: Relationship between indirect sourcing share and firm size

Note: This table reports OLS estimates of the relationship between log sales and indirect sourcing share, as well as the number of direct suppliers and buyers of Turkish manufacturing firms between 2012 and 2019. Here, indirect sourcing refers to the purchases of manufacturing firms from wholesalers, direct sourcing refers to the purchases of manufacturing firms from other manufacturing firms, and indirect sourcing share is the ratio of indirect sourcing to the sum of indirect and direct sourcing. Industry fixed effect controls for the 4-digit NACE industry that each manufacturing firm is associated with. * 10%, ** 5%, *** 1% significance levels. Standard errors clustered at the province level are in parentheses.

Table 1 reports OLS regressions of a firm's indirect sourcing share—purchases from wholesalers divided by total purchases—on log sales. Column 1 shows that a one log-point increase in sales is associated with a 1.9 p.p. decline in the indirect sourcing share; the estimate remains stable after controlling for industry, province, and year fixed effects (Column 2). Relatedly, Columns 3 and 4 of the table report OLS regressions of the number of direct suppliers and buyers on sales: a one log-point

increase in sales is associated with increases of 0.55 (suppliers) and 0.44 (buyers) log points.

Fact 3: Wholesale trade in Turkey is highly concentrated, and competition is local

Panel A: National-Level Sales Concentration (2012)		
Top Percentile of Firms	**Sales Share (%)**	
Top 1% of Firms	43.4%	
Top 5% of Firms	69.6%	
Top 10% of Firms	81.1%	
Panel B: Market-Level Sales Concentration (2012)		
Largest Firms in Market	**Median Share (%)**	**Mean Share (%)**
Top 5 Firms	87.3%	79.8%
Top 10 Firms	98.0%	88.7%
Top 20 Firms	100.0%	94.3%
Each market is a unique province-industry (4-digit NACE) combination		

Table 2: Wholesale Sales Concentration in 2012: National and Market-Level Shares

Note: Panel A reports the share of total national wholesale sales accounted for by the top 1%, 5%, and 10% of wholesalers in 2012. Panel B reports the median and mean share of market-level sales accounted for by the top 5, 10, and 20 firms, where a market is defined as a unique province-industry (NACE 4-digit) combination.

A defining feature of wholesale trade in Turkey is its high degree of sales concentration. In 2012, the top 1% of wholesale firms accounted for 43% of total national wholesale sales, while the top 5% accounted for nearly 70% (Panel A of Table 2). These patterns closely mirror those documented for the U.S. wholesale sector by Ganapati (2024), suggesting extreme concentration is a structural feature of modern wholesale intermediation rather than a country-specific anomaly.

Concentration is also pronounced at the market level. Panel B of Table 2 shows that within province-industry markets, the top five firms captured 87% of wholesale sales at the median and nearly 80% on average in 2012. Thus, a small number of intermediaries dominate both local and national wholesale trade in Turkey. This local concentration is all the more relevant considering most firms source almost exclusively from local wholesalers: the median local wholesale supplier share—a firm's purchases from wholesalers in its own province divided by its total wholesale purchases—is 99%.[9]

While high sales concentration does not by itself imply market power, these figures offer a complementary perspective on market structure—especially when paired with direct measures of pricing power such as markups. This motivates the next fact, which turns to markup-based estimates of wholesalers' market power over time.

Fact 4: Wholesalers charge higher markups than manufacturers, and these rose over time

To estimate firm-level markups, I follow De Loecker and Warzynski (2012) and Edmond, Midrigan

[9]Proximity is a key advantage of indirect sourcing (Grant and Startz, 2022); U.S. wholesalers likewise sell predominantly to nearby destinations (Ganapati, 2024).

and Xu (2023), who show that the markup of firm i in sector s and year t is:

$$\mu_{it}(s) = \frac{p_{it}(s)y_{it}(s)}{w_t l_{it}(s)}\,\alpha_t^l(s), \qquad \alpha_t^l(s) = \frac{w_t l_{it}(s)}{w_t l_{it}(s) + r_t k_{it}(s) + x_{it}(s)}\,\text{RTS},$$

where $p_{it}(s)y_{it}(s)/(w_t l_{it}(s))$ is the inverse labor share in sales, and $\alpha_t^l(s)$ is a sector–year-specific output elasticity of labor. I estimate $\alpha_t^l(s)$ using firms' cost-minimization conditions following Edmond, Midrigan and Xu (2023), where $r_t k_{it}(s)$ and $x_{it}(s)$ denote capital rental and materials, and RTS is returns to scale.[10][11] I assume constant returns to scale (RTS = 1) following their baseline.

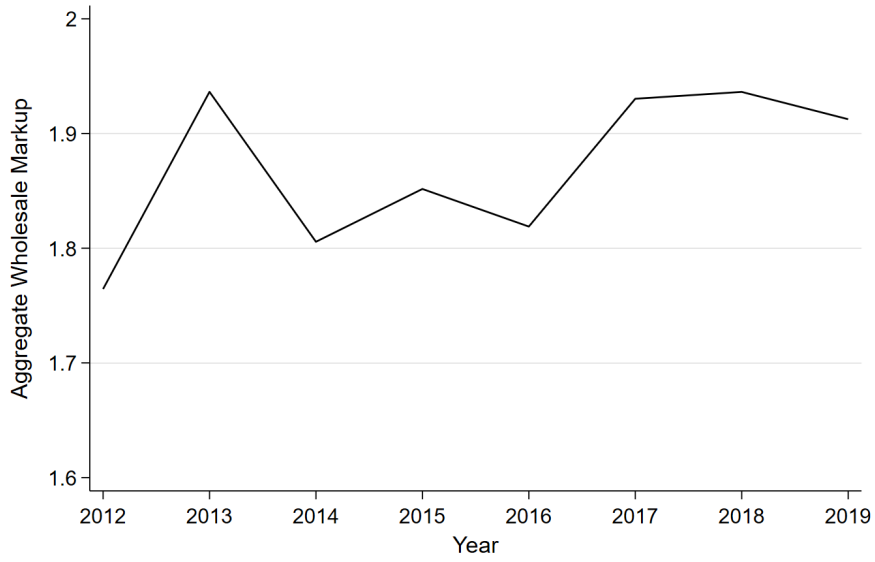

Figure 2: Evolution of Aggregate Wholesale Markup for Turkey

Note: This chart plots the aggregate wholesale markup in Turkey between 2012 and 2019. Aggregate wholesale markup is measured as the cost-weighted average of wholesalers' markups.

Using these firm-level markups, I compute the aggregate wholesale markup as a cost-weighted average, which is plotted over the sample period in Figure 2. Wholesalers exhibit substantial market power: in 2012 the aggregate wholesale markup was 1.76, exceeding manufacturing's 1.56. Moreover, the wholesale markup rose from 1.76 in 2012 to 1.91 in 2019.

Summary. These patterns suggest that forming direct buyer–supplier relationships entails sizable fixed costs that only sufficiently large firms can absorb. Together with the high markups charged by wholesalers, this supports the view that firms face a trade-off between indirect and direct sourcing—indirect sourcing has lower fixed costs but higher variable costs, whereas direct sourcing requires higher fixed costs but yields lower variable costs—consistent with Grant and Startz (2022), who find that traders relying on

[10] Quantities and prices are not observed separately in the Turkish data, so production functions cannot be consistently estimated from revenue alone (Bond et al., 2021).

[11] The Turkish data do not report capital rental costs; I borrow sector–year-specific rental rates from the U.S. BLS and combine them with firms' capital stocks to construct $r_t k_{it}(s)$.

wholesalers tend to be smaller and face higher per-unit prices but lower fixed costs than those sourcing directly. I incorporate this trade-off in the model by assuming that search costs are required for firms building direct connections, while wholesalers charge a markup to cover the pooling of search costs to connect buyers and suppliers.

Motivated by the high level of concentration in wholesale trade, I assume that wholesalers compete oligopolistically à la Cournot. Modeling the endogenous market structure of wholesale trade also allows for the joint determination of indirect trade share and wholesale markup, and an investigation of their evolution over time.

3 Production Network with Wholesale Intermediation

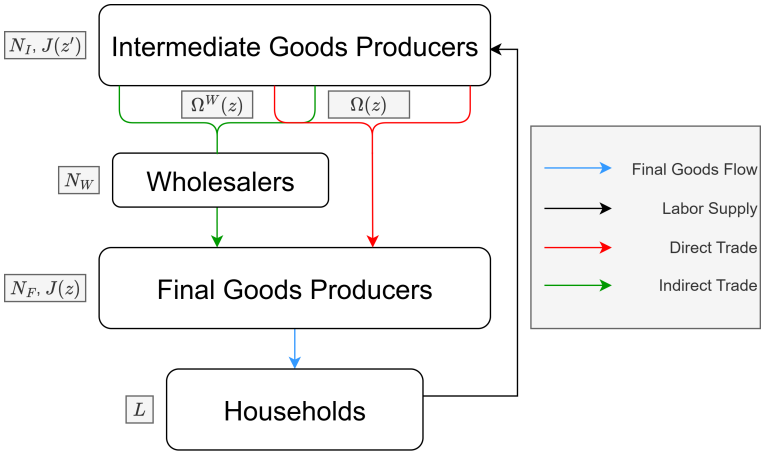

Figure 3: Graphical Illustration of the Production Network (Single Location)

Note: Intermediate goods producers may sell directly to final-good firms or indirectly via wholesalers; wholesale intermediation arises endogenously in equilibrium.

The model builds on the production network framework of Demir et al. (2024) and its spatial extension by Arkolakis, Huneeus and Miyauchi (2023), where firms endogenously form buyer–supplier relationships by posting costly ads in matching markets. My key departures are: buyers and suppliers can trade *indirectly* via wholesalers even when they are not matched directly, and the wholesale market structure and markups are endogenized.

There is an exogenous measure L of households. Each household supplies one unit of labor and receives a wage w. A continuum of intermediate goods producers (measure N_I) produce differentiated varieties using only labor; a continuum of final goods producers (measure N_F) produce differentiated final goods using bundles of intermediate goods. Final goods are then consumed by the households.

There are two modes of trading intermediate goods, illustrated in Figure 3. First, firms can engage in direct trade (red arrows) by posting ads in matching markets and forming direct buyer-supplier matches.

Intermediate goods traded directly are assumed to be more productive, potentially capturing the gains from customization.[12] Alternatively, firms can trade indirectly through wholesalers (green arrows) and economize on the search costs of building buyer-supplier networks by outsourcing this task to wholesalers. There is a discrete number of wholesalers N_W. These wholesalers are exogenously matched with final goods producers, and the costs of reaching buyers are assumed to be covered by fixed entry costs (e.g., setting up physical premises).[13] To offer indirect trade, wholesalers must pay an additional cost to search for intermediate goods suppliers. Once they are matched, the wholesaler buys the variety from the supplier and resells it to many buyers at a markup.[14] Wholesalers compete à la Cournot when reselling to buyers. This assumption allows markups to be determined endogenously and ensures that markups decrease with increased competition (i.e., a larger number of wholesalers).

The choice between direct and indirect trade reflects a trade-off: direct trade requires search costs but yields lower unit costs; indirect trade avoids search costs but entails higher unit costs.[15]

N_I and N_F are pinned down by free entry conditions, implying zero aggregate profits for intermediate and final producers. N_W is discrete and also satisfies a free entry condition; discreteness allows aggregate post-entry profit to exceed aggregate entry cost, with the excess rebated to households. Household income thus includes labor earnings and rebated wholesaler profits. Intermediate and final producers draw productivity z post-entry from $J(z)$ with density $j(z)$; the measure of type-z intermediates is $N_I j(z)$ (each a distinct variety). The same holds for final producers; wholesalers are homogeneous.

I denote $\Omega(z)$ as the set of intermediate goods each type-z final goods producer can purchase directly, if there exists a direct match with the supplier. Likewise, $\Omega^W(z)$ denotes the set of intermediate goods each type-z final goods producer can purchase indirectly through wholesalers if the supplier is matched with the wholesalers. Therefore, $\{\Omega(\cdot), \Omega^W(\cdot)\}$ represent the production network. In what follows, I will start by describing the sourcing and pricing decisions of firms and wholesalers taking the network as given. I will then proceed to discuss the endogenous formation of the network.

[12] For example, when buying electronic control boards or chips directly, the supplier can tailor settings and screening (custom pass/fail tests and chip parameters)—power limits, timing tolerances (how much early or late a signal can be and still be read correctly), and thermal safeguards—to fit the product's case and cooling.

[13] Prominent wholesalers bear sizable *local* fixed/sunk costs (e.g., regional service centers, cold-chain hubs, storage terminals, compliance labs), allowing nearby buyers to avoid upstream travel, negotiation, and compliance costs.

[14] Wholesalers bundle thousands of SKUs from many suppliers (e.g., metals, industrial supplies, chemicals), pooling relationship and inventory-management costs so downstream buyers can source multiple inputs on one account and shipment.

[15] This trade-off is common in practice: steel mills (ArcelorMittal, Nucor) and chemical majors (Dow, BASF) often serve a broad tail of smaller manufacturers via service centers and distributors rather than one-off contracts. On the *buyer* side, small and mid-sized firms face fixed costs to locate, vet, and negotiate with each upstream producer (engineering audits, minimum-order negotiations, legal/compliance checks, per-SKU logistics setup). On the *supplier* side, mills and chemical producers incur relationship-specific costs to serve many small accounts (credit, fragmented invoicing, tailored packaging/shipment sizes, spec certification, sales/technical support). Wholesalers aggregate these costs and standardize terms, making indirect trade attractive despite higher per-unit prices.

3.1 Sourcing and Pricing Decisions given Network

Households. Each household consumes a CES bundle of all the final goods varieties:

$$\left[N_F \int_Z c^H(z)^{\frac{\sigma-1}{\sigma}} \, j(z) \, dz \right]^{\frac{\sigma}{\sigma-1}}$$

Income I comes from wage earnings and wholesalers' net profit Π^W:

$$I = w\,L + \Pi^W \tag{1}$$

Utility maximization yields the price index:

$$P^H = \left[N_F \int_Z p_F(z)^{1-\sigma} \, j(z) \, dz \right]^{\frac{1}{1-\sigma}} \tag{2}$$

and demand for final goods producers:

$$p_F(z)\, c^H(z) = p_F(z)^{1-\sigma} \, D_H \;,\; D_H \equiv \frac{I}{P^{H\,1-\sigma}}$$

where $p_F(z)$ is the price charged by a type-z final goods producer.

Final Goods Producers. Consider a final goods producer with productivity z. The firm's production function is $z\,Y_I(z)$, where $Y_I(z)$ is a CES aggregate of intermediates sourced either directly or indirectly:

$$Y_I(z) = \left\{ \int_{\omega \in \Omega(z)} y_I(z,\omega)^{\frac{\sigma-1}{\sigma}} \phi_c^{\frac{1}{\sigma}} \, d\omega + \int_{\nu \in \Omega^W(z) \setminus \Omega(z)} y^W(z,\nu)^{\frac{\sigma-1}{\sigma}} \, d\nu \right\}^{\frac{\sigma}{\sigma-1}} \tag{3}$$

where $y_I(z,\omega)$ is the quantity of variety $\omega \in \Omega(z)$ that is sourced directly by the type-z buyer from its supplier; $y^W(z,\nu)$ is the quantity of variety $\nu \in \Omega^W(z) \setminus \Omega(z)$ that is sourced indirectly by the type-z buyer through wholesalers; $\phi_c \geq 1$ captures the customization productivity gains available only for directly traded inputs; and σ is the elasticity of substitution across intermediate goods varieties.[16] Notice that the firm only sources a variety ν indirectly if its producer is matched with the wholesalers, i.e. if $\nu \in \Omega^W(z)$, and if that producer is not directly matched with the firm, i.e. if $\nu \notin \Omega(z)$. This is because directly traded inputs are both more productive due to customization benefits and cheaper. The latter advantage will become clear when we discuss wholesalers' pricing decisions. Cost minimization

[16] Equation (3) highlights two departures from Arkolakis, Huneeus and Miyauchi (2023). First, wholesalers only resell intermediate goods without any transformation and do not create new varieties in the process, so indirect trade competes with and can be cannibalized by direct sales if a direct match is formed. Second, I do not impose a Cobb–Douglas aggregation between directly and indirectly traded bundles; instead, the same CES elasticity σ applies across and within bundles. Removing the Cobb–Douglas restriction lets the model endogenously generate compositional shifts between direct and indirect trade—crucial for studying disintermediation's welfare implications.

then yields the marginal cost of production for the type-z final goods producer:

$$c(z) = \left\{ \int_{\omega \in \Omega(z)} p_I(\omega)^{1-\sigma} \, \phi_c \, d\omega + \int_{\nu \in \Omega^W(z) \setminus \Omega(z)} p^W(\nu)^{1-\sigma} \, d\nu \right\}^{\frac{1}{1-\sigma}} \tag{4}$$

where $p_I(\omega)$ is the price of a directly traded input variety ω, and $p^W(\nu)$ is the price of an input variety ν that the firm pays when sourcing it from wholesalers.

The firm is matched exogenously to all households and competes monopolistically against other final goods producers. Thus, the firm sets $p_F(z)$ to charge a constant markup over its marginal cost:

$$p_F(z) = \frac{\sigma}{\sigma - 1} \frac{1}{z} c(z)$$

Intermediate Goods Producers. Consider an intermediate goods producer with productivity z'. The firm has a production function that is linear in labor input: $z' \, l$. Since the firm competes against other intermediate goods producers monopolistically when supplying to any directly matched final goods producer, it charges a constant markup over its marginal cost:

$$p_I(z') = \frac{\sigma}{\sigma - 1} \frac{1}{z'} w$$

where $p_I(z')$ is the price charged by a type z' intermediate goods producer when selling to any directly matched final goods producer. The firm also sells indirectly through wholesalers to any final goods producer that is not directly matched with it, if it is matched with wholesalers. I will revisit the price $p_W(z')$ it sets when selling to wholesalers in that case after discussing wholesalers' pricing decision below.

Wholesalers. Consider an intermediate goods variety ν that has been matched with the wholesalers. The finite number of N_W homogeneous wholesalers together compete monopolistically against other intermediate goods varieties when reselling this variety ν to any final goods producer that is not directly matched with the supplier of ν. Therefore, the wholesale sector faces an isoelastic demand with demand elasticity $-\sigma$. Now, these N_W wholesalers compete against each other à la Cournot in reselling variety ν, and therefore charge a markup over the cost of sourcing this variety from the supplier of ν—$p_W(\nu)$—that is decreasing in N_W:

$$p^W(\nu) = \mu^W \, p_W(\nu), \; \mu^W \equiv \frac{N_W \, \sigma}{N_W \, \sigma - 1}$$

The supplier of ν, knowing that demand from indirect sales is proportional to $p^W(\nu)^{-\sigma}$ and thus $p_W(\nu)^{-\sigma}$, effectively faces an isoelastic demand with demand elasticity $-\sigma$ and charges the same markup

when selling to wholesalers, i.e.

$$p_W(\nu) = p_I(\nu) = \frac{\sigma}{\sigma - 1} \frac{1}{z(\nu)} w$$

Consequently,

$$p^W(z') = \mu^W p_W(z') = \mu^W p_I(z') > p_I(z')$$

Wholesalers' double marginalization makes indirectly traded inputs relatively more expensive.[17]

3.2 Production Network Formation

Now, I describe how the production network $\{\Omega(\cdot), \Omega^W(\cdot)\}$ is formed endogenously through (1) search and matching between intermediate and final goods producers; and (2) supplier search by wholesalers.

3.2.1 Direct Network

Intermediate and final goods producers participate in a matching market to build direct connections $\{\Omega(\cdot)\}$ by posting ads. The number of ads posted by an intermediate goods producer with productivity z' is denoted as $v(z')$ (v stands for visibility), while the number of ads posted by a final goods producer with productivity z is denoted as $m(z)$ (m stands for material).

The total measures of ads searching for buyers (V) and suppliers (M) are:

$$M = N_F \int_Z m(z)\, j(z)\, dz \qquad (5) \qquad\qquad V = N_I \int_Z v(z)\, j(z)\, dz \qquad (6)$$

Following Arkolakis, Huneeus and Miyauchi (2023), I assume there to be a Cobb-Douglas matching function that determines the number of matches generated from the ads:

$$\widetilde{M} = \kappa V^{\lambda_V} M^{\lambda_M}$$

where κ governs matching efficiency.

Denote the success rate of ads searching for buyers θ^v and suppliers θ^m as:

$$\theta^v = \frac{\widetilde{M}}{V} = \kappa V^{\lambda_V - 1} M^{\lambda_M} \qquad (7) \qquad\qquad \theta^m = \frac{\widetilde{M}}{M} = \kappa V^{\lambda_V} M^{\lambda_M - 1} \qquad (8)$$

[17] This is consistent with Motivational Fact 5 in Appendix D.1, which finds no evidence of additional markdowns when selling to wholesalers. While resale price maintenance (RPM) could, in principle, mitigate this inefficiency, implementing it is legally risky in Turkey. Fixed and minimum RPM are per se unlawful; maximum or recommended prices are only permitted if they don't become de-facto fixed/minimum and, above the 30% market-share threshold, lose the block-exemption safe harbor and face case-by-case scrutiny. In practice, the stronger the supplier, the greater the risk that a "maximum" becomes a focal point, and is more likely to be challenged by authority. In my environment each intermediate goods producer is the sole source of its variety, so a binding, enforceable max-RPM is especially unlikely. Absent effective coordination, the upstream supplier posts its preferred price and wholesalers add their own markup—exactly the double-marginalization structure I model.

The measure of type z' suppliers matched to a type z buyer is:

$$\overline{m}(z, z') = m(z)\,\theta^m\,\frac{N_I\,v(z')\,j(z')}{V}$$

that is, the buyer's ads times their success rate times the share of seller ads accounted for by type z'.

Similarly, the measure of type z buyers matched to a type z' supplier is:

$$\overline{v}(z', z) = v(z')\,\theta^v\,\frac{N_F\,m(z)\,j(z)}{M}$$

3.2.2 Indirect Network

While the N_W wholesalers are matched with all final goods producers after paying the fixed entry cost, they must exert effort s to search for suppliers and start trading intermediate goods. I model wholesalers' supplier search as one-sided matching. To allow congestion, only a fraction θ^W of effort converts into successful matches:

$$\theta^W \equiv \frac{(N_W\,s)^{\lambda_W}}{N_W\,s}\ ,\ \lambda_W \leq 1$$

Define $S \equiv s\,\theta^W$. The number of varieties matched with each wholesaler is assumed to be given by $S\,N_I$, which has a natural upper bound: $S\,N_I \leq N_I$. Consequently, $S \leq 1$ can be interpreted as the share or probability a variety is matched with and traded by wholesalers. I further assume that each variety matched with wholesalers is randomly drawn from the firm distribution $J(\cdot)$. Thus, the measure of type z' intermediate goods that are matched with wholesalers is given by $S\,N_I\,j(z')$. Note that the set of matched varieties is identical across all wholesalers; there is no variation across them in this regard.

I assume that final goods producers do not incur search costs when sourcing indirectly through wholesalers. This is a normalization that captures the notion that firms effectively *outsource* the costly task of forming buyer-supplier networks to wholesalers, who bear these search costs as part of their fixed investment.[18] By paying this search cost once per matched variety, wholesalers can *pool* supplier relationships and resell to many downstream buyers. Buyers thus form a single link to the wholesaler rather than separate links to each supplier, gaining access to a broad set of input varieties without duplicating search effort.

Pooling, however, requires large upfront investments in entry and supplier search, so profitability hinges on scale. Limited entry follows, and with fewer wholesalers, Cournot competition generates a wholesale markup $\mu^W > 1$ on indirectly traded inputs ("double marginalization"). In equilibrium, indirect trade therefore helps firms avoid search and matching costs but entails higher per-unit prices.

[18] Alternatively, introducing a small search cost to be paid by firms to connect with wholesalers would be innocuous if such search cost is low enough that every firm pay it, and can therefore be subsumed in the fixed entry cost of firms.

3.2.3 Optimal Search by Firms and Wholesalers

We can now express the production network $\{\Omega(\cdot), \Omega^W(\cdot)\}$ in terms of the measure of direct and indirect matches. Specifically, we can expand the CES aggregate of intermediates (3) by their type z':

$$Y_I(z) = \left\{ \int_Z \underbrace{y_I(z,z')^{\frac{\sigma-1}{\sigma}} \, \phi_c^{\frac{1}{\sigma}} \, \overline{m}(z,z')}_{\text{directly sourced}} + \underbrace{y^W(z,z')^{\frac{\sigma-1}{\sigma}} \, S\left[N_I \, j(z') - \overline{m}(z,z')\right]}_{\text{indirectly sourced}} dz' \right\}^{\frac{\sigma}{\sigma-1}} \tag{9}$$

where $N_I j(z')$ denotes the total measure of type z' intermediates; hence $N_I j(z') - \overline{m}(z,z')$ are those not directly matched. Multiplying by S (the share matched with wholesalers) gives the measure of intermediate varieties matched with wholesalers that are not directly matched with a type z final goods producer.

We can also expand the unit cost of intermediate goods bundle (4) by their type z':

$$c(z, m(z)) = \left[m(z) \, \theta^m \, \phi_c \, c_m^{1-\sigma} + S \, N_I \, c_W^{1-\sigma} - S \, m(z) \, \theta^m \, \mu^{W\,1-\sigma} \, c_m^{1-\sigma} \right]^{\frac{1}{1-\sigma}} \tag{10}$$

where $c_m = \left[\int_Z p_I(z')^{1-\sigma} \, \frac{N_I \, v(z') \, j(z')}{V} \, dz' \right]^{\frac{1}{1-\sigma}}$ is the CES aggregator of the prices of directly traded intermediate goods, using a search effort weighted probability density function of productivities; and $c_W = \left[\int_Z p^W(z')^{1-\sigma} \, j(z') \, dz' \right]^{\frac{1}{1-\sigma}}$ the CES aggregator of the prices of indirectly traded intermediate goods, using the unweighted probability density function of productivities. Intuitively, the first term inside the square bracket of (10) captures the unit cost of a bundle of directly traded inputs (raised by power $1-\sigma$), which decreases with the mass of direct suppliers ($m(z)\,\theta^m$, due to the love-of-variety effect) and the customization productivity gains (ϕ_c). The second and the third terms capture the unit cost of a bundle of indirectly traded inputs (raised by power $1-\sigma$), which decreases with the number of varieties matched with wholesaler ($S\,N_I$) *net of* cannibalization by direct matches ($S\,m(z)\,\theta^m$).[19] Given this unit cost, the final goods producer's revenue $x_H(z, m(z))$ and variable profit $\pi_F(z, m(z))$ can be written as:

$$x_H(z, m(z)) \equiv \left(\frac{\sigma}{\sigma - 1} \, \frac{c(z, m(z))}{z} \right)^{1-\sigma} D_H \,, \qquad \pi_F(z, m(z)) \equiv \frac{1}{\sigma} \, x_H(z, m(z)) \tag{11}$$

We can write the revenue of a type z' intermediate goods producer from direct sales $x_m(z', v(z'))$ as

$$x_m(z', v(z')) \equiv p_I(z')^{1-\sigma} \, v(z') \, \theta^v \, D_m \tag{12}$$

[19] Direct matches cannibalize indirect matches in sourcing in the sense that a direct match to a supplier who is also matched with wholesalers represents less than a pure variety gain to the buyer.

where D_m is the average demand per direct buyer.[20]Moreover, the revenue from indirect sales $x_W(z') - x_{Wm}(z', v(z'))$ is given by:

$$x_W(z') \equiv p_I(z')^{1-\sigma} S D_W\,,\ x_{Wm}(z', v(z')) \equiv p_I(z')^{1-\sigma} v(z')\,\theta^v\, S\,\mu^{W-\sigma}\,\phi_c^{-1} D_m \tag{13}$$

D_W is the total demand from indirect sales to all downstream buyers, which is multiplied by S, the probability of being matched with wholesalers, to represent the expected demand from indirect sales. But this demand is cannibalized by direct sales, as captured by $x_{Wm}(z', v(z'))$, which increases with the number of direct buyers $v(z')\,\theta^v$. However, the average demand of these cannibalized sales is less than the average demand per direct buyer due to wholesaler double marginalization ($\mu^{W-\sigma}$), and the lack of customization for indirect sales (ϕ_c^{-1}). Define variable profit of a type z' intermediate goods producer as:

$$\pi_I(z', v(z')) \equiv \frac{1}{\sigma}\left(x_m(z', v(z')) + x_W(z') - x_{Wm}(z', v(z'))\right)$$

Next, the total sales of wholesalers in reselling this type z' variety, *conditional on it being adopted*, is just $(\mu^W/S)\,(x_W(z') - x_{Wm}(z'))$. Thus, the expected per-variety profit Π_W for a wholesaler when each matched variety is drawn randomly from $j(\cdot)$ is:

$$\Pi_W \equiv \int_Z \frac{1}{N_W}\,\frac{1}{N_W\,\sigma}\,(\mu^W/S)\,(x_W(z') - x_{Wm}(z'))\,j(z')\,dz' \tag{14}$$

Therefore, the total resale profit of a wholesaler is given by $\pi_W(s) \equiv s\,\theta^W\,N_I\,\Pi_W$.

Now, I assume that posting v ads for buyers incurs a convex search cost $f_I(v) \equiv w\,f_v\,v^\beta/\beta$, posting m ads for suppliers incurs a convex search cost $f_F(m) \equiv w\,f_m\,m^\beta/\beta$, while exerting a level s of search effort incurs a convex search cost $f_W(s) \equiv w\,f_W\,(s\,N_I/N_W)^{\beta_W}/\beta_W$. All these costs are paid in units of labor, with $f_v, f_m, f_W > 0$ and $\beta, \beta_W > 1$. Wholesalers' search cost rises with the number of varieties matched ($s\,N_I$) and falls with N_W. The latter assumption is imposed to capture potential knowledge spillover among wholesalers that reduces information friction inhibiting the matching with upstream varieties. As discussed in 4.1, this implies efficient wholesaler entry in the decentralized equilibrium when there is no congestion in wholesalers' search for suppliers ($\lambda_W = 1$).

The structure of the optimization problems behind the choice of v, m, s is very similar: intermediate goods producers trade off the cost of posting ads for more direct buyers, higher demand, and therefore higher variable profits; likewise, final goods producers trade off the cost of posting ads for more direct suppliers, lower unit cost, and again higher variable profits; lastly, wholesalers trade off the search cost for more matched varieties and higher resale profits. Therefore, we can write their optimization problems

[20]Detailed derivations of the demand shifters are provided in Appendix B.1.

generically as:

$$\max_{a} \qquad \pi_i(z,a) - f_i(a)\,,\ i = I, F, W$$

And the first order conditions are:

$$m(z) = \Pi_m\, z^{\frac{\sigma-1}{\beta-1}}, \qquad \Pi_m \equiv \left[\frac{1}{w\, f_m}\,\frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}\left(\phi_c - S\,\mu^{W\,1-\sigma}\right)\theta^m\, c_m^{1-\sigma}\, D_H\right]^{\frac{1}{\beta-1}} \tag{15}$$

$$v(z') = \Pi_v\, z'^{\frac{\sigma-1}{\beta-1}}, \qquad \Pi_v \equiv \left[\frac{1}{w\, f_v}\,\frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} w^{1-\sigma}\left(1 - S\,\mu^{W-\sigma}\,\phi_c{}^{-1}\right)\theta^v\, D_m\right]^{\frac{1}{\beta-1}} \tag{16}$$

$$s = \left[\left(\frac{N_W}{N_I}\right)^{\beta_W}\,\frac{1}{w\, f_W}\,\theta^W\, N_I\,\Pi_W\right]^{\frac{1}{\beta_W-1}} \tag{17}$$

Intuitively, optimal $m(z)$, $v(z')$, and s rise with the marginal increase in variable profit associated with more ad posting or search effort relative to their marginal costs, and more productive firms post more ads. A marginal increase in $m(z)$ leads to θ^m more direct suppliers, which reduces the unit cost, and raises variable profit. But the increase is dampened by the cannibalization of inputs that have already been matched with wholesalers and therefore could have been sourced indirectly. Likewise, a marginal increase in $v(z')$ results in θ^v more direct buyers, which raises demand and therefore variable profit. But again, the increase is dampened by the cannibalization of indirect sales. Lastly, a marginal increase in s generates $\theta^W\, N_I$ additional matched varieties and raises resale profit for wholesalers.[21]

3.3 Equilibrium

Free entry pins down N_I, N_F, and the discrete N_W. Aggregate post-entry profits weakly exceed aggregate entry costs, with equality for intermediate and final goods producers. Entry costs are paid in labor, and their levels are controlled by the parameters F_I, F_F, and F_W. The last equilibrium condition is the labor market clearing condition, which states that the total supply of labor is equal to the total demand for it, which consists of labor demand for intermediate goods production, for posting search ads, for financing wholesalers' search effort, and for financing entry of firms and wholesalers. Details of these equilibrium conditions are provided in Appendix B.2.

The general equilibrium is defined by the set of endogenous variables $\{I, P^H, \theta^v, \theta^m, S, N_I, N_F, N_W\}$ that solve (1), (56), (7), (8), (17), (51), (53), and (55).

[21] There is an implicit timing assumption leading up to wholesalers' optimal search effort s: in the first stage, manufacturing firms choose the number of ads, taking the probability of being matched with wholesalers S as given. Each firm is infinitesimal and ignores its impact on S. In the second stage, a finite number N_W of wholesalers choose their search effort s, taking the realized direct network and the success rate θ^W as given. A stationary rational-expectations equilibrium requires that firms' beliefs about S coincide with the one implied by wholesalers' optimal search effort. This timing assumption greatly simplifies the problem by ruling out strategic use of s by wholesalers to influence firms' search decisions.

4 Theoretical Results

This section presents the core theoretical results on how wholesale market power distorts production network formation and shapes the aggregate welfare implications of disintermediation. I begin with the social planner's problem to identify the inefficiencies in the decentralized equilibrium. Next, I compare the decentralized allocation to the first-best to quantify the resulting losses in aggregate productivity and welfare, showing that both rise with the wholesale markup. Finally, I characterize the endogenous response of wholesale market structure to improvements in direct trade technology.

4.1 Social Planner's Problem

The planner chooses quantities of directly and indirectly traded inputs $y_I(z,z')$ and $y^W(z,z')$, firms' ads $m(z)$ and $v(z)$, wholesalers' search effort s, and entry levels N_I, N_F, and N_W to maximize aggregate household consumption:

$$\max_{\{N_I,N_F,N_W,m(z),v(z),s,y_I,y^W\}} L\left[N_F\int_Z c^H(z)^{\frac{\sigma-1}{\sigma}}j(z)dz\right]^{\frac{\sigma}{\sigma-1}}.$$

Consumption of each final good z depends on a CES aggregator over inputs:

$$c^H(z)=z\left\{\int_Z\left[y_I(z,z')^{\frac{\sigma-1}{\sigma}}\phi_c^{\frac{1}{\sigma}}\overline{m}(z,z')+y^W(z,z')^{\frac{\sigma-1}{\sigma}}S\left(N_Ij(z')-\overline{m}(z,z')\right)\right]dz'\right\}^{\frac{\sigma}{\sigma-1}}.$$

Subject to labor resource constraint:

$$\begin{aligned}
L &= L_P + L_A + L_S + L_E,\\
L_A &\equiv \int_Z\left[N_F\,f_m\,\frac{m(z)^\beta}{\beta}+N_I\,f_v\,\frac{v(z)^\beta}{\beta}\right]j(z)\,dz\\
L_S &\equiv N_W\,f_W\,\frac{\left(\frac{s\,N_I}{N_W}\right)^{\beta_W}}{\beta_W}\\
L_E &\equiv F_I\,N_I + F_F\,N_F\; + F_W\,F_W
\end{aligned}$$

where L_P, L_A, L_S, and L_E are labor used for intermediate goods production, posting search ads, financing wholesalers' search effort, and firm/wholesaler entry respectively.

The focus is the gap between the decentralized allocation and the first-best due to two distortions: (i) markup wedges from wholesale market power, and (ii) congestion externalities in matching. I compare the planner's and decentralized first order conditions to identify these inefficiencies.

Wedge in input quantity. Combining the planner's first order conditions in $y_I(z,z')$ and $y^W(z,z')$ yields:

$$\frac{y_I(z,z')}{y^W(z,z')} = \phi_c$$

whereas the decentralized first order conditions imply:

$$\frac{y_I(z,z')}{y^W(z,z')} = \mu^W \phi_c$$

Conditional on the matches, the quantity of indirectly traded inputs is inefficiently low as wholesale markup inflates the price of indirectly traded ones.

Wedges in ad posting. Wholesaler's double marginalization distorts not only the intensive margin but also match formation. The planner's first order condition in $v(z)$ is:

$$\frac{N_F\, f_m\, m(z)^{\beta-1}\, j(z)}{P^H(1)} = \frac{L_P}{P^H(1)} \frac{\overline{\psi}(1)\, j_{ZM}(z)}{m(z)} \frac{1}{\sigma-1} + \frac{L_P}{P^H(1)} \frac{\overline{\psi}(1)\, j_M(z)}{m(z)} (\lambda_M - 1) \frac{1}{\sigma-1} \tag{18}$$

where

$$j_M(z) \equiv \frac{N_F\, m(z)\, j(z)}{M}, \qquad j_V(z) \equiv \frac{N_I\, v(z)\, j(z)}{V}, \qquad j_{ZM}(z) \equiv \frac{z^{\sigma-1}\, j_M(z)}{\int_Z z^{\sigma-1}\, j_M(z) dz}$$

$$\overline{P^H}(x) \equiv \left\{ (\phi_c - x\, S)\, \widetilde{M} \left[\int_Z z^{\sigma-1}\, j_M(z) dz\right] \left[\int_Z z^{\sigma-1}\, j_V(z) dz\right] + x\, N_F\, N_I\, S \left(\mathbb{E}\left[z^{\sigma-1}\right]\right)^2 \right\}^{\frac{1}{1-\sigma}}$$

$$\overline{\psi}(x) \equiv \frac{(\phi_c - x\, S)\, \widetilde{M} \left[\int_Z z^{\sigma-1}\, j_M(z) dz\right] \left[\int_Z z^{\sigma-1}\, j_V(z) dz\right]}{\overline{P^H}(x)^{1-\sigma}}$$

$j_M(\cdot)$ and $j_V(\cdot)$ are search effort weighted probability density function of productivities. $j_{ZM}(\cdot)$ is a final goods producers' direct sales weighted probability density function of productivities, taking into account endogenous direct match formation. $\overline{P^H}(1)$ is the planner's counterpart to the aggregate price index, and therefore $1/\overline{P^H}(1)$ is the aggregate productivity. $\overline{\psi}(1)$ is the direct trade share net of cannibalized indirect trade in planner's allocation.

The left hand side of 18 says the marginal social cost of $m(z)$ is given by the product of the required additional labor and aggregate productivity, yielding the amount of aggregate consumption forgone. The right hand side captures the marginal social benefit of $m(z)$. The first term is the production labor times the marginal increase in aggregate productivity, which is proportional to the sales share given by $\overline{\psi}(1)\, j_{ZM}(z)/m(z)$, as well as the constant $1/(\sigma-1)$ that scales the marginal increase in aggregate productivity depending on how substitutable final goods varieties are. The second term on the right is the production labor times the marginal decrease in aggregate productivity due to congestion externality in direct matching (when $\lambda_M < 1$, the empirically relevant case).

In contrast, firm's first order condition in $m(z)$ in the decentralized equilibrium can be written as:

$$\frac{N_F\, f_m\, m(z)^{\beta-1}\, j(z)}{\overline{P^H}(1)} = \frac{L_P}{\overline{P^H}(1)} \frac{\overline{\psi}(\mu^{W^{1-\sigma}})\, j_{ZM}(z)}{m(z)} \frac{1}{\sigma-1} \frac{\sigma}{\sigma-1} \frac{\overline{P^H}(\mu^{W^{1-\sigma}})^{1-\sigma}}{\overline{P^H}(\mu^{W^{-\sigma}})^{1-\sigma}} \tag{19}$$

While the marginal private cost (MPC) coincides with the marginal social cost (MSC), the marginal private benefit (MPB) of posting ads is higher than the marginal social benefit (MSB), leading to an inefficiently high number of ads being posted. $\overline{\psi}(\mu^{W^{1-\sigma}}) > \overline{\psi}(1)$ reflects how wholesalers' markup raises the price of indirectly traded inputs relative to direct ones, understating the extent of cannibalization and inflating the private return to ads. Moreover, firms fail to internalize how additional ads reduce market tightness and ad success for all, again overstating MPB. Lastly, there is also a misalignment between private and social incentives to form matches: under CES, alignment requires the monopolistic markup with the firm capturing a $1/(\sigma-1)$ share of the social production cost (Dhingra and Morrow, 2019), whereas in the decentralized equilibrium, final goods sales embed three markup layers—intermediate, wholesale, and final—exceeding the monopolistic markup needed to align private and social incentives, leading to excessive ad posting. In particular, the last fraction $\overline{P^H}(\mu^{W^{1-\sigma}})^{1-\sigma}/\overline{P^H}(\mu^{W^{-\sigma}})^{1-\sigma}$ captures the trade share weighted average of wholesale markup, i.e. $1\times\Omega(\mu^{W^{-\sigma}})+\mu^W\times(1-\Omega(\mu^{W^{-\sigma}}))$, where $\Omega(\mu^{W^{-\sigma}})$ is the direct trade share, exclusive of wholesale markup, defined via:

$$\Omega(x) \equiv \frac{\phi_c\, \widetilde{M}\, \left[\int_Z z^{\sigma-1}\, j_M(z)dz\right]\, \left[\int_Z z^{\sigma-1}\, j_V(z)dz\right]}{\overline{P^H}(x)^{1-\sigma}}$$

The same logic applies to $v(z)$. Except that there is no misalignment between private and social incentives to form matches as intermediate goods sales only embed one layer of monopolistic markup.

Wedges in firm entry. The planner's first order conditions in N_F and N_I are:

$$\left[F_F + \int_Z f_m \frac{m(z)^\beta}{\beta}\, j(z)\, dz\right] \frac{1}{\overline{P^H}(1)} = \frac{L_P}{\overline{P^H}(1)} \frac{1}{\sigma-1} \left[\frac{1}{N_F} + \frac{\overline{\psi}(1)}{N_F}\,(\lambda_M - 1)\right] \tag{20}$$

$$\left[F_I + \int_Z f_v \frac{v(z)^\beta}{\beta}\, j(z)\, dz + f_W \frac{(s\, N_I/N_W)^{\beta_W-1}}{\beta_W}\right] \frac{1}{\overline{P^H}(1)} = \frac{L_P}{\overline{P^H}(1)} \frac{1}{\sigma-1} \left[\frac{1}{N_I} + \frac{\overline{\psi}(1)}{N_I}\,(\lambda_V - 1)\right] \tag{21}$$

While the firms' free entry conditions in the decentralized equilibrium can be written as:

$$\left[F_F + \int_Z f_m \frac{m(z)^\beta}{\beta}\, j(z)\, dz\right] \frac{1}{\overline{P^H}(1)} = \frac{L_P}{\overline{P^H}(1)} \frac{1}{N_F} \frac{1}{\sigma-1} \frac{\sigma}{\sigma-1} \frac{\overline{P^H}(\mu^{W^{1-\sigma}})^{1-\sigma}}{\overline{P^H}(\mu^{W^{-\sigma}})^{1-\sigma}} \tag{22}$$

$$\left[F_I + \int_Z f_v \frac{v(z)^\beta}{\beta}\, j(z)\, dz\right] \frac{1}{\overline{P^H}(1)} = \frac{L_P}{\overline{P^H}(1)} \frac{1}{N_I} \frac{1}{\sigma-1} \tag{23}$$

Similar to $m(z)$, there is an excessive entry of final goods producers due to their failure to internalize the congestion externality in ad posting, and the misalignment between private and social incentives to

form matches due to the three markup layers embedded in final goods sales. There is also an excessive entry of intermediate goods producers, because they fail to internalize both the congestion externality and the effect of their additional entry on raising the search cost for wholesalers.

Wedges in wholesalers' search effort The planner's first order condition in wholesalers' search effort s is:

$$N_W f_W \left(\frac{s N_I}{N_W}\right)^{\beta_W - 1} \frac{N_I}{N_W} \frac{1}{\overline{P^H}(1)} = \frac{L_P}{\overline{P^H}(1)} (1 - \Omega(1)) \frac{1}{s} \frac{1}{\sigma - 1} \lambda_W \tag{24}$$

which equates the MSC of exerting search effort, given by the product of forgone labor and aggregate productivity, with the MSB, given by the product of production labor and the marginal increase in aggregate productivity, taking into account the congestion effect of additional search (when $\lambda_W < 1$).

In contrast, wholesalers' first order condition in s in the decentralized equilibrium can be written as:

$$N_W f_W \left(\frac{s N_I}{N_W}\right)^{\beta_W - 1} \frac{N_I}{N_W} \frac{1}{\overline{P^H}(1)} = \frac{L_P}{\overline{P^H}(1)} (1 - \Omega(\mu^{W-\sigma})) \frac{1}{s} \frac{1}{\sigma - 1} \frac{1}{N_W \sigma - 1} \frac{\sigma}{\sigma - 1} \tag{25}$$

$\Omega(\mu^{W-\sigma}) > \Omega(1)$ captures how wholesaler's markup raises the relative price of indirectly traded intermediate goods, reducing the marginal increase in resale profit from additional matches with suppliers for wholesalers. In contrast, uninternalized congestion in search inflates MPB. Lastly, there is a misalignment between private and social incentives to form indirect matches, because under Cournot, and factoring in the additional layer of monopolistic markup charged by intermediate goods producers, wholesalers capture a share $\frac{1}{N_W(\sigma-1)} \frac{\sigma}{\sigma-1}$ of the social production cost, which can be either higher or lower than the share $\frac{1}{\sigma-1}$ required for efficiency. Overall, whether s is too high or too low is ambiguous.

Wedge in wholesaler entry. Lastly, the social and decentralized optimality conditions for N_W are:

$$\frac{F_W}{\overline{P^H}(1)} \leq (\beta_W - 1) f_W \frac{(s N_I / N_W)^{\beta_W}}{\beta_W} \frac{1}{\overline{P^H}(1)} - \frac{L_P}{\overline{P^H}(1)} (1 - \Omega(1)) (1 - \lambda_W) \frac{1}{N_W} \frac{1}{\sigma - 1} \tag{26}$$

$$\frac{F_W}{\overline{P^H}(1)} \leq (\beta_W - 1) f_W \frac{(s N_I / N_W)^{\beta_W}}{\beta_W} \frac{1}{\overline{P^H}(1)} \tag{27}$$

So the only inefficiency in wholesaler entry arises from their failure to internalize congestion externality.

To summarize, two main forces drive inefficiency: (1) wholesaler double marginalization, which misaligns the relative price of direct vs indirect inputs with relative marginal costs and distorts intensive and extensive margins as well as entry, and (2) search and matching congestion externalities that agents fail to internalize, leading to excessive search.

4.2 Aggregate Productivity and Welfare

While the social planner analysis identifies the markup wedge from wholesale market power, it does not alone show how this distortion maps into losses in aggregate productivity and welfare. To make this link explicit, I compare the decentralized equilibrium to the first-best along both dimensions[22].

Proposition 1 (Aggregate productivity under exogenous network)**.** *The first-best aggregate productivity $A_{efficient}$ and its decentralized counterpart A, defined as the ratio of total production of the consumption aggregate to labor employed net of fixed costs, are*

$$A_{efficient} = (A_D^{\sigma-1} + A_I^{\sigma-1})^{\frac{1}{\sigma-1}}, \qquad A = \left[\left(\frac{\mu_D}{\mu}\right)^{-\sigma} A_D^{\sigma-1} + \left(\frac{\mu_I}{\mu}\right)^{-\sigma} A_I^{\sigma-1}\right]^{\frac{1}{\sigma-1}}$$

where A_D and A_I (μ_D and μ_I) are aggregate productivity (markup) for directly and indirectly traded inputs

$$\begin{aligned}
A_D &\equiv \left\{\phi_c \frac{\widetilde{M}}{MV}\left[N_F \int_Z z^{\sigma-1} m(z) j(z) dz\right]\left[N_I \int_Z z^{\sigma-1} v(z) j(z) dz\right]\right\}^{\frac{1}{\sigma-1}} \\
A_I &\equiv \left(S\left\{N_F N_I \left[\mathbb{E}(z^{\sigma-1})\right]^2 - \frac{\widetilde{M}}{MV}\left[N_F \int_Z z^{\sigma-1} m(z) j(z) dz\right]\left[N_I \int_Z z^{\sigma-1} v(z) j(z) dz\right]\right\}\right)^{\frac{1}{\sigma-1}} \\
\mu_D &\equiv \left(\frac{\sigma}{\sigma-1}\right)^2, \qquad \mu_I \equiv \left(\frac{\sigma}{\sigma-1}\right)^2 \mu^W
\end{aligned}$$

and the aggregate markup μ equals the aggregate price index P^H divided by aggregate marginal cost w/A

$$\mu \equiv \frac{P^H}{w/A} = P^H A = \left[\frac{1}{\mu_D}\Omega + \frac{1}{\mu_I}(1-\Omega)\right]^{-1},$$

where $\Omega \equiv \frac{A_D^{\sigma-1}}{A_D^{\sigma-1} + (\mu^W)^{1-\sigma} A_I^{\sigma-1}}$ is direct trade share inclusive of wholesale markup.

Proposition 1 shows that, holding matches and entry at their planner levels, decentralized productivity falls short of the first-best due to dispersion in markups between direct and indirect inputs. As in Edmond, Midrigan and Xu (2015), the aggregate markup μ is a revenue-weighted harmonic mean across these inputs. This dispersion reflects the wholesale-markup-induced relative price distortion, which reduces allocative efficiency at the intensive margin and vanishes only when μ^W approaches 1.

Corollary 1.1 in Appendix B.5 establishes that, to first order, the productivity gap rises with markup dispersion, which is measured as the variance of log markup $\Omega^*(1-\Omega^*)\left(\ln \mu^W\right)^2$ weighted by the efficient direct trade share Ω^*. In particular, conditional on the same efficient direct trade share, the productivity gap rises with wholesale markup. We can observe these patterns in Figure 4. Notice that for the same level of wholesale markup, as the efficient direct trade share rises (by raising κ), the productivity gap first widens, peaks at an intermediate share, and then declines. Intuitively, a relative price distortion that induces underuse of indirect inputs causes little productivity loss when indirect inputs

[22]The proof of Proposition 1 is provided in Appendix B.4.

Figure 4: Aggregate Productivity Gap (Exogenous Matching)

Note: The figure shows how the gap in aggregate productivity between the decentralized equilibrium and the first-best varies with the efficient direct trade share under exogenous matching

are not productive—i.e., when the efficient direct trade share is high. Conversely, when the efficient direct trade share is low, direct inputs are not productive, and therefore the underuse of indirect inputs induced by the relative price distortion is limited, and the productivity loss remains small.[23] Lastly, the higher the wholesale markup, the larger the relative price distortion, and therefore the greater the markup dispersion as well as aggregate productivity loss conditional on the same Ω^*.

Wholesale markups also distort the extensive margin. The next result combines intensive and extensive margins to show the welfare loss.[24]

Proposition 2 (Aggregate welfare with endogenous direct matching)**.** *In a simplified model with a single productivity level $z = 1$, $\beta = 2$, $\lambda_M = \lambda_V = 1$, and a tax on supplier search $\frac{\sigma-1}{\sigma}$, let $\widetilde{L}$ be labor net of fixed entry and wholesalers' search costs.*[25] *Define the misallocation wedge*

$$\Delta_{misallocation} = \frac{S}{\phi_c - S} - \frac{(\mu^W)^{-\sigma} S}{\phi_c - (\mu^W)^{1-\sigma} S} > 0$$

[23] *Intuition for the peak below one-half.* The productivity gap is driven by a *mismatch* between the efficient (Ω^*) and decentralized (Ω) direct trade shares. Wholesale double marginalization tilts spending toward direct inputs, so $\Omega > \Omega^*$. The mismatch is zero at the extremes and largest when the planner wants relatively more indirect trade while the market is pulled toward direct trade—i.e., for $\Omega^* < 1/2$. With endogenous matching, the wholesale wedge also induces excessive ad posting and too many direct matches, pushing Ω further above Ω^* where the planner prefers fewer direct trades. The peak of the welfare-relevant dispersion then tends to occur at an even lower Ω^* than under exogenous matches.

[24] The proof of Proposition 2 is provided in Appendix B.6.

[25] The tax removes firms' double marginalization as in Proposition 4.

Aggregate welfare $\mathcal{W} = L_p \times A$ can be written as efficient vs decentralized components:

$$L_{p,\text{efficient}} = \frac{\sigma-1}{\sigma}\widetilde{L} + \lambda_p \frac{S}{\phi_c - S}, \qquad L_{p,\text{decentralised}} \approx L_{p,\text{efficient}} - \lambda_p \Delta_{misallocation},$$

$$A_{\text{efficient}} = (A_D^{\sigma-1} + A_I^{\sigma-1})^{\frac{1}{\sigma-1}},$$

$$A_{\text{decentralised}} \approx \left[\left(\tfrac{\mu_D}{\mu}\right)^{-\sigma}\left(A_D^{\sigma-1} + \lambda_D \Delta_{misallocation}\right) + \left(\tfrac{\mu_I}{\mu}\right)^{-\sigma}\left(A_I^{\sigma-1} - \lambda_I \Delta_{misallocation}\, S\right)\right]^{\frac{1}{\sigma-1}}$$

with positive coefficients

$$\lambda_p = \frac{\sigma-1}{\sigma}\frac{(f_m f_v)^{1/2}(N_F N_I)^{1/2}}{\kappa}, \qquad \lambda_D = \frac{\sigma-1}{\sigma}\phi_c N_F N_I, \qquad \lambda_I = \frac{\sigma-1}{\sigma}N_F N_I$$

and productivity terms

$$A_D = \left(\phi_c \widetilde{M}\right)^{\frac{1}{\sigma-1}}, \qquad A_I = \left[S(N_F N_I - \widetilde{M})\right]^{\frac{1}{\sigma-1}}, \qquad \widetilde{M} = \kappa\left(\frac{N_F N_I}{f_v f_m}\right)^{\frac{1}{2}}\frac{1}{\sigma}\widetilde{L} - \frac{\sigma-1}{\sigma}N_F N_I \frac{S}{\phi_c - S}$$

When we abstract from the allocation of labor to firm entry and wholesalers' search for suppliers, the remaining labor is used either for forming direct matches or for carrying out production. The optimal allocation of labor between direct match formation and production requires the social planner to balance the marginal benefit of an additional match—captured by the increase in aggregate productivity due to the love of variety—against the marginal cost of posting ads. Here, welfare is given by the labor allocated to production multiplied by aggregate productivity. I analyze the efficient level of welfare by examining these two components separately.

When there is no wholesale intermediation ($S = 0$), the efficient amount of labor allocated to production simplifies to $\frac{\sigma-1}{\sigma}\widetilde{L}$, which is reminiscent of the efficient allocation of labor to production after accounting for firm entry costs in standard monopolistic competition models. When wholesale intermediation is present ($S > 0$), the efficient labor allocated to production increases. This is because additional direct matches can now cannibalize indirect trade, reducing the net benefit of each new match below the pure gain from an additional variety. The extent of this reduction depends on two key factors: (i) the number of indirect matches, given by $S\,N_F\,N_I$, and (ii) the cost of posting ads relative to the productivity gains from the resulting direct matches, captured by $\left(\frac{f_m\,f_v}{N_I\,N_F}\right)^{\frac{1}{2}}\frac{1}{(\phi_c - S)\,\kappa}$.

Second, the first-best aggregate productivity increases with the number of direct matches, which is itself determined by the labor allocated to direct match formation, scaled by the productivity of the matches formed relative to their cost. $A_{\text{efficient}}$ also increases with the number of indirect matches.

In the decentralized equilibrium, the first-best is restored as $\mu^W \to 1$. For $\mu^W > 1$, two forces lower welfare. First, double marginalization creates markup dispersion across direct and indirect inputs, distorting relative prices and pushing too much expenditure toward direct inputs conditional on matches.

Second, because indirect inputs are relatively costlier, firms perceive less cannibalization and post too many ads, generating too many direct matches and too little production labor. Corollary 2.1 in Appendix B.7 shows the welfare gap rises, to first order, with markup dispersion. In particular, conditional on the same efficient direct trade share, the welfare gap rises with wholesale markup.

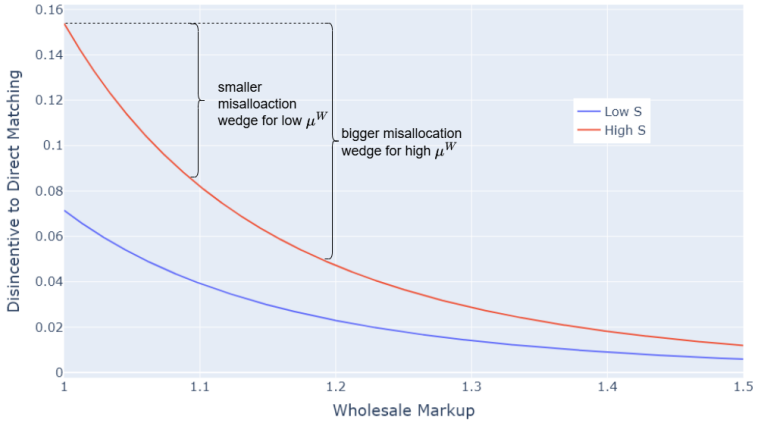

Figure 5: Misallocation Wedge (Endogenous Direct Matching)

Note: The figure plots the disincentive to direct matching as a function of the wholesale markup μ^W for different S. The vertical gap from the $\mu^W = 1$ baseline is the misallocation wedge, which rises with both μ^W and S

Figure 5 plots the disincentive to direct matching, $\frac{(\mu^W)^{-\sigma}S}{\phi_c-(\mu^W)^{1-\sigma}S}$, against μ^W. For any S, the gap from the $\mu^W = 1$ level is the misallocation wedge, which increases in μ^W. The wedge is larger at higher S because wholesale markups understate the productivity of indirectly traded inputs, and this understatement bites more when the share of suppliers that are matched with wholesalers is higher.

4.3 Disintermediation and Rising Wholesale Markup

Wholesale double marginalization distorts allocation and lowers productivity and welfare, and the misallocation rises with the wholesale markup. What determines the markup, and how does it evolve? The next proposition shows that disintermediation—greater direct sourcing—reduces the number of wholesalers and raises wholesale markups[26].

Proposition 3. *Suppose* $N_W \geq \frac{\beta_W-1}{\beta_W\sigma}(1-\Omega)\frac{\sigma-1}{\sigma}$. *The number of wholesalers is a function of the direct trade share (inclusive of wholesale markup)* $\Omega \equiv \frac{X_m}{X_m+X^W-X_m^W}$*:*

$$N_W \leq \left[\frac{(\beta_W-1)(\sigma-1)}{\beta_W\sigma^2}(1-\Omega)\frac{L}{F_W}\right]^{\frac{1}{2}}$$

Holding local labor L_i *and the entry cost shifter* F_{Wi} *fixed,* N_{Wi} *declines as the direct share* Ω_i *increases*

[26]The proof of proposition 3 is provided in Appendix B.3.

Intuitively, an improvement direct trade technology (e.g. ϕ_c increases) encourages firms to post more search ads and form more direct matches, lowers reliance on wholesalers, and increases Ω. Lower wholesale demand reduces post-entry profits in the wholesale sector, so fewer wholesalers enter and N_W falls. Since $\mu^W = \frac{N_W \sigma}{N_W \sigma - 1}$, a lower N_W raises the wholesale markup.

However, a higher wholesale markup does not imply higher misallocation mechanically. What matters is markup *dispersion*. As firms shift toward direct sourcing, the share of trade subject to wholesale markups shrinks; once direct trade is sufficiently high, dispersion can fall and allocative efficiency can improve. Thus, disintermediation and misallocation need not move monotonically.[27]

The net effect depends on initial reliance on wholesale trade, the shock's magnitude, how much wholesale trade declines, and the markup response. These competing forces make the effect of technology-driven disintermediation ambiguous ex ante.

This motivates the empirical analysis. Using variation in fiber rollout across Turkish provinces—a shock that plausibly improves direct trade—I test whether disintermediation occurred, and then whether the model's predictions hold in the data, including changes in wholesale market structure and markups. The results validate the theoretical mechanisms and underpin a quantitative evaluation: I use the estimated effects to infer the shock magnitude, and assess whether the expansion of fiber internet ultimately improved or worsened allocative efficiency, and its overall effect on aggregate welfare.

5 Empirical Analysis

I begin by outlining the empirical context of Turkey's fiber internet expansion, including the institutional background, data sources, and empirical strategy. I then present the main finding.

5.1 Empirical Setting

5.1.1 Background of Fiber Internet Expansion in Turkey

Over the past decade, Turkey has witnessed a rapid deployment of fiber-optic infrastructure, driven in large part by a transformative policy introduced on October 3, 2011, by the Information and Communication Technologies Authority (ICTA). This policy exempted fiber access services from regulatory obligations for five years or until fiber internet subscribers constituted 25% of the fixed broadband base, whichever came first. By reducing regulatory burdens and offering a "regulatory holiday", the government incentivized operators like Türk Telekom to accelerate investments in fiber networks. A critical condition of this policy required Türk Telekom to provide wholesale fiber services to Internet Service

[27] This mirrors Epifani and Gancia (2011): asymmetric liberalization can first raise misallocation by widening markup dispersion across sectors, then lower it as more sectors open. Here, the direct trade share plays the analogous role: at low direct shares, disintermediation can widen dispersion; at high shares, further disintermediation compresses dispersion and improves allocation.

Providers (ISPs) on non-discriminatory terms. The impact of this initiative was significant: Figure D.1 in Appendix D.2 depicts the evolution of the total length of fiber cable deployed in Turkey, which almost doubled between 2012 and 2019 - increasing from 210,286 km in 2012 to 390,816 km in 2019. Meanwhile, the total number fiber internet subscribers increased by five-fold, from 645,092 in 2012 to 3,213,298 in 2019.

I build on Demir, Javorcik and Panigrahi (2023) by exploiting the same temporal variation in fiber internet access across Turkish provinces to study intra-national trade of manufacturing inputs, but I shift the focus from firms' direct sourcing patterns to indirect trade intermediated by wholesalers. Specifically, I estimate the causal effect of fiber expansion on (i) the share of manufacturing trade intermediated by wholesalers, and (ii) wholesaler market structure and market power. These new empirical findings validate the model's mechanism, and inform the inference of shocks used in the quantitative exercise.

5.1.2 Data

Five datasets are used for the empirical analysis: (1) Data on fiber internet infrastructure in Turkey; (2) Turkish firm-level data and firm-to-firm transaction data; (3) Map of oil and natural gas pipeline network in Turkey; (4) Turkish administrative economic data (5) Data on capital rental rate released by the US Bureau of Labor Statistics (BLS). I will describe the details of each data set in this section.

Data on Fiber Internet Infrastructure in Turkey. The Information and Communication Technologies Authority (ICTA) of Turkey releases annual data on the adoption of telecommunications technologies across Turkish provinces since 2007. It also releases annual data on the total length of fiber optic cable deployed in each province. The first year in which the length of fiber optic cable is reported is 2012. Following Demir, Javorcik and Panigrahi (2023), I will make use of the length of fiber optic cable to construct a measure of fiber intensity in each province, which will then be used to derive a measure of fiber internet connectivity across province pairs.

Turkish Firm-Level Data and Firm-to-Firm Transaction Data. The details of these data sets are discussed in section 2.

Map of Oil and Natural Gas Pipeline Network in Turkey. To construct an instrument for fiber connectivity, I digitize the map of oil and natural pipeline network of Turkey's state-owned energy distributor BOTAŞ, as in Demir, Javorcik and Panigrahi (2023), to measure the distance between each province to the closest pipeline. BOTAŞ publishes the map of their pipeline network on their website. To capture the effect of pre-existing pipeline network, prior to the policy change that catalyzed the rollout of

fiber optic cable in 2011, I digitize the map of BOTAŞ oil and natural pipeline network at the beginning of 2011.

Turkish Administrative Economic Data. Data such as population, area, and GDP, at the province and district level, are also used throughout the analysis.

Capital Rental Rate from the US Bureau of Labor Statistics. I borrow the capital rental rate across sectors over the years, published by the US Bureau of Labor Statistics (BLS), to measure the capital rental cost of Turkish firms, as such estimates are not available for Turkey.

5.1.3 Specification

I follow the empirical strategy of Demir, Javorcik and Panigrahi (2023) to measure fiber connectivity between pairs of provinces $I_{ud,t}$ as the minimum fiber intensity between the origin province ($I_{u,t}$) and destination province ($I_{d,t}$):

$$I_{ud,t} = \min\left\{I_{u,t}, I_{d,t}\right\}, \qquad I_{i,t} = \log\left(1 + \frac{L_{it}}{A_i}\right)$$

where L_{it} denotes the length of fiber optic cable (in kilometers) deployed in province i at time t, and A_i is the area (in square kilometers) of province i.

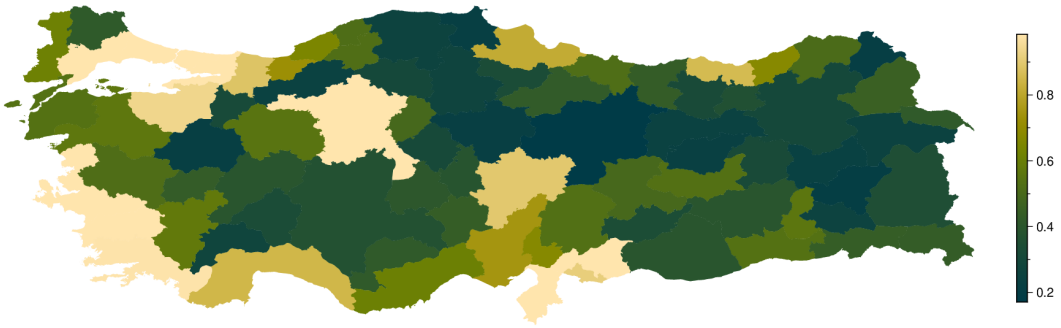

Figure 6: Change in standardized fiber intensity between 2012-2019

Note: The figure shows the change in standardized fiber intensity across provinces between 2012 and 2019. Fiber intensity is standardized by subtracting the mean and dividing by the standard deviation across provinces over the sample period. Light colors indicate provinces with larger increases in deployment.

Figure 6 shows the change in standardized fiber intensity across Turkish provinces from 2012 to 2019. The median province experienced an increase of 0.44 in standardized fiber intensity, with an interquartile range of 0.35.[28]

[28] Standardized by subtracting the mean and dividing by the standard deviation across provinces over the sample period.

The key empirical specification estimates the following equation on bilateral inter-provincial trade:

$$y_{ud,t} = \beta I_{ud,t} + \alpha_{u,t} + \alpha_{d,t} + \alpha_{ud} + \epsilon_{ud,t}$$

where $y_{ud,t}$ represents trade-related outcomes between province pair ud in year t, and $I_{ud,t}$ is their fiber connectivity. The specification includes origin-year fixed effects ($\alpha_{u,t}$), destination-year fixed effects ($\alpha_{d,t}$), and origin-destination pair fixed effects (α_{ud}). The coefficient of interest is β, capturing the impact of fiber connectivity on trade outcomes.

To address potential endogeneity concerns, such as provinces investing in digital infrastructure due to unobserved growth expectations, I adopt an instrumental variable approach. Following Demir, Javorcik and Panigrahi (2023), I exploit the historical placement of oil and gas pipelines by BOTAŞ, Turkey's state-owned energy distributor, to construct an IV for fiber connectivity. Specifically, fiber optic cables were originally laid alongside existing oil and gas pipelines for internal pipeline monitoring purposes, long before their commercial broadband use. A decision by the Turkish government to grant internet providers access to BOTAŞ's fiber optic infrastructure accelerated the rollout of fiber internet. I instrument fiber connectivity $I_{ud,t}$ with the maximum distance of the two provinces in a pair to the nearest oil pipeline interacted with year fixed effects, $Z_{ud} \times \mathbf{1}\{t\}$, where $Z_{ud} = \max\{Z_u, Z_d\}$ and Z_i is the population-weighted average distance of province i's districts to the nearest oil pipeline. The idea is that fiber cable construction is cumulative: province pairs with smaller Z_{ud} benefit from their proximity to the pipeline fiber infrastructure and experience larger gains in connectivity each year, so the (negative) effect of distance on connectivity grows in magnitude over time. I show this in the year-by-year first stage (Figure D.2)—slopes are negative and become more negative over time.

The exclusion restriction requires that pipeline proximity interacted with time dummies affects provincial trade outcomes solely through its effect on fiber connectivity. For it to fail after absorbing origin-year, destination-year, and time-invariant pair heterogeneity, any confounder must satisfy both:

1. **Be pair-specific and time-varying.** Only a *pair-specific, time-varying* component of a confounder can survive the fixed effects ($\alpha_{ud}, \alpha_{u,t}, \alpha_{d,t}$) and thus possibly bias the IV.

2. **Have time-varying correlation with Z_{ud}.** Exclusion can be violated only if the *year-by-year covariance* between the confounder C_{udt} and Z_{ud} *changes over time*. If $\mathrm{Cov}(C_{udt}, Z_{ud})$ is constant across years, the part aligned with Z_{ud} is time-invariant and removed by the pair fixed effects, so the interacted instrument $Z_{ud} \times \mathbf{1}\{t\}$ is orthogonal to the residual in every year.

Several considerations support this assumption. First, the pipeline network predates fiber internet expansion and was originally laid out according to factors such as natural resource endowments, terrain, and engineering feasibility—factors inherently stable and invariant over the relatively short sample

period. Therefore, the inclusion of province-pair fixed effects absorbs any provincial characteristics that influence both pipeline proximity and the level of trade-related outcomes, substantially mitigating concerns that unobserved, time-invariant confounders might violate the exclusion restriction.

Second, the sample period coincides with a policy-driven, unusually rapid *cumulative* fiber rollout along existing oil pipeline. For the exclusion restriction to fail, there would have to exist a *pair-specific, time-varying* mechanism whose *year-by-year covariance with* Z_{ud} also changes over this window—i.e., a bilateral force that intensifies specifically (and increasingly) for pairs whose bottleneck province lies closer to the pipeline. Absent a clear, context-specific mechanism, the emergence of such a confounder is unlikely.

In addition to the bilateral specification, I also run province-level regressions:

$$y_{i,t} = \beta I_{i,t} + \alpha_i + \alpha_t + \epsilon_{i,t}$$

controlling for province fixed effects (α_i) and year fixed effects (α_t), using province distance to pipeline interacted with year dummies as the instrument.

Lastly, firm-level regressions investigate fiber intensity's effect on firm-level outcomes:

$$y_{\omega,t} = \beta I_{i,t} + \alpha_\omega + \alpha_i + \alpha_t + \epsilon_{\omega,t}$$

including firm fixed effects (α_ω), province fixed effects (α_i), and year fixed effects (α_t), again employing province distance to pipelines interacted with year dummies as the instrument.

5.2 Empirical Results

5.2.1 Impact of Fiber Internet Expansion on Disintermediation

Finding 1: Province pairs with faster internet connectivity experience a relative decline in the share of indirect trade

Table 3 presents the results of regressing inter-provincial trade outcomes on standardized fiber connectivity. Here, indirect trade share is measured as the share of total bilateral manufacturing trade in which the downstream buyer is a wholesaler.[29] To assess the magnitude, note that the median province pair experienced a 0.99 standard deviation increase in fiber connectivity over the sample period, with

[29]This choice is motivated by Fact 3: firms source almost exclusively from *local* wholesalers. Because shipment-level tracking is unavailable, two natural proxies differ in what they miss: (i) upstream manufacturers' sales to wholesalers in the downstream province (used here), and (ii) downstream manufacturers' purchases from wholesalers in the upstream province. With strong locality on the buyer side, proxy (ii) systematically *omits* upstream-origin goods that are first shipped to downstream wholesalers and then resold locally; proxy (i) therefore has smaller bias for bilateral indirect trade flows. To limit potential *overstatement* under proxy (i) when wholesalers sell to retailers/final consumers, I exclude wholesalers trading consumer-goods (NACE 4641–4649) and wholesale on a fee/contract basis (NACE 4611–4619), which may also include intermediaries operating through online platforms.

an interquartile range of 0.64. This implies that a province pair at the 75th percentile would see a 33 percentage point lower share of indirect trade relative to one at the 25th percentile—a substantial economic effect.[30]

	Indirect Trade Share	Log Direct Trade Flow	Log Indirect Trade Flow	Log Direct Extensive	Log Direct Intensive
Panel A: OLS					
Std Fiber Connectivity	-0.015	0.379***	0.300***	0.303***	0.076*
	(0.012)	(0.070)	(0.064)	(0.045)	(0.041)
Panel B: 2SLS					
Std Fiber Connectivity	-0.517***	3.099***	0.232	1.654***	1.444**
	(0.198)	(1.083)	(0.645)	(0.569)	(0.667)
Origin Province-Year FE	✓	✓	✓	✓	✓
Destination Province-Year FE	✓	✓	✓	✓	✓
Origin-Destination FE	✓	✓	✓	✓	✓
Observations	39,995	35,107	34,620	35,107	35,107

Table 3: Impact of Fiber Internet Expansion on Inter-Provincial Trade

Note: This table reports OLS (Panel A) and 2SLS (Panel B) estimates of the relationship between fiber connectivity and inter-provincial trade flows. The dependent variables are shown in the column headers. The fiber connectivity measure is standardized by subtracting its mean and dividing by its standard deviation over the sample period. The 2SLS regressions instrument fiber connectivity with the maximum distance of the two provinces in a pair to the nearest oil pipeline, interacted with year dummies. All regressions include origin-year, destination-year, and origin-destination fixed effects. * 10%, ** 5%, *** 1% significance levels. Standard errors clustered at the province-pair level are reported in parentheses.

Finding 2: Disintermediation is driven by relative increases in the extensive and intensive margins of direct trade flow

Columns 2 and 3 of Table 3 show that the relative decline in the indirect trade share for province pairs experiencing faster growth in fiber connectivity is primarily driven by a relative increase in the size of their direct trade flow. In contrast, the size of their indirect trade flow does not exhibit a statistically significant relative change. Columns 4 and 5 decompose direct trade flow into its extensive margin (the number of direct buyer-supplier matches) and intensive margin (the average trade flow per match), revealing that both margins play a significant role in driving the relative increase in direct trade flow—each contributing approximately half of the total effect.

5.2.2 Impact of Fiber Internet Expansion on Wholesale Trade Concentration and Markups

With evidence in hand that fiber internet expansion facilitates disintermediation, I now turn to testing two key predictions of the model: that disintermediation leads to an increase in wholesale trade concentration and in wholesale markups. The empirical results confirm both predictions.

Finding 3: Provinces with faster growth in fiber internet intensity experience a relative decline in the number of wholesalers, with the surviving wholesalers gaining market share

[30] I report a series of robustness checks in Appendix D.3, and find that the pattern of disintermediation holds at both the firm level and the province level.

	Concentration			Markup		
	Log Wholesaler Number (Level: Province)	**Wholesaler Market Share** (Level: Firm)	**Wholesaler Market Share** (Level: Firm)	**Agg Wholesale Markup** (Level: Province)	**Wholesaler Markup** (Level: Firm)	**Wholesaler Markup** (Level: Firm)
Panel A: OLS						
Std Fiber Intensity	-0.0480**	0.0004*		0.0166	0.0189***	
	(0.0188)	(0.0002)		(0.0486)	(0.0054)	
Std Fiber Intensity (Firm-specific)			0.0001			-0.0001
			(0.0001)			(0.0041)
Panel B: 2SLS						
Std Fiber Intensity	-0.3375*	0.0010*		0.6137*	0.0227***	
	(0.2076)	(0.0006)		(0.3549)	(0.0089)	
Std Fiber Intensity (Firm-specific)			0.0077*			0.1243**
			(0.0054)			(0.0528)
Province FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Firm FE		✓	✓		✓	✓
Observations	648	139,063	121,289	648	139,063	121,289

Table 4: Impact of Fiber Internet Expansion on Wholesale Concentration and Markups (OLS and 2SLS)

Notes: Columns 1–2 and 4–5 use province-level standardized fiber intensity. Columns 3 and 6 use the firm-specific fiber intensity measure constructed as a sales-share-weighted average of standardized fiber intensity across provinces. 2SLS instruments fiber intensity with distance to the nearest oil pipeline interacted with year dummies. All regressions include province and year fixed effects; firm fixed effects are added in firm-level columns. * 10%, ** 5%, *** 1% significance levels. Standard errors (in parentheses) are clustered at the province level.

First, I examine the impact of fiber internet expansion on wholesale trade concentration. Column 1 of Table 4 shows that provinces with faster fiber growth experienced a statistically significant decline in the number of wholesalers, consistent with the model's prediction that disintermediation induces wholesaler exits. Column 2 shows that this decline in wholesaler count is accompanied by an increase in the average market share of the surviving wholesalers, indicating an increase in concentration.[31] Column 3 confirms that the result holds qualitatively, despite a lack of statistical significance, when using a firm-specific fiber intensity measure constructed as a sales-share-weighted average of standardized fiber intensity across provinces, suggesting the effect is robust across different measures of internet exposure.

Finding 4: Provinces with faster growth in fiber internet intensity experience a relative increase in aggregate wholesale markups

Table 4 also presents the relationship between fiber internet expansion and wholesale markups. Column 4 shows that provinces with faster growth in fiber intensity experienced a statistically significant increase in aggregate wholesale markups, measured as the cost-weighted average of firm-level markups. This supports the model's prediction that disintermediation leads to greater market power among surviving wholesalers. Columns 5 and 6 confirm that this pattern also holds at the firm level, using either province-level fiber intensity or a firm-specific measure as the regressor. The magnitudes

[31] Market share is defined as the share of sales accounted for by each wholesaler relative to the total sales of all wholesalers in the province.

are economically meaningful and consistent across specifications.[32]

Together, these results confirm that fiber internet expansion led to disintermediation in Turkish production networks, and that this disintermediation reshaped wholesale market structure—reducing the number of wholesalers, raising concentration, and increasing markups. These findings align closely with the model's predictions and provide empirical support for the underlying mechanisms. Importantly, the observed changes in direct trade shares and wholesale markups across provinces provide a basis for calibrating the magnitude of the shock to direct trade technology in the quantitative model.

6 Quantitative Exercise

This section presents two quantitative exercises with the calibrated model: (i) a decomposition of the welfare costs of inefficiencies, and (ii) an evaluation of the welfare impact of fiber internet expansion. I begin with the first exercise to establish the methodology for measuring the welfare cost of misallocation induced by wholesale markups; I then apply this metric to quantify how that cost changed with fiber expansion in the second exercise. Before presenting the results, I first detail the model's calibration.

6.1 Calibration

This section presents the calibration of the single-location model described in Section 3 using indirect inference. While all parameters jointly matter for matching every targeted moment, certain parameters are more tightly linked to specific moments; I therefore organize the discussion around these natural pairings. I set the wage $w = 1$ as the numeraire. I then choose the unit of labor so that $L = 1$. For entry, the masses of final and intermediate producers (N_F, N_I) are equilibrium outcomes under free entry; I fix the units in which varieties are counted by normalizing the fixed entry costs to $F_F = F_I = 1$. In the convex search costs, the level shifters f_m, f_v are not separately identified from matching efficiency κ; I set $f_m = f_v = 1$ and calibrate κ. I assume firm productivity z follows a Pareto distribution with scale parameter 1 and shape parameter α. Table 5 reports the calibrated parameter values, the moment each is most closely associated with, and the corresponding model fit relative to the data. The targeted moments are based on 2012 data.

Convexity of matching effort (β, β_W) Using the first order condition in $m(z)$ (15), I can show that a final goods producer's number of direct suppliers is proportional to its total direct purchases raised to the

[32] Table D.7 in Appendix D.3 reports regressions of firm-level manufacturing markups on fiber intensity and its interaction with the indirect sales share. The results indicate that firms with higher indirect sales shares do not experience a disproportionately larger decline in markups. This eases the concern that rising wholesale market power may have been accompanied by an increase in wholesale markdowns, which would offset the rise in markups and leave the total markup on indirectly traded inputs unchanged. As such, the empirical evidence supports the view that the modeling assumption of identical and constant markups across sales channels is unlikely to materially distort the welfare predictions.

Parameter	Value	Moment	Data	Model
β	2.15	Elasticity of supplier # w.r.t. direct purchases	0.465	0.465
β_W	3.55	Elasticity of supplier # w.r.t. sales	0.282	0.282
σ	4.35	Trade elasticity	-4.88	-4.90
λ_V	0.80	Extensive margin direct trade elasticity	-2.51	-2.53
λ_M	0.80	Extensive margin direct trade elasticity	-2.51	-2.53
λ_W	0.62	Extensive margin indirect trade elasticity	-1.28	-1.24
κ	0.0293	Aggregate direct trade share	0.47	0.48
f_W	160,000	Share of suppliers selling to wholesalers	0.26	0.26
F_W	0.014	Aggregate wholesale markup	1.10	1.13
α	6.27	Direct match dispersion	1.98	1.58
ϕ_c	1.20	Intensive margin distance elasticity difference	0.06	0.06

Table 5: Parameter Calibration

Note: This table summarizes the calibrated values of model parameters and their targeted moment fit.

power $1/\beta$:

$$m(z)\,\theta^m = \theta^m \left[\frac{1}{w\,f_m}\,\frac{1}{\sigma-1}\left(\phi_c - S\,\mu^{W\,1-\sigma}\right) \underbrace{\frac{\sigma-1}{\sigma}\,x_H(z)}_{\text{total purchases}}\ \underbrace{\frac{m(z)\,c_m^{1-\sigma}}{c(z,m(z))^{1-\sigma}}}_{\text{direct purchase share}} \right]^{\frac{1}{\beta}}$$

Accordingly, I calibrate β by regressing the log number of direct suppliers of each firm on its log total direct purchases, controlling for province and year fixed effects. Similarly, a wholesaler's number of suppliers is proportional to its total sales raised to the power $1/\beta_W$ (by (17)), so I calibrate β_W by regressing the log number of suppliers of each wholesaler on its log total sales, again controlling for province and year fixed effects. The calibrated $\beta = 2.15$ is lower than $\beta_W = 3.55$, reflecting that the elasticity of the number of direct suppliers with respect to the direct purchases of manufacturing firms is higher than the elasticity of the number of suppliers with respect to wholesalers' total sales.

Trade and matching elasticities (σ, λ_V, λ_M, λ_W) The elasticity of substitution σ and the matching-function elasticities λ_V, λ_M, and λ_W are calibrated to match trade elasticities estimated from Turkish international trade flows. I follow Fontagné, Guimbard and Orefice (2022) and use variation in Turkey's import tariffs over the sample period. Specifically, I estimate

$$\log y_{opt} = \eta \log \tau_{opt} + \gamma_{ot} + \delta_{pt} + \xi_{op} + \epsilon_{opt},$$

where y_{opt} is the outcome for exporting country o, HS6 product p, in year t, and τ_{opt} is the corresponding Turkish import tariff. I consider two outcome variables: (i) the aggregate import value of product p from country o in year t, and (ii) the number of unique Turkish importers that import a positive amount of product p from country o in year t. The former yields an estimate of the overall trade elasticity; the latter

is intended to capture the extensive-margin trade elasticity.[33] Tariff data come from MAcMAP-HS6 (CEPII); I use the observations within my sample period (2013, 2016, and 2019).

To mitigate endogeneity concerns, I include exporter-year, HS6-year, and exporter-HS6 fixed effects. Identification thus relies on within-pair tariff variation over time. Product-specific time trends and exporter-specific economic shocks that could affect supply are controlled for by the HS6-year and exporter-year fixed effects, respectively. Time-invariant characteristics of each exporter-HS6 pair are likewise absorbed by the pair fixed effects.

	Total Trade	**Direct Trade**	**Indirect Trade**	**Direct Ext. Margin**	**Indirect Ext. Margin**
Tariff	-4.875***	-4.422***	-5.773***	-2.509**	-1.276
	(1.433)	(1.669)	(2.263)	(1.052)	(1.235)
Exporter-Year FE	✓	✓	✓	✓	✓
HS6-Year FE	✓	✓	✓	✓	✓
Exporter-HS6 FE	✓	✓	✓	✓	✓
Observations	147,721	119,105	84,237	119,105	84,237

Table 6: Tariff Regressions

Note: Each observation is weighted by the value of the dependent variable. * 10%, ** 5%, *** 1% significance levels. Standard errors clustered at the HS6 product level are in parentheses.

Table 6 reports the regression results. Column (1) regresses total Turkish imports of an HS6 product from a given exporter in a given year on the associated tariff. Columns (2) and (3) split total trade into direct trade (imports by Turkish manufacturing firms) and indirect trade (imports by Turkish wholesalers). Columns (4) and (5) use the extensive margins of direct and indirect trade, respectively, as dependent variables. The results show that the ratio of extensive-margin to intensive-margin trade elasticities is roughly 1:1 for direct trade, but only about 1:3 for indirect trade. Intuitively, conditional on β and β_W, the parameters λ_V, λ_M, and λ_W govern the extensive-to-intensive margin ratio for direct and indirect trade, whereas σ pins down the level of the elasticities. The calibrated $\sigma = 4.35$ is in line with the consensus in the literature. The calibrated $\lambda_V = 0.80$ (with λ_M assumed equal) implies increasing returns to scale in direct matching as well as congestion externalities. The lower extensive-to-intensive margin ratio for indirect trade implies $\lambda_W = 0.62$, smaller than λ_V and λ_M even though β_W is already calibrated to be larger than β.

Matching efficiency/cost (κ, f_W) The direct matching efficiency κ is calibrated to match the aggregate direct trade share, while the level of wholesalers' search cost f_W is calibrated to match the observed

[33] Ideally, the extensive margin would be measured using the total number of importer–exporter matches. However, the identity of the foreign exporting firm is not reported in the Turkish customs data. My measure coincides with the number of importer–exporter matches if each Turkish importer sources a given HS6 product from only one exporter in a specific country-year, which is not implausible given the high level of disaggregation.

share of manufacturing firms that are selling to wholesalers. Recall that in the model, S captures the share of upstream varieties traded by wholesalers. I therefore measure the corresponding moment in the data analogously, as the share of upstream manufacturing firms that sell to wholesalers. I compute this share for every province pair, and obtain the aggregate share as the average province-pair share weighted by its trade share.

Wholesaler entry cost (F_W) The wholesaler entry cost F_W is calibrated to match the aggregate wholesale markup in Turkey. Specifically, I follow the procedure outlined in Section 2 to compute firm-level markups using the production approach of De Loecker and Warzynski (2012). I then compute the aggregate wholesale markup as the cost-weighted average of firm-level wholesale markups. The aggregate wholesale markup for 2012 is 1.76, which is much higher than the monopolistic markup admissible in this model given the calibrated value of σ, namely $\sigma/(\sigma-1) = 1.30$ for $\sigma = 4.35$.

Instead of matching the level of aggregate wholesale markup per se, I calibrate F_W to match the *net* aggregate wholesale markup (0.76)[34] *relative* to the highest province-level net aggregate wholesale markup observed in the sample (2.288). This yields a targeted aggregate wholesale markup of $1 + (0.76/2.288) \times 0.3 = 1.10$, where 0.3 is the net markup implied by the model's monopolistic markup $(1.30 - 1 = 0.30)$. This targeted aggregate wholesale markup lies between the wholesale markup levels prevailing when $N_W = 3$ (1.08) and $N_W = 2$ (1.13). I interpret this as indicating that the profitability of the wholesale sector is not sufficient to accommodate the entry of three wholesalers within the relevant local competitive boundary. I therefore calibrate F_W so that $N_W = 2$, which yields a calibrated aggregate wholesale markup of 1.13.[35]

Firm productivity distribution shape parameter (α) The shape parameter α of the Pareto distribution from which firm productivity is drawn determines the dispersion of firm size and therefore the dispersion of the number of direct matches each firm has in the model. The lower α is, the heavier the Pareto tail and the greater the dispersion. I compute the dispersion of the number of direct matches as its coefficient of variation, which is measured to be 1.98 in the data. I calibrate α to be 6.27, which yields a model counterpart of 1.58. I could not lower α further to more closely match the observed dispersion, as the

[34] *Net markup* is defined as the markup in excess of one, i.e. $\mu - 1$. Hence, a markup of 1.76 corresponds to a net markup of 0.76.

[35] While the data contain thousands of wholesalers, they do not all compete head-to-head. Manufacturing firms' sourcing from wholesalers is highly localized (Fact 3), with a median local share of 99%. Moreover, wholesalers operate in specialized sub-industries and are therefore not necessarily close substitutes for one another. Any concentration measure based on administratively defined or arbitrary market boundaries (e.g., "all wholesalers in Turkey" or even "all wholesalers in a province") is thus tenuous. Instead, the markup measure reveals the degree of *effective* competition within each economically meaningful market cell—defined by proximity, product scope, and the fixed costs of direct sourcing—what I refer to as the *relevant local competitive boundary*. Fact 4 reinforces this segmentation: wholesale sales are extremely concentrated at both national and province-industry levels (Table 2). Calibrating $N_W = 2$ therefore does *not* claim there are only two wholesalers in Turkey; rather, it captures that within the relevant local/product market boundary, competition is effectively duopolistic.

existence of the expectation $\mathbb{E}\left[z^{\frac{\beta(\sigma-1)}{\beta-1}}\right]$ requires $\alpha > \frac{\beta(\sigma-1)}{\beta-1} = 6.26$. This expectation is needed to pin down the aggregate direct trade flow in the model.

Customization productivity gains (ϕ_c) The difference in the intensive margin between direct and indirect trade flows helps identify ϕ_{cud}. From the gravity equation in Appendix B.10 derived for the spatial extension of the model, the multilateral resistance term in the intensive margin of direct trade flow is $\tau_{ud}^{1-\sigma}\,\phi_{cud}$, whereas for potential indirect trade flow it is $\tau_{ud}^{1-\sigma}$. The difference between the two thus identifies ϕ_{cud}. Specifically, the difference in the intensive margin distance elasticity times log distance yields $\log \phi_{cud}$.

In practice, only the actual indirect trade flow, $X_{Wud} - X_{Wmud}$, is observed, not X_{Wud}. Approximating the former with the latter overestimates ϕ_{cud}, as an increase in distance lowers X_{Wmud}, dampening the observed decline in indirect trade. An alternative approximation,

$$X_{Wud} - X_{Wmud} + \left(\frac{\sigma}{\sigma-1}\right)^{-\sigma} S_{ud} X_{mud} \;>\; X_{Wud} \quad (\text{since } \phi_{cud} > 1),$$

provides a lower bound for ϕ_{cud}. I estimate ϕ_{cud} using both approaches and take the average of the two.[36] I obtain an aggregate measure of ϕ_c by computing the average of these province-pair ϕ_{cud}, weighted by their trade shares. The calibrated $\phi_c = 1.20$ implies that directly sourced intermediate goods are 20% more productive than the same variety sourced indirectly in the model.

6.2 Decomposing the Welfare Cost of Inefficiencies

With the calibrated model in hand, we are now ready to conduct our first quantitative exercise. The goal is to decompose the welfare cost of two key sources of inefficiency—(1) resource misallocation induced by wholesale markups; and (2) congestion externalities in search and matching—measured as the welfare gain from eliminating them. Proposition 4 in Appendix B.8 shows that a wholesale subsidy equal to the inverse of the wholesale markup, together with a set of taxes/subsidies on search and matching, as well as firm entry, is required to restore the first-best allocation from the decentralized equilibrium. The former corrects resource misallocation induced by wholesale markups, while the latter policies address congestion externalities and double marginalization by manufacturing firms. I present the welfare increase from implementing these policies one by one.

Table 7 reports the results for the baseline parameterization.[37] I report the *level* of selected variables—for three cases: (i) the decentralized equilibrium, (ii) the equilibrium with only the wholesale subsidy, and

[36] The intensive margin of potential indirect trade distance elasticity using the upper-bound approach is 0.175, while that using the lower-bound approach is 0.167; the two are fairly close.

[37] I also report the results for different alternative parameterizations in Appendix C.1 to highlight the key mechanisms generating inefficiency and to illustrate the sensitivity of the welfare decomposition to perturbations of parameter values.

(iii) the efficient allocation with the full set of optimal policies. Note that the welfare in the decentralized equilibrium is normalized to be 1.

Table 7: Welfare Decomposition by Policy (Baseline)

	Ω	μ^W	$\sigma^2(\mu)$	$\tilde{M}$	S	A	L_A	L_S	L_E	L_P	Welfare
Baseline											
Decentralized	0.4800	1.1299	0.0037	4.6653	0.2630	0.3852	0.0799	0.0129	0.3465	0.5608	1.0000
Wholesale subsidy	0.1800	1.0000	0.0000	2.2730	0.2539	0.3914	0.0261	0.0194	0.3895	0.5650	1.0237
Efficient	0.2200	1.0000	0.0000	0.0435	0.3740	0.3556	0.0230	0.0270	0.2818	0.6682	1.1000

Notes: Column 4 reports $\tilde{M}$ multiplied by 10,000. The last column reports welfare, normalized to 1 in the decentralized equilibrium for each parameter set.

Implementing wholesale subsidy. In the decentralized equilibrium, the wholesale markup equals 1.13, which generates a dispersion in markups between directly traded inputs, $(\sigma/(\sigma-1))^2$, and indirectly traded inputs, $(\sigma/(\sigma-1))^2 \mu^W$. Column 3 reports this dispersion to be 0.0037, measured as the variance of log markups.[38] Recall Propositions 1 and 2: such dispersion distorts the relative prices of directly and indirectly traded inputs, leading to excessive use of the former on the intensive margin, as well as excessive creation of direct matches. This is evident in the table: when a wholesale subsidy is implemented to eliminate the markup dispersion induced by wholesale markups, the number of direct matches (Column 4) falls by more than 50%, while labor allocated to ad posting (Column 7) falls by 67%.

Eliminating markup dispersion also raises aggregate productivity (Column 6), despite the reduction in direct match formation. This happens partly because inputs are used more efficiently on the intensive margin, and partly because the number of indirect matches increases. Although the share of suppliers matched with wholesalers (Column 5) decreases slightly, the increase in the number of firms more than offsets this decline, generating a rise in the number of indirect matches ($S N_F N_I$), as reflected in the increase in labor allocated to firm entry (Column 9). Labor allocated to wholesalers' search (Column 8) also increases, despite the decrease in S, due to the larger number of intermediate goods producers (recall that the wholesalers' search cost rises with both search effort and the number of intermediate goods producers). Finally, some labor is reallocated away from ad posting toward production (Column 10). Overall, the wholesale subsidy yields a welfare gain of 2.4%.

Implementing match and entry taxes/subsidies. If we further implement the matching and entry taxes/subsidies described in Proposition 4, welfare increases by an additional 7.2%.[39] These additional instruments address congestion externalities in search and matching, as well as double marginalization by manufacturers, by reallocating labor away from match and firm formation toward production. The

[38] The variance of log markups is given by $\Omega(1-\Omega)(\ln \mu_D - \ln \mu_I)^2 = \Omega(1-\Omega)(\ln \mu^W)^2$.

[39] I decompose welfare change due to an incremental policy change as $\ln(\widehat{Welfare}_1) - \ln(\widehat{Welfare}_0)$, where $\widehat{Welfare}_0$ ($\widehat{Welfare}_1$) is the proportional welfare change from the decentralized equilibrium without (with) the incremental policy.

effect of these congestion taxes is substantial: in the efficient allocation, the number of direct matches is less than 1% of that in the decentralized equilibrium. This reallocation shifts a sizable amount of labor from forming direct matches to production. Aggregate productivity falls as a result of the reallocation, but much less than the decline in the number of matches. This more muted drop occurs because the congestion taxes are size-dependent and primarily target ad-posting labor involved in forming direct matches between less productive firms. The most productive links are preserved, so aggregate productivity decreases by a much smaller extent.

While these congestion taxes/subsidies promise large welfare gains, they are much harder to implement in practice (they are size-dependent and require knowledge of the efficient allocation). As discussed in the next section, these congestion externalities also tend to remain stable following disintermediation, and therefore do not materially affect its overall welfare impact.

6.3 Evaluating the Welfare Impact of Fiber Internet Expansion

This subsection presents the main quantitative result of the paper: quantifying the overall welfare impact of technology-induced disintermediation. Specifically, I calibrate shocks in the model to replicate the episode of fiber-internet expansion in Turkey discussed in Section 5, using it as a case study of technology-induced disintermediation. I conduct the counterfactual in the spatial extension of the baseline model developed in Appendix A, to capture heterogeneity across provinces—in both the speed of fiber-internet rollout and the importance of wholesale trade—and thus provide a richer welfare evaluation. As shown in Figure 7, the importance of indirect trade varies significantly across regions, with an interquartile range of 32 percentage points across Turkish provinces in 2012.[40]

The model counterfactuals are solved using the exact-hat algebra (Dekle, Eaton and Kortum, 2007), with the full system of hat algebra equations provided in Appendix B.9. In addition to observed trade flows, computing the counterfactuals requires knowledge of S_{ud}, which I measure as the share of firms in location u that sell to wholesalers in location d, mirroring the calibration of the single location model. I use the same set of parameterization for the structural parameters $\{\lambda_V, \lambda_M, \lambda_W, \beta, \beta_W, \sigma\}$ in the spatial extension as the ones calibrated using the single location model.

6.3.1 Shock inference

To discipline the counterfactual, I calibrate a pair of shocks that reproduce the relative changes in (i) the indirect trade share, (ii) direct trade flow, and (iii) indirect trade flow reported in Columns 1–3 of

[40] Figure 7 also shows that provinces in eastern Turkey—far from western hubs such as Istanbul and Ankara—tend to exhibit higher indirect trade share. Table D.2 confirms this by showing that interprovincial direct trade has a higher distance elasticity than indirect trade, and Table D.3 shows that this difference holds for both the extensive margin (number of matches) and the intensive margin (trade flow per match). These findings suggest that the benefits of direct trade decline more sharply with distance, underscoring the role of wholesalers in facilitating economic integration across Turkish provinces.

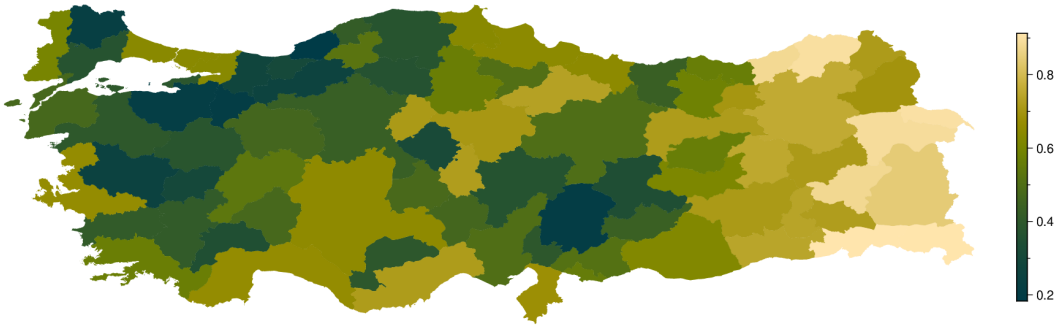

Figure 7: Aggregate indirect trade share across provinces in 2012

Note: This map shows the aggregate indirect trade share across Turkish provinces in 2012. Here, indirect trade refers to total sales of manufacturing goods to wholesalers, direct trade refers to sales to manufacturing firms, and the aggregate indirect trade share of each province is defined as the ratio of indirect trade to the sum of direct and indirect trade, aggregating across upstream provinces.

Table 3. Formally, for each origin–destination province pair (u, d) I postulate proportional shocks to the customization productivity gains of direct trade, ϕ_{cud}, and the level of wholesalers' search cost, f_{Wud}, of the form

$$\hat{\phi}_{cud} = 1 + a_{\phi}(\Delta I_{ud} - s), \qquad \hat{f}_{Wud} = 1 + a_{f_W}(\Delta I_{ud} - s),$$

where $\Delta I_{ud} \equiv I_{ud,2019} - I_{ud,2012}$ is the growth in fiber connectivity between 2012 and 2019. The three scalars (a_{ϕ}, a_{f_W}, s) are chosen so that the model exactly matches the three targeted moments above.

Identification. Higher values of a_{ϕ} and a_{f_W} generate greater differential changes in customization productivity and wholesalers' search cost across province pairs with varying fiber internet connectivity growth, thereby increasing the relative changes in both direct and indirect trade. Meanwhile, s determines the overall level of the shocks, which cannot be identified solely based on the relative changes in direct and indirect trade flows. Instead, I exploit the fact that the direct trade share is bounded above by 1. A higher shock level increases the direct trade share across province pairs, causing more province pairs to reach this upper bound. This, in turn, reduces the variation in changes in the bilateral indirect trade share across province pairs with different levels of fiber connectivity growth. By matching the relative change in the bilateral indirect trade share, I identify the appropriate value of s.

The calibrated values are $a_{\phi} = 1.22$, $a_{f_W} = -1.58$, $s = 0.95$, implying trade-weighted average shocks of $\hat{\phi}_{cud} = 1.37$ and $\hat{f}_{Wud} = 0.53$.

Table 8 compares the model's implied relative changes of various variables with the data. The three targeted moments are matched by construction. Among untargeted moments, the model reproduces the qualitative rise in both the extensive and intensive margins of direct trade, but—owing to the strong extensive-margin elasticity built into the baseline calibration—*over-predicts* the increase in matches and *under-predicts* the rise in trade per match.

Moment (relative change)	Model	Data	Data (95% CI)
Targeted moments			
Indirect trade share	-0.517	-0.517	[-0.905, -0.129]
Direct trade flow	3.076	3.099	[0.976, 5.222]
Indirect trade flow	0.234	0.232	[-1.032, 1.496]
Untargeted moments			
Direct trade (extensive)	2.292	1.654	[0.539, 2.760]
Direct trade (intensive)	0.784	1.444	[0.137, 2.751]
Number of wholesalers	-0.696	-0.337	[-0.744, 0.069]
Number of manufacturing firms	-0.234	-0.724	[-1.130, -0.318]
Wholesale markup	0.453	0.617	[-0.077, 1.311]

Table 8: Simulation moment match

Notes: The first block reports the *targeted* moments used for inferring the shocks; the second block shows additional, untargeted moments.

Why do we shock only customization productivity? The customization productivity shock, $\hat{\phi}_{cud}$, lowers the relative unit cost of direct sourcing.[41] This change operates on the intensive margin directly *and*, through firms' optimal search decisions, on the extensive margin. In the calibrated model a 1 % rise in ϕ_{cud} elicits a large—perhaps too large—search response, so that matching the total change in direct trade flow already pushes the extensive margin up more than the data. Introducing an additional shock to direct matching efficiency (which affects only the extensive margin) would widen this discrepancy: the model would fit the total direct trade flow but overstate the number of matches and understate the average trade flow even further. For that reason, I restrict attention to $(\hat{\phi}_{cud}, \hat{f}_{Wud})$ and leave the matching efficiency parameter unchanged.

Untargeted-moment performance. The simulation also reproduces, at least qualitatively, other empirical patterns: a decline in the number of wholesalers, a contraction—though smaller than observed—in the number of manufacturing firms, and an increase in wholesale markups in provinces with faster fiber rollout.

Interpreting the high s. The calibrated shift $s = 0.95$ is close to the median change in fiber connectivity, implying that roughly half of province pairs experience negative net shocks to the productivity of direct trade. This seemingly counter-intuitive result can reflect relative obsolescence: as digital search becomes the norm, regions that lag in adopting modern platforms may end up worse off than before. Firms that persist with paper directories such as the Yellow Pages face shrinking coverage and outdated information,

[41] We may interpret this shock by building on the chips example mentioned in Section 3. With cloud-based manufacturing systems (real-time logs from test benches and the production line), the factory can share heat readings (how hot parts run), failure codes (what went wrong), and voltage noise (unwanted electrical fluctuation) right away; fiber-optic internet enables this rapid exchange, so the supplier updates parameters or provides better-matched parts in the next lot—cutting test failures, avoiding overheating, and speeding final assembly.

while firms embracing digital marketplaces benefit from better search tools and network effects. Falling behind the technological frontier can thus translate into lower effective direct trade productivity, even amid nationwide infrastructure upgrades.

6.3.2 Welfare Impact of Fiber Internet Expansion

Table 9 reports the percentage changes of aggregate variables relative to the pre-shock decentralized equilibrium in the spatial model—for three cases: (i) the pre-shock equilibrium with wholesale subsidy, (ii) the post-shock equilibrium without wholesale subsidy, and (iii) the post-shock equilibrium with wholesale subsidy. Province-level variables are weighted by pre-shock province nominal GDP, and province-pair-level variables are weighted by pre-shock province-pair trade shares. For comparison, Table 10 reports percentage changes of aggregate variables relative to the pre-shock decentralized equilibrium in the single-location model. In addition to the three cases shown in the spatial model, I also report the pre-shock and post-shock efficient equilibria for the single-location model. Note that these efficient cases are not solvable in the spatial extension, as they require knowledge of province-specific firm productivity distributions and the efficient allocation. The latter is not known because we can only solve the spatial model in proportional changes using hat algebra, without the knowledge of all the bilateral frictions.[42]

Table 9: Welfare Decomposition (Spatial Model): Percentage Changes Relative to the Pre-Shock Decentralized Equilibrium

	Ω	μ^W	$\sigma^2(\mu)$	$\tilde{M}$	S	A	L_A	L_S	L_E	L_P	Welfare
Pre Shock											
Wholesale subsidy	−42.0	−9.4	−97.7	−43.6	−2.3	−4.1	−47.4	30.5	13.2	6.1	1.6
Post Shock											
Decentralized	53.2	8.9	164.9	48.5	15.0	3.8	67.3	−12.5	−14.8	0.9	4.7
Wholesale subsidy	−4.9	−5.4	−69.5	−8.1	17.4	4.3	−7.5	64.0	0.7	3.4	7.8

Note: This table reports percentage changes of aggregate variables relative to the pre-shock decentralized equilibrium in the spatial model. Province-level variables are weighted using pre-shock province nominal GDP, while province-pair-level variables are weighted using pre-shock province-pair trade share.

Table 10: Welfare Decomposition (Single-Location Model): Percentage Changes Relative to the Pre-Shock Decentralized Equilibrium

	Ω	μ^W	$\sigma^2(\mu)$	$\tilde{M}$	S	A	L_A	L_S	L_E	L_P	Welfare
Pre Shock											
Wholesale subsidy	−63.0	−11.5	−100.0	−51.3	−3.5	1.6	−67.3	51.0	12.4	0.7	2.4
Efficient	−54.0	−11.5	−100.0	−99.1	42.2	−7.7	−71.3	110.4	−18.7	19.2	10.0
Post Shock											
Decentralized	63.0	14.9	197.3	29.5	17.0	1.2	77.5	−19.8	−20.4	2.0	3.3
Wholesale subsidy	−34.0	−11.5	−100.0	−27.0	9.1	6.5	−39.8	25.8	6.5	1.1	7.7
Efficient	−12.0	−11.5	−100.0	−98.4	46.1	−2.1	−38.8	53.5	−23.2	18.6	16.1

Note: Percentage changes are relative to the pre-shock decentralized equilibrium in the single-location model.

[42] In Appendix C.2, I rerun the internet counterfactual in the single-location model using different sets of parameters and discuss the sensitivity of the results to parameter perturbation.

The effect of wholesale subsidy. I first discuss the effect of implementing wholesale subsidy in the spatial model to evaluate the welfare cost of misallocation induced by wholesale markup, and draw a comparison against the single location model. Following Baqaee and Farhi (2019) and Arkolakis, Huneeus and Miyauchi (2023), I define the change in aggregate welfare as the nominal GDP weighted change of welfare of each location:

$$\Delta \log \mathcal{W} \equiv \sum_i I_i \left(\Delta \log I_i - \Delta \log P_i^H\right)$$

As the last columns of Tables 9 and 10 show, implementing the optimal wholesaler subsidy results in a smaller welfare gain in the spatial model.

There are two underlying reasons. First, as Column 2 indicates, wholesale subsidy results in a smaller decline in net aggregate wholesale markup in the spatial model, implying that the level of wholesale markup in the decentralized equilibrium of the spatial model is lower. This is due to the discreteness of the number of wholesalers in the single location model, which prevents precise calibration of F_W to match aggregate wholesale markups, as discussed in Section 6.1. This discreteness also explains why wholesale subsidy fails to fully eliminate the dispersion of markups in the spatial model ($\widehat{\sigma^2(\mu)} = 0.0231 > 0$): as the number of locations grows, it becomes more likely that incremental subsidy increases trigger additional wholesaler entry, causing discrete drops in wholesale markup. Consequently, eliminating markup dispersion entirely through subsidy becomes increasingly challenging.

Second, as Figure 7 shows, there is a substantial dispersion of indirect trade share across provinces in Turkey. This heterogeneity in indirect trade share implies a lower aggregate markup dispersion in the spatial model, even if the aggregate wholesale markups are identical. This is because the dispersion of markups would be low when indirect trade share is either very high or very low, which means, according to Corollaries 1.1 and 2.1, wholesale markup in these provinces would also result in a smaller degree of misallocation. In fact, the dispersion of markups in the decentralized equilibrium of the spatial model is only around 0.0032, compared to 0.0037 in the single location model.

To summarize, the discreteness of N_W and the dispersion of indirect trade share both contribute to a lower dispersion of markups in the spatial model, explaining why the welfare cost of misallocation driven by wholesale markup is smaller.

Technological impact of fiber internet expansion in the decentralized equilibrium. Next, I discuss the impact of fiber internet expansion. Relative to the pre-shock decentralized equilibrium in the spatial model, direct trade share increases significantly in the decentralized equilibrium post-shock (Table 9, Column 1). This rise results from increased customization productivity in direct trade and subsequent growth in labor allocated to ad posting (Column 7), thus leading to more direct match formation (Column

4). Collectively, these changes boost aggregate productivity by 3.8% (Column 6). Overall, fiber internet expansion increases welfare by 4.7% in the decentralized equilibrium of the spatial model.

Notably, fiber internet expansion yields greater welfare increases in the decentralized equilibrium of the spatial model compared to the single-location model. This difference arises from dispersion of shocks across provinces and substantial higher-order effects in this model. By relaxing the Cobb-Douglas assumption and allowing the direct trade share to evolve endogenously, incremental increases in the customization productivity in this model would influence an increasing share of trade, specifically, share of trade that happens directly, thereby amplifying welfare gains from positive shocks (Baqaee and Farhi, 2019). This non-linearity can be most clearly observed in Figure D.4 in Appendix D.4, which plots a decomposition of the first-order effects of the incremental internet shock (derived in Proposition 6 in Appendix B.12). Specifically, I split the full internet shock into a geometric product of n incremental shocks, so that, for example, each incremental shock to ϕ_c is equal to $\widehat{\phi_c}^{1/n}$. This figure shows that not only is the overall welfare impact of each incremental shock positive, but it is also increasing as the step increases. This acceleration of welfare impact coupled with the dispersion of fiber internet rollout across provinces imply that those provinces with the fastest internet rollout would experience disproportionately larger welfare gains, and contribute to a larger increase in aggregate welfare.

Impact of fiber internet expansion on allocative efficiency. Disintermediation, nevertheless, has led to an increase in wholesale markup (Column 2) and also an increase in markup dispersion (Column 3), as fewer wholesalers remain profitable following reduced demand for intermediation. To understand how fiber internet expansion has impacted allocative efficiency by increasing wholesale markup, I adopt the following decomposition of welfare changes resulting from the shock into the change arising due to changes in the welfare level without wholesale markup distortion as well as due to the change in allocative efficiency:

$$\Delta \log \mathcal{W}_{\text{decentralized}} = \underbrace{\Delta \log \mathcal{W}_{\text{subsidy}}}_{\substack{\text{change in welfare} \\ \text{with wholesale subsidy}}} + \underbrace{\left(\Delta \log \mathcal{W}_{\text{decentralized}} - \Delta \log \mathcal{W}_{\text{subsidy}}\right)}_{\substack{\text{change in allocative efficiency} \\ \text{due to rising wholesale markup}}}$$

The first term on the right-hand side is the change in welfare without wholesale markup distortion, when wholesale subsidy is implemented to restore the efficient relative price, while the second term gives the change in allocative efficiency due to rising wholesale markup. Using this decomposition, fiber internet expansion worsened allocative efficiency by 1.4%. The welfare gains from fiber internet expansion would therefore have been 30% higher were optimal wholesale subsidy implemented. This significant difference highlights the importance of using *complementary competition policy* to fully realize the welfare potential of infrastructure investment.

In the single location model, we are able to simulate the efficient equilibrium both before and after the shock, which allows us to further decompose the change in allocative efficiency into contributions from congestion externalities and manufacturer double marginalization:

$$\Delta \log \mathcal{W}_{\text{decentralized}} = \underbrace{\Delta \log \mathcal{W}_{\text{efficient}}}_{\substack{\text{change in first-best welfare}}} + \underbrace{\left(\Delta \log \mathcal{W}_{\text{decentralized}} - \Delta \log \mathcal{W}_{\text{subsidy}}\right)}_{\substack{\text{change in allocative efficiency} \\ \text{due to rising wholesale markup}}}$$

$$+ \underbrace{\left(\Delta \log \mathcal{W}_{\text{subsidy}} - \Delta \log \mathcal{W}_{\text{efficient}}\right)}_{\substack{\text{change in allocative efficiency} \\ \text{due to congestion \& manufacturer DM}}}$$

It turns out that allocative efficiency also worsened due to congestion externalities and manufacturer double marginalization, but only by 0.4%. The relative stability of the welfare cost from these additional inefficiencies suggests that they do not materially affect the overall welfare impact of fiber internet expansion in the way that wholesale markup does.

Distributional consequences of fiber internet expansion. As shown in Figure 6, there is a significant dispersion in the speed of fiber internet rollout across Turkish provinces. However, productivity gains shared via trade substantially mute welfare disparities (Figure D.5). Crucially, the presence of wholesale trade, which tends to decline more slowly with distance[43], helps transmit productivity gains concentrated in western Turkey to less-developed eastern provinces. We can visualize this phenomenon, for example, by noticing the much sharper decline in the direct purchases of other provinces from Istanbul as a share of their own GDP (Figure D.11), relative to the much flatter decline in their total purchases from Istanbul relative to GDP (Figure D.12), which includes indirect purchases through wholesalers.

If we compare the distribution of fiber internet rollout across provinces (Figure 6) against their indirect trade share (Figure 7), we would uncover an interesting negative correlation between the two: that fiber internet tended to expand faster in provinces with low pre-shock indirect trade share. This negative correlation has probably limited the extent of disintermediation in the spatial model, relative to the single location model, as reflected in the more muted increase in the aggregate direct trade share following the internet shock (Column 1 of Table 9 and 10). Consequently, fiber internet has caused smaller increases in the aggregate wholesale markup and markup dispersion in the spatial model, and therefore dampening the increase in the degree of misallocation induced by rising wholesale markup (-1.4% in the spatial model, relative to -1.8% in the single location model).

This negative correlation between the speed of fiber internet rollout and pre-shock indirect trade share also implies that disintermediation has not caused much dampening of the extent of gains sharing across provinces as the distant provinces in the east are still relying heavily on indirect trade, which

[43] As reflected by the smaller distance elasticity of indirect trade (Table D.2).

can be seen from the relatively constant share of indirect trade among those provinces, as depicted in Figure D.8. On the flip side, western provinces that are closer to the provinces experiencing more rapid rollout have seen disproportionately greater degree of disintermediation, compared to eastern provinces that have received similar shocks. Lastly, this amplification (dampening) of disintermediation across provinces have also led to disproportionately larger (smaller) increases in wholesale markup in the west (east), as Figure D.10 shows.

7 Conclusion

This paper studies how wholesale market power shapes the efficiency of production network formation and the welfare effects of technology-induced disintermediation. I develop a model featuring endogenous intermediation, wholesaler entry and exit, and markups to show that wholesale market power distorts relative input prices and misallocates production resources across both the intensive and extensive margins. The model predicts that when technological improvements make direct trade more efficient, firms substitute away from wholesalers, leading some wholesalers to exit. Fewer wholesalers raise market concentration and markups, which exacerbate misallocation and offset part of the welfare gains from disintermediation.

Empirically, I exploit Turkey's staggered fiber internet rollout as a natural experiment to test these mechanisms. The evidence confirms the model's key predictions: provinces with faster fiber expansion experience larger relative declines in the share of intermediated trade and in the number of wholesalers, alongside relative increases in wholesale markups.

Quantitative analysis calibrated to the observed changes in trade flows shows that endogenous increases in wholesale markups reduce the welfare gains from fiber-induced disintermediation by about 30%. Taken together, the results underscore the need for complementary competition policies—such as wholesale subsidies—to mitigate markup-induced distortions and fully realize the benefits of digital infrastructure investments. More broadly, they caution against policies that seek to "cut out the middleman" through the promotion of technology adoption: such efforts can backfire when disintermediation triggers the exit of marginal wholesalers and consolidates market power in the hands of a few surviving wholesalers.

References

Akerman, Anders, Edwin Leuven, and Magne Mogstad. 2022. "Information Frictions, Internet, and the Relationship between Distance and Trade." *American Economic Journal: Applied Economics*, 14(1): 133–63.

Alessandria, George, Joseph P. Kaboski, and Virgiliu Midrigan. 2010. "Inventories, Lumpy Trade, and Large Devaluations." *American Economic Review*, 100(5): 2304–39.

Amiti, Mary, and Jozef Konings. 2007. "Trade Liberalization, Intermediate Inputs, and Productivity: Evidence from Indonesia." *American Economic Review*, 97(5): 1611–1638.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare. 2012. "New Trade Models, Same Old Gains?" *American Economic Review*, 102(1): 94–130.

Arkolakis, Costas, Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare. 2019. "The Elusive Pro-Competitive Effects of Trade." *The Review of Economic Studies*, 86(1): 46–80.

Arkolakis, Costas, Federico Huneeus, and Yuhei Miyauchi. 2023. "Spatial Production Networks." *Working Paper*.

Atkin, David, Amit K. Khandelwal, and Adam Osman. 2017. "Exporting and Firm Performance: Evidence from a Randomized Experiment*." *The Quarterly Journal of Economics*, 132(2): 551–615.

Baqaee, David Rezza, and Emmanuel Farhi. 2019. "The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem." *Econometrica*, 87(4): 1155–1203.

Bartkus, Viva Ona, Wyatt Brooks, Joseph P. Kaboski, and Carolyn Pelnik. 2022. "Big fish in thin markets: Competing with the middlemen to increase market access in the Amazon." *Journal of Development Economics*, 155: 102757.

Blum, Bernardo S., Sebastian Claro, Kunal Dasgupta, Ignatius J. Horstmann, and Marcos A. Rangel. 2023. "Wholesalers in International Production Networks and Their Effects on Aggregate Productivity." *Working Paper*.

Bond, Steve, Arshia Hashemi, Greg Kaplan, and Piotr Zoch. 2021. "Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data." *Journal of Monetary Economics*, 121: 1–14.

Dekle, Robert, Jonathan Eaton, and Samuel Kortum. 2007. "Unbalanced Trade." *American Economic Review*, 97(2): 351–355.

De Loecker, Jan, and Frederic Warzynski. 2012. "Markups and Firm-Level Export Status." *American Economic Review*, 102(6): 2437–71.

Demir, Banu, Ana Cecília Fieler, Daniel Yi Xu, and Kelly Kaili Yang. 2024. "O-Ring Production Networks." *Journal of Political Economy*, 132(1): 200–247.

Demir, Banu, Beata Javorcik, and Piyush Panigrahi. 2023. "Breaking Invisible Barriers: Does Fast Internet Improve Access to Input Markets?" *Working Paper*.

Dhingra, Swati, and John Morrow. 2019. "Monopolistic Competition and Optimum Product Diversity under Firm Heterogeneity." *Journal of Political Economy*, 127(1): 196–232.

Dhyne, Emmanuel, Ayumu Ken Kikkawa, Toshiaki Komatsu, Magne Mogstad, and Felix Tintelnot. 2022. "Foreign Demand Shocks to Production Networks: Firm Responses and Worker Impacts." National Bureau of Economic Research Working Paper 30447.

Dhyne, Emmanuel, Ayumu Ken Kikkawa, Xianglong Kong, Magne Mogstad, and Felix Tintelnot. 2023. "Endogenous production networks with fixed costs." *Journal of International Economics*, 145: 103841.

Donaldson, Dave, and Richard Hornbeck. 2016. "Railroads and American Economic Growth: A "Market Access" Approach *." *The Quarterly Journal of Economics*, 131(2): 799–858.

Eaton, Jonathan, Samuel S Kortum, and Francis Kramarz. 2022. "Firm-to-Firm Trade: Imports, Exports, and the Labor Market." National Bureau of Economic Research Working Paper 29685.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu. 2015. "Competition, Markups, and the Gains from International Trade." *American Economic Review*, 105(10): 3183–3221.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu. 2023. "How Costly Are Markups?" *Journal of Political Economy*, 131(7): 1619–1675.

Epifani, Paolo, and Gino Gancia. 2011. "Trade, markup heterogeneity and misallocations." *Journal of International Economics*, 83(1): 1–13.

Fernandes, Ana M., Aaditya Mattoo, Huy Nguyen, and Marc Schiffbauer. 2019. "The internet and Chinese exports in the pre-ali baba era." *Journal of Development Economics*, 138: 57–76.

Fontagné, Lionel, Houssein Guimbard, and Gianluca Orefice. 2022. "Tariff-based product-level trade elasticities." *Journal of International Economics*, 137: 103593.

Ganapati, Sharat. 2024. "The Modern Wholesaler: Global Sourcing, Domestic Distribution, and Scale Economies." *Working Paper*.

Goldberg, Pinelopi Koujianou, Amit Kumar Khandelwal, Nina Pavcnik, and Petia Topalova. 2010. "Imported Intermediate Inputs and Domestic Product Growth: Evidence from India." *The Quarterly Journal of Economics*, 125(4): 1727–1767.

Grant, Matthew, and Meredith Startz. 2022. "Cutting Out the Middleman: The Structure of Chains of Intermediation." National Bureau of Economic Research Working Paper 30109.

Hulten, Charles R. 1978. "Growth Accounting with Intermediate Inputs." *The Review of Economic Studies*, 45(3): 511–518.

Huneeus, Federico. 2018. "Production Network Dynamics and the Propagation of Shocks." *Working Paper*.

Iacovone, Leonardo, and David McKenzie. 2022. "Shortening Supply Chains: Experimental Evidence from Fruit and Vegetable Vendors in Bogota." *Economic Development and Cultural Change*, 71(1): 111–149.

Kasahara, Hiroyuki, and Joe Rodrigue. 2005. "Does the use of imported intermediates increase productivity? Plant-Level Evidence." EPRI Working Paper.

Krugman, Paul. 1980. "Scale Economies, Product Differentiation, and the Pattern of Trade." *American Economic Review*, 70(5): 950–959.

Krugman, Paul R. 1979. "Increasing returns, monopolistic competition, and international trade." *Journal of International Economics*, 9(4): 469–479.

Leamer, Edward E. 2007. "A Flat World, a Level Playing Field, a Small World after All, or None of the above? A Review of Thomas L. Friedman's "The World is Flat"." *Journal of Economic Literature*, 45(1): 83–126.

Malgouyres, Clément, Thierry Mayer, and Clément Mazet-Sonilhac. 2021. "Technology-induced trade shocks? Evidence from broadband expansion in France." *Journal of International Economics*, 133: 103520.

Manova, Kalina, Andreas Moxnes, and Oscar Perelló. 2024. "Trade Intermediation in Global Production Networks." *Working Paper*.

Perelló, Oscar. 2024. "Trade Intermediation and Resilience in Global Sourcing." *Working Paper*.

Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Nicholas Trachter. 2021. "Diverging Trends in National and Local Concentration." *NBER Macroeconomics Annual*, 35: 115–150.

A Spatial Extension

This appendix presents the spatial extension of the single-location model in Section 3.

A.1 Environment and New Notation

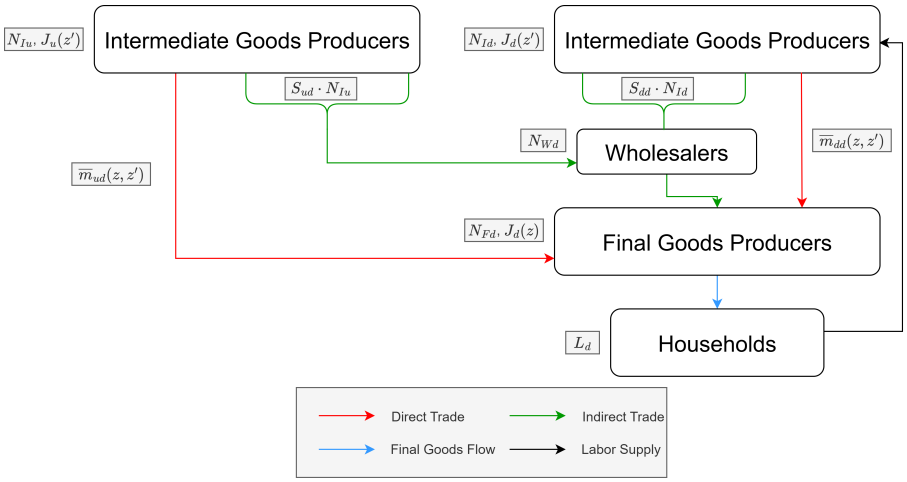

Figure A.1: Graphical Illustration of the Production Network

Note: This figure illustrates the spatial production network in the model. Firms trade directly or indirectly across provinces.

There is now a finite set of locations $\mathcal{N} = \{1, \ldots, N\}$. Each location i has exogenous labor supply L_i and location-specific wage w_i. Final goods are non-tradable (consumed locally), while intermediate goods are tradable across locations subject to iceberg trade costs $\tau_{ud} \geq 1$ when shipped from upstream location u to downstream location d. Aggregate nominal GDP is now chosen as the numeraire so that $\sum_i I_i = 1$.

The measures of entrants are location-specific: N_{Ii} intermediate producers, N_{Fi} final producers, and N_{Wi} wholesalers. Productivity draws are from location-specific distributions $J_i(z)$ with density $j_i(z)$. Wholesalers are local; a location d wholesaler can search for varieties produced in any upstream location u and resell them only to location d final producers. This assumption is motivated by the Motivational Fact 3 that indirect sourcing is mostly intermediated by local wholesalers. Moreover, there are separate matching markets for each upstream-downstream location pair.

All objects indexed by location (or ordered pairs of locations) are new relative to the single-location model. When an object coincides with its single-location analogue after suppressing subscripts, it is not re-defined.

A.2 Final Goods Producers

A type z final goods producer in location i earns revenue

$$p_{Fi}(z)^{1-\sigma} D_{Hi},$$

and uses a CES bundle of intermediate inputs sourced directly or indirectly from all upstream locations u:

$$Y_{Ii}(z) = \left\{ \sum_{u \in \mathcal{N}} \int_Z y_{Iui}(z')^{\frac{\sigma-1}{\sigma}} \phi_{cui}^{\frac{1}{\sigma}} \overline{m}_{ui}(z,z') + y_{ui}^W(z')^{\frac{\sigma-1}{\sigma}} S_{ui} [N_{Iu} j_u(z') - \overline{m}_{ui}(z,z')] dz' \right\}^{\frac{\sigma}{\sigma-1}}.$$

Relative to Section 3, direct matches and wholesalers' matches with suppliers are now indexed by location pairs (u,i).

The corresponding unit cost function becomes

$$c_i(z) = \left\{ \sum_{u \in \mathcal{N}} \left[m_{ui}(z)\, \theta_{ui}^m\, \phi_c\, c_{mui}^{1-\sigma} + S_{ui}\, N_{Iu}\, c_{Wui}^{1-\sigma} - S_{ui}\, m_{ui}(z)\, \theta^m\, \mu_d^{W\,1-\sigma}\, c_{mui}^{1-\sigma} \right] \right\}^{\frac{1}{1-\sigma}}$$

$$c_{mui} = \left[\int_Z p_{Iui}(z')^{1-\sigma} \frac{N_{Iu}\, v_{ud}(z')\, j_u(z')}{V_{ud}}\, dz' \right]^{\frac{1}{1-\sigma}}, \qquad c_{Wui} = \left[\int_Z p_{ui}^W(z')^{1-\sigma}\, j_u(z')\, dz' \right]^{\frac{1}{1-\sigma}}$$

Posting ads is now location-pair specific with cost $\sum_u w_i f_{mui} m_{ui}^\beta / \beta$, yielding the first order condition:

$$m_{ui}(z) = \Pi_{mui}\, z^{\frac{\sigma-1}{\beta-1}}, \qquad \Pi_{mui} \equiv \left[\frac{1}{w_i f_{mui}} \frac{1}{\sigma} \left(\frac{\sigma}{\sigma-1} \right)^{1-\sigma} (\phi_{cui} - S_{ui}\, \mu_i^{W\,1-\sigma})\, \theta_{ui}^m\, c_{mui}^{1-\sigma}\, D_{Hi} \right]^{\frac{1}{\beta-1}}.$$

A.3 Intermediate Goods Producers

A type z' intermediate producer in location i can now sell their goods to all downstream locations d, and sets a destination-specific price

$$p_{Iid}(z') = \frac{\sigma}{\sigma-1} \frac{w_i}{z'} \tau_{id},$$

and pays location-pair specific search costs $\sum_d w_i f_{vid} v_{id}^\beta / \beta$. Profits across all destinations are

$$\max_{\{p_{Iid}, v_{id}\}} \sum_{d \in \mathcal{N}} p_{Iid}^{1-\sigma} (v_{id}\, \theta_{id}^v\, D_{mid} + S_{id} D_{Wid} - v_{id}\, \theta^v\, S_{id}\, \mu_d^{W-\sigma}\, \phi_{cid}^{-1}\, D_{mid})$$
$$- \sum_{d \in \mathcal{N}} \frac{1}{z} w_i \tau_{id} \left[p_{Iid}^{-\sigma} (v_{id} D_{mid} + S_{id} D_{Wid} - v_{id} S_{id} D_{Wmid}) \right] - \sum_{d \in \mathcal{N}} w_i f_{vid} \frac{v_{id}^\beta}{\beta}. \tag{28}$$

The ad-posting first order condition becomes

$$v_{id}(z') = \Pi_{vid}\, z'^{\frac{\sigma-1}{\beta-1}}, \qquad \Pi_{vid} \equiv \left[\frac{1}{w_i\, f_{vid}}\,\frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} w_i^{1-\sigma}\,(1 - S_{id}\,\mu_d^{W-\sigma}\,\phi_{cid}{}^{-1})\,\theta_{id}^{v}\,D_{mid}\right]^{\frac{1}{\beta-1}}$$

A.4 Direct Matching Across Locations

Each location pair (u, d) features its own matching market with total ads

$$V_{ud} = N_{Iu}\int_Z v_{ud}(z)\, j_u(z)dz, \tag{29}$$

$$M_{ud} = N_{Fd}\int_Z m_{ud}(z)\, j_d(z)dz, \tag{30}$$

and matching function $\widetilde{M}_{ud} = \kappa_{ud} V_{ud}^{\lambda_V} M_{ud}^{\lambda_M}$. Ad success rates are

$$\theta_{ud}^{v} = \kappa_{ud} V_{ud}^{\lambda_V - 1} M_{ud}^{\lambda_M} \tag{31}$$

$$\theta_{ud}^{m} = \kappa_{ud} V_{ud}^{\lambda_V} M_{ud}^{\lambda_M - 1}. \tag{32}$$

The mass of suppliers of type z' from u matched to a distinct type z buyer from d is:

$$\overline{m}_{ud}(z, z') = m_{ud}(z)\theta_{ud}^{m}\frac{N_{Iu} v_{ud}(z')\, j_u(z')}{V_{ud}}$$

And the mass of buyers of type z from location d that are matched to a distinct type z' supplier from location u is:

$$\overline{v}_{ud}(z', z) = v_{ud}(z')\theta_{ud}^{v}\frac{N_{Fd} m_{ud}(z)\, j_d(z)}{M_{ud}}$$

A.5 Wholesalers

In each location d, there are N_{Wd} wholesalers, who compete locally à la Cournot, resulting in a location-specific wholesale markup that depends on the number of local wholesalers:

$$p_{ud}^{W}(z) = \frac{N_{Wd}\sigma}{N_{Wd}\sigma - 1} p_{Iud}(z), \qquad \mu_d^{W} \equiv \frac{N_{Wd}\sigma}{N_{Wd}\sigma - 1}. \tag{33}$$

Wholesalers choose pair-specific search efforts s_{ud}, facing congestion $\theta_{ud}^{W} = \frac{(N_{Wd} s_{ud})^{\lambda_W}}{N_{Wd} s_{ud}}$, and incurring costs $w_d f_{Wud}(s_{ud} N_{Iu}/N_{Wd})^{\beta_W}/\beta_W$. The first order condition for search effort is

$$s_{ud} = \left[\left(\frac{N_{Wd}}{N_{Iu}}\right)^{\beta_W}\frac{1}{w_d\, f_{Wd}}\,\theta_{ud}^{W}\, N_{Iu}\,\Pi_{Wud}\right]^{\frac{1}{\beta_W - 1}}, \tag{34}$$

where

$$\Pi_{Wud} \equiv \int_Z \frac{1}{N_{Wd}} \frac{1}{N_{Wd}\,\sigma} \left(\mu_d^W / S_{ud}\right) \left(x_{Wud}(z') - x_{Wmud}(z')\right) j_u(z')\, dz'$$

with $S_{ud} = s_{ud}\theta_{ud}^W$ the share of location u intermediate goods varieties matched with wholesalers in d.

A.6 Households

Local household i earns

$$I_i = w_i L_i + \Pi_i^W, \tag{35}$$

faces local price index

$$P_i^H = \left[N_{Fi} \int_Z p_{Fi}(z)^{1-\sigma} j_i(z) dz \right]^{\frac{1}{1-\sigma}}, \tag{36}$$

and demands $p_{Fi}(z)c_i^H(z) = p_{Fi}(z)^{1-\sigma} D_{Hi}$ with

$$D_{Hi} = I_i / P_i^{H\,1-\sigma} \tag{37}$$

A.7 Free Entry, Labor Market Clearing, and Trade Balance

Location-specific free entry conditions (compare to (51)-(55)) become

$$w_i F_{Ii} N_{Ii} = \sum_d \left(\frac{1}{\sigma} X_{mid} + \frac{1}{\sigma} X_{Wid} - \frac{1}{\sigma} X_{Wmid} \right) - \sum_d N_{Ii} \int_Z w_i f_{vid} \frac{v_{id}(z)^\beta}{\beta} j_i(z) dz, \tag{38}$$

$$w_i F_{Fi} N_{Fi} = \frac{1}{\sigma} X_{Hi} - \sum_u N_{Fi} \int_Z w_i f_{mui} \frac{m_{ui}(z)^\beta}{\beta} j_i(z) dz, \tag{39}$$

$$w_i F_{Wi} N_{Wi} \leq N_{Wi} \sum_u \left[s_{ui} \theta_{ui}^W N_{Iu} \Pi_{Wui} - w_i f_{Wui} \frac{(s_{ui} N_{Iu} / N_{Wi})^{\beta_W}}{\beta_W} \right]. \tag{40}$$

Labor market clearing in each i:

$$\begin{aligned} L_i = &\sum_d \frac{1}{w_i} \left(\frac{\sigma - 1}{\sigma} X_{mid} + \frac{\sigma - 1}{\sigma} X_{Wid} - \frac{\sigma - 1}{\sigma} X_{Wmid} \right) + \sum_d N_{Ii} \int_Z f_{vid} \frac{v_{id}(z)^\beta}{\beta} j_i(z) dz \\ &+ \sum_u N_{Fi} \int_Z f_{mui} \frac{m_{ui}(z)^\beta}{\beta} j_i(z) dz + \sum_u f_{Wui} \frac{(s_{ui} N_{Iu} / N_{Wi})^{\beta_W}}{\beta_W} N_{Wi} + F_{Ii} N_{Fi} + F_{Fi} N_{Ii} + F_{Wi} N_{Wi}. \end{aligned} \tag{41}$$

Combining labor market clearing, household budget constraints and free entry implies bilateral trade balance net of wholesale profits:

$$\sum_u (X_{mui} + X_{Wui} - X_{Wmui}) = \sum_d (X_{mid} + X_{Wid} - X_{Wmid}). \tag{42}$$

A.8 Spatial Equilibrium

A spatial equilibrium is a set of endogenous variables $\{w_i, I_i, P_i^H, \theta_{ud}^v, \theta_{ud}^m, S_{ud}, N_{Ii}, N_{Fi}, N_{Wi}\}$ satisfying (41), (35), (36), (31), (32), (34), (38), (39), and (40). Setting $N = 1$, $w_i = w$, $\tau_{ud} = 1$, and dropping location subscripts collapses all expressions to the single-location model in Section 3.

B Derivations

B.1 Detailed Derivation for Final Goods Producers' Cost Minimization

The cost minimization problem of a type z buyer is:

$$\min_{\{y_I(z,z')\},\{y^W(z,z')\}} \int_Z p_I(z')y_I(z,z')\overline{m}(z,z') + p^W(z')y^W(z,z')S[N_I j(z') - \overline{m}(z,z')]dz'$$

$$\text{s.t. } Y_I(z) = \left\{ \int_Z y_I(z,z')^{\frac{\sigma-1}{\sigma}} \phi_c^{\frac{1}{\sigma}} \overline{m}(z,z') + y^W(z,z)^{\frac{\sigma-1}{\sigma}} S[N_I j(z') - \overline{m}(z,z')]dz' \right\}^{\frac{\sigma}{\sigma-1}} \geq Y$$

First order condition with respect to $y_I(z')$ is:

$$p_I(z')\overline{m}(z,z') = \lambda(z)\, Y^{\frac{1}{\sigma}}\, y_I(z,z')^{-\frac{1}{\sigma}} \phi_c^{\frac{1}{\sigma}} \overline{m}(z,z') \tag{43}$$

First order condition with respect to $y^W(z')$ is:

$$p^W(z')\, S\, [N_I j(z') - \overline{m}(z,z')] = \lambda(z)\, Y^{\frac{1}{\sigma}} y^W(z,z')^{-\frac{1}{\sigma}} S[N_I j(z') - \overline{m}(z,z')] \tag{44}$$

where $\lambda(z)$ is the Lagrange multiplier of the constraint. Using the constraint, adding terms, and integrate yield:

$$C(z) \equiv \int_Z p_I(z')y_I(z,z')\overline{m}(z,z') + p^W(z')y^W(z,z')S[N_I j(z') - \overline{m}(z,z')]dz' = \lambda(z)\, Y \tag{45}$$

Also, define $c(z) \equiv \frac{C(z)}{Y}$. Now, substitute λ into (43) using (45):

$$p_I(z')y_I(z,z') = \phi_c \left[\frac{p_I(z')}{c(z)} \right]^{1-\sigma} C(z)$$

Now, $C(z)$ is the total cost of production of a type z final goods producer, i.e.:

$$C(z) = \frac{\sigma - 1}{\sigma} x_H(z)$$

Therefore, the sales of a supplier z to a matched buyer z' is:

$$p_I(z')y_I(z,z') = \phi_c \left[\frac{p_I(z')}{c(z)}\right]^{1-\sigma} \frac{\sigma-1}{\sigma} x_H(z) \tag{46}$$

Similarly, the sales of a type z' intermediate goods variety to a buyer z through wholesalers is:

$$p^W(z')y^W(z,z') = \left[\frac{p^W(z')}{c(z)}\right]^{1-\sigma} \frac{\sigma-1}{\sigma} x_H(z) \tag{47}$$

Lastly, to derive $c(z)$, raise both sides of (43) and (44) to power $1-\sigma$, adding the two and using the constraint:

$$\begin{aligned}
\lambda(z)^{1-\sigma} &= \int_Z p_I(z')^{1-\sigma}\phi_c\overline{m}(z,z') + {p^W(z')}^{1-\sigma} S[N_I j(z') - \overline{m}(z,z')]dz' \\
c(z) = \lambda(z) &= \left\{\sum_{u\in N}\left[m(z)\theta^m\phi_c c_m^{1-\sigma} + S\,N_I\,c_W^{1-\sigma} - S\,m(z)\theta^m {\mu^W}^{1-\sigma} c_m^{1-\sigma}\right]\right\}^{\frac{1}{1-\sigma}}
\end{aligned} \tag{48}$$

where

$$\begin{aligned}
c_m &= \left[\int_Z p_I(z')^{1-\sigma}\frac{N_I v(z') j(z')}{V}dz'\right]^{\frac{1}{1-\sigma}} \\
c_W &= \left[\int_Z {p^W(z')}^{1-\sigma} j(z')dz'\right]^{\frac{1}{1-\sigma}}
\end{aligned}$$

The total direct sales of an intermediate goods producer of type z' to all the matched final goods producers $x_m(z', v(z'))$ is therefore:

$$\begin{aligned}
x_m(z') &= \int_Z \phi_c \left[\frac{p_I(z')}{c(z)}\right]^{1-\sigma} \frac{\sigma-1}{\sigma} x_H(z)\, v(z')\theta^v \frac{N_F m(z) j(z)}{M} dz \\
&= p_I(z')^{1-\sigma} v(z')\,\theta^v\, D_m
\end{aligned}$$

where

$$D_m = \phi_c \frac{\sigma-1}{\sigma} \int_Z \frac{x_H(z)}{c(z)^{1-\sigma}} \frac{N_F m(z) j(z)}{M} dz \tag{49}$$

Moreover, the *expected* total indirect sales of an intermediate goods producer of type z' to all the *unmatched* final goods producers through wholesalers $x_W(z') - x_{Wm}(z', v(z'))$is:

$$\begin{aligned}
x_W(z') - x_{Wm}(z', v(z')) &= \frac{1}{\mu^W} S \int_Z \left[\frac{p^W(z')}{c(z)}\right]^{1-\sigma} \frac{\sigma-1}{\sigma} x_H(z) \left[N_F j(z) - v(z')\theta^v \frac{N_F\, m(z)\, j(z)}{M}\right] dz \\
&= p_I(z')^{1-\sigma}\, S\, D_W - p_I(z')^{1-\sigma}\, v(z')\,\theta^v\, S\, {\mu^W}^{-\sigma}\, \phi_c^{-1} D_m
\end{aligned}$$

where

$$D_W = \mu^{W-\sigma}\,\frac{\sigma-1}{\sigma}\,N_F\int_Z \frac{x_H(z)}{c(z)^{1-\sigma}}\,j(z)\,dz \tag{50}$$

B.2 Equilibrium Conditions

Free entry pins down N_I, N_F, and the discrete N_W. Aggregate post-entry profits weakly exceed aggregate entry costs, with equality for intermediate and final goods producers. Entry costs are paid in labor, and their levels are controlled by the parameters F_I, F_F, and F_W.

Free entry condition for intermediate goods producers

$$w\,F_I\,N_I = \frac{1}{\sigma}X_m + \frac{1}{\sigma}X_W - \frac{1}{\sigma}X_{Wm} - N_I\int_Z w\,f_{v'}\,\frac{v(z')^\beta}{\beta}\,j(z')\,dz' \tag{51}$$

where

$$X_i \equiv N_I\int_Z x_i(z')\,j(z')\,dz',\ \ i = m, W, Wm \tag{52}$$

Free entry condition for final goods producers

$$w\,F_F\,N_F = \frac{1}{\sigma}X_H - N_F\int_Z w\,f_m\,\frac{m(z)^\beta}{\beta}\,j(z)\,dz \tag{53}$$

where

$$X_H \equiv N_F\int_Z x_H(z)\,j(z)\,dz = N_F\int_Z p_F(z)^{1-\sigma}\,D_H\,j(z)\,dz = P^{H\,1-\sigma}\,D_H = I = w\,L + \Pi^W \tag{54}$$

is total sales of final goods producers.

Free entry condition for wholesalers

$$w\,F_W\,N_W \le N_W\left[s\,\theta^W\,N_I\,\Pi_W - w\,f_W\,\frac{(s\,N_I/N_W)^{\beta_W}}{\beta_W}\right] \tag{55}$$

The last equilibrium condition is the labor market clearing condition, which states that the total supply of labor is equal to the total demand for it, which consists of labor demand for intermediate goods production, for posting search ads, for financing wholesalers' search effort, and for financing entry of firms and wholesalers.

Labor market clearing condition

Labor Supply =Production Labor + Ads Labor + Wholesalers' Search Labor + Entry Cost Labor

$$\begin{aligned} L =& \frac{1}{w}\left(\frac{\sigma-1}{\sigma}X_m + \frac{\sigma-1}{\sigma}X_W - \frac{\sigma-1}{\sigma}X_{Wm}\right) \\ &+ N_I \int_Z f_v \frac{v(z)^\beta}{\beta}\, j(z)\, dz + N_F \int_Z f_m \frac{m(z)^\beta}{\beta}\, j(z)\, dz \\ &+ f_W \frac{(s\, N_I/N_W)^{\beta_W}}{\beta_W} N_W + F_I\, N_F + F_F\, N_I + F_W\, N_W \end{aligned} \tag{56}$$

B.3 Endogenous Entry

Using wholesalers' first order condition, we can rewrite the sum of their search cost as:

$$\begin{aligned} & w\, f_W \frac{(s\, N_I/N_W)^{\beta_W}}{\beta_W} \\ =& \frac{1}{\beta_W}\, S\, N_I\, \Pi_W \end{aligned}$$

Substitute it back to (40) yields:

$$\begin{aligned} wF_W N_W \leq& N_W \frac{\beta_W - 1}{\beta_W} S\, N_I\, \Pi_W \\ wF_W \leq& \frac{\beta_W - 1}{\beta_W \sigma}\, N_W^{-2} \sum_{u\in\mathcal{N}} (X^W - X_m^W) \\ wF_W\, {N_W}^2 \leq& \frac{\beta_W - 1}{\beta_W \sigma}\, (1-\Omega)\, (X_m + X^W - X_m^W) \\ wF_W\, {N_W}^2 \leq& \frac{\beta_W - 1}{\beta_W \sigma}\, (1-\Omega)\, \frac{\sigma-1}{\sigma}\, (wL + \Pi^W) \\ N_W \leq& \left[\frac{(\beta_W - 1)(\sigma-1)}{\beta_W \sigma^2}\, (1-\Omega)\, \frac{wL + \Pi^W}{wF_W}\right]^{\frac{1}{2}} \end{aligned}$$

where

$$\Omega \equiv \frac{X_m}{X_m + X^W - X_m^W}$$

is the share of direct trade inclusive of wholesale profits.

We can further tighten this inequality. From

$$\Pi^W \;\equiv\; \frac{\beta_W - 1}{\beta_W \sigma}\, (1-\Omega)\, \frac{\sigma-1}{\sigma}\, (wL + \Pi^W)\, \frac{1}{N_W} \;-\; wF_W N_W,$$

multiplying both sides by N_W and rearranging gives the identity

$$wF_W N_W^2 \;=\; \frac{\beta_W - 1}{\beta_W \sigma}\, (1-\Omega)\, \frac{\sigma-1}{\sigma}\, (wL + \Pi^W) \;-\; N_W\, \Pi^W.$$

Hence,

$$wF_W N_W^2 \;-\; \frac{\beta_W - 1}{\beta_W \sigma}\,(1-\Omega)\,\frac{\sigma-1}{\sigma}\,wL \;=\; \Pi^W\left(\frac{\beta_W - 1}{\beta_W \sigma}\,(1-\Omega)\,\frac{\sigma-1}{\sigma} \;-\; N_W\right).$$

Therefore, under $\Pi^W \geq 0$ and

$$N_W \;\geq\; \frac{\beta_W - 1}{\beta_W \sigma}\,(1-\Omega)\,\frac{\sigma-1}{\sigma},$$

we obtain the tighter bound

$$wF_W N_W^2 \;\leq\; \frac{\beta_W - 1}{\beta_W \sigma}\,(1-\Omega)\,\frac{\sigma-1}{\sigma}\,wL,$$

which implies

$$N_W \;\leq\; \left[\frac{(\beta_W - 1)(\sigma-1)}{\beta_W \sigma^2}\,(1-\Omega)\,\frac{L}{F_W}\right]^{\frac{1}{2}}.$$

B.4 Proof of Proposition 1

Proof. $A_{\text{efficient}}$ is derived using the expression for P^H derived for social planner's problem in Appendix 4.1, and the fact that $A_{\text{efficient}} = \frac{1}{P^H}$. I proceed to solve for A.

Step 1. Average productivity in the decentralized equilibrium Define average labor productivity as output (welfare) per unit of production labor,

$$A \;=\; \frac{\text{Welfare}}{L_p} \;=\; \frac{I}{P^H}\,\frac{1}{L_p}.$$

Step 2. Relating labor to revenues Because intermediate goods firms charge a markup $\sigma/(\sigma-1)$, variable cost absorbs the fraction $(\sigma-1)/\sigma$ of revenue:

$$L_p = \frac{\sigma-1}{\sigma}\left(X_m + X_W - X_{Wm}\right).$$

Step 3. Income in terms of revenues The household budget constraint (54) and the identical markup used by final goods firms imply

$$I = \frac{\sigma}{\sigma-1}\sum_{u\in\mathcal{N}}\left(X_m + X_W - X_{Wm}\right).$$

Step 4. Trade flows and the price index Aggregating firm-level revenues and eliminating demand

shifters gives

$$L_p = \left(\frac{\sigma}{\sigma-1}\right)^{-2\sigma}\Big\{\big[\phi_c - (\mu^W)^{-\sigma}S\big]\frac{\widetilde{M}}{MV}\big[N_F\int z^{\sigma-1}m(z)j(z)\,dz\big]\big[N_I\int z^{\sigma-1}v(z)j(z)\,dz\big] + (\mu^W)^{-\sigma}S\,N_FN_I\big(\mathbb{E}[z^{\sigma-1}]\big)^2\Big\},$$

$$P^H = \left(\frac{\sigma}{\sigma-1}\right)^{2}\Big\{\big[\phi_c - (\mu^W)^{1-\sigma}S\big]\frac{\widetilde{M}}{MV}\big[N_F\int z^{\sigma-1}m(z)j(z)\,dz\big]\big[N_I\int z^{\sigma-1}v(z)j(z)\,dz\big] + (\mu^W)^{1-\sigma}S\,N_FN_I\big(\mathbb{E}[z^{\sigma-1}]\big)^2\Big\},$$

$$I = \left(\frac{\sigma}{\sigma-1}\right)^{2} L_p \quad \frac{\{\big[\phi_c - (\mu^W)^{1-\sigma}S\big]\frac{\widetilde{M}}{MV}\big[N_F\int z^{\sigma-1}m(z)j(z)\,dz\big]\big[N_I\int z^{\sigma-1}v(z)j(z)\,dz\big] + (\mu^W)^{1-\sigma}S\,N_FN_I\big(\mathbb{E}[z^{\sigma-1}]\big)^2\Big\}}{\{\big[\phi_c - (\mu^W)^{-\sigma}S\big]\frac{\widetilde{M}}{MV}\big[N_F\int z^{\sigma-1}m(z)j(z)\,dz\big]\big[N_I\int z^{\sigma-1}v(z)j(z)\,dz\big] + (\mu^W)^{-\sigma}S\,N_FN_I\big(\mathbb{E}[z^{\sigma-1}]\big)^2\Big\}}$$

Step 5. Solving for A Substitute out L_p, P^H, and I from A, and introduce the aggregate markup

$$\mu \equiv \frac{P^H}{w/A} = P^H A.$$

Solving for A gives the expression stated in Proposition 1. □

B.5 Corollary 1.1

Corollary 1.1 (Productivity gap and dispersion, first-order approximation with varying Ω^*)**.** *In the single-location economy of Proposition 1, define*

$$\Theta(\mu^W, \Omega^*) \;\equiv\; \frac{A(\mu^W, \Omega^*)}{A_{efficient}(\Omega^*)} \in (0, 1],$$

where $A(\mu^W, \Omega^*)$ and $A_{efficient}(\Omega^*)$ are given in Proposition 1, $\mu_D = (\sigma/(\sigma-1))^2$, $\mu_I = \mu_D\,\mu^W$, and $\Omega^* \equiv A_D^{\sigma-1}/(A_D^{\sigma-1} + A_I^{\sigma-1})$ is the efficient direct-trade share (which may vary with primitives such as A_D). Then:

(i) ***No-markup benchmark:*** $\Theta(1, \Omega^*) = 1$ *for any* $\Omega^* \in (0, 1)$.

(ii) ***Monotone widening with*** μ^W ***(fixed*** A_D, A_I***):***

$$\frac{\partial\,\Theta(\mu^W, \Omega^*)}{\partial \mu^W} < 0 \quad \textit{for all } \mu^W > 1.$$

(iii) ***First-order dispersion result (varying*** Ω^****):*** *Let $(\bar{\mu}^W, \bar{\Omega}^*)$ be a baseline. Define the efficient-weight*

dispersion measure

$$\mathcal{V}^* \;\equiv\; \Omega^*(1-\Omega^*)\big(\ln\mu_I - \ln\mu_D\big)^2 = \Omega^*(1-\Omega^*)\big(\ln\mu^W\big)^2.$$

Then, using the standard quadratic approximation to $\ln\Theta$ *around* $(\bar{\mu}^W, \bar{\Omega}^*)$*, its* first-order *differential satisfies*

$$d\ln\Theta \;\approx\; -\frac{\sigma}{2(\sigma-1)}\, d\mathcal{V}^*.$$

Hence, for small changes in primitives that alter both μ^W *and* Ω^**, the productivity wedge increases approximately with markup dispersion measured using the efficient direct-trade share.*

Proof. **(i)-(ii): Monotonicity in μ^W with A_D, A_I fixed.** Keep A_D and A_I fixed and treat μ^W as the sole variable.

Step 1 (restating A). From Proposition 1,

$$A = \left[\left(\tfrac{\mu_D}{\mu}\right)^{-\sigma} A_D^{\sigma-1} + \left(\tfrac{\mu_I}{\mu}\right)^{-\sigma} A_I^{\sigma-1}\right]^{\frac{1}{\sigma-1}}, \qquad \mu_I = \mu_D \mu^W, \qquad \mu_D = \left(\tfrac{\sigma}{\sigma-1}\right)^2.$$

Step 2 (compact ratio). Define

$$U(\mu^W) = A_D^{\sigma-1} + A_I^{\sigma-1}\,(\mu^W)^{-(\sigma-1)}, \qquad V(\mu^W) = A_D^{\sigma-1} + A_I^{\sigma-1}\,(\mu^W)^{-\sigma},$$

so that

$$\Theta(\mu^W) = \left[\frac{U(\mu^W)^\sigma}{V(\mu^W)^{\sigma-1}\,(A_D^{\sigma-1} + A_I^{\sigma-1})}\right]^{\frac{1}{\sigma-1}}.$$

Step 3 (sign of the derivative). Let $g = \log\Theta$. Then

$$g'(\mu^W) = \sigma\frac{U'}{U} - (\sigma-1)\frac{V'}{V} = \sigma(\sigma-1)A_I^{\sigma-1}\,(\mu^W)^{-(\sigma+1)}\left[-\frac{\mu^W}{U} + \frac{1}{V}\right].$$

For $\mu^W > 1$, $-\mu^W V + U = A_D^{\sigma-1}(1-\mu^W) < 0$, hence $g'(\mu^W) < 0$ and thus $\Theta'(\mu^W) < 0$. Moreover, $\Theta(1) = 1$, so the productivity gap expands strictly as μ^W rises. This proves parts (i) and (ii).

(iii): First-order dispersion result with varying Ω^*. Now allow A_D (hence $\Omega^* \equiv A_D^{\sigma-1}/(A_D^{\sigma-1} + A_I^{\sigma-1})$) to vary. Define the dispersion measure (using efficient weights)

$$\mathcal{V}^* \;\equiv\; \Omega^*(1-\Omega^*)\big(\ln\mu_I - \ln\mu_D\big)^2 = \Omega^*(1-\Omega^*)\big(\ln\mu^W\big)^2.$$

Consider a baseline $(\bar{\mu}^W, \bar{\Omega}^*)$ and small changes $(\Delta\mu^W, \Delta\Omega^*)$. A quadratic approximation around the

baseline gives

$$\ln \Theta(\mu^W, \Omega^*) \;\approx\; -\frac{\sigma}{2(\sigma-1)}\,\mathcal{V}^*.$$

Taking *first-order differentials* of this approximation yields

$$d\ln\Theta \;\approx\; -\frac{\sigma}{2(\sigma-1)}\,d\mathcal{V}^*.$$

Thus, to a first-order (linear) approximation, any increase in markup dispersion measured with the efficient direct-trade share lowers $\ln\Theta$ and widens the productivity gap. □

B.6 Proof of Proposition 2

Proof. Under the simplifying assumptions of proposition 2, the social planner's first order conditions for m and v, derived in Section 4, can be rewritten as:

$$\begin{aligned} f_m\, m &= L_p\, \frac{1}{\sigma-1}\, \left[(\phi_c - S)\,\kappa\, N_F\, m\, N_I\, v + N_F\, N_I\, S\right]^{-1}\, (\phi_c - S)\,\kappa\, N_I\, v \\ f_v\, v &= L_p\, \frac{1}{\sigma-1}\, \left[(\phi_c - S)\,\kappa\, N_F\, m\, N_I\, v + N_F\, N_I\, S\right]^{-1}\, (\phi_c - S)\,\kappa\, N_F\, m \end{aligned}$$

Combining these first order conditions solves for m and v:

$$\begin{aligned} m &= \left\{ \frac{1}{\sigma\, N_F\, f_m} \left[\widetilde{L} - (\sigma-1) \left(\frac{f_m\, f_v}{N_I\, N_F} \right)^{\frac{1}{2}} \frac{N_F\, N_I}{(\phi_c - S)\,\kappa}\, S \right] \right\} \\ v &= \left\{ \frac{1}{\sigma\, N_I\, f_v} \left[\widetilde{L} - (\sigma-1) \left(\frac{f_m\, f_v}{N_I\, N_F} \right)^{\frac{1}{2}} \frac{N_F\, N_I}{(\phi_c - S)\,\kappa}\, S \right] \right\} \end{aligned}$$

these give us the total labor used for posting ads, which can be subtracted from $\widetilde{L}$ to arrive at $L_{p,\text{efficient}}$. Substituting out m and v solves for P^H, the inverse of which is $A_{\text{efficient}}$.

Following the same procedure, we can solve for m and v in the decentralized equilibrium:

$$
\begin{aligned}
m =& \left(\left\{ \sigma - 1 + \left[\frac{\phi_c - (\mu^W)^{1-\sigma} S}{\phi_c - (\mu^W)^{-\sigma} S} \frac{1}{2} + \frac{1}{2} \right] \right\}^{-1} \frac{\phi_c - (\mu^W)^{1-\sigma} S}{\phi_c - (\mu^W)^{-\sigma} S} \frac{1}{N_F f_m} \right. \\
& \left. \left[\widetilde{L} - (\sigma - 1) \left[\frac{\phi_c - (\mu^W)^{1-\sigma} S}{\phi_c - (\mu^W)^{-\sigma} S} \frac{f_m f_v}{N_I N_F} \right]^{\frac{1}{2}} \frac{(\mu^W)^{-\sigma} N_F N_I}{(\phi_c - (\mu^W)^{1-\sigma} S) \kappa} S \right] \right) \\
\approx & \left\{ \frac{1}{\sigma N_F f_m} \left[\widetilde{L} - (\sigma - 1) \left(\frac{f_m f_v}{N_I N_F} \right)^{\frac{1}{2}} \frac{(\mu^W)^{-\sigma} N_F N_I}{(\phi_c - (\mu^W)^{1-\sigma} S) \kappa} S \right] \right\} \\
v =& \left(\left\{ \sigma - 1 + \left[\frac{\phi_c - (\mu^W)^{1-\sigma} S}{\phi_c - (\mu^W)^{-\sigma} S} \frac{1}{2} + \frac{1}{2} \right] \right\}^{-1} \frac{\phi_c - (\mu^W)^{1-\sigma} S}{\phi_c - (\mu^W)^{-\sigma} S} \frac{1}{N_I f_v} \right. \\
& \left. \left[\widetilde{L} - (\sigma - 1) \left[\frac{\phi_c - (\mu^W)^{1-\sigma} S}{\phi_c - (\mu^W)^{-\sigma} S} \frac{f_m f_v}{N_I N_F} \right]^{\frac{1}{2}} \frac{(\mu^W)^{-\sigma} N_F N_I}{(\phi_c - (\mu^W)^{1-\sigma} S) \kappa} S \right] \right) \\
\approx & \left\{ \frac{1}{\sigma N_I f_v} \left[\widetilde{L} - (\sigma - 1) \left(\frac{f_m f_v}{N_I N_F} \right)^{\frac{1}{2}} \frac{(\mu^W)^{-\sigma} N_F N_I}{(\phi_c - (\mu^W)^{1-\sigma} S) \kappa} S \right] \right\}
\end{aligned}
$$

where the approximation makes use of the fact that $\frac{\phi_c - (\mu^W)^{1-\sigma} S}{\phi_c - (\mu^W)^{-\sigma} S} \approx 1$. Again, these expressions for ads give us the total labor used for posting ads, which can be subtracted from $\widetilde{L}$ to arrive at $L_{p,\text{decentralized}}$. Lastly, we can substitute out m and v from the aggregate productivity expression derived in Proposition 1 for the decentralized equilibrium to arrive at $A_{\text{decentralized}}$. □

B.7 Corollary 2.1

Corollary 2.1 (Welfare gap and dispersion, first-order approximation with varying Ω^*)**.** *Under the assumptions of Proposition 2, define*

$$
\Xi(\mu^W, \Omega^*) \equiv \frac{\mathcal{W}(\mu^W, \Omega^*)}{\mathcal{W}_{efficient}(\Omega^*)} \in (0, 1], \qquad \mu^W \geq 1,
$$

where Ω^* is the efficient direct-trade share (which may vary). Then:

(i) ***No-markup benchmark:*** $\Xi(1, \Omega^*) = 1$ *for any* $\Omega^* \in (0, 1)$.

(ii) ***Monotone widening with μ^W (fixed A_D, A_I):***

$$
\frac{\partial \, \Xi(\mu^W, \Omega^*)}{\partial \mu^W} < 0 \quad \textit{for all } \mu^W > 1.
$$

(iii) ***First-order dispersion result (varying Ω^*):*** *Using the same dispersion measure*

$$
\mathcal{V}^* \equiv \Omega^* (1 - \Omega^*) \big(\ln \mu_I - \ln \mu_D \big)^2,
$$

a first-order differential of the standard quadratic approximation to $\ln \Xi$ *gives*

$$d \ln \Xi \;\approx\; -\kappa_\sigma \, d\mathcal{V}^*,$$

where $\kappa_\sigma > 0$ *depends on* σ *and baseline allocations (see Proposition 2). Consequently, for small changes, the welfare wedge* $\mathcal{W}_{\textit{efficient}} - \mathcal{W}$ *increases approximately with markup dispersion measured using the efficient direct-trade share.*

Proof. Technology, matching parameters, and the numbers of firms (N_F, N_I) are fixed; only μ^W varies.

1. Log-derivative of decentralized welfare. $\mathcal{W} = L_p(\mu^W)\; X(\mu^W)^{1/(\sigma-1)}$ with

$$X(\mu^W) = \left(\tfrac{\mu_D}{\mu}\right)^{-\sigma} \phi_c \, \widetilde{M} + \left(\tfrac{\mu_I}{\mu}\right)^{-\sigma} S \left(N_F N_I - \widetilde{M}\right),$$

and $L_p(\mu^W)$ and $\widetilde{M}(\mu^W)$ given in Proposition 2. Because $\mathcal{W}_{\text{efficient}}$ is constant, sign $d\Xi/d\mu^W$ = sign $d \ln \mathcal{W}/d\mu^W$. Thus

$$\frac{d \ln \mathcal{W}}{d\mu^W} = \frac{1}{L_p} \frac{dL_p}{d\mu^W} + \frac{1}{\sigma - 1} \frac{1}{X} \frac{dX}{d\mu^W}.$$

2. Sign of $dL_p/d\mu^W$**.** Denote $a(\mu^W) \equiv \phi_c - \mu^{W\,1-\sigma} S, \quad b(\mu^W) \equiv \phi_c - \mu^{W\,-\sigma} S$. Both a and b *increase* in μ^W, but b increases faster because its exponent is $-\sigma < 1 - \sigma$. The critical ratio inside L_p is $\rho(\mu^W) \equiv a/b$. Standard differentiation gives

$$\frac{d\rho}{d\mu^W} = \rho \, \frac{S}{\mu^W} \frac{\sigma \, b - (\sigma - 1)\, a}{b^2} > 0;$$

the inequality uses $b > a > 0$. Inspecting the closed form of L_p one sees it is *decreasing* in ρ and thus in μ^W; hence $dL_p/d\mu^W < 0$.

3. Sign of $dX/d\mu^W$**.** First, $\widetilde{M}(\mu^W)$ is proportional to $\rho^{-1/2}$, so $\widetilde{M}$ *falls* when μ^W rises. Second, $\mu_I = \mu_D \mu^W$ grows linearly in μ^W, while μ grows more slowly (being a convex combination of μ_D and μ_I), so $(\mu_I/\mu)^{-\sigma}$ falls. Putting these together, each term of X declines, giving $dX/d\mu^W < 0$.

4. Overall sign in μ^W**.** Both pieces of $d \ln \mathcal{W}/d\mu^W$ are negative, so the derivative is negative and $\Xi(\mu^W)$ is strictly decreasing on $(1, \infty)$.

5. First-order dispersion result (efficient weights, varying Ω^***).** Define the *efficient* direct-trade share $\Omega^* \equiv A_D/(A_D + A_I)$ and the dispersion measure

$$\mathcal{V}^* \;\equiv\; \Omega^*(1 - \Omega^*)\left(\ln \mu_I - \ln \mu_D\right)^2 = \Omega^*(1 - \Omega^*)\left(\ln \mu^W\right)^2,$$

where $\mu_D = (\sigma/(\sigma - 1))^2$ and $\mu_I = \mu_D \, \mu^W$. Following Proposition 2, a standard quadratic (second-

order) approximation to $\ln \Xi$ around a baseline $(\bar{\mu}^W, \bar{\Omega}^*)$ takes the generic form

$$\ln \Xi(\mu^W, \Omega^*) \;\approx\; -\kappa_\sigma\, \Omega^*(1-\Omega^*)\big(\ln \mu^W\big)^2 \;=\; -\kappa_\sigma\, \mathcal{V}^*,$$

for some $\kappa_\sigma > 0$ that depends on σ and the baseline allocation (through the coefficients in Proposition 2). Taking *first-order differentials* of this approximation yields

$$d \ln \Xi \;\approx\; -\kappa_\sigma\, d\mathcal{V}^*.$$

Thus, to a first-order approximation (i.e. for small joint changes in μ^W and Ω^*), an *increase* in the markup-dispersion measure $\mathcal{V}^*$ reduces $\ln \Xi$ and therefore widens the welfare gap. This establishes the approximate monotonic relationship between the welfare gap and dispersion measured with efficient weights. □

B.8 Optimal Policy

A planner needs instruments that correct both the relative price distortion and congestion. The next proposition characterizes an optimal combination of a wholesale subsidy τ^W, size-dependent ad-posting taxes $\{\tau_v^M(z), \tau_m^M(z)\}$, a tax/subsidy on wholesalers' supplier search τ_s^M, entry taxes on firms τ_I^E, τ_F^E, and a tax/subsidy on wholesaler entry τ_W^E that decentralize the efficient allocation.

Proposition 4 (Optimal Policy)**.** *The optimal gross taxes and subsidies restoring the first-best from the decentralized equilibrium satisfy*

$$\tau^W = (\mu^{W^*})^{-1}\,,\quad \tau_v^M = \frac{z^{\sigma-1}}{z^{\sigma-1} + (\lambda_V - 1)\,\frac{V_Z}{V}}\,,\quad \tau_m^M = \frac{\sigma}{\sigma-1}\,\frac{z^{\sigma-1}}{z^{\sigma-1} + (\lambda_M - 1)\,\frac{M_Z}{M}}$$

$$\tau_s^M = \frac{1}{N_W{}^* \lambda_W}\,,\quad \tau_I^E = \frac{1 - \frac{\psi^*}{\beta}}{1 - \lambda_V\,\frac{\psi^*}{\beta} + \psi^*\,(\lambda_V - 1) - (1-\Omega^*)\left(\frac{S^*}{\theta^{W^*}}\right)^{1-\lambda_W}\lambda_W}$$

$$\tau_F^E = \frac{\sigma}{\sigma-1}\,\frac{1 - \frac{\psi^*}{\beta}}{1 - \lambda_M\,\frac{\psi^*}{\beta} + \psi^*\,(\lambda_M - 1)}\,,\qquad \tau_W^E = \frac{(\beta_W - 1)\,\frac{1}{N_W{}^*\lambda_W}}{\beta_W\left(\frac{2\lambda_W - 1}{\lambda_W}\right) - 1}$$

where

$$\Omega \equiv \frac{X_m}{X_m + X^W - X_m^W}\,,\qquad \psi \equiv \frac{X_m - X_m^W}{X_m + X^W - X_m^W}$$

$$V_Z \equiv \int_Z z^{\sigma-1}\, v(z)\, dz\,,\qquad M_Z \equiv \int_Z z^{\sigma-1}\, m(z)\, dz$$

and asterisks denote efficient allocations. Lump-sum transfers finance the instruments.

It is straightforward to verify that the planner's first order conditions and the optimality conditions in the decentralized equilibrium coincide when the taxes and subsidies satisfy the expressions in Proposition 4.

Wholesale subsidy. The optimal wholesale subsidy equals the inverse of the wholesale markup under efficient wholesaler entry, exactly offsetting double marginalization and realigning the relative price of direct and indirect inputs with relative marginal costs, eliminating misallocation on both the intensive and extensive margins.

Tax on direct matching. When $\lambda_M < 1$ (or $\lambda_V < 1$), congestion in matching requires a tax on ads. The tax is size dependent—it declines with productivity z—because each ad imposes the same congestion externality regardless of who posts it, leading to less productive firms crowding out more productive ones. τ_m^M includes an extra factor $\sigma/(\sigma-1)$ that corrects excessive ads induced by the two monopolistic markups of intermediates and finals, which inflate the marginal benefit of ads.

Tax on firm entry. Congestion also induces excessive entry, requiring an entry tax. Like τ_m^M, τ_F^E carries the factor $\sigma/(\sigma-1)$ to correct for double marginalization by intermediates and finals. The remaining term is the ratio between the marginal private benefit and the marginal social benefit to enter. Both private and social marginal benefits fall as the net direct share ψ^* rises: while higher ψ^* raises search costs relative to variable profits to firms, it raises search cost relative to the marginal increase in aggregate productivity to the social planner. However, the marginal social benefit falls more because the planner also internalizes congestion in matching. With $\beta > 1$, congestion unambiguously lowers the marginal social benefit of entry, so the optimal tax must fully internalize these externalities[44]

Tax/subsidy on wholesalers' supplier search. The optimal τ_s^M corrects two forces: misalignment between wholesalers' and the planner's gains from adding varieties, which calls for a subsidy of $1/N_W{}^*$, and congestion in searching, which calls for a tax of $1/\lambda_W$.[45] The net policy can be a tax or a subsidy depending on which force dominates.

[44]The optimal tax on buyer ads τ_v^M features the additional term $-(1-\Omega^*)\left(\frac{S^*}{\theta^{W*}}\right)^{1-\lambda_W}\lambda_W$ in the denominator, raising the tax to correct intermediate producers' failure to internalize that their entry raises wholesalers' search costs.

[45]Recall the discussion in Section 4: wholesaler are earning a share $\frac{1}{N_W}\frac{\sigma}{\sigma-1}\frac{\sigma}{\sigma-1}$ of the social production cost. With wholesale subsidy, this share increases to $\tau^W\frac{1}{N_W}\frac{\sigma}{\sigma-1}\frac{\sigma}{\sigma-1} = \frac{1}{N_W}\frac{\sigma}{\sigma}\frac{\sigma}{\sigma-1}$. Aligning with the socially optimal share of $\frac{1}{\sigma-1}$ therefore requires a subsidy equal to $\frac{1}{N_W{}^*}$.

Tax/subsidy on wholesaler entry. Like τ_I^E and τ_F^E, the wholesaler entry instrument τ_W^E equals the ratio of the marginal private benefit to the marginal social benefit of entry:

$$\text{MPB}(N_W) = (\beta_W - 1)\,\tau_s^M \times \text{Search Cost} \qquad \text{MSB}(N_W) = \left[\beta_W \left(\frac{2\lambda_W - 1}{\lambda_W}\right) - 1\right] \times \text{Search Cost}$$

The marginal private benefit of wholesale entry is the variable profit net of search cost, while the marginal social benefit is the marginal increase in aggregate productivity net of search cost as well as the cost of congestion when $\lambda_W < 1$.[46]

The subtle but critical insight is that while the tax/subsidy of wholesalers' supplier search τ_s^M is designed to correct inefficiencies in searching, it creates an indirect distortion in wholesaler entry. Specifically, *conditional on achieving the efficient search effort*, a subsidy (tax) on wholesalers' supplier search gives wholesalers an excessively low (high) marginal private benefit to enter. Without a corrective tax/subsidy on wholesaler entry, the equilibrium entry would be inefficiently low (high). This conditional perspective is key to understanding why the tax/subsidy on wholesalers' supplier search τ_s^M appears explicitly in the entry tax formula ($\frac{1}{N_W{}^* \lambda_W}$).

B.9 Hat-Algebra for Counterfactuals

To evaluate the counterfactual aggregate welfare, we need to know $\widehat{w_i}$, $\widehat{P_i^H}$, $\widehat{S_{ud}}$, $\widehat{N_{Ii}}$, $\widehat{N_{Fi}}$, and $\widehat{N_{Wi}}$, which can be obtained by rearranging the following system of equilibrium conditions in terms of hat variables:

$$\begin{aligned}
I_i =& \frac{\sigma}{\sigma - 1} \sum_{u \in \mathcal{N}} \left(X_{mui} + X_{ui}^W - X_{mui}^W\right) \\
I_i =& \frac{\sigma}{\sigma - 1} \sum_{d \in \mathcal{N}} \left(X_{mid} + X_{id}^W - X_{mid}^W\right) + \frac{1}{N_{Wi}(\sigma - 1)} \sum_{u \in \mathcal{N}} (X_{ui}^W - X_{mui}^W) - \sum_{d \in \mathcal{N}} \frac{1}{N_{Wd}(\sigma - 1)} (X_{id}^W - X_{mid}^W) \\
I_i =& w_i\, L_i \left[1 - \frac{(\beta_W - 1)(\sigma - 1)}{\beta_W\, \sigma^2} (1 - \Omega_i) \left(\frac{1}{N_{Wi}} - \frac{N_{Wi}}{\widetilde{N_{Wi}}^2}\right)\right]^{-1} \\
S_{ud} =& \left(\frac{X_{ud}^W - X_{mud}^W}{\sigma\, w_d\, f_{Wud}}\right)^{\frac{\lambda_W}{\beta_W}} N_{Iu}^{-\lambda_W}\, N_{Wd}^{\lambda_W (2 - \frac{2}{\beta_W} - \frac{1}{\lambda_W})} \\
N_{Fi} =& \frac{1}{\beta\sigma} (\beta - \psi_i) \frac{I_i}{F_{Fi\, w_i}} \\
N_{Ii} =& \frac{\sigma - 1}{\beta\sigma^2} \xi_i (\beta - \overline{\psi}_i^R) \frac{I_i}{F_{Ii\, w_i}} \\
N_{Wi} \leq& \left[\frac{(\beta_W - 1)(\sigma - 1)}{\beta_W \sigma^2} (1 - \Omega_i) \frac{I_i}{F_{Wi\, w_i}}\right]^{\frac{1}{2}}
\end{aligned}$$

where the first equation combines the household's budget constraint, labor market clearing condition, free entry conditions of firms and trade balance condition, and was derived in section **??**; the second equation is similar to the first one except the trade balance condition is not used; the third equation gives household's income as the sum of wage income and net profit from wholesalers; the fourth equation

[46] Specifically, the cost of congestion is a fraction $(1 - \lambda_W)/\lambda_W$ of the marginal increase in aggregate productivity.

is wholesalers' first order condition for search effort (34); the fifth equation comes from the free entry condition of final goods producers and their first order condition; the sixth equation comes from the free entry condition of intermediate goods producers and their first order condition; and the last equation comes from the free entry condition of wholesalers and their first order condition. The last three equations are derived in the appendix B.3. These can be rearranged in terms of hat variables:

$$\widehat{I_i} = \sum_{u\in N}\left[\frac{\frac{C_{2ui}}{\phi_{cui}}X_{mui}}{\sum_{l\in N}\left(\frac{C_{2li}}{\phi_{cli}}X_{mli}+X_{li}^W\right)}\widehat{\frac{C_{2ui}}{\phi_{cui}}}\,\widehat{X_{mui}}\right]+\sum_{u\in N}\left\{\left[\frac{X_{ui}^W-X_{mui}^W}{\sum_{l\in N}\left(\frac{C_{2li}}{\phi_{cli}}X_{mli}+X_{li}^W\right)}+\frac{\left(1-\frac{C_{2ui}}{\phi_{cui}}\right)X_{mui}}{\sum_{l\in N}\left(\frac{C_{2li}}{\phi_{cli}}X_{mli}+X_{li}^W\right)}\right]\widehat{X_{ui}^W}\right\}$$

$$\widehat{I_i} = \sum_{d\in N}\left[\frac{\frac{C_{2id}}{\phi_{cid}}X_{mid}}{\sum_{l\in N}\left(\frac{C_{2li}}{\phi_{cli}}X_{mli}+X_{li}^W\right)}\widehat{\frac{C_{2id}}{\phi_{cid}}}\,\widehat{X_{mid}}\right]+\sum_{d\in N}\left\{\left[\frac{X_{id}^W-X_{mid}^W}{\sum_{l\in N}\left(\frac{C_{2li}}{\phi_{cli}}X_{mli}+X_{li}^W\right)}+\frac{\left(1-\frac{C_{2id}}{\phi_{cid}}\right)X_{mid}}{\sum_{l\in N}\left(\frac{C_{2li}}{\phi_{cli}}X_{mli}+X_{li}^W\right)}\right]\widehat{X_{id}^W}\right\}$$

$$+\sum_{u\in N}\left\{\frac{1}{N_{Wi}\,\sigma}\left[\frac{X_{ui}^W-X_{mui}^W}{\sum_{l\in N}\left(\frac{C_{2li}}{\phi_{cli}}X_{mli}+X_{li}^W\right)}+\frac{\left(1-\frac{C_{2ui}}{\phi_{cui}}\right)X_{mui}}{\sum_{l\in N}\left(\frac{C_{2li}}{\phi_{cli}}X_{mli}+X_{li}^W\right)}\right]\frac{\widehat{X_{ui}^W}}{\widehat{N_{Wi}}}\right\}$$

$$-\sum_{u\in N}\left[\frac{1}{N_{Wi}\,\sigma}\frac{\left(1-\frac{C_{2ui}}{\phi_{cui}}\right)X_{mui}}{\sum_{l\in N}\left(\frac{C_{2li}}{\phi_{cli}}X_{mli}+X_{li}^W\right)}\widehat{\mu_i^W}^{1-\sigma}\,\frac{\widehat{S_{ui}}}{\widehat{\phi_{cui}}}\,\frac{\widehat{X_{mui}}}{\widehat{N_{Wi}}}\right]$$

$$-\sum_{d\in N}\left\{\frac{1}{N_{Wd}\,\sigma}\left[\frac{X_{id}^W-X_{mid}^W}{\sum_{l\in N}\left(\frac{C_{2li}}{\phi_{cli}}X_{mli}+X_{li}^W\right)}+\frac{\left(1-\frac{C_{2id}}{\phi_{cid}}\right)X_{mid}}{\sum_{l\in N}\left(\frac{C_{2li}}{\phi_{cli}}X_{mli}+X_{li}^W\right)}\right]\frac{\widehat{X_{id}^W}}{\widehat{N_{Wd}}}\right\}$$

$$+\sum_{d\in N}\left[\frac{1}{N_{Wd}\,\sigma}\frac{\left(1-\frac{C_{2id}}{\phi_{cid}}\right)X_{mid}}{\sum_{l\in N}\left(\frac{C_{2li}}{\phi_{cli}}X_{mli}+X_{li}^W\right)}\widehat{\mu_d^W}^{1-\sigma}\,\frac{\widehat{S_{id}}}{\widehat{\phi_{cid}}}\,\frac{\widehat{X_{mid}}}{\widehat{N_{Wd}}}\right]$$

$$\widehat{I_i} = \widehat{w_i}\,\widehat{L_i}\left\{\left[1-\frac{(\beta_W-1)(\sigma-1)}{\beta_W\,\sigma^2}(1-\Omega_i)\left(\frac{1}{N_{Wi}}-\frac{N_{Wi}}{\widehat{N_{Wi}}^2}\right)\right]^{-1}\right.$$

$$\left.-\frac{\frac{(\beta_W-1)(\sigma-1)}{\beta_W\,\sigma^2}(1-\Omega_i)\left(\frac{1}{N_{Wi}}-\frac{N_{Wi}}{\widehat{N_{Wi}}^2}\right)}{\left[1-\frac{(\beta_W-1)(\sigma-1)}{\beta_W\,\sigma^2}(1-\Omega_i)\left(\frac{1}{N_{Wi}}-\frac{N_{Wi}}{\widehat{N_{Wi}}^2}\right)\right]}\left(\frac{1}{1-\Omega_i}-\frac{\Omega_i}{1-\Omega_i}\widehat{\Omega_i}\right)\left(\frac{N_{Wi}^{-1}}{N_{Wi}^{-1}-\frac{N_{Wi}}{\widehat{N_{Wi}}^2}}\widehat{N_{Wi}}^{-1}-\frac{\frac{N_{Wi}}{\widehat{N_{Wi}}^2}}{N_{Wi}^{-1}-\frac{N_{Wi}}{\widehat{N_{Wi}}^2}}\frac{\widehat{N_{Wi}}}{\widehat{N_{Wi}}^2}\right)\right\}^{-1}$$

$$\widehat{S_{ud}} = (\widehat{w_d}\,\widehat{f_{Wud}})^{-\frac{\lambda_W}{\beta_W}}\,\widehat{N_{Iu}}^{-\lambda_W}\,\widehat{N_{Wd}}^{\lambda_W(2-\frac{2}{\beta_W}-\frac{1}{\lambda_W})}\left(\widehat{X_{ud}^W-X_{mud}^W}\right)^{\frac{\lambda_W}{\beta_W}}$$

$$\widehat{N_{Fi}} = \left(\frac{\beta}{\beta-\psi_i}-\frac{\psi_i}{\beta-\psi_i}\widehat{\psi_i}\right)\frac{\widehat{I_i}}{\widehat{w_i}\,\widehat{F_{Fi}}}$$

$$\widehat{N_{Ii}} = \widehat{\xi_i}\left(\frac{\beta}{\beta-\overline{\psi_i^R}}-\frac{\overline{\psi_i^R}}{\beta-\overline{\psi_i^R}}\widehat{\psi_i^R}\right)\frac{\widehat{I_i}}{\widehat{w_i}\,\widehat{F_{Ii}}}$$

$$\widehat{N_{Wi}} \le \left[\left(\frac{1}{1-\Omega_i}-\frac{\Omega_i}{1-\Omega_i}\widehat{\Omega_i}\right)\frac{\widehat{I_i}}{\widehat{w_i}\,\widehat{F_{Wi}}}\right]^{\frac{1}{2}}$$

where

$$\widehat{\psi_i} = \left\{ \sum_{u \in N} \left[\frac{\frac{C_{2ui}}{\phi_{cui}} X_{mui}}{\sum_{l \in N} \left(\frac{C_{2li}}{\phi_{cli}} X_{mli} \right)} \frac{\widehat{C_{2ui}}}{\widehat{\phi_{cui}}} \widehat{X_{mui}} \right] \right\}$$

$$\left\{ \sum_{u \in N} \left[\frac{\frac{C_{2ui}}{\phi_{cui}} X_{mui}}{\sum_{l \in N} \left(\frac{C_{2li}}{\phi_{cli}} X_{mli} + X_{li}^W \right)} \frac{\widehat{C_{2ui}}}{\widehat{\phi_{cui}}} \widehat{X_{mui}} \right] + \sum_{u \in N} \left[\left(\frac{X_{ui}^W - X_{mui}^W}{\sum_{l \in N} \left(\frac{C_{2li}}{\phi_{cli}} X_{mli} + X_{li}^W \right)} + \frac{\left(1 - \frac{C_{2ui}}{\phi_{cui}}\right) X_{mui}}{\sum_{l \in N} \left(\frac{C_{2li}}{\phi_{cli}} X_{mli} + X_{li}^W \right)} \right) \widehat{X_{ui}^W} \right] \right\}^{-1}$$

$$\widehat{\psi_i^R} = \left\{ \sum_{d \in N} \left[\frac{C_{1id} X_{mid}}{\sum_{l \in N} (C_{1il} X_{mil})} \widehat{C_{1id}} \widehat{X_{mid}} \right] \right\}$$

$$\left\{ \sum_{d \in N} \left[\frac{C_{1id} X_{mid}}{\sum_{l \in N} (C_{1il} X_{mil} + X_{Wil})} \widehat{C_{1id}} \widehat{X_{mid}} \right] + \sum_{d \in N} \left[\left(\frac{X_{Wid} - X_{Wmid}}{\sum_{l \in N} (C_{1il} X_{mil} + X_{Wil})} + \frac{(1 - C_{1id}) X_{mid}}{\sum_{l \in N} (C_{1il} X_{mil} + X_{Wil})} \right) \widehat{X_{Wid}} \right] \right\}^{-1}$$

$$\widehat{\xi_i} = (\mu_i^W \widehat{\mu_i^W})^{-1} \xi_i^{-1} + \left[\frac{1}{N_{Wi} \sigma} \frac{\sum_{u \in N} X_{mui}}{\sum_{u \in N} \left(X_{mui} + X_{ui}^W - X_{mui}^W \right)} \xi_i^{-1} \frac{1}{\widehat{N_{Wi}}} \sum_{u \in N} \left(\frac{X_{mui}}{\sum_{l \in N} X_{mli}} \widehat{X_{mui}} \right) \right.$$

$$\left. \left\{ \sum_{u \in N} \left[\frac{\frac{C_{2ui}}{\phi_{cui}} X_{mui}}{\sum_{l \in N} \left(\frac{C_{2li}}{\phi_{cli}} X_{mli} + X_{li}^W \right)} \frac{\widehat{C_{2ui}}}{\widehat{\phi_{cui}}} \widehat{X_{mui}} \right] + \sum_{u \in N} \left[\left(\frac{X_{ui}^W - X_{mui}^W}{\sum_{l \in N} \left(\frac{C_{2li}}{\phi_{cli}} X_{mli} + X_{li}^W \right)} + \frac{\left(1 - \frac{C_{2ui}}{\phi_{cui}}\right) X_{mui}}{\sum_{l \in N} \left(\frac{C_{2li}}{\phi_{cli}} X_{mli} + X_{li}^W \right)} \right) \widehat{X_{ui}^W} \right] \right\}^{-1} \right]$$

$$\widehat{\Omega_i} = \left\{ \sum_{u \in N} \left[\frac{X_{mui}}{\sum_{l \in N} (X_{mli})} \widehat{X_{mui}} \right] \right\}$$

$$\left\{ \sum_{u \in N} \left[\frac{\frac{C_{2ui}}{\phi_{cui}} X_{mui}}{\sum_{l \in N} \left(\frac{C_{2li}}{\phi_{cli}} X_{mli} + X_{li}^W \right)} \frac{\widehat{C_{2ui}}}{\widehat{\phi_{cui}}} \widehat{X_{mui}} \right] + \sum_{u \in N} \left[\left(\frac{X_{ui}^W - X_{mui}^W}{\sum_{l \in N} \left(\frac{C_{2li}}{\phi_{cli}} X_{mli} + X_{li}^W \right)} + \frac{\left(1 - \frac{C_{2ui}}{\phi_{cui}}\right) X_{mui}}{\sum_{l \in N} \left(\frac{C_{2li}}{\phi_{cli}} X_{mli} + X_{li}^W \right)} \right) \widehat{X_{ui}^W} \right] \right\}^{-1}$$

$$\widehat{X_{ud}^W - X_{mud}^W} = \left[1 + \frac{\left(\mu_d^W\right)^{1-\sigma} S_{ud} \phi_{cud}^{-1} X_{mud}}{X_{ud}^W - X_{mud}^W} \right] \widehat{X_{ud}^W} - \frac{\left(\mu_d^W\right)^{1-\sigma} S_{ud} \phi_{cud}^{-1} X_{mud}}{X_{ud}^W - X_{mud}^W} \left(\widehat{\mu_d^W}\right)^{1-\sigma} \widehat{S_{ud}} \widehat{\phi_{cud}}^{-1} \widehat{X_{mud}}$$

$$\widehat{X_{mud}} = \widehat{K_{mud}} \widehat{\theta_{ud}^m}^{\frac{\beta}{\beta-1}} \widehat{N_{Fd}} \widehat{w_d}^{\frac{1}{1-\beta}} I_d^{\frac{\beta}{\beta-1}} \widehat{P_d^H}^{\gamma} \widehat{w_u}^{-\gamma} \widehat{C_{2ud}}^{\frac{1}{\beta-1}}$$

$$\widehat{X_{ud}^W} = \left(\widehat{\mu_d^W}\right)^{1-\sigma} \widehat{w_u}^{1-\sigma} \widehat{\tau_{ud}}^{1-\sigma} \widehat{S_{ud}} \widehat{N_{Iu}} \widehat{N_{Fd}} \widehat{I_d} \left(\widehat{P_d^H}\right)^{\sigma-1}$$

$$\widehat{\theta_{ud}^m} = \left\{ \widehat{\kappa_{ud}} \widehat{N_{Iu}}^{\frac{\beta-1}{\beta} \lambda_V} \widehat{w_u}^{-\left[\frac{1}{\beta} + \frac{\gamma}{\beta}(\lambda_V + \lambda_M - 1)\right]} \widehat{\tau_{ud}}^{-\frac{\gamma}{\beta}(\lambda_V + \lambda_M - 1)} \widehat{C_{1ud}}^{\frac{\lambda_V}{\beta}} \widehat{C_{2ud}}^{\left[\frac{\lambda_V}{\beta(\beta-1)} + \frac{\lambda_M - 1}{\beta - 1}\right]} \widehat{f_{vud}}^{-\frac{\lambda_V}{\beta}} \widehat{f_{mud}}^{-\left[\frac{\lambda_V}{\beta(\beta-1)} + \frac{\lambda_M - 1}{\beta - 1}\right]} \right.$$

$$\left. \widehat{\phi_{cud}}^{\frac{\lambda_V}{\beta}} \widehat{N_{Fd}}^{\left(\frac{\lambda_V}{\beta} + \lambda_M - 1\right)} \widehat{w_d}^{\frac{-\lambda_V}{\beta(\beta-1)} - \frac{\lambda_M - 1}{\beta - 1}} \widehat{I_d}^{\frac{\lambda_V}{\beta-1} + \frac{\lambda_M - 1}{\beta - 1}} \widehat{L_d}^{\left(\frac{\lambda_V}{\beta-1} - \frac{\lambda_M - 1}{\beta - 1}\right)} \widehat{P_d^H}^{\frac{\gamma}{\beta}(\lambda_V + \lambda_M - 1)} \right\}^{\frac{\beta-1}{\beta - \lambda_V - \lambda_M}}$$

$$\widehat{K_{mud}} = \widehat{\tau_{ud}}^{-\gamma} \widehat{\phi_{cud}} \widehat{f_{mud}}^{\frac{1}{1-\beta}}$$

$$\widehat{C_{1ud}} = C_{1ud}^{-1} - C_{1ud}^{-1} \left(\mu_d^W\right)^{-\sigma} S_{ud} \phi_{cud}^{-1} \left(\widehat{\mu_d^W}\right)^{-\sigma} \widehat{S_{ud}} \widehat{\phi_{cud}}^{-1}$$

$$\widehat{C_{2ud}} = C_{2ud}^{-1} \phi_{cud} \widehat{\phi_{cud}} - C_{2ud}^{-1} \left(\mu_d^W\right)^{1-\sigma} S_{ud} \left(\widehat{\mu_d^W}\right)^{1-\sigma} \widehat{S_{ud}}$$

$$\widehat{\mu_i^W} = \frac{\widehat{N_{Wi}} (N_{Wi} \sigma - 1)}{\widehat{N_{Wi}} N_{Wi} \sigma - 1}$$

$$\widehat{N_{Wi}} \equiv \left[\frac{(\beta_W - 1)(\sigma - 1)}{\beta_W \sigma^2} (1 - \Omega_i) \frac{L_i}{F_{Wi} w_i} \right]^{\frac{1}{2}}$$

$$\widehat{\widehat{N_{Wi}}} = \left[\left(\frac{1}{1 - \Omega_i} - \frac{\Omega_i}{1 - \Omega_i} \widehat{\Omega_i} \right) \frac{\widehat{L_i}}{\widehat{F_{Wi}}} \right]^{\frac{1}{2}}$$

B.10 Gravity Equation

The total revenue of intermediate goods producers from direct sales is:

$$\begin{aligned}
X_{mud} =& N_{Iu} \int_Z x_{mud}(z) j_u(z) dz \\
=& N_{Iu} \int_Z \left[\left(\frac{\sigma}{\sigma - 1} w_u \tau_{ud} \right)^{\beta(1-\sigma)} \left(\frac{C_{1ud}}{w_u f_{vud} \sigma} \right) \theta_{ud}^{v\,\beta} D_{mud}^{\beta} \right]^{\frac{1}{\beta-1}} z^{\gamma} j_u(z) dz \\
=& N_{Iu} \mathbb{E}_u [z^{\gamma}] \left[\left(\frac{\sigma}{\sigma - 1} w_u \tau_{ud} \right)^{\beta(1-\sigma)} \left(\frac{C_{1ud}}{w_u f_{vud} \sigma} \right) \theta_{ud}^{v\,\beta} D_{mud}^{\beta} \right]^{\frac{1}{\beta-1}}
\end{aligned}$$

where $C_{1ud} \equiv 1 - \mu_d^{W-\sigma} S_{ud} \phi_{cud}^{-1}$ summarizes the marginal increase in variable profit for a intermediate goods producer, due to an increase in demand associated with an additional direct supplier.

Now we can substitute out D_{mud} (which can be solved for by combining the spatial counterparts of (10), (49), and (16)):

$$\begin{aligned}
X_{mud} =& N_{Iu} \mathbb{E}_u [z^{\gamma}] \left[\left(\frac{\sigma}{\sigma - 1} w_u \tau_{ud} \right)^{\beta(1-\sigma)} \left(\frac{C_{1ud}}{w_u f_{vud} \sigma} \right) \right]^{\frac{1}{\beta-1}} \theta_{ud}^{m\,\frac{\beta}{\beta-1}} \phi_{cud} \\
& \left(\frac{N_{Fd}}{N_{Iu}} \right) \left(\frac{\sigma - 1}{\sigma} \right)^{1+\gamma} \left(\frac{f_{vud}}{f_{mud}} \right)^{\frac{1}{\beta-1}} \left(\frac{C_{2ud}}{C_{1ud}} \right)^{\frac{1}{\beta-1}} \left(\frac{w_u}{w_d} \right)^{\frac{1}{\beta-1}} D_{Hd}^{\frac{\beta}{\beta-1}} \frac{\mathbb{E}_d [z^{\gamma}]}{\mathbb{E}_u [z^{\gamma}]} \left(\frac{\mathbb{E}_u [z^{\gamma}]}{\mathbb{E}_u \left[z^{\frac{\gamma}{\beta}} \right]} \right)^{\frac{\beta}{\beta-1}}
\end{aligned} \tag{57}$$

where $C_{2ud} \equiv \phi_{cud} - \mu_d^{W\,1-\sigma} S_{ud}$ summarizes the marginal increase in variable profit for a final goods producer, due to a reduction in unit cost associated with an additional direct supplier.

We can also rewrite X_{mud} as a product of the number of direct matches $\widetilde{M}_{ud}$ and the average trade flow. To this end, we first derive $\widetilde{M}_{ud}$ using the definition of θ_{ud}^m:

$$\begin{aligned}
\widetilde{M}_{ud} =& \theta_{ud}^m M_{ud} \\
=& \theta_{ud}^m Fd \left[\frac{1}{w_d f_{mud} \sigma} \left(\frac{\sigma}{\sigma - 1} \right)^{1-\sigma} D_{Hd} \left(\theta_{ud}^m \phi_{cud} c_{mud}^{1-\sigma} - S_{ud} \theta_{ud}^m \mu^{W\,1-\sigma} c_{mud}^{1-\sigma} \right) \right]^{\frac{1}{\beta-1}} \mathbb{E}_d \left[z^{\frac{\gamma}{\beta}} \right] \\
=& \theta_{ud}^{m\,\frac{\beta}{\beta-1}} Fd \left[\frac{1}{w_d f_{mud} \sigma} \left(\frac{\sigma}{\sigma - 1} \right)^{1-\sigma} D_{Hd} \right]^{\frac{1}{\beta-1}} \mathbb{E}_d \left[z^{\frac{\gamma}{\beta}} \right] \\
& \left[C_{2ud} \left(\frac{\sigma}{\sigma - 1} w_u \tau_{ud} \right)^{1-\sigma} \frac{\mathbb{E}_u [z^{\gamma}]}{\mathbb{E}_u \left[z^{\frac{\gamma}{\beta}} \right]} \right]^{\frac{1}{\beta-1}}
\end{aligned}$$

where the second line makes use of equation (30). It can then be shown that:

$$X_{mud} = \widetilde{M}_{ud} \left(\frac{\sigma}{\sigma - 1} \right)^{1-2\sigma} w_u^{1-\sigma} \tau_{ud}^{1-\sigma} \phi_{cud} w_d L_d P_d^{H\,\sigma-1} \frac{\mathbb{E}_d [z^{\gamma}]}{\mathbb{E}_d \left[z^{\frac{\gamma}{\beta}} \right]} \frac{\mathbb{E}_u [z^{\gamma}]}{\mathbb{E}_u \left[z^{\frac{\gamma}{\beta}} \right]} \tag{58}$$

Location d wholesalers' total sales of location u intermediate goods ($X^W_{ud} - X^W_{mud}$) is:

$$\begin{aligned}
&X^W_{ud} - X^W_{mud} \\
=&\mu^W_d \, X_{Wud} - \mu^W_d \, X_{Wmud} \\
=&\mu^W_d \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} (w_u \, \tau_{ud})^{1-\sigma} \, S_{ud} \, D_{Wud} \, N_{Iu} \, \mathbb{E}_u \left[z^{\sigma-1}\right] - \mu_d^{W\,1-\sigma} \, S_{ud} \, \phi_{cud}{}^{-1} \, X_{mud}
\end{aligned}$$

Now,

$$\begin{aligned}
D_{Wud} =&\mu_d^{W-\sigma} \, \frac{\sigma-1}{\sigma} \, N_{Fd} \int_Z \frac{x_{Hd}(z')}{c_d(z')^{1-\sigma}} j_d(z') dz' \\
=&\mu_d^{W-\sigma} \, \frac{\sigma-1}{\sigma} \, N_{Fd} \int_Z \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} z'^{\sigma-1} D_{Hi} j_d(z') dz' \\
=&\mu_d^{W-\sigma} \left(\frac{\sigma}{\sigma-1}\right)^{-\sigma} N_{Fd} \, D_{Hd} \, \mathbb{E}_d \left[z^{\sigma-1}\right]
\end{aligned}$$

where the first and second lines make use of the spatial counterparts of (50) and (11). Substitute D_{Wud} out then yields:

$$\begin{aligned}
X^W_{ud} - X^W_{mud} =&\mu_d^{W\,1-\sigma} \left(\frac{\sigma}{\sigma-1}\right)^{1-2\sigma} (w_u \, \tau_{ud})^{1-\sigma} \, S_{ud} \, D_{Hd} \, N_{Fd} \, N_{Iu} \, \mathbb{E}_d \left[z^{\sigma-1}\right] \, \mathbb{E}_u \left[z^{\sigma-1}\right] \\
&- \mu_d^{W\,1-\sigma} \, S_{ud} \, \phi_{cud}{}^{-1} \, X_{mud}
\end{aligned} \tag{59}$$

B.11 Sufficient Statistics for Welfare Change

In this section, I analyze how the welfare of each location i responds to exogenous shocks. Specifically, I focus on shocks to the efficiency and productivity of direct matching, κ_{ud} and ϕ_{cud}, while assuming no shocks occur in location i itself to simplify the exposition. This focus is motivated by the idea that recent technological progress may have improved direct matching technologies and, in doing so, reduced the relevance of wholesale intermediation. The key objective of this analysis is to understand the welfare implications of disintermediation. Following the approach of Arkolakis, Costinot and Rodríguez-Clare (2012), I express welfare changes in terms of a set of sufficient statistics, summarized in the following proposition:

Proposition 5. *For any exogenous shocks to $\{\kappa_{ud}\}$ and $\{\phi_{cud}\}$ satisfying $\widehat{\kappa_{ii}} = 1$ and $\widehat{\phi_{cud}} = 1$, the change in location i welfare is:*

$$\widehat{\frac{I_i}{P_i^H}} = \widehat{T_i} \left(\frac{\widehat{\Omega_{ii}}}{\widehat{N_{Fi}} \, \widehat{M_{ii}}} \, \widehat{\Lambda_{ii}}\right)^{\frac{1}{1-\sigma}} \tag{60}$$

where $\Omega_{ud} \equiv \frac{X_{mud}}{X_{ud}}$, $\Lambda_{ud} \equiv \frac{X_{ud}}{\sum_{l \in \mathcal{N}} X_{ld}}$, $\overline{M}_{ud} \equiv \frac{\widetilde{M}_{ud}}{N_{Fd}}$, $X_{ud} \equiv X_{mud} + X_{ud}^W - X_{mud}^W$

$$I_i = T_i \, w_i \, L_i \,, \qquad T_i \equiv \left[1 - \frac{(\beta_W - 1)\,(\sigma - 1)}{\beta_W \, \sigma^2} \, (1 - \Omega_i) \left(\frac{1}{N_{Wi}} - \frac{N_{Wi}}{\widetilde{N_{Wi}}^2} \right) \right]^{-1}$$

Proof. First, use hat to denote the proportional change of any variable: $\widehat{x} = \frac{x'}{x}$, where x' is the value of x after shock. Also, define the following weights:

$$\Lambda_{ud} \equiv \frac{X_{mud} + X_{ud}^W - X_{mud}^W}{\sum_{l \in \mathcal{N}} \left(X_{mld} + X_{ld}^W - X_{mld}^W \right)}$$

is the share of location d's expenditure on intermediate goods accounted for by location u.

$$\Omega_{ud} \equiv \frac{X_{mud}}{X_{mud} + X_{mud}^W - X_{mud}^W}$$

is the share of direct trade in location d's purchases of intermediate goods from location u.

Rewrite the final consumption price index in terms of the above shares:

$$\begin{aligned}
(P_i^H)^{1-\sigma} =& N_{Fi} \int_Z \left(\frac{\sigma}{\sigma - 1} \right)^{1-\sigma} \left\{ \sum_{u \in \mathbb{N}} \left[m_{ui}(z) \, \theta_{ui}^m \, c_{mui}^{1-\sigma} + S_{ui} \, N_{Iu} \, c_{Wui}^{1-\sigma} - m_{ui}(z) \, \theta_{ui}^m \, S_{ui} \, {\mu_i^W}^{1-\sigma} \, c_{mui}^{1-\sigma} \right] \right\} z^{\sigma-1} \, j_i(z) dz \\
(P_i^H)^{1-\sigma} =& N_{Fi} \left(\frac{\sigma}{\sigma - 1} \right)^{1-\sigma} \left\{ \int_Z \left[m_{ii}(z) \, \theta_{ii}^m \, c_{mii}^{1-\sigma} + S_{ii} \, N_{Ii} \, c_{Wii}^{1-\sigma} - m_{ii}(z) \, \theta_{ii}^m S_{ii} \, {\mu_i^W}^{1-\sigma} \, c_{mii}^{1-\sigma} \right] z^{\sigma-1} \, j_i(z) dz \right\} \Lambda_{ii}^{-1} \\
(P_i^H)^{1-\sigma} =& N_{Fi} \left(\frac{\sigma}{\sigma - 1} \right)^{1-\sigma} \int_Z m_{ii}(z) \, \theta_{ii}^m \, c_{mii}^{1-\sigma} \, z^{\sigma-1} \, j_i(z) dz \, \Omega_{ii}^{-1} \, \Lambda_{ii}^{-1} \\
(P_i^H)^{1-\sigma} =& N_{Fi} \left(\frac{\sigma}{\sigma - 1} \right)^{1-\sigma} \Pi_{mii} \, \theta_{ii}^m \, c_{mii}^{1-\sigma} \, \mathbb{E}_i \left[z^\gamma \right] \, \Omega_{ii}^{-1} \, \Lambda_{ii}^{-1} \\
\widehat{(P_i^H)^{1-\sigma}} =& \widehat{N_{Fi}} \, \widehat{\Pi_{mii}} \, \widehat{\theta_{ii}^m} \, \widehat{c_{mii}}^{1-\sigma} \, \widehat{\Omega_{ii}}^{-1} \, \widehat{\Lambda_{ii}}^{-1}
\end{aligned}$$

Now, recall the definition of θ_{ii}^m:

$$\begin{aligned}
\widetilde{M}_{ii} =& \theta_{ii}^m \, M_{ii} \\
\widehat{\widetilde{M}_{ii}} =& \widehat{\theta_{ii}^m} \, \widehat{M_{ii}} = \widehat{\theta_{ii}^m} \, \widehat{N_{Fi}} \, \widehat{\Pi_{mii}} = \widehat{N_{Fi}} \, \widehat{\overline{M}_{ii}} \\
\widehat{\overline{M}_{ii}} \equiv& \widehat{\theta_{ii}^m} \, \widehat{\Pi_{mii}}
\end{aligned}$$

Also, from the spatial counterpart of (10) we know that

$$\widehat{c_{mii}}^{1-\sigma} = \widehat{w_i}^{1-\sigma}$$

Therefore,

$$\widehat{(P_i^H)^{1-\sigma}} = \widehat{N_{Fi}}\,\widehat{\overline{M}_{ii}}\,\widehat{w_i}^{1-\sigma}\,\widehat{\Omega_{ii}}^{-1}\,\widehat{\Lambda_{ii}}^{-1}$$

$$\widehat{\frac{I_i}{P_i^H}} = \widehat{T_i}\,\widehat{\frac{w_i}{P_i^H}} = \widehat{T_i}\left(\frac{\widehat{\Omega_{ii}}}{\widehat{N_{Fi}}\,\widehat{\overline{M}_{ii}}}\,\widehat{\Lambda_{ii}}\right)^{\frac{1}{1-\sigma}}$$

□

Similar to Arkolakis, Huneeus and Miyauchi (2023), this welfare change expression departs from that of Arkolakis, Costinot and Rodríguez-Clare (2012) as the number of direct matches within location i might change when the formation of production network is endogenous. As argued by Arkolakis, Huneeus and Miyauchi (2023), $\widehat{\overline{M}_{ii}}$ appears in the expression as it affects the aggregate productivity of final goods producers in location i through a love-of-variety effect.

However, the welfare change expression above further departs from that of Arkolakis, Huneeus and Miyauchi (2023) for two reasons. First, even without within location shocks, the change in the total number of direct matches depends not only on the change in the average number of matches per firm, $\widehat{\overline{M}_{ii}}$, but also on the change in the number of final goods producers, $\widehat{N_{Fi}}$. Specifically, using the free entry condition, we can derive the following relationship[47]:

$$\widehat{N_{Fi}} = \frac{\beta}{\beta - \psi_i} - \frac{\psi_i}{\beta - \psi_i}\widehat{\psi_i} \tag{61}$$

where

$$\psi_i \equiv \frac{\sum_{u \in \mathcal{N}} \left(X_{mud} - X_{mud}^W\right)}{\sum_{u \in \mathcal{N}} \left(X_{mud} + X_{ud}^W - X_{mud}^W\right)}$$

I refer to ψ_i as the net direct purchase share, which represents the share of location i's final goods producers' total intermediate input purchases accounted for by direct trade, net of cannibalized indirect purchases. Equation (61) states that an increase in the net direct purchase share for location i leads to a decline in the number of final goods producers. The intuition is that a higher net direct purchase share raises the total fixed costs of searching for suppliers relative to revenue, thereby reducing net profit margins. This, in turn, increases entry barriers and discourages the entry of new final goods producers. As a result, the model predicts a decline in the number of final goods producers following disintermediation, which dampens the welfare gains from improvements in direct matching technology.

Second, the welfare change also depends on $\widehat{\Omega_{ii}}$, which represents the change in the share of location i's domestic intermediate goods expenditure accounted for by direct trade. This term appears in the welfare change expression because changes in the number of direct matches, $\widehat{\widetilde{M}_{ii}}$, do not fully capture

[47] A detailed derivation is provided in Appendix B.3

the endogenous network formation effects of shocks on welfare. Intuitively, a change in $\widetilde{M}_{ii}$ does not necessarily lead to a proportional change in aggregate productivity—and thus in the price index—for several reasons that are reflected in $\widehat{\Omega_{ii}}$. First, direct trade accounts for only a fraction of total expenditure on domestic intermediate goods. Therefore, the impact of changes in $\widetilde{M}_{ii}$ on the price index must be scaled accordingly. For example, holding everything else constant, a given $\widehat{\widetilde{M}_{ii}}$ leads to a smaller $\widehat{\Omega_{ii}}$ when Ω_{ii} is large, implying that welfare gains are larger when direct trade already plays a greater role. Second, increases in $\widetilde{M}_{ii}$ may cannibalize indirect trade, raising the direct trade share even if S_{ii} increases proportionally with $\widetilde{M}_{ii}$. This substitution dampens welfare gains, since replacing existing indirect matches with direct ones yields less benefit than forming entirely new matches. Third, wholesalers may adjust their search effort decisions in response to shocks, thereby changing the set of varieties available through indirect trade. As a result, even absent cannibalization, the direct trade share may shift if the evolution of wholesalers' product offerings diverges from the change in direct matches. Lastly, wholesaler markups may respond endogenously to shocks if they influence the number of wholesalers, thereby affecting the prices of indirectly traded inputs.

Taken together, if shocks lead to disintermediation in location i's domestic trade—i.e. $\widehat{\Omega_{ii}} > 1$—the resulting decline in the importance of indirect trade dampens welfare gains.

Despite these differences, the welfare change expression in (60) shares the same exponent as that in Arkolakis, Costinot and Rodríguez-Clare (2012). However, the inverse of this exponent differs from the trade elasticity (i.e., the elasticity of total trade flow with respect to iceberg trade costs) in the model due to the endogenous formation of production networks. This suggests that corrections must be applied to the trade elasticity estimated using the gravity equation to avoid biasing the inference of this exponent. In particular, such corrections require knowledge of the share of direct trade, as direct and indirect trade flows have different trade elasticities. The aggregate trade elasticity is a weighted average of the two, where the weights are determined by their respective shares in total trade. The detailed derivation of the trade elasticity is discussed in Appendix B.10.

B.12 Aggregate Welfare Change

While the sufficient statistic for welfare changes presented in the previous section offers a valuable tool for potentially measuring welfare using only observed data, it is limited in that it evaluates welfare at the level of individual locations and does not fully unpack the distinct channels through which welfare is affected. To address this limitation, this section derives the first-order effects of exogenous shocks on aggregate welfare, following the approach of Hulten (1978), as a complement to the sufficient statistics framework. I also compare these results with those of Arkolakis, Huneeus and Miyauchi (2023), highlighting how endogenizing the share of wholesale trade and the market structure of the wholesale sector leads to key

differences in the welfare implications.

Similar to the previous section, I consider shocks to the direct matching efficiency $\{\kappa_{ud}\}$ and the productivity gains from customization $\{\phi_{cud}\}$, this time without imposing any restrictions on the shocks. Following Baqaee and Farhi (2019) and Arkolakis, Huneeus and Miyauchi (2023), I define the change in aggregate welfare as follows:

$$\begin{aligned} d\log \mathcal{W} &\equiv \sum_i I_i \left(d\log I_i - d\log P_i^H\right) \\ &= -\sum_i I_i \, d\log P_i^H) \end{aligned}$$

where the second equality follows from the choice of setting the aggregate nominal GDP as numeraire:

$$\sum_i I_i \, d\log I_i = d\left(\sum_i I_i\right) = 0$$

Now, using equation (54) and the final goods producers' optimal pricing decision—which implies that their total cost of production is a fraction $\frac{\sigma-1}{\sigma}$ of their sales—we obtain:

$$I_i = \frac{\sigma}{\sigma-1} \sum_{u\in\mathcal{N}} \left(X_{mui} + X_{ui}^W - X_{mui}^W\right) \tag{62}$$

We can then log-linearize equation (62) to derive the following decomposition:

Proposition 6. *Suppose there are shocks to the direct matching efficiency $\{\kappa_{ud}\}$ and to the productivity gains from customization $\{\phi_{cud}\}$. The first-order effect on aggregate welfare is:*

$$\begin{aligned} d\log \mathcal{W} = &\underbrace{\frac{1}{\sigma-1}\sum_{u\in\mathcal{N}}\sum_{d\in\mathcal{N}} \frac{\sigma}{\sigma-1} X_{mud} \, d\log \phi_{cud}}_{\textit{technological effect}} + \underbrace{\frac{1}{\sigma-1}\sum_{u\in\mathcal{N}}\sum_{d\in\mathcal{N}} \frac{\sigma}{\sigma-1} X_{mud} \, d\log \overline{M}_{ud}}_{\textit{direct match effect}} \\ &+ \underbrace{\frac{1}{\sigma-1}\sum_{u\in\mathcal{N}}\sum_{d\in\mathcal{N}} \frac{\sigma}{\sigma-1} \left(X_{ud}^W - X_{mud}^W\right) d\log \overline{M}_{ud}^W}_{\textit{indirect match effect}} - \underbrace{\sum_{u\in\mathcal{N}}\sum_{d\in\mathcal{N}} \frac{\sigma}{\sigma-1} X_{ud} \, d\log w_u}_{\textit{wage effect}} \\ &+ \underbrace{\frac{1}{\sigma-1}\sum_{u\in\mathcal{N}}\sum_{d\in\mathcal{N}} \frac{\sigma}{\sigma-1} X_{mud}^W \, d\log N_{Iu} - \frac{1}{\sigma-1}\sum_{u\in\mathcal{N}}\sum_{d\in\mathcal{N}} \frac{\sigma}{\sigma-1} X_{mud}^W \, d\log \overline{M}_{ud}}_{\textit{cannibalization effect}} \\ &+ \underbrace{\frac{1}{\sigma-1}\sum_{i\in\mathcal{N}} I_i \, d\log N_{Fi}}_{\textit{endogenous entry effect}} - \underbrace{\sum_{i\in\mathcal{N}} (1-\Omega_i) \, I_i \, \rho_i \, d\log \Omega_i}_{\textit{wholesale markup effect}} \end{aligned}$$

where $\overline{M}_{ud}^W \equiv S_{ud} \, N_{Iu}$, $\rho_i \equiv \frac{d\log \mu_i^W}{d\log \Omega_i}$

Proof. Substitute out X_{mui} and $X^W_{ui} - X^W_{mui}$ from equation (62) using equations (57) and (59), and using equation (37) to substitute out D_{Hi} yields the following log-linearized equation:

$$
\begin{aligned}
& d\log I_i \\
=& \sum_{u\in\mathcal{N}}\left(\frac{X_{mui}}{\sum_{l\in\mathcal{N}} X_{li}}\, d\log X_{mui}\right) + \sum_{u\in\mathcal{N}}\left(\frac{X^W_{ui}}{\sum_{l\in\mathcal{N}} X_{li}}\, d\log X^W_{ui}\right) - \sum_{u\in\mathcal{N}}\left(\frac{X^W_{mui}}{\sum_{l\in\mathcal{N}} X_{li}}\, d\log X^W_{mui}\right) \\
=& \sum_{u\in\mathcal{N}}\left[\frac{X_{mui}}{\sum_{l\in\mathcal{N}} X_{li}}\left(d\log N_{Fi} + d\log \overline{M}_{ui} + (1-\sigma)d\log w_u + d\log\phi_{cui} + d\log I_i + (\sigma-1)d\log P^H_i\right)\right] \\
&+ \sum_{u\in\mathcal{N}}\left[\frac{X^W_{ui}}{\sum_{l\in\mathcal{N}} X_{li}}\left(d\log N_{Fi} + d\log \overline{M}^W_{ui} + (1-\sigma)d\log w_u + d\log I_i + (\sigma-1)d\log P^H_i + (1-\sigma)d\log \mu^W_i\right)\right] \\
&- \sum_{u\in\mathcal{N}}\left[\frac{X^W_{mui}}{\sum_{l\in\mathcal{N}} X_{li}}\left(d\log\left(1-\frac{C_{2ui}}{\phi_{cui}}\right) + d\log N_{Fi} + d\log \overline{M}_{ui} + (1-\sigma)d\log w_u + d\log\phi_{cui} + d\log I_i + (\sigma-1)d\log P^H_i\right)\right]
\end{aligned}
$$

where $X_{ud} \equiv X_{mud} + X^W_{ud} + X^W_{mud}$; $\overline{M}_{ud} \equiv \frac{\widetilde{M}_{ud}}{N_{Fd}}$ and $\overline{M}^W_{ud} \equiv S_{ud}\, N_{Iu}$ represent the average number of direct and (potential) indirect matches per final goods producer.

Simplify the above equation yields:

$$
\begin{aligned}
& -(\sigma-1)d\log P^H_i \\
=& d\log N_{Fi} + \sum_{u\in\mathcal{N}}\left[\frac{X_{mui}}{\sum_{l\in\mathcal{N}} X_{li}}\left(d\log \overline{M}_{ui} + (1-\sigma)d\log w_u + d\log\phi_{cui}\right)\right] \\
&+ \sum_{u\in\mathcal{N}}\left[\frac{X^W_{ui} - X^W_{mui}}{\sum_{l\in\mathcal{N}} X_{li}}\left(d\log \overline{M}^W_{ui} + (1-\sigma)d\log w_u + (1-\sigma)d\log \mu^W_i\right)\right] \\
&- \sum_{u\in\mathcal{N}}\left[\frac{X^W_{mui}}{\sum_{l\in\mathcal{N}} X_{li}}\left(-d\log N_{Iu} + d\log \overline{M}_{ui}\right)\right]
\end{aligned}
$$

which can be rewritten in vector form:

$$
\begin{aligned}
& -(\sigma-1)d\log \boldsymbol{P^H} \\
=& d\log \boldsymbol{N_F} + (1-\sigma)\boldsymbol{\chi}' d\log \boldsymbol{w} + \left(\boldsymbol{\chi_m}' \odot d\log \overline{\boldsymbol{M}}'\right)\mathbf{1} + \left(\boldsymbol{\chi_m}' \odot d\log \boldsymbol{\phi_c}'\right)\mathbf{1} \\
&+ \left[\left(\boldsymbol{\chi^{W\prime}} - \boldsymbol{\chi^{W\prime}_m}\right) \odot d\log \overline{\boldsymbol{M}}^{\boldsymbol{W\prime}}\right]\mathbf{1} + (1-\sigma)(\mathbf{1}-\boldsymbol{\Omega})' d\log \boldsymbol{\mu^W} + \boldsymbol{\chi^{W\prime}_m} d\log N_I - \left(\boldsymbol{\chi^{W\prime}_m} \odot d\log \overline{\boldsymbol{M}}'\right)\mathbf{1}
\end{aligned}
$$

where $\boldsymbol{\Omega}$ is a $|\mathcal{N}| \times |\mathcal{N}|$ diagonal matrix whose (i,i)-th element are Ω_i; $\boldsymbol{\chi}$, $\boldsymbol{\chi_m}$, $\boldsymbol{\chi^W}$, $\boldsymbol{\chi^W_m}$, $d\log\boldsymbol{\phi_c}$, $d\log\overline{\boldsymbol{M}}$, and $d\log\overline{\boldsymbol{M}}^{\boldsymbol{W}}$ are $|\mathcal{N}| \times |\mathcal{N}|$ matrices with (i,j)-th element χ_{ij}, χ_{mij}, χ^W_{ij}, χ^W_{mij}, $d\log\phi_{cij}$, $d\log\overline{M}_{ij}$, and $d\log\overline{M}^W_{ij}$ respectively; $d\log\boldsymbol{P^H}$, $d\log\boldsymbol{w}$, $d\log\boldsymbol{N_I}$, $d\log\boldsymbol{N_F}$, and $d\log\boldsymbol{\mu^W}$, are $|\mathcal{N}|\times 1$ column vectors with i-th element $d\log P^H_i$, $d\log w_i$, $d\log N_{Ii}$, $d\log N_{Fi}$, and $d\log\mu^W_i$ respectively; $\mathbf{1}$ is a $|\mathcal{N}| \times 1$ column vector whose entries are all 1 . Note that $\odot$ denotes element-wise multiplication

between matrices. Also:

$$\Omega_i \equiv \frac{\sum_{u\in\mathcal{N}} X_{mui}}{\sum_{u\in\mathcal{N}} (X_{ui})}\ ,\ x_{ud} \equiv X_{mud} + X^W_{ud} - X^W_{mud}\ ,\ \chi_{ud} \equiv \frac{X_{ud}}{\sum_{l\in\mathcal{N}} (X_{ld})}$$

$$\chi_{mud} \equiv \frac{X_{mud}}{\sum_{l\in\mathcal{N}} (X_{mld})}\ ,\ \chi^W_{ud} \equiv \frac{X^W_{ud}}{\sum_{l\in\mathcal{N}} \left(X^W_{ld}\right)}\ ,\ \chi^W_{mud} \equiv \frac{X^W_{mud}}{\sum_{l\in\mathcal{N}} \left(X^W_{mld}\right)}$$

Define $\boldsymbol{I}$ to be a $|\mathcal{N}| \times 1$ column vector with i-th element I_i. The change in aggregate welfare is therefore:

$$\begin{aligned}
d\log \mathcal{W} =& -\boldsymbol{I}' d\log \boldsymbol{P}^{\boldsymbol{H}} \\
=& \underbrace{\frac{1}{\sigma-1}\sum_{u\in\mathcal{N}}\sum_{d\in\mathcal{N}} \frac{\sigma}{\sigma-1} X_{mud}\, d\log \phi_{cud}}_{\text{technological effect}} + \underbrace{\frac{1}{\sigma-1}\sum_{u\in\mathcal{N}}\sum_{d\in\mathcal{N}} \frac{\sigma}{\sigma-1} X_{mud}\, d\log \overline{M}_{ud}}_{\text{direct match effect}} \\
&+ \underbrace{\frac{1}{\sigma-1}\sum_{u\in\mathcal{N}}\sum_{d\in\mathcal{N}} \frac{\sigma}{\sigma-1} (X^W_{ud} - X^W_{mud})\, d\log \overline{M}^W_{ud}}_{\text{indirect match effect}} - \underbrace{\sum_{u\in\mathcal{N}}\sum_{d\in\mathcal{N}} \frac{\sigma}{\sigma-1} X_{ud}\, d\log w_u}_{\text{wage effect}} \\
&+ \underbrace{\frac{1}{\sigma-1}\sum_{u\in\mathcal{N}}\sum_{d\in\mathcal{N}} \frac{\sigma}{\sigma-1} X^W_{mud}\, d\log N_{Iu} - \frac{1}{\sigma-1}\sum_{u\in\mathcal{N}}\sum_{d\in\mathcal{N}} \frac{\sigma}{\sigma-1} X^W_{mud}\, d\log \overline{M}_{ud}}_{\text{cannibalization effect}} \\
&+ \underbrace{\frac{1}{\sigma-1}\sum_{i\in\mathcal{N}} I_i\, d\log N_{Fi}}_{\text{endogenous entry effect}} + \underbrace{\sum_{i\in\mathcal{N}} (1-\Omega_i)\, I_i\, d\log \mu^W_i}_{\text{wholesale markup effect}}
\end{aligned}$$

□

Proposition 6 establishes that, in addition to the technological effect and match effects present in Arkolakis, Huneeus and Miyauchi (2023), there are five extra first-order effects on aggregate welfare, which I refer to as the wage effect, cannibalization effect, endogenous entry effect, wholesale markup effect, and net wholesale profit effect.

First, the wage effect captures how changes in wages influence production costs and, consequently, the final consumption price index in downstream locations. The appropriate weight for changes in wages across upstream locations u in measuring their impact on downstream price index is their wholesale profit-inclusive total exports. Since trade balance holds only when wholesale profits are excluded in this model, and nominal GDP is proportional to wholesale profit-inclusive total imports rather than exports (as shown in equation (62)), changes in wages are not weighted by the nominal GDP of their respective upstream locations. As a result, even when aggregate nominal GDP is chosen as the numeraire, the

wage effect does not necessarily equal zero. However, this effect is generally small and quantitatively negligible, as total imports and total exports are highly correlated in the data.

Second, the cannibalization effect captures the welfare loss resulting from the cannibalization of existing indirect matches by direct matches, which undermines the welfare gains from increases in the total number of direct matches as they no longer represent pure variety gains. Given the number of indirect matches, defined as $\overline{M}_{ud}^{W} \equiv S_{ud}\, N_{Iu}$, the strength of the cannibalization effect increases with the number of direct matches and the share of suppliers matched with wholesalers. The latter decreases with N_{Iu} for a given $\overline{M}_{ud}^{W}$. The cannibalization effect is therefore expected to dampen the welfare gains from shocks that induce an increase in the number of direct matches.

On the other hand, the endogenous entry effect captures how an increase in the number of final goods producers improves aggregate welfare by enhancing consumer welfare through a love-of-variety effect. As discussed in the previous section, the number of final goods producers decreases with the net direct purchase share of a location. Consequently, this additional effect is likely to dampen the aggregate welfare gains from shocks that cause disintermediation.

In addition, the wholesale markup effect reflects changes in the final consumption price index driven by changes in wholesale markups. First, notice that in a model with either constant direct purchase share, $d \log \Omega_i = 0$, or constant wholesale markup, $\rho_i = 0$, the wholesale markup effect would be equal to 0. Now, to determine the sign of the elasticity of wholesale markup with respect to direct purchase share, ρ_i, recall from equation (33) that wholesale markups strictly decrease with the number of wholesalers under the assumption of Cournot competition. Using the free entry condition for wholesalers, we can derive the following relationship between the number of wholesalers and the direct purchase share[48]:

$$\widehat{N_{Wi}} \leq \left[\left(\frac{1}{1-\Omega_i} - \frac{\Omega_i}{1-\Omega_i}\, \widehat{\Omega_i} \right) \frac{\widehat{L_i}}{\widehat{F_{Wi}}} \right]^{\frac{1}{2}} \tag{63}$$

Equation (63) shows that the number of wholesalers decreases with the direct purchase share of a location. Intuitively, as the direct purchase share rises, the demand for wholesale trade declines, reducing the aggregate profit of wholesalers. As a result, fewer wholesalers can operate profitably in the market. This implies that $\rho_i > 0$: disintermediation potentially induced by recent technological progress is likely to reduce the number of wholesalers and increase wholesale markups, thereby dampening the welfare gains.

Lastly, the net wholesale markup effect captures the change in household income due to a change in net wholesale profit, which is transferred to the household.

It is important to emphasize that the decomposition presented in this section implies that, conditional

[48] A detailed derivation is provided in Appendix B.3

on the same shocks and the **same change in the number of matches**, the current model features additional effects that are likely to dampen the aggregate welfare gains from fiber internet expansion. This exercise is similar in spirit to the comparison conducted by Arkolakis et al. (2019), who assess welfare changes in a trade model with variable markups against the ACR formula to evaluate whether trade liberalization generates pro-competitive gains.

However, this decomposition does **not** imply that the current model must predict a smaller aggregate welfare gain than Arkolakis, Huneeus and Miyauchi (2023) under identical shocks, even if all additional first-order effects are negative. There are two key reasons for this. First, the first-order effects cannot be expressed analytically in terms of exogenous shocks alone, so it is not possible to rule out offsetting interactions among these effects. In particular, the decomposition does not reveal the precise extent to which worsening allocative efficiency—caused by rising wholesale markups—dampens welfare gains from improved direct matching technology. To address this, I follow the methodology of Edmond, Midrigan and Xu (2015) and decompose welfare changes from improvements in direct matching technology into two components: changes in the first-best level of welfare and changes due to allocative inefficiency. The results of this alternative decomposition are presented in Section 6.

Second, the departure from the Cobb-Douglas assumption in Arkolakis, Huneeus and Miyauchi (2023) allows for potentially large higher-order effects. As shown by Baqaee and Farhi (2019), such effects tend to amplify welfare gains from positive shocks and could, in principle, generate larger welfare improvements in the current model.

C Sensitivity Analysis for Counterfactuals

C.1 Decomposing the Welfare Cost of Inefficiencies

Low Ω parameterization. In the first sensitivity analysis, I consider a "low Ω" parameterization in which κ is reduced by 28%, lowering the decentralized direct trade share to 25%. This exercise underscores that the welfare cost of wholesale markups depends not only on their level but, crucially, on the *dispersion* of markups they induce. The dispersion of markups—measured as the variance of log markups—can be written as $\Omega\,(1-\Omega)\,(\ln\mu_D - \ln\mu_I)^2 \;=\; \Omega\,(1-\Omega)\,(\ln\mu^W)^2$. Holding μ^W fixed, this dispersion falls when Ω is either very high or very low. Recall Corollaries 1.1 and 2.1: to first order, the degree of misallocation induced by wholesale markups is approximately increasing in markup dispersion when the latter is measured using the *efficient* direct trade share. In this "low Ω" parameterization, the dispersion of markups declines both at the decentralized share and at the efficient share, which explains why the welfare gain from implementing the wholesale subsidy is smaller in this case. Notice that the welfare gain from implementing congestion taxes/subsidies remains similar.

Table C.1: Welfare Decomposition by Policy and Parameterization

	Ω	μ^W	$\sigma^2(\mu)$	$\tilde{M}$	S	A	L_A	L_S	L_E	L_P	Welfare
Baseline											
Decentralized	0.4800	1.1299	0.0037	4.6653	0.2630	0.3852	0.0799	0.0129	0.3465	0.5608	1.0000
Wholesale subsidy	0.1800	1.0000	0.0000	2.2730	0.2539	0.3914	0.0261	0.0194	0.3895	0.5650	1.0237
Efficient	0.2200	1.0000	0.0000	0.0435	0.3740	0.3556	0.0230	0.0270	0.2818	0.6682	1.1000
Low Ω											
Decentralized	0.2500	1.1299	0.0028	2.3099	0.2642	0.3891	0.0427	0.0189	0.3850	0.5534	1.0000
Wholesale subsidy	0.0620	1.0000	0.0000	0.7586	0.2530	0.3888	0.0089	0.0222	0.4056	0.5633	1.0172
Efficient	0.0650	1.0000	0.0000	0.0122	0.4062	0.3506	0.0065	0.0328	0.2889	0.6718	1.0939
Low μ^W											
Decentralized	0.4800	1.0610	0.0009	5.2940	0.2374	0.3889	0.0771	0.0064	0.3408	0.5757	1.0000
Wholesale subsidy	0.3200	1.0000	0.0000	3.9752	0.2337	0.3904	0.0482	0.0082	0.3649	0.5787	1.0092
Efficient	0.2900	1.0000	0.0000	0.0606	0.3764	0.3636	0.0297	0.0248	0.2747	0.6708	1.0892
High S											
Decentralized	0.4800	1.1299	0.0037	5.1786	0.2960	0.3984	0.0777	0.0129	0.3487	0.5607	1.0000
Wholesale subsidy	0.1700	1.0000	0.0000	2.3825	0.2861	0.4052	0.0235	0.0197	0.3920	0.5648	1.0246
Efficient	0.1900	1.0000	0.0000	0.0425	0.4257	0.3676	0.0188	0.0280	0.2847	0.6684	1.0999
Low λ_V, λ_M											
Decentralized	0.4800	1.1299	0.0037	4.5892	0.2631	0.3852	0.0788	0.0130	0.3476	0.5606	1.0000
Wholesale subsidy	0.2000	1.0000	0.0000	2.4926	0.2539	0.3918	0.0285	0.0190	0.3872	0.5652	1.0255
Efficient	0.4400	1.0000	0.0000	0.0399	0.3360	0.3630	0.0450	0.0192	0.2675	0.6682	1.1233
Low λ_W											
Decentralized	0.4800	1.1299	0.0037	4.6628	0.2631	0.3853	0.0798	0.0129	0.3466	0.5608	1.0000
Wholesale subsidy	0.1800	1.0000	0.0000	2.2534	0.2547	0.3917	0.0258	0.0195	0.3898	0.5649	1.0244
Efficient	0.2200	1.0000	0.0000	0.0444	0.3633	0.3600	0.0227	0.0249	0.2878	0.6647	1.1077

Notes: Column 4 reports $\tilde{M}$ multiplied by 10,000. The last column reports welfare, normalized to 1 in the decentralized equilibrium for each parameter set. The "Low Ω" case reduces κ by 28%, shrinking Ω to 25% in the decentralized equilibrium. "Low μ^W" reduces F_W by 72%, doubles the number of wholesalers N_W, and raises κ by 14% to keep Ω similar. "High S" halves f_W and raises κ by 11%. "Low λ_V, λ_M" sets $\lambda_V = \lambda_M = 0.75$ and lowers κ by 24%. "Low λ_W" sets $\lambda_W = 0.57$ and raises f_W by 38%.

Given this observation, in the remaining sets of parameterizations I recalibrate either κ or f_W to hold Ω constant, in order to isolate the pure welfare effects of perturbing other parameters.

Low μ^W parameterization. Next, I simulate the model using a "low μ^W" parameterization, which doubles the number of wholesalers by reducing the wholesaler entry cost shifter F_W by 72%, thereby shrinking the wholesale markup to around 1.06. Relative to the baseline, the dispersion of markups is lowered by a factor of 4. Consequently, the implementation of the wholesale subsidy results in a more modest decline in the number of direct matches, and the increase in aggregate productivity is smaller. Welfare rises by only 0.9%, which is 40% of the magnitude observed in the baseline. These results confirm the prediction of Corollaries 1.1 and 2.1: the degree of misallocation induced by wholesale markups is increasing in their level.

High S parameterization. The next parameterization halves the level of wholesalers' supplier search cost f_W, raising the equilibrium share of suppliers matched with wholesalers S. As a result, the wholesale subsidy induces a slightly greater reduction in the number of direct matches and causes a larger reduction in the amount of labor allocated to ad posting. This is consistent with Proposition 2, which implies that the amount of excessive labor allocated to the creation of direct matches increases with the misallocation wedge, itself increasing in the share of suppliers matched with wholesalers S. The intuition is that a higher S implies a greater extent of cannibalization of indirect matches by direct matches, and consequently a greater *understatement* of such cannibalization caused by wholesale markups.

Low λ_V, λ_M parameterization. I now switch gears to investigate the sensitivity of the welfare cost of congestion externalities to the matching-function elasticities. In particular, I lower λ_V and λ_M by 0.05 to 0.75. In this parameterization, implementing the congestion taxes/subsidies now raises welfare by 9.1%—1.9% greater than in the baseline. This is in line with the congestion wedge derived in Section 4.1 by comparing the optimality conditions of the social planner against those in the decentralized equilibrium, which scales with $(\lambda_V - 1)$ and $(\lambda_M - 1)$. Moreover, the welfare gain from introducing the wholesale subsidy increases slightly by around 0.2%. Intuitively, wholesale markup interacts with congestion externality in direct matching by raising the share of direct trade, thereby amplifying the impact of its congestion. As a result, this interaction effect tends to raise the welfare gain from eliminating wholesale markup. Now, lower λ_V and λ_M make direct trade more congested and result in a stronger interaction effect, yielding an even larger welfare gain from the elimination of wholesale markup.

Low λ_W parameterization. Lastly, I examine the sensitivity of the welfare decomposition to a reduction in λ_W—which governs the congestion externality in wholesalers' supplier search—by 0.05, to 0.57. As

in the case of lowering λ_V and λ_M, implementing congestion taxes/subsidies now increases welfare by an additional 0.7% relative to the baseline, reflecting a stronger congestion externality in wholesalers' supplier search. This effect, however, occurs only when $N_W = 1$, as in the efficient allocation. Recall from Proposition 4 that the optimal tax/subsidy on wholesalers' supplier search is $\tau_s^M = (\frac{\mu^W - 1}{\mu^W}/\sigma)/\lambda_W = \frac{1}{N_W^* \lambda_W}$. When $N_W > 1$—as in the equilibrium with only the wholesale subsidy ($N_W = 2$)—there is a misalignment between the wholesalers' incentive to search for an additional variety and that of the social planner. Wholesalers charge a markup below the monopolistic level required to align these incentives, leading to inefficiently low search effort. Correcting this distortion requires a subsidy equal to $1/N_W$. When $N_W = 2$, the congestion externality offsets this misalignment exactly when λ_W approaches 0.5. Hence, in this alternative parameterization, the lower λ_W actually reduces the overall inefficiency in wholesalers' supplier search. Consequently, the welfare gain from introducing the wholesale subsidy increases slightly, by about 0.07%. Intuitively, the wholesale markup interacts with inefficiency in wholesalers' supplier search by lowering the share of indirect trade, thereby dampening the impact of this inefficiency. When λ_W is lower, the overall inefficiency in wholesalers' supplier search is smaller, so this dampening effect is weaker, leading to a larger welfare gain from removing the wholesale markup.

C.2 Fiber Internet Expansion Counterfactuals

In this section, I discuss how the internet counterfactual may lead to different welfare and efficiency consequences under different sets of parameterization. Table C.2 reports the *proportional changes* of different variables relative to the pre-shock decentralized equilibrium of the post-shock equilibrium and equilibria across different policy regimes within each set of parameterization. It also reports the *levels* of those variables in each of the pre-shock decentralized equilibria.

Low Ω parameterization. I rerun the counterfactual using a "low Ω" parameterization, in which κ is reduced by 28%, lowering the decentralized direct trade share to $\Omega = 25\%$. Relative to the baseline, the direct trade share experiences a smaller proportional increase following the shock. This occurs partly because the proportional change in the indirect trade share is smaller in this parameterization, and as a result, the number of wholesalers and the wholesale markup remain unchanged. The stability of the wholesale markup also implies that misallocation from wholesaler double marginalization remains essentially the same post-shock. In fact, allocative efficiency worsens by only 0.5%, with this slight decline likely attributable to increased *misallocation* markup dispersion following an increase in the efficient direct trade share.

Low μ^W parameterization. Next, I simulate the counterfactual using a "low μ^W" parameterization, which doubles the number of wholesalers by reducing the wholesaler entry cost shifter F_W by 72%,

Table C.2: Welfare Decomposition (Single-Location Model): Percentage Changes Relative to the Pre-Shock Decentralized Equilibrium across Parameterizations

	Ω	μ^W	$\sigma^2(\mu)$	$\tilde{M}$	S	A	L_A	L_S	L_E	L_P	Welfare
Baseline (Pre Shock)											
Wholesale subsidy	−63.0	−11.5	−100.0	−51.3	−3.5	1.6	−67.3	51.0	12.4	0.7	2.4
Efficient	−54.0	−11.5	−100.0	−99.1	42.2	−7.7	−71.3	110.4	−18.7	19.2	10.0
Baseline (Post Shock)											
Decentralized	63.0	14.9	197.3	29.5	17.0	1.2	77.5	−19.8	−20.4	2.0	3.3
Wholesale subsidy	−34.0	−11.5	−100.0	−27.0	9.1	6.5	−39.8	25.8	6.5	1.1	7.7
Efficient	−12.0	−11.5	−100.0	−98.4	46.1	−2.1	−38.8	53.5	−23.2	18.6	16.1
Low Ω (Pre Shock)											
Wholesale subsidy	−76.0	−11.5	−100.0	−67.2	−4.2	−0.1	−79.2	17.4	5.3	1.8	1.7
Efficient	−75.0	−11.5	−100.0	−99.5	53.7	−9.9	−84.8	73.1	−25.0	21.4	9.4
Low Ω (Post Shock)											
Decentralized	53.0	0.0	25.2	31.4	12.8	3.0	56.8	−19.4	−6.5	0.8	3.8
Wholesale subsidy	−48.0	−11.5	−100.0	−41.6	8.1	4.0	−53.4	8.7	2.7	2.0	6.0
Efficient	−36.0	−11.5	−100.0	−98.9	65.3	−5.8	−57.3	54.0	−26.6	21.1	14.0
Low μ^W (Pre Shock)											
Wholesale subsidy	−34.0	−5.7	−100.0	−24.9	−1.6	0.4	−37.5	27.7	7.1	0.5	0.9
Efficient	−40.0	−5.7	−100.0	−98.9	58.5	−6.5	−61.4	285.1	−19.4	16.5	8.9
Low μ^W (Post Shock)											
Decentralized	33.0	2.1	66.6	18.1	15.3	5.7	39.0	−7.8	−9.6	0.5	6.3
Wholesale subsidy	−0.6	−5.7	−100.0	−2.5	10.9	6.0	−3.1	−1.7	−0.5	0.7	6.7
Efficient	6.4	−5.7	−100.0	−98.2	61.2	−0.4	−24.0	162.3	−24.7	16.0	15.5
High S (Pre Shock)											
Wholesale subsidy	−65.0	−11.5	−100.0	−54.0	−3.4	1.7	−69.7	52.2	12.4	0.7	2.5
Efficient	−60.0	−11.5	−100.0	−99.2	43.8	−7.7	−75.8	116.5	−18.3	19.2	10.0
High S (Post Shock)											
Decentralized	65.0	14.9	190.3	31.3	16.5	1.0	82.1	−23.0	−20.9	2.1	3.1
Wholesale subsidy	−36.0	−11.5	−100.0	−28.7	9.1	6.5	−41.7	26.8	6.6	1.1	7.7
Efficient	−15.0	−11.5	−100.0	−98.5	46.4	−2.2	−41.8	56.7	−22.7	18.6	16.0
Low λ_V, λ_M (Pre Shock)											
Wholesale subsidy	−59.0	−11.5	−100.0	−45.7	−3.5	1.7	−63.8	46.0	11.4	0.8	2.6
Efficient	−6.5	−11.5	−100.0	−99.1	27.7	−5.8	−42.8	47.3	−23.0	19.2	12.3
Low λ_V, λ_M (Post Shock)											
Decentralized	63.0	14.9	208.0	27.9	17.6	1.0	77.8	−17.0	−20.2	2.0	2.9
Wholesale subsidy	−32.0	−11.5	−100.0	−24.2	9.0	6.6	−37.7	23.0	5.9	1.1	7.8
Efficient	33.0	−11.5	−100.0	−98.9	31.4	0.7	−10.3	−4.3	−28.7	19.3	20.2
Low λ_W (Pre Shock)											
Wholesale subsidy	−63.0	−11.5	−100.0	−51.7	−3.2	1.7	−67.7	51.3	12.5	0.7	2.4
Efficient	−55.0	−11.5	−100.0	−99.0	38.1	−6.5	−71.6	93.4	−17.0	18.5	10.8
Low λ_W (Post Shock)											
Decentralized	60.0	14.9	212.4	28.9	22.9	1.9	74.1	−14.1	−19.6	1.9	3.8
Wholesale subsidy	−33.0	−11.5	−100.0	−26.2	8.3	6.3	−38.7	24.8	6.2	1.1	7.5
Efficient	−14.0	−11.5	−100.0	−98.4	43.8	−1.3	−40.0	42.8	−21.8	18.2	16.7

Note: Within each set of parameterization, this table reports *percentage changes* relative to the pre-shock decentralized equilibrium of the post-shock equilibrium and equilibria across different policy regimes. The "Low Ω" case reduces κ by 28%, shrinking Ω to 25% in the decentralized equilibrium. "Low μ^W" reduces F_W by 72%, doubles the number of wholesalers N_W, and raises κ by 14% to keep Ω similar. "High S" halves f_W and raises κ by 11%. "Low λ_V, λ_M" sets $\lambda_V = \lambda_M = 0.75$ and lowers κ by 24%. "Low λ_W" sets $\lambda_W = 0.57$ and raises f_W by 38%.

thereby lowering the wholesale markup to around 1.06. Relative to the baseline, the post-shock increase in the wholesale markup is much more modest, rising by 2% instead of 15%. Together with the fact that the efficient direct trade share is likely to have increased beyond the level corresponding to maximum misallocation—which is especially likely given the high level of pre-shock efficient direct trade share implied by the low level of wholesale markup, this explains why misallocation actually improves by 0.5% post-shock. This alternative parameterization highlights that rising wholesale markups do not necessarily exacerbate misallocation, since what ultimately matters for aggregate efficiency is the dispersion of markups—not their level. Whether disintermediation exacerbates misallocation depends on the horse race between rising markups and declining indirect trade shares, and their net effect on markup dispersion.

High S parameterization. The next parameterization halves the level of wholesalers' supplier search cost f_W, raising the equilibrium share of suppliers matched with wholesalers S. Recall from our earlier discussion that a higher S implies greater cannibalization of indirect matches by direct matches, and consequently a greater *understatement* of such cannibalization caused by wholesale markups. This explains why the pre-shock degree of misallocation is higher. Following the same logic, the degree of misallocation also worsens more with the increase in the wholesale markup post-shock when S is higher, by 1.9% compared with 1.8% in the baseline.

Low λ_V, λ_M parameterization. This parameterization raises the extent of direct matching congestion by lowering λ_V and λ_M to 0.75. As in the pre-shock case, stronger congestion intensifies the interaction between congestion externalities and wholesale markup, increasing the welfare gain from the wholesale subsidy. By the same logic, the allocative-efficiency loss from a higher wholesale markup after the shock is also amplified: the post-shock welfare cost of wholesale markup rises by 2.1%, compared with 1.8% in the baseline.

Low λ_W parameterization. Here I lower λ_W to 0.57 to examine how the post-shock change in allocative efficiency responds to stronger congestion in wholesales' supplier search. As discussed earlier, the overall inefficiency in wholesalers' supplier search reflects both congestion and the misalignment created when wholesalers charge a markup below the monopolistic level. Since $N_W = 1$ after the internet shock, the markup-related misalignment disappears, leaving only congestion. A lower λ_W raises the congestion externality and therefore the overall distortion to wholesalers' supplier search, which then strengthens the markup's dampening effect on the inefficiency associated with wholesalers' supplier search by reducing the share of indirect trade. The result is a more muted increase in the welfare gain from removing the markup: 1.1% post-shock.

D Tables and Figures

D.1 Tables and Figures for Motivational Facts

Fact 5: Manufacturing firms do not face additional markdowns when selling to wholesalers

	Province Aggregate Manufacturing Markup	Firm-Level Manufacturing Markup
Indirect Sales Share	-0.0151	0.0063*
	(0.1291)	(0.0036)
Province FE	✓	✓
Industry FE		✓
Year FE	✓	✓
Log Sales Control		✓
Observations	648	304,831
R-squared	0.5921	0.0297

Table D.1: Relationship between indirect sales share and manufacturing markups at aggregate and firm level

Note: Column (1) reports OLS estimates from regressions of aggregate manufacturing markups of Turkish provinces on their indirect sales share between 2012 and 2019. Column (2) reports analogous regressions at the firm level. Aggregate markup is measured as the cost-weighted average of firm markups. Indirect trade refers to sales to wholesalers; direct trade refers to sales to manufacturing firms. Indirect sales share is the ratio of indirect trade to the sum of direct and indirect trade. * 10%, ** 5%, *** 1% significance levels. Robust standard errors are reported in parentheses.

An additional motivational fact explores whether manufacturing firms face additional markdowns when selling to wholesalers, compared to their direct sales to other manufacturing firms. To investigate this, I examine the relationship between manufacturing markups and the share of indirect sales—defined as the ratio of sales to wholesalers over the sum of sales to wholesalers and direct buyers.

Table D.1 presents regression results at both the aggregate and firm levels. Column 1 shows estimates from regressing province-level aggregate manufacturing markups on aggregate indirect sales share, controlling for province and year fixed effects. The estimated coefficient is negative but statistically insignificant, suggesting no systematic relationship between higher exposure to wholesalers and lower provincial manufacturing markups.

Column 2 reports firm-level regressions of individual manufacturing firm markups on their own indirect sales share. This specification includes province, industry, and year fixed effects, as well as controls for firm size (log sales). The estimated coefficient is small and positive, and statistically significant at conventional levels. This further reinforces the lack of evidence that selling through wholesalers reduces markups—if anything, the results suggest a weakly positive correlation.

Taken together, these findings indicate that wholesalers do not appear to exert systematically greater buyer power than other types of buyers, at least not in a way that consistently reduces manufactur-

ing markups. This may seem counterintuitive, as wholesalers are often thought to possess bargaining advantages due to scale. However, it is plausible that large direct buyers—such as downstream manufacturers—possess comparable negotiating leverage. In such cases, suppliers may face similar pricing pressure regardless of whether they sell to wholesalers or other firms, resulting in no systematic markup differences.

	Total	**Direct**	**Indirect**
Log Distance	-1.241***	-1.339***	-1.153***
	(0.011)	(0.013)	(0.013)
Origin Province-Year FE	✓	✓	✓
Destination Province-Year FE	✓	✓	✓
Same Province-Year FE	✓	✓	✓
Observations	39,995	35,107	34,620
R-squared	0.736	0.707	0.707

Table D.2: Distance Elasticity of Direct versus Indirect Trade

Note: This table reports OLS estimates of the relationship between log distance and bilateral manufacturing trade flows between Turkish provinces. The dependent variables are indicated in the column headers. Here, indirect trade refers to the sales of upstream manufacturing firms to wholesalers in the downstream province, while direct trade refers to the sales of upstream manufacturing firms to manufacturing firms in the downstream province. * 10%, ** 5%, *** 1% significance levels. Standard errors are reported in parentheses.

	Extensive Margin		**Intensive Margin**	
	Direct	Indirect	Direct	Indirect
Log Distance	-1.108***	-0.987***	-0.230***	-0.167***
	(0.006)	(0.006)	(0.010)	(0.010)
Origin Province-Year FE	✓	✓	✓	✓
Destination Province-Year FE	✓	✓	✓	✓
Same Province-Year FE	✓	✓	✓	✓
Observations	35,107	34,620	35,107	34,620
R-squared	0.866	0.858	0.289	0.390

Table D.3: Intensive and Extensive Margin Distance Elasticity of Direct versus Indirect Trade

Note: This table reports OLS estimates of the relationship between log distance and bilateral manufacturing trade flows between Turkish provinces. The dependent variables are indicated in the column headers. Here, indirect trade refers to the sales of upstream manufacturing firms to wholesalers in the downstream province, while direct trade refers to the sales of upstream manufacturing firms to manufacturing firms in the downstream province. Extensive margin refers to the number of matches, while intensive margin refers to the average trade flow per match. * 10%, ** 5%, *** 1% significance levels. Standard errors are reported in parentheses.

D.2 Tables and Figures for Fiber Internet Expansion Empirics

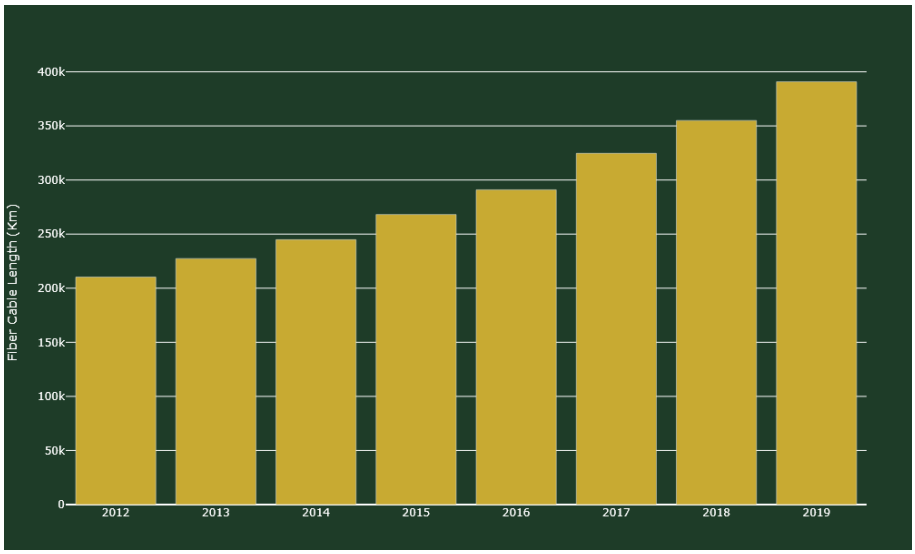

Figure D.1: Evolution of the total length of fiber cable deployed in Turkey between 2012-2019

Note: This figure shows the evolution of the total length of fiber cable deployed in Turkey between 2012 and 2019. Data is sourced from the Turkish Information and Communication Technologies Authority (BTK).

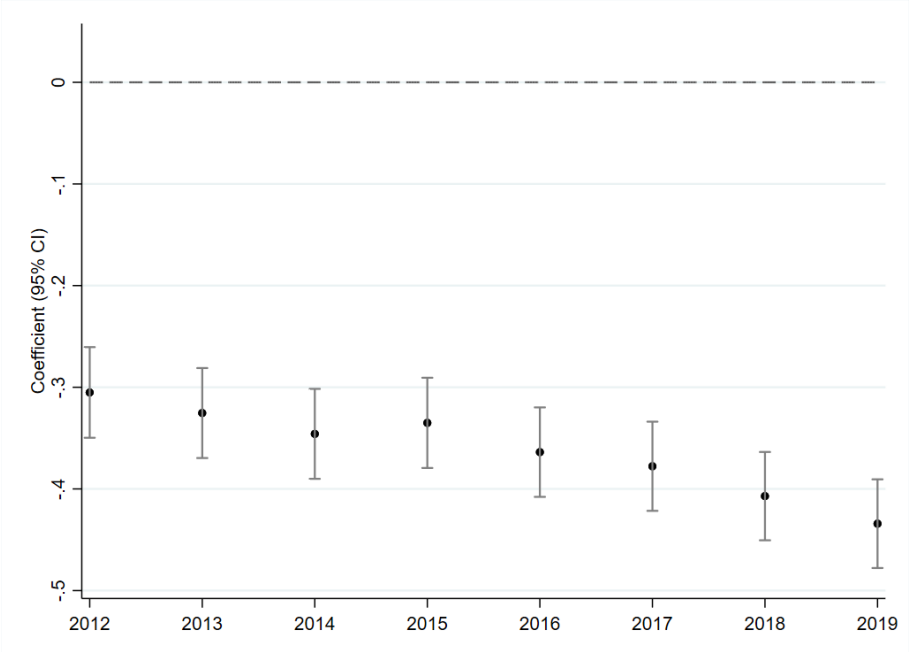

Figure D.2: First-stage coefficient estimates of distance to oil pipeline interacted with year dummies

Note: This figure presents first-stage coefficients from IV regressions of standardized fiber intensity on distance to the nearest pipeline interacted with year dummies, controlling for origin-year, destination-year, and origin-destination pair fixed effects. Standard errors are clustered at the province-pair level.

Figure D.3: Oil and gas pipeline network across Turkish provinces

Note: This map displays the oil and gas pipeline network maintained by BOTAŞ, the state-owned energy enterprise.

D.3 Robustness Checks for Fiber Internet Regressions

Table D.4 shows that the pattern of disintermediation also holds at the firm level. Firms operating in provinces with faster fiber rollout experience statistically significant declines in the share of both their sourcing and sales accounted for by indirect trade (Columns 1 and 2). However, the estimated magnitudes are smaller than those in the inter-provincial specification. One explanation is that firm-level regressions effectively assign more weight to firms located in densely populated and economically advanced provinces like Istanbul and Ankara, where fiber intensity is already high and grew rapidly. If the relationship between fiber expansion and disintermediation is concave, the marginal effect of additional fiber in those regions would be lower, dampening the average estimated effect.

Despite this, the firm-level results remain meaningful. Column 1 suggests that a typical firm in Istanbul, where standardized fiber intensity increased by approximately 3.5 units over the period, experienced a 4.1 percentage point decline in indirect sourcing share. Columns 3 and 4 confirm that these patterns are robust to controlling for firm-level labor share, implying that disintermediation is not simply a result of greater outsourcing or labor substitution.

Table D.5 reports results using a firm-specific measure of fiber exposure—constructed as a weighted average of provincial fiber intensity, with weights based on each firm's sales distribution across provinces. These results are qualitatively similar, reinforcing the conclusion that digital infrastructure expansion facilitates disintermediation in production network. Lastly, Table D.6 presents the results of regressing the aggregate indirect trade share of each province on fiber intensity. The 2SLS estimate is negative and statistically significant, indicating that fiber internet expansion does not just reallocate indirect trade

	Indirect Sourcing Share	Indirect Sales Share	Indirect Sourcing Share	Indirect Sales Share
Panel A: OLS				
Std Fiber Intensity	-0.0064***	-0.0061***	-0.0064***	-0.0061***
	(0.0008)	(0.0012)	(0.0014)	(0.0012)
Panel B: 2SLS				
Std Fiber Intensity	-0.0118***	-0.0076***	-0.0119***	-0.0076***
	(0.0045)	(0.0025)	(0.0045)	(0.0026)
Firm FE	✓	✓	✓	✓
Province FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Labor Share Control			✓	✓
Observations	733,358	651,320	733,299	651,276

Table D.4: Impact of Fiber Internet Expansion on Disintermediation (Firm-Level)

Note: This table reports OLS (Panel A) and 2SLS (Panel B) estimates of the relationship between fiber intensity and firm-level indirect trade shares. The dependent variables are listed in the column headers. Fiber intensity is standardized by subtracting its mean and dividing by its standard deviation over the sample period. The 2SLS regressions use distance to the nearest oil pipeline as an instrument for fiber intensity, interacted with year dummies. All specifications include firm, province, and year fixed effects. Columns 3 and 4 additionally control for firm-level labor share. * 10%, ** 5%, *** 1% significance levels. Standard errors clustered at the province-pair level are reported in parentheses.

	Indirect Sourcing Share	Indirect Sales Share	Indirect Sourcing Share	Indirect Sales Share
Panel A: OLS				
Std Fiber Intensity	0.0327	-0.0004	0.0327***	-0.0004
(Firm-specific)	(0.0446)	(0.0018)	(0.0014)	(0.0018)
Panel B: 2SLS				
Std Fiber Intensity	-0.0626***	-0.0416***	-0.0632***	-0.0420***
(Firm-specific)	(0.0223)	(0.0139)	(0.0224)	(0.0140)
Firm FE	✓	✓	✓	✓
Province FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Labor Share Control			✓	✓
Observations	733,358	651,320	733,299	651,276

Table D.5: Impact of Fiber Internet Expansion on Disintermediation (Firm-Specific Fiber Intensity)

Note: This table reports OLS (Panel A) and 2SLS (Panel B) estimates of the relationship between firm-specific fiber intensity and indirect trade shares at the firm level. The dependent variables are listed in the column headers. The fiber intensity variable is constructed as a firm-specific weighted average of province-level fiber intensity, using each firm's sales distribution across provinces as weights. Fiber intensity is standardized by subtracting its mean and dividing by its standard deviation over the sample period. The 2SLS specifications use distance to the nearest oil pipeline, interacted with year dummies, as an instrument. All regressions include firm, province, and year fixed effects; Columns 3 and 4 additionally control for labor share. * 10%, ** 5%, *** 1% significance levels. Standard errors are clustered at the province level and reported in parentheses.

	Aggregate Indirect Trade Share
Panel A: OLS	
Std Fiber Intensity	-0.016*
	(0.009)
Panel B: 2SLS	
Std Fiber Intensity	-0.239**
	(0.110)
Province FE	✓
Year FE	✓
Observations	648

Table D.6: Impact of Fiber Internet Expansion on Disintermediation (Aggregate)

Note: This table reports OLS and 2SLS estimates of the relationship between fiber intensity and aggregate indirect trade share of a province. Fiber intensity is standardized by subtracting its mean, divided by its standard deviation, over the sample period. The instrumental variable used for the 2SLS regression is the maximum distance of the province to the nearest oil pipeline, interacted with year dummies. * 10%, ** 5%, *** 1% significance levels. Robust standard errors are reported in parentheses.

flows of a province to its trading partners with higher fiber intensity, but also shrinks the aggregate share of trade flow intermediated through wholesalers. To understand the magnitude, the median province saw an increase of 0.39 in the standardized fiber intensity, with an interquartile range of 0.31. This implies that a province at the 75th percentile of fiber intensity growth would experience a relative decline in the aggregate indirect trade share by 7.4 percentage points compared to one at the 25th percentile—again, an economically significant impact.

	OLS		2SLS	
	Markup	**Markup**	**Markup**	**Markup**
Std Fiber Intensity	0.0126***	0.0985***	0.0081	0.1882***
	(0.0042)	(0.0169)	(0.0063)	(0.0680)
Intensity × Indirect Sales Share	-0.0013	0.0009	-0.0031	-0.0009
	(0.0013)	(0.0013)	(0.0026)	(0.0024)
Intensity × Log Sales		-0.0047***		-0.0100***
		(0.0009)		(0.0040)
Firm FE	✓	✓	✓	✓
Province FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Indirect Sales Share Control	✓	✓	✓	✓
Log Sales Control		✓		✓
Observations	260,868	260,868	260,868	260,868

Table D.7: Impact of Fiber Internet Expansion on Manufacturing Firms' Markup

Note: This table reports OLS and 2SLS estimates of the relationship between fiber intensity and manufacturing markup at the firm level. Fiber intensity is standardized by subtracting its mean, divided by its standard deviation, over the sample period. The instrumental variable used for the 2SLS regression is the maximum distance of the province to the nearest oil pipeline, interacted with year dummies. * 10%, ** 5%, *** 1% significance levels. Standard errors are reported in parentheses.

D.4 Tables and Figures for Counterfactuals

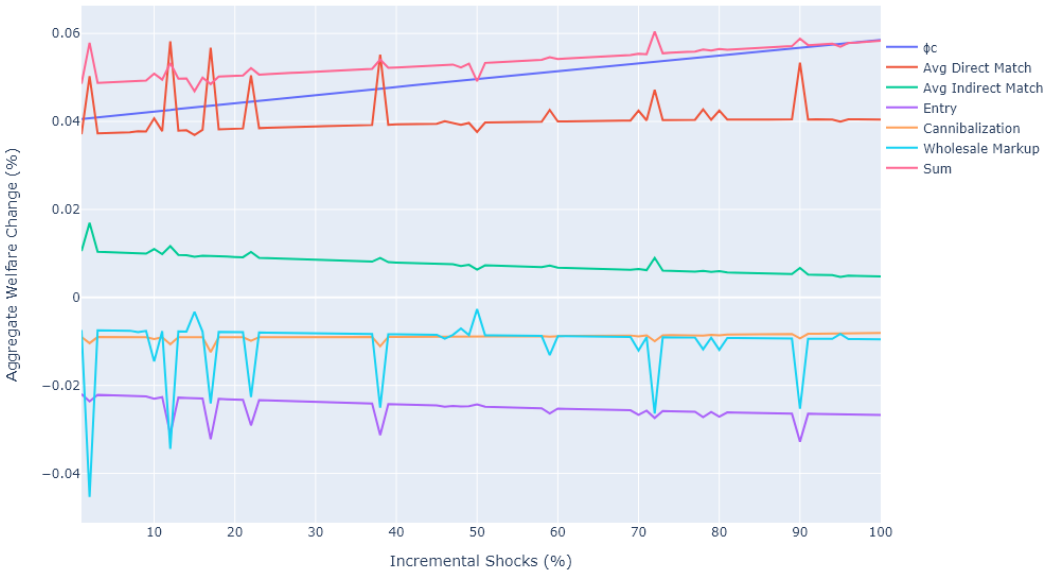

Figure D.4: Stepwise First Order Effects of Internet Shock

Note: The figure reports the stepwise decomposition of first-order welfare effects of internet expansion. Channels include changes in customization productivity, firm entry, direct and indirect matches, indirect match cannibalization, wholesale markups, and net wholesaler profits. All components are expressed in percentage change of aggregate welfare.

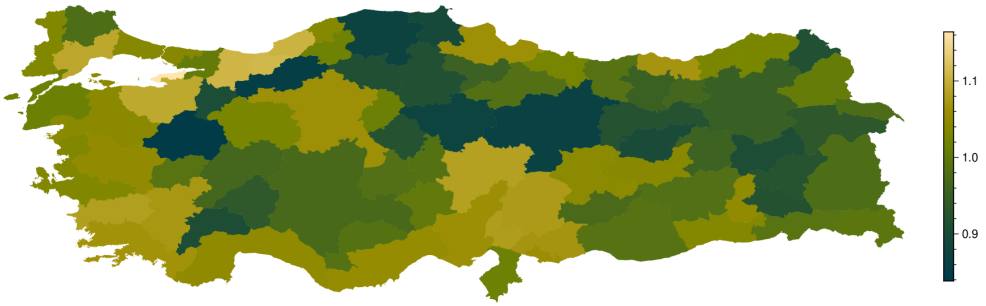

Figure D.5: Proportional Change in Welfare

Note: This map reports the model-implied proportional change in welfare by province following the expansion of fiber internet.

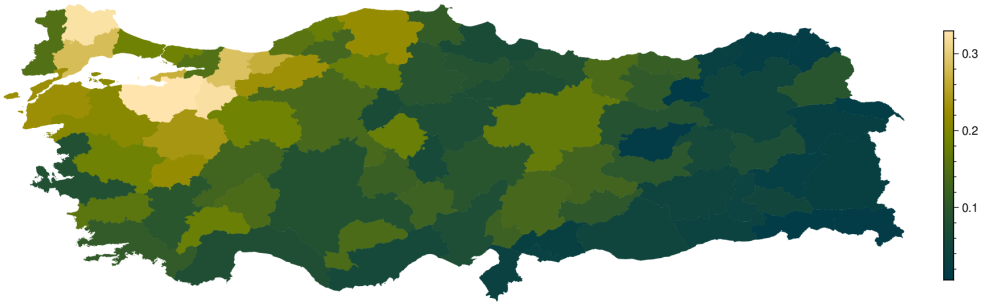

Figure D.6: Ratio of direct purchases from Istanbul to province-level GDP

Note: The figure displays the ratio of direct purchases from Istanbul to local GDP across provinces. Direct purchases refer to trade with manufacturing suppliers located in Istanbul. The ratio proxies for dependence on Istanbul-based sourcing through direct links.

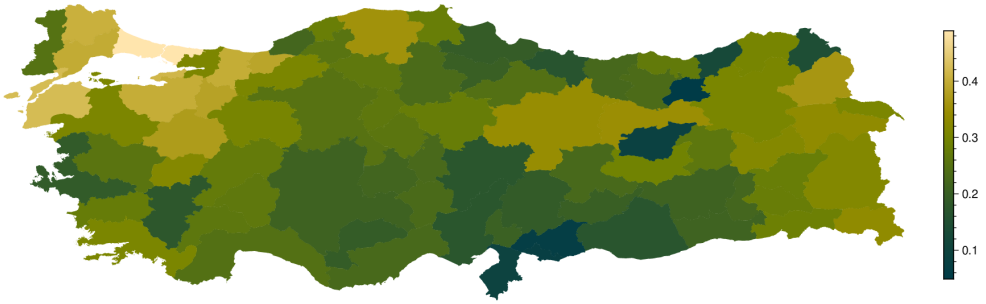

Figure D.7: Ratio of total purchases from Istanbul to province-level GDP

Note: The figure displays the total (direct and indirect) purchases from Istanbul-based suppliers as a ratio to local GDP. Indirect purchases include trade intermediated through wholesalers. The measure reflects the overall centrality of Istanbul in provincial sourcing patterns.

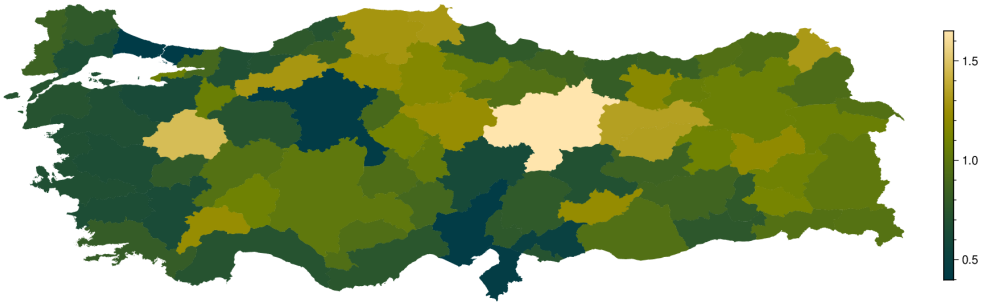

Figure D.8: Proportional Change in Indirect Trade Share

Note: This map shows the proportional change in the share of indirect trade, defined as purchases routed through wholesalers, following the expansion of fiber internet.

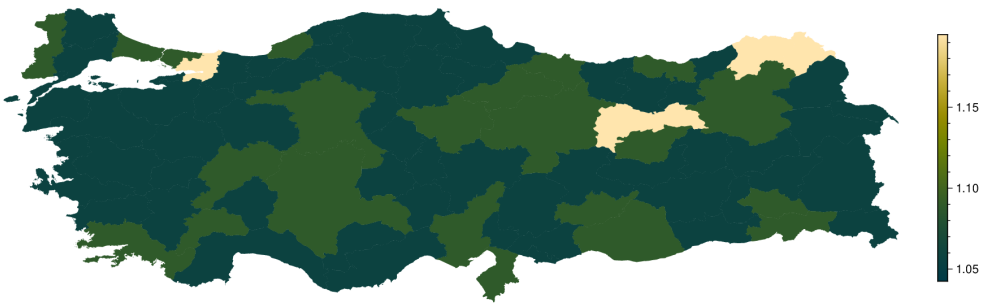

Figure D.9: Wholesale Markup (Pre-Shock)

Note: The figure reports model-implied wholesale markups across provinces before the internet expansion.

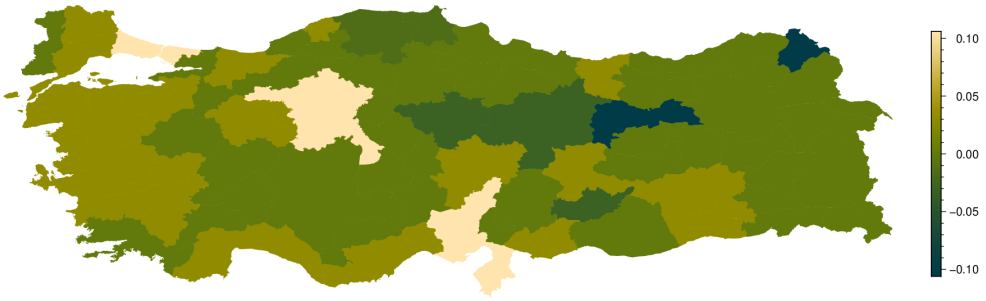

Figure D.10: Proportional Change in Wholesale Markup

Note: This figure presents the proportional change in wholesale markups across provinces following internet expansion.

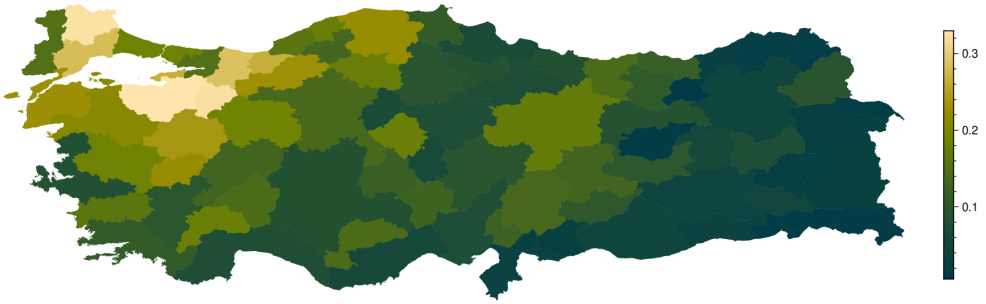

Figure D.11: Ratio of Direct Purchases from Istanbul to Province GDP

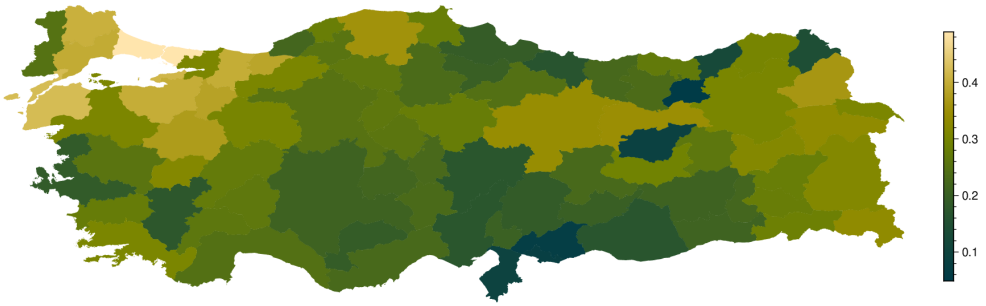

Figure D.12: Ratio of Total Purchases from Istanbul to Province GDP