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

Yiu Hing Barron Tsai

September 2025

Most Recent Version

Abstract

Policy debates often start from the premise that wholesale markups simply inflate prices, making technologies or regulations that "cut out the middleman" appear unambiguously welfare-improving. This paper shows the trade-off is more nuanced. Wholesalers help firms avoid the fixed costs of forming direct buyer-supplier links, but when direct-trade technology improves, demand for intermediation falls; marginal wholesalers exit, survivors gain share, and markups rise — distorting relative input prices and misallocating resources — partially offsetting gains from disintermediation. I develop a production network model with endogenous intermediation, wholesaler entry/exit, and markups, and test its 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 elasticities, endogenous markup increases reduce welfare gains from fiber-induced disintermediation by 1.4 p.p. ($\approx$30%), highlighting the need for complementary competition policy.

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

1 Introduction

Wholesalers are an integral part of the production network and account for more than half of the domestic trade of manufacturing goods in Turkey (e.g., in 2012). They help firms avoid the fixed costs of building many direct buyer-supplier links by pooling these costs — firms connect once to the intermediary and gain access to a broad network.[1] However, in this theoretical view, pooling fixed costs across firms 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]

Despite the nontrivial policy trade-off implied by wholesalers' technology efficiency and market power, policy debates often start from the premise that wholesale markups simply inflate prices and that welfare rises if technology or regulation encourages buyers and sellers to transact directly — i.e., cut out the middleman.[3][4] Yet a single-minded emphasis on disintermediation can backfire in general equilibrium: as demand for intermediation falls, some wholesalers exit, survivors gain market share, and their markups can rise. With upstream production adjusting endogenously — and with some firms still relying on intermediaries — higher wholesale markups can distort relative input prices further, widen markup dispersion, and exacerbate resource misallocation, potentially offsetting gains from disintermediation.

This paper quantifies the extent to which increases in wholesale markups — induced by endogenous changes in market structure — offset the gains from technology-induced disintermediation.

I address this question with a model of production network formation featuring endogenous indirect trade shares, wholesale market structure, and markups. My framework builds upon the single-quality version of 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

[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).

[3]Examples include China's Two-Invoice System flattening pharmaceutical distribution to curb drug prices; India's regulated agricultural marketplaces (APMC "mandis"), which require first sale in designated yards through *licensed* traders with state-regulated commissions; and Bangladesh's ban on delivery-order traders in edible oils; see also small-scale interventions showing partial-equilibrium gains from technology adoption that facilitates disintermediation, such as Bartkus et al. (2022) (Amazon fishing cooperatives) and Iacovone and McKenzie (2022) (mobile-enabled direct sourcing of fruit and vegetables in Colombia).

[4]By contrast, Blum et al. (2023) argue for promoting the wholesale sector because wholesalers deepen production network and amplify the gains from lower trade costs; Perelló (2024) emphasizes improving access to intermediation to enhance supply-chain resilience.

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 adopting upstream input varieties. This assumption captures the notion that wholesalers help firms economize on the fixed costs associated with establishing direct supplier relationships, consistent with my empirical observations in Turkey that smaller firms, lacking the necessary scale to afford 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 fixed search and matching costs but benefit from lower per-unit input costs, whereas firms opting for wholesalers avoid search-related fixed 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.[5] 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

[5]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.

productivity and welfare. Together, these findings reveal a more nuanced view of the welfare implications of technology-induced disintermediation: while improved direct matching 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 technology-induced disintermediation raises wholesale markups. I assemble a province–pair panel from Turkey's VAT records, which report the value of all domestic 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 BOTAŞ 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 BOTAŞ's fiber optic infrastructure accelerated expansion. Because the pipeline network predates the internet rollout and was planned for energy logistics, proximity to these corridors 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 roll-out — 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, 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. ($\approx$30%). This underscores the value of complementary policies — such as wholesale subsidies — to mitigate markup distortions and fully harness the gains from digital infrastructure.

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.[6] 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.[7]

[6]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.

[7]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

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, 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.[8]

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)[9], 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

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.

[8] 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.

[9] For example, internet roll-out 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).

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 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 different institutions, creating a comprehensive database of all formal firms in the country. 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 (approximately 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 (sales from manufacturing firms, identified by 2-digit NACE codes 10-33, to other manufacturing firms) and indirect trade flows (sales from manufacturing firms through wholesale intermediaries, identified by 2-digit NACE code 46). I will also make use of each firm's sales, wage bill, costs of goods sold, and capital stock, contained in the income statements.

Throughout the analyses in this paper, I focus exclusively on the direct and indirect trade of manufacturing firms, resulting in a sample of 138,503 manufacturing firms and 73,543 wholesalers, for example, for the year 2019.[10] 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 over time

Figure 1 plots the aggregate indirect trade share in the domestic trade of manufacturing goods in

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

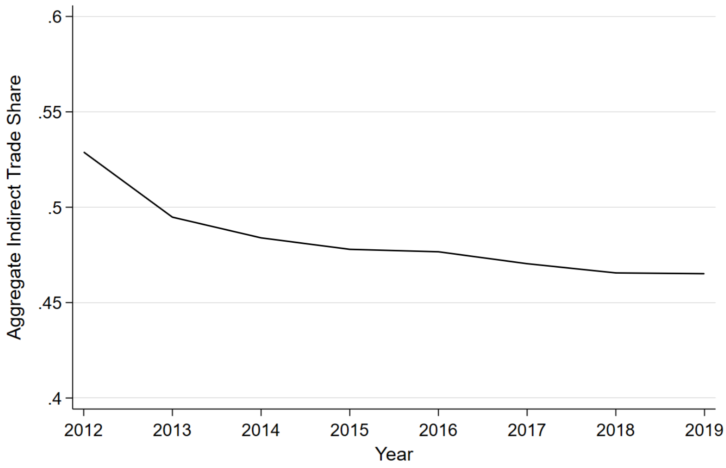


Figure 1: Evolution of Aggregate Indirect Trade Share for Turkey

Note: This chart plots the aggregate indirect trade share in Turkey between 2012 and 2019.

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 that indirect trade through wholesalers accounted for 53% of the domestic manufacturing goods trade in Turkey in 2012, which has been declining steadily over the sample period, reaching 46% in 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

Standard errors clustered at the province level in parentheses

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. Standard errors are reported 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), underscoring that 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 5 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%.[11] [12]

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

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

[12] This pattern motivates measuring province-pair indirect trade as upstream manufacturers' sales to wholesalers located in the downstream province. Because shipment-level tracking is unavailable, neither candidate proxy for province-pair indirect trade — (i) upstream manufacturers' sales to wholesalers in the downstream province, or (ii) downstream manufacturers' purchases from wholesalers in the upstream province — can recover true production origin or final-use location. However, given the strong locality of wholesale purchases, proxy (ii) would systematically miss upstream-origin goods that are resold by downstream wholesalers; proxy (i) is therefore the less biased approximation. To limit any overstatement from 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.

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, that rose over time

To estimate firm-level markups, I follow De Loecker and Warzynski (2012), 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.[13][14] I assume constant returns to scale (RTS = 1) following their baseline.

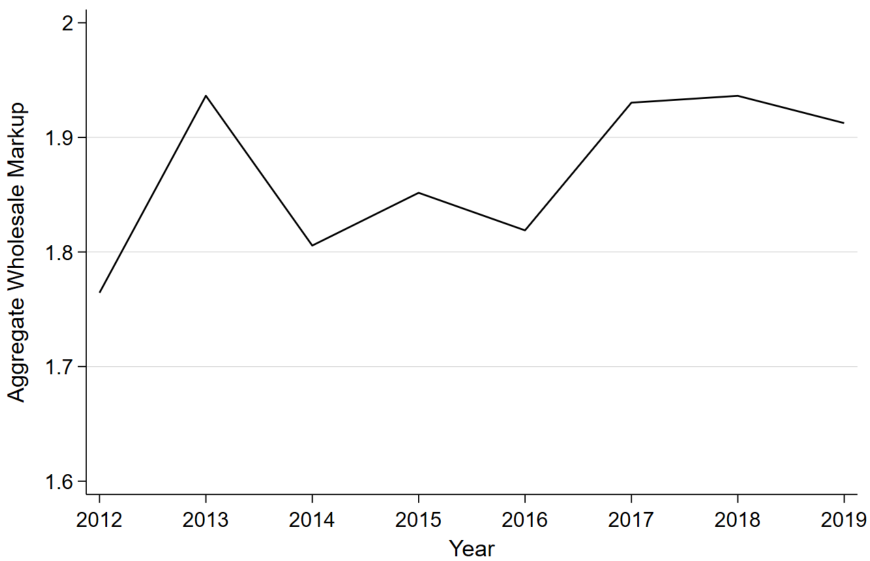


Figure 2: Evolution of Aggregate Wholesale Markup for Turkey

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.

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

[14]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 $r_t k_{it}(s)$.

Summary. These patterns suggest that forming direct buyer–supplier relationships entails sizable fixed costs that only sufficiently large firms can absorb. The high markups charged by wholesalers then 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 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 fixed costs are required for firms building direct connections, while wholesalers would charge a markup to cover the pooling of fixed 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 overtime trend.

3 Production Network with Wholesale Intermediation

Given the pivotal role of wholesale intermediation in manufacturing trade (over 50% of domestic trade flows) and the significant, rising market power documented above, I develop a production network model with endogenous wholesale intermediation, market structure, and markups. The framework clarifies how wholesale market power affects the efficiency of network formation and enables a quantitative assessment of the welfare consequences of disintermediation. This section presents a single-location version; Appendix A provides a spatial extension to incorporate heterogeneity in wholesale intensity and fiber roll-out for a richer aggregate welfare evaluation.

3.1 Model Overview

The model builds on the single-quality 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: firms can trade *indirectly* via wholesalers even without a direct match, and the wholesale market structure and markups are endogenized.

There is an exogenous labor supply L. 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 inputs. The elasticity of substitution σ is assumed to be identical across intermediate and final varieties. Households earn wage w (the numeraire)

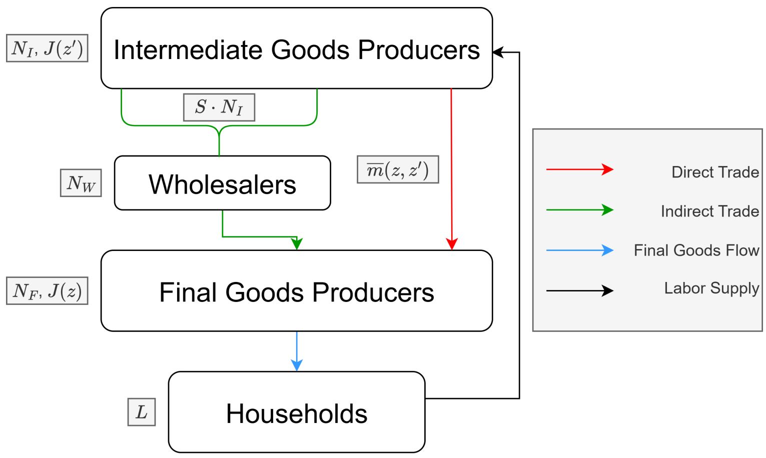


Figure 3: Graphical Illustration of the Production Network (Single Location)

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

and consume final goods.

There are two modes of trade, illustrated in Figure 3. First, firms can engage in direct trade by posting ads in matching markets and forming direct buyer-supplier matches. Once a match is formed, sellers (intermediate goods producers) post prices, and buyers (final goods producers) choose purchase quantities and pay posted prices. These direct trade flows are shown as red arrows in the figure.

Alternatively, firms can trade indirectly through wholesalers, shown by the green arrows in Figure 3. 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).[15] This assumption aligns with the stylized fact that smaller firms, which lack the scale to absorb fixed costs of direct sourcing, rely more on wholesalers.

To offer indirect trade, wholesalers must pay an additional fixed cost to adopt and distribute intermediate goods. Once adopted, the wholesaler buys the same variety from the producer and resells it to many buyers at a markup. This setup captures the idea that wholesalers economize on the fixed costs of relationship formation by aggregating demand and supply, and charge a markup to cover their fixed costs.[16] Wholesalers compete à la Cournot when reselling to buyers. This assumption allows markups to be determined endogenously — even when wholesalers carry identical varieties — 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 fixed search/matching costs but yields lower unit costs (and a customization gain that makes the same input more productive

[15]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.

[16]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.

in direct trade); indirect trade avoids fixed costs but entails higher unit costs due to wholesale markup.[17]

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.

In what follows, Section 3.2 describes the price-setting and ad-posting decisions of intermediate and final producers, taking as given demand shifters, input-bundle costs, and wages. Section 3.3 details the trading arrangement between firms: direct-match formation, wholesalers' product adoption and pricing, and buyers' cost minimization, which together determine demand shifters and input costs. Sections 3.4 and 3.5 present the household problem and the equilibrium conditions that close the model.

3.2 Pricing and Search Decisions

3.2.1 Problem of Final Goods Producers

Consider a final goods producer with productivity z. The firm is matched exogenously to all households and competes monopolistically against other final goods producers. Its revenue is $p_F^{1-\sigma} D_H$, where p_F is the firm's price, D_H is an endogenous demand shifter from household utility maximization, and σ is the elasticity of substitution across varieties.

The firm produces using a CES aggregate of intermediates sourced either directly or indirectly:

$$z\,Y_I(z) = z\left\{\int_{\omega\in\Omega(z)} y(\omega)^{\frac{\sigma-1}{\sigma}}\,\phi(\omega)^{\frac{1}{\sigma}}\,d\omega\right\}^{\frac{\sigma}{\sigma-1}}$$

where ω indexes varieties from set $\Omega(z)$ that the type z producer has access to. The productivity shifter $\phi(\omega)$ equals 1 for indirectly traded varieties and $\phi_c \geq 1$ for directly traded ones. $\Omega(z)$ is determined by the firm's direct matches and wholesalers' product adoption. Expanding by intermediate type z':

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

[17]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.

Here, $y_I(z')$ and $y^W(z')$ are the quantities of type z' inputs sourced directly and indirectly. $\overline{m}(z,z')$ is the measure of type z' intermediates directly matched to a type z final goods producer (endogenous via search/matching; Section 3.3.1). The term $N_I j(z')$ is 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 adopted by wholesalers; Section 3.3.3) gives the measure available indirectly. Because wholesalers' resale price exceeds the producer's direct price (double marginalization) and directly sourced inputs yield a customization gain $\phi_c^{1/\sigma} \geq 1$, it is strictly dominant to buy directly whenever a direct match exists.

Equation (1) highlights two departures from Arkolakis, Huneeus and Miyauchi (2023). First, varieties sold by wholesalers are the same intermediate varieties, so indirect trade competes with and can be cannibalized direct sales if a match is formed (as opposed to introducing entirely new varieties). 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.

Given the measures of direct/indirect matches and posted prices, the final producer chooses input quantities to minimize costs (Section 3.3.2). The resulting cost function for the intermediate bundle is

$$c(z) = \left[m(z)\, c_m^{1-\sigma} + S\, c_W^{1-\sigma} - m(z)\, S\, c_{Wm}^{1-\sigma}\right]^{\frac{1}{1-\sigma}}$$

Here, c_m is the cost of directly sourced inputs, scaled by $m(z)$ (the number of ads the firm posts to find direct suppliers) to capture variety gains. Posting m ads incurs a convex fixed cost $w\, f_m\, m^\beta/\beta$, where $f_m > 0$ and $\beta > 1$ govern the level and the curvature of the ad posting cost respectively. Moreover, direct matches also cannibalize indirect sourcing through wholesalers, which is captured by $m(z)\, S\, c_{Wm}^{1-\sigma}$. Meanwhile, c_W represents the cost function for potential indirect sourcing, before accounting for cannibalization. Both c_W and c_{Wm} are multiplied by S, the share of intermediate goods adopted by wholesalers, reflecting the variety gains associated with more product adoption.

The type z firm then chooses p_F and m to maximize profit:

$$\max_{p_F,\, m} \quad p_F^{1-\sigma} D_H \;-\; \frac{1}{z}\, c(z)\big(p_F^{-\sigma} D_H\big) \;-\; w f_m \frac{m^\beta}{\beta} \tag{2}$$

The FOC in p_F yields the standard monopolistic price: $p_F(z) = \frac{\sigma}{\sigma-1}\,\frac{1}{z}\, c(z)$, and the FOC in m is:[18]

$$m(z) = \Pi_{H1}\, z^{\frac{\gamma}{\beta}}, \qquad \Pi_{H1} \equiv \left[\frac{1}{w\, f_m\, \sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} D_H\left(c_m^{1-\sigma} - S\, c_{Wm}^{1-\sigma}\right)\right]^{\frac{1}{\beta-1}}, \qquad \gamma \equiv \frac{\beta\,(\sigma-1)}{\beta-1} \tag{3}$$

[18] Derivation in Appendix B.1.

Intuitively, $m(z)$ rises with the unit-cost reduction from an extra direct match net of cannibalization ($c_m^{1-\sigma} - S\, c_{Wm}^{1-\sigma}$), relative to the cost of posting ads. It also increases with demand D_H, and with productivity z. The firm's direct sourcing share :

$$\frac{m(z)\, c_m^{1-\sigma}}{m(z)\, c_m^{1-\sigma} + S\, c_W^{1-\sigma} - m(z)\, S\, c_{Wm}^{1-\sigma}},$$

which is increasing in $m(z)$. Hence more productive firms post more ads, form more direct matches, and have higher direct shares — consistent with Motivational Fact 3.

Finally, define the type z firm's revenue as: $x_H(z) \equiv p_F(z)^{1-\sigma} D_H$.

3.2.2 Problem of Intermediate Goods Producers

Consider an intermediate goods producer with productivity z. The firm hires l units of local labor at wage w to produce its unique intermediate variety with linear technology: $z\, l$.

The variety can be sold directly or indirectly. The producer takes as given some demand from wholesalers without incurring any search cost, as its variety could potentially be adopted and traded by wholesalers. To sell directly, the firm must form matches with final goods producers by posting v ads at convex cost $w\, f_v\, v^\beta/\beta$, with $f_v > 0$ and $\beta > 1$. The firm internalizes that a direct match cannibalizes wholesale sales of the same variety. Revenue from direct sales is:

$$p_I^{1-\sigma}\, v\, D_m$$

where D_m is a demand shifter for direct sales scaled by v to reflect higher demand from additional direct matches. Revenue from indirect sales is:

$$p_I^{1-\sigma}\, S\, D_W \;-\; p_I^{1-\sigma}\, v\, S\, D_{Wm}$$

where D_W is a pre-cannibalization demand shifter for wholesale sales. This demand shifter is multiplied by S, the share of intermediate goods adopted by wholesalers, to reflect the expected demand faced by the intermediate goods producer through indirect trade. The second term captures indirect sales cannibalized by direct matches. These demand shifters and wholesalers' adoption are taken as given by the producer but are determined endogenously by search and matching, wholesalers' optimization, and buyers' cost minimization.

A type z intermediate goods producer chooses price p_I and ads v to maximize profit:

$$\max_{p_I,v} \; p_I^{1-\sigma}\big(v\,D_m + S\,D_W - v\,S\,D_{Wm}\big) \\ - \; \frac{w}{z}\Big[p_I^{-\sigma}\big(v\,D_m + S\,D_W - v\,S\,D_{Wm}\big)\Big] \; - \; w\,f_v\,\frac{v^{\beta}}{\beta} \tag{4}$$

where the first line is revenue, followed by production cost and the search cost.

FOC w.r.t. p_I and v are:[19]

$$p_I(z) = \frac{\sigma}{\sigma-1}\,\frac{w}{z}, \qquad v(z) = \left[\frac{x_m(z)}{w\,f_v\,\sigma} - \frac{x_{Wm}(z)}{w\,f_v\,\sigma}\right]^{\frac{1}{\beta}} \tag{5}$$

where $x_m(z) \equiv p_I(z)^{1-\sigma}\,v(z)\,D_m$ is revenue from direct sales, $x_{Wm}(z) \equiv p_I(z)^{1-\sigma}\,v(z)\,S\,D_{Wm}$ is indirect sales cannibalized by direct sales, and $x_W(z) \equiv p_I(z)^{1-\sigma}\,S\,D_W$ is expected indirect sales before cannibalization.

Intermediate goods producers charge the same price when selling directly or via wholesalers. The wholesale sector competes monopolistically against other input varieties — mirroring the environment in direct sales — so producers internalize the effect of their price on wholesale quantities and set the same monopolistic markup. The model therefore does not feature channel-specific markups.[20]

3.3 Trading

In this section, I discuss the organization of trade between firms. I first model the search and matching market for direct trade, which delivers the total mass of direct matches. Next I examine the cost minimization problem of final goods producers, who choose quantities from directly matched suppliers and, when no direct match exists but a variety is adopted by wholesalers, purchase it indirectly. Final goods producers take wholesalers' adoption and pricing as given, to which I return at the end of the section.

3.3.1 Search and Matching

There is a single direct matching market. The total measures of ads by buyers (M) and sellers (V) are:

[19] A detailed derivation is provided in Appendix B.2.

[20] This is consistent with Motivational Fact 6 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 or enforced ceiling, 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.

$$M = N_F \int_Z m(z)\, j(z)\, dz \quad (6) \qquad V = N_I \int_Z v(z)\, j(z)\, dz \quad (7)$$

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 for sellers θ^v and buyers θ^m as:

$$\theta^v = \frac{\widetilde{M}}{V} = \kappa\, V^{\lambda_V - 1} M^{\lambda_M} \quad (8) \qquad \theta^m = \frac{\widetilde{M}}{M} = \kappa\, V^{\lambda_V} M^{\lambda_M - 1} \quad (9)$$

The mass 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 mass of type z' buyers matched to a type z supplier is:

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

3.3.2 Final Goods Producers' Cost Minimization

Next, I study the cost minimization problem of a type z' buyer, taking as given the number of direct matches formed $\{\overline{m}(z', z)\}$, the share of products adopted by wholesalers $\{S\}$, as well as the prices charged by the direct supplier $\{p_I\}$ and the wholesalers $\{p^W\}$, to derive demand shifters for intermediate goods producers and wholesalers, and also the unit cost of a bundle of intermediate inputs.

Formally, the cost minimization problem of a type z' buyer is:

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

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

Recall $y_I(z)$ is the quantity of a distinct type z input variety purchased directly by the type z' buyer, and $y^W(z)$ is the quantity of a distinct type z input variety purchased through wholesalers when no direct match exists and the variety is adopted. The prices for directly and indirectly sourced varieties are denoted by p_I and p^W, respectively. The type z' buyer chooses $\{y_I, y^W\}$ to minimize the cost of producing one unit of final good.

As shown in Appendix B.3, FOCs w.r.t. $y_I(z)$ and $y^W(z)$ and the constraint yield the direct sales of a type z intermediate goods producer to a type z' final goods producer:

$$p_I(z)\, y_I(z) = \phi_c \left[\frac{p_I(z)}{c(z')}\right]^{1-\sigma} \frac{\sigma-1}{\sigma}\, x_H(z') \tag{10}$$

multiplied by the measure of direct matches for the type z intermediate producer yields total direct sales:

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

where the demand shifter is:

$$D_m \equiv \phi_c \frac{\sigma-1}{\sigma} \frac{\theta^v}{M} N_F \int_Z \frac{x_H(z)}{c(z)^{1-\sigma}}\, m(z)\, j(z)\, dz \tag{11}$$

Similarly, the indirect sales of a type z intermediate goods variety through wholesalers to a type z' buyer — conditional on being adopted by the wholesalers and that the buyer and the supplier do not have a direct match — is :

$$p^W(z)\, y^W(z) = \left[\frac{p^W(z)}{c(z')}\right]^{1-\sigma} \frac{\sigma-1}{\sigma}\, x_H(z') \tag{12}$$

which will be used in the next section to derive final producers' demand for indirectly traded varieties.

Combining the two FOCs with the constraint delivers the unit cost of the intermediate bundle for the type z' final goods producer:

$$c(z') = \left[m(z')\, c_m^{1-\sigma} + S\, c_W^{1-\sigma} - m(z')\, S\, c_{Wm}^{1-\sigma}\right]^{\frac{1}{1-\sigma}} \tag{13}$$

where

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

The optimal ad posting conditions (5) and (3) can be combined with (11) and (13) to solve for the ad success rates, with details in Appendix B.4.

3.3.3 Profit Maximization of Wholesalers

I now characterize wholesalers' profit maximization to determine the set of adopted varieties, resale prices, and wholesalers' demand for intermediates

There is a discrete number N_W of homogeneous wholesalers that adopt intermediate varieties and resell them to final goods producers. Each wholesaler chooses adoption effort s, trading off the profit earned from reselling the adopted varieties against the cost of adoption, as well as the resale price of each adopted variety $p^W(z)$. Wholesalers compete in Cournot when reselling the same adopted variety, which generates markup endogenously. The Cournot assumption is motivated by Motivational Fact 4, where few wholesalers dominate local markets, and it is essential for rationalizing Motivational Fact 5 — the rising wholesale markups accompanying disintermediation observed in Turkey.

We can write the inverse demand for reselling a type z variety to a type z' final goods producer (when the two firms are not directly matched with each other) as:

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

total quantity of the type z variety resold is the sum across N_W wholesaler: $y^W(z) = \sum_{i=1}^{N_W} y^{Wi}(z)$.

Conditional on adoption, wholesaler i chooses $y^{Wi}(z)$ to maximize profit:

$$\max_{y^{Wi}(z)} p^W(z)\, y^{Wi}(z) - p_I(z)\, y^{Wi}(z)$$

FOC[21]

$$p^W(z) = \frac{N_W\, \sigma}{N_W\, \sigma - 1} p_I(z) = \frac{N_W\, \sigma}{N_W\, \sigma - 1} \frac{\sigma}{\sigma - 1} \frac{w}{z} \tag{14}$$

so wholesalers charge an additional markup that decreases in N_W.

Given optimal pricing, the profit for wholesaler i from reselling an adopted type z variety to an unmatched type z' buyer is:

$$\begin{aligned} & p^W(z)\, y^{Wi}(z) - p_I(z)\, y^{Wi}(z) \\ =& \frac{1}{N_W} \frac{1}{N_W\, \sigma} \left[\frac{N_W\, \sigma}{N_W\, \sigma - 1} p_I(z) \right]^{1-\sigma} \frac{\sigma - 1}{\sigma} \frac{x_H(z')}{c(z')^{1-\sigma}} \end{aligned}$$

[21] A detailed derivation is provided in Appendix B.5

Aggregating over all unmatched buyers gives:

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

where $\overline{D^W}(z) \equiv D^W - v(z) D_m^W$ and

$$
D^W = \frac{\sigma-1}{\sigma}\,N_F \int_Z \frac{x_H(z')}{c(z')^{1-\sigma}}\,j(z')\,dz' \tag{15}
$$

$$
D_m^W = \frac{\sigma-1}{\sigma}\,\frac{\theta^v}{M}\,N_F \int_Z \frac{x_H(z')}{c(z')^{1-\sigma}}\,m(z')\,j(z')\,dz' = \phi_c^{-1}\,D_m < D_m \tag{16}
$$

I denote wholesalers' demand shifters by a superscript W to distinguish them from D_W and D_{Wm}, which are the shifters faced by intermediate producers arising from wholesale trade. If each adopted variety is drawn randomly from $j(z)$, the expected per-variety profit for a wholesaler is:

$$
\Pi_W \equiv \int_Z \frac{1}{N_W} \frac{1}{N_W\,\sigma} \left[\frac{N_W\,\sigma}{N_W\,\sigma-1}\,p_I(z)\right]^{1-\sigma} \overline{D^W}(z)\,j(z)\,dz
$$

I model adoption as one-sided matching. Analogous to firms' search for direct matches, wholesalers exert effort s to adopt varieties. To allow congestion, only a fraction θ^W of effort converts into adoption:

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

so $s\,\theta^W\,N_I$ is the number of varieties adopted by each homogeneous wholesaler. All wholesalers adopt the same set of varieties; there is no variation across them in this regard.

A wholesaler chooses s to maximize profit, trading off the profit made from additional adopted varieties against the cost of adoption:

$$
\max_s\ s\,\theta^W\,N_I\,\Pi_W - w\,f_W\,\frac{(s\,N_I/N_W)^{\beta_W}}{\beta_W}
$$

with $f_W > 0$ and $\beta_W > 1$. Adoption cost rises with the number of varieties adopted $(s\,N_I)$ and falls with N_W. The latter assumption is imposed to capture potential knowledge spillover among wholesalers that reduces information friction inhibiting the adoption of upstream varieties. As discussed in 4.1, this implies efficient wholesaler entry in the decentralized equilibrium when there is no congestion in adoption $(\lambda_W = 1)$. FOC is:

$$
S = \left(\frac{X^W - X_m^W}{\sigma\,w\,f_W}\right)^{\frac{\lambda_W}{\beta_W}} N_I^{-\lambda_W}\,N_W^{\lambda_W\left(2-\frac{2}{\beta_W}-\frac{1}{\lambda_W}\right)} \tag{17}
$$

where

$$
\begin{aligned}
S &\equiv s\,\theta^W \\
X^W \equiv N_W\, x^W &= S\, N_I \int_Z \left[\frac{N_W\,\sigma}{N_W\,\sigma - 1}\, p_I(z)\right]^{1-\sigma} D^W\, j(z)\, dz \\
X_m^W \equiv N_W\, x_m^W &= S\, N_I \int_Z \left[\frac{N_W\,\sigma}{N_W\,\sigma - 1}\, p_I(z)\right]^{1-\sigma} v(z)\, D_m^W\, j(z)\, dz
\end{aligned}
$$

so $x^W - x_m^W$ is the revenue for each wholesaler and $X^W - X_m^W$ is the aggregate across wholesalers.

Having solved wholesalers' profit maximization problem, we may now proceed to derive the total sales of intermediate goods producers to wholesalers, given by $X_W - X_{Wm}$:

$$
X_k \equiv N_I \int_Z x_k(z)\, j(z)\, dz \quad , \quad k = W, Wm
$$

and $x_W(z) - x_{Wm}(z)$ is the revenue of a type z intermediate producer from indirect sales. Conditional on adoption, the total quantity of type z resold is:

$$
p^W(z)^{-\sigma}\, \overline{D^W}(z) = \left[\frac{N_W\,\sigma}{N_W\,\sigma - 1}\, p_I(z)\right]^{-\sigma} \overline{D^W}(z)
$$

so sales of type z intermediate producer to wholesalers, conditional on it being adopted, are:

$$
\begin{aligned}
&p_I(z) \left[\frac{N_W\,\sigma}{N_W\,\sigma - 1}\, p_I(z)\right]^{-\sigma} \overline{D^W}(z) \\
&= p_I(z)^{1-\sigma}\, D_W - p_I(z)^{1-\sigma}\, v(z)\, D_{Wm}
\end{aligned}
$$

with

$$
D_W = \left(\frac{N_W\,\sigma}{N_W\,\sigma - 1}\right)^{-\sigma} D^W \tag{18}
$$

$$
D_{Wm} = \left(\frac{N_W\,\sigma}{N_W\,\sigma - 1}\right)^{-\sigma} D_m^W = \left(\frac{N_W\,\sigma}{N_W\,\sigma - 1}\right)^{-\sigma} \phi_c^{-1}\, D_m < D_m \tag{19}
$$

Thus an intermediate producer internalizes both the effect of its price on wholesale quantities and the cannibalization from its ad posting.

There is a natural upper bound on adoption: $S\, N_I \leq N_I$. Consequently, $S \leq 1$ can be interpreted as the share or probability a variety is adopted and traded by wholesalers, and is therefore multiplied with D_W and D_{Wm} to capture expected wholesale demand in the intermediate producers' problem in 3.2.

Finally, it can be shown that:

$$X^W = \frac{N_W\,\sigma}{N_W\,\sigma - 1}\,X_W = \mu^W\,X_W \tag{20}$$

$$X^W_m = \frac{N_W\,\sigma}{N_W\,\sigma - 1}\,X_{Wm} = \mu^W\,X_{Wm} \tag{21}$$

where $\mu^W \equiv \frac{N_W\,\sigma}{N_W\,\sigma-1}$ is the wholesale markup

3.4 Household's Problem

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 wages and local wholesalers' net profit Π^W:

$$I = w\,L + \Pi^W \tag{22}$$

Utility maximization yields the price index:[22]

$$P^H = \left[N_F \int_Z p_F(z)^{1-\sigma}\,j(z)\,dz\right]^{\frac{1}{1-\sigma}} \tag{23}$$

and demand for final goods producers:

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

3.5 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' product adoption, and for financing entry of firms and wholesalers. Details of these equilibrium conditions are provided in Appendix B.7.

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 (22), (73), (8), (9), (17), (69), (70), and (72).

[22] A detailed derivation is provided in Appendix B.6

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: a markup wedge from double marginalization in wholesale trade and a congestion externality from uncoordinated match formation. I then show how the first-best allocation can be restored from the decentralized equilibrium using a combination of a wholesale subsidy and taxes/subsidies on match and firm formation.

Next, I compare the decentralized allocation to the first-best benchmark 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: as firms increasingly bypass intermediaries, some wholesalers exit, raising concentration and market power. The resulting increase in markups can exacerbate misallocation and widen the gap between decentralized and efficient allocations.

4.1 Social Planner's Problem

I study the social planner's problem to identify inefficiencies in the model. 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' adoption effort s, and entry levels N_I, N_F, and N_W to maximize aggregate household consumption[23]. 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 isolate these sources of inefficiency.

Wedge in input quantity. Combining the planner's FOC 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 FOC implies:

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

Conditional on the matches, the quantity of indirectly traded inputs is inefficiently low as whole- saler's markup inflates the price of indirectly traded inputs relative to the directly traded ones.

[23] A detailed setup is in Appendix B.9.

Wedge in ad posting. Double marginalization distorts not only the intensive margin but also match formation. The FOC in $v(z)$ can be written as

$$\text{MSC}(v(z)) = \text{MPC}(v(z)) = \text{MPB}(v(z)) = \text{MSB}(v(z)) \underbrace{\Delta_{markup}(v(z))}_{>1} + \underbrace{\Delta_{congestion}(v(z))}_{\geq 0}$$

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. $\Delta_{markup}(v(z))$ reflects how wholesalers' markup lowers the price of directly traded inputs relative to indirect ones, inflating the private return to ads. $\Delta_{congestion}(v(z))$ captures the failure to internalize how additional ads reduce market tightness and ad success for all, again overstating MPB. The same logic applies to $m(z)$.

Wedge in product adoption. Wholesalers' markup also distorts adoption incentives:

$$\text{MSC}(s) = \text{MPC}(s) = \text{MPB}(s) = \text{MSB}(s) \underbrace{\Delta_{markup}(s)}_{\geq 1/\leq 1} \underbrace{\Delta_{congestion}(s)}_{\leq 1}$$

$\Delta_{markup}(s)$ captures how wholesaler's markup raises the relative price of indirectly traded intermediate good, reducing wholesalers' revenue from adopted varieties[24] and dampening the private return to adoption. It also captures misalignment between private and social incentives to add varieties: 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 under Cournot wholesalers capture only $1/[N_W(\sigma - 1)]$, further lowering MPB relative to MSB except when $N_W = 1$. $\Delta_{congestion}(s)$ reflects uninternalized congestion in adoption that inflates MPB.

Wedge in firm and wholesaler entry. For entry, both intermediate producers and wholesalers fail to internalize the congestion externalities that lower match and adoption success rate. In addition, intermediate producers also fail to internalize that their entry raises wholesalers' adoption costs:

$$\text{MSC}(N_W) = \text{MPC}(N_W) = \text{MPB}(N_W) = \text{MSB}(N_W) + \underbrace{\Delta_{congestion}(N_W)}_{\geq 0}$$

$$\text{MSC}(N_I) = \text{MPC}(N_I) = \text{MPB}(N_I) = \text{MSB}(N_I) + \underbrace{\Delta_{congestion}(N_I)}_{\geq 0} + \underbrace{\Delta_{AdoptionCost}(N_I)}_{<0}$$

[24]This is not fully offset by the induced rise in the price index, since indirect trade is only a fraction of total trade and the markup applies only to indirect inputs.

For N_F the FOC implies

$$\text{MSC}(N_F) = \text{MPC}(N_F) = \text{MPB}(N_F) = \text{MSB}(N_F)\ \underbrace{\Delta_{markup}(N_F)}_{>1} + \underbrace{\Delta_{congestion}(N_F)}_{\geq 0}$$

Here $\Delta_{markup}(N_F)$ arises because 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 entry. $\Delta_{congestion}(N_F)$ again reflects uninternalized matching congestion.

To summarize, two main forces drive inefficiency: (1) 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) congestion externalities from ads and adoption that agents fail to internalize, leading to excessive ad posting and adoption.

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 wholesale adoption 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[25]

Proposition 1 (Optimal Policy)**.** *The optimal gross taxes and subsidies restoring the first-best from the decentralized equilibrium satisfy*

$$\tau^W = (\mu^{W*})^{-1}\ ,\ \tau_v^M = \frac{z^{\sigma-1}}{z^{\sigma-1} + \lambda_V - 1}\ ,\ \tau_m^M = \frac{\sigma}{\sigma - 1}\,\frac{z^{\sigma-1}}{z^{\sigma-1} + \lambda_M - 1}\ ,\ \tau_s^M = \frac{1}{N_W{}^*\,\lambda_W}$$

$$\tau_I^E = \frac{1 - \frac{\psi^*}{\beta}}{(1 - \frac{\psi^*}{\beta}) + \left[\psi^*(\lambda_V - 1) - \frac{\psi^*}{\beta}(\lambda_V - 1)\frac{V^*}{V_Z{}^*}\right] - (1 - \Omega^*)\left(\frac{S^*}{\theta^{W*}}\right)^{1-\lambda_W}\lambda_W}$$

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

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.

Wholesale subsidy. The optimal wholesale subsidy equals the inverse of the efficient wholesale markup, exactly offsetting double marginalization and realigning the relative price of direct and indirect inputs

[25]The proof of Proposition 1 is in Appendix B.10.

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 private incentive and the social incentive to enter. Both private and social incentives fall as the net direct share ψ^* rises, since higher ψ^* raises search costs relative to variable profits by ψ^*/β, but the social incentive falls more because the planner internalizes congestion in matching. With $\beta > 1$, congestion unambiguously lowers the net social benefit of entry, so the optimal tax must fully internalize these externalities[26]

Tax/subsidy on wholesale product adoption. 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 $(\mu^W - 1)/(\mu^W\sigma) = 1/{N_W}^*$, and congestion in adoption, which calls for a tax of $1/\lambda_W$. The net policy can be a tax or a subsidy depending on which force dominates.

Tax/subsidy on wholesaler entry. Like τ_I^E and τ_F^E, the wholesaler entry instrument τ_W^E equals the private-to-social incentive ratio for entry:

$$\text{Private: } \frac{\beta_W - 1}{\beta_W}\tau_s^M \times \text{Adoption Cost} \qquad \text{Social: } \frac{\beta_W\left(\frac{2\lambda_W - 1}{\lambda_W}\right) - 1}{\beta_W} \times \text{Adoption Cost}$$

The private incentive is variable profit net of adoption cost, while the social incentive also subtracts the congestion effect when $\lambda_W < 1$[27].

The subtle but critical insight is that while the product adoption tax/subsidy τ_s^M is designed to correct inefficiencies in product adoption, it creates an indirect distortion in wholesaler entry. Specifically, *conditional on achieving the efficient adoption rate*, an adoption subsidy (tax) gives wholesalers an excessively low (high) private incentive to enter. Without a corrective tax/subsidy on wholesaler entry, the

[26]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' adoption costs.

[27]Specifically, a fraction $(1-\lambda_W)/\lambda_W$ of variable profit.

equilibrium entry would be inefficiently low (high). This conditional perspective is key to understanding why the product adoption tax/subsidy rate τ_s^M appears explicitly in the entry tax formula ($\frac{1}{N_W{}^* \lambda_W}$).

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[28].

Proposition 2 (Aggregate productivity under exogenous network)**.** *The first-best aggregate productivity $A_{\textit{efficient}}$ and its decentralized counterpart A, defined as the ratio of total production of the consumption aggregate to labor employed net of fixed costs, are*

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

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

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

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

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

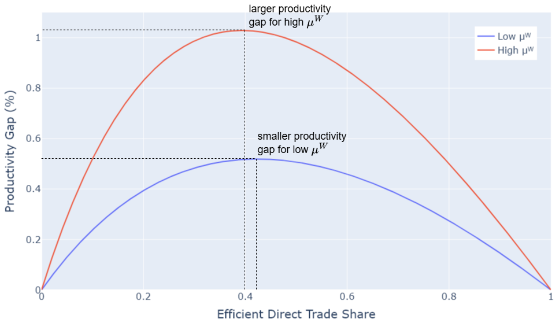
Proposition 2 shows that, holding matches and entry at their planner levels, decentralized productivity falls short of the first-best due to dispersion in markups between direct and indirect inputs. As in Edmond, Midrigan and Xu (2015), the aggregate markup μ is a revenue-weighted harmonic mean across these inputs. This dispersion reflects the wholesale-markup-induced relative price distortion, which reduces allocative efficiency at the intensive margin and vanishes only when $\mu^W \to 1$.

Corollary 2.1 in Appendix B.12 establishes that the productivity gap rises with the wholesale markup and, to first order, with a measure of markup dispersion weighted by the *efficient* direct trade share.

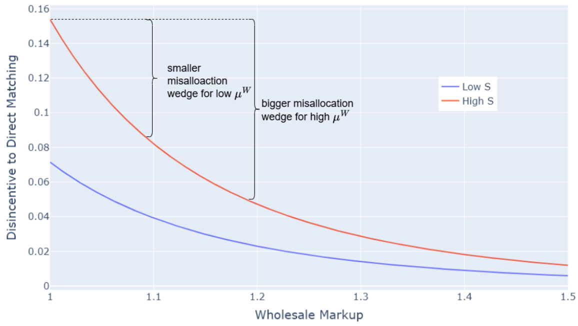
Figure 4a shows that, as κ increases and the efficient direct share rises, the productivity gap first widens, peaks at an intermediate share, and then declines. This pattern is summarized by an exact index — *misallocation markup dispersion*[29] — defined using the efficient direct trade share. Dispersion is

[28] The proof of Proposition 2 is provided in Appendix B.11.

[29] Define $\mathcal{D}_\sigma(\mu^W,\Omega^*) \equiv -\frac{1}{\sigma}\ln\left[\Omega^*(\frac{\mu}{\mu_D})^\sigma + (1-\Omega^*)(\frac{\mu}{\mu_I})^\sigma\right]$, with $\mu_D = (\sigma/(\sigma-1))^2$, $\mu_I = \mu_D\,\mu^W$, $\mu = (\frac{\Omega}{\mu_D} + \frac{1-\Omega}{\mu_I})^{-1}$, and $\Omega = \frac{\Omega^*}{\Omega^*+(\mu^W)^{1-\sigma}(1-\Omega^*)}$. It links exactly to the productivity wedge via $\ln\Theta = -(\sigma/(\sigma-1))\,\mathcal{D}_\sigma$



(a) Aggregate Productivity Gap (Exogenous Matching)



(b) Misallocation Wedge (Endogenous Direct 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

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 4: Aggregate Productivity Gap and Misallocation Wedge

minimized when the direct share is very low or very high and peaks at $\Omega^* < 1/2$.

Intuition for the peak below one-half. The wedge 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 $\Omega^* < 1/2$[30].

For any direct trade share, a higher μ^W strictly raises dispersion — and thus the productivity gap — with both collapsing only when $\mu^W = 1$. A larger markup magnifies the relative price distortion and shifts the peak toward lower Ω^*.

Wholesale markups also distort the extensive margin. The next result combines intensive and extensive margins to show the welfare loss[31].

Proposition 3 (Aggregate welfare with endogenous direct matching)**.** *In a simplified single-location economy with $z = 1$, $\beta = 2$, $\lambda_M = \lambda_V = 1$, and a tax on supplier search $\frac{\sigma-1}{\sigma}$[32], let $\widetilde{L}$ be labor net of fixed entry and adoption costs. Define the misallocation wedge*

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

[30] 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.

[31] The proof of Proposition 3 is provided in Appendix B.13.

[32] The tax removes firms' double marginalization as in Proposition 1.

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

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

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

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

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

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

When we abstract from the allocation of labor to firm entry and wholesalers' product adoption, 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. However, this effect is dampened, as indicated by the coefficient $1/\sigma$, as a larger number of indirect matches discourages direct match formation, thereby muting the increase in aggregate productivity.

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 3.1 in Appendix B.14 shows the welfare gap rises with μ^W and, to first order, with markup dispersion.

Figure 4b 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 effective productivity of indirectly traded inputs, and this understatement bites more when a larger share of varieties is sourced indirectly.

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[33].

Proposition 4. *The number of wholesalers is a function of the direct share $\Omega \equiv \frac{X_m}{X_m+X^W-X_m^W}$:*

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

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

Intuitively, a fall in the fixed cost of direct matching 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.

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 share plays the analogous role: at low direct shares, disintermediation can widen dispersion; at high shares, further disintermediation compresses dispersion and improves allocation.

[33]The proof of proposition 3 is provided in Appendix B.8.

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 roll-out 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 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.

This paper follows the empirical strategy of Demir, Javorcik and Panigrahi (2023), leveraging temporal variation in fiber internet access across Turkish provinces to examine the effect of digital infrastructure on wholesale intermediation. The empirical contribution of this paper mainly lies in the investigation of the causal impact of fiber internet expansion on a distinct set of outcome variables associated with wholesale intermediation, contrasting with the focus of Demir, Javorcik and Panigrahi (2023) on manufacturing firms' direct sourcing. These new empirical findings lend support to the model's mechanism, and inform the inference of shocks for the quantitative exercise.

5.1.2 Data

Five sets of data are used for the empirical analysis: (1) ICTA data on fiber internet infrastructure; (2) Turkish firm-level data and firm-to-firm transaction data; (3) Map of BOTAŞ oil and natural gas pipeline network; (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.

ICTA Data on Fiber Internet Infrastructure. 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 BOTAŞ Oil and Natural Gas Pipeline Network. To construct an instrument for fiber connectivity, I digitize the map of BOTAŞ oil and natural pipeline network, 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 roll-out 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 ($I_{d,t}$):

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

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.

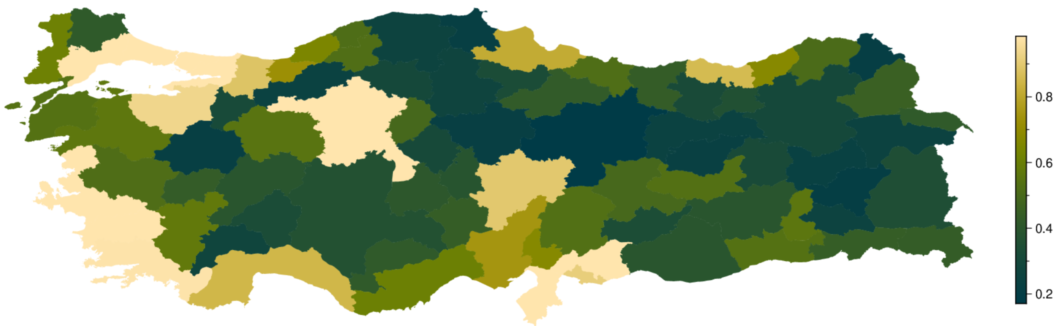


Figure 5: Change in standardized fiber intensity between 2012-2019

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 5 shows the change in standardized fiber intensity[34] 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.

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

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

state-owned energy distributor, as an instrumental variable 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 roll-out of fiber internet. In particular, provinces closer to existing pipelines naturally experienced earlier and cumulatively greater fiber roll-out. Thus, 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 to capture this cumulative dynamic:

$$Z_{ud} = \max\{Z_u, Z_d\}, \qquad Z_i = \sum_{j \in i} \frac{\text{Population}_{j,2011}}{\text{Population}_{i,2011}} \times Z_j^D$$

where each province's distance, Z_i, is calculated as the population-weighted average distance of its districts to the nearest oil pipeline. Z_j^D is the distance from district j to the nearest pipeline, using 2011 district population shares as weights.

Figure D.2 in Appendix D.2 presents the first-stage regression estimates, regressing fiber connectivity on pipeline distance interacted with year dummies. The coefficients are negative and become increasingly negative over time, consistent with the expectation that provinces closer to pipelines experienced faster and cumulative growth in fiber connectivity.

Regarding the exclusion restriction, it requires that proximity to historical pipeline infrastructure influences provincial trade outcomes solely through its effect on fiber connectivity. 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. Crucially, 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.

The plausibility of the exclusion restriction is further supported by visual evidence. Figure D.3 in Appendix D.2 overlays the BOTAŞ pipeline network, while Figure 5 shows the actual change in fiber intensity. Although some provinces located near pipelines experienced rapid fiber growth, others — especially in the eastern region — did not. This lack of mechanical co-movement suggests that pipeline proximity is not simply a proxy for latent economic fundamentals or province-level development priorities. Instead, it acts as a legacy infrastructure constraint that shapes the cost and feasibility of subsequent fiber deployment.

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

	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. Standard errors clustered at the province-pair level are reported in parentheses.

Table 3 presents the results of regressing inter-provincial trade outcomes on standardized fiber connectivity. The 2SLS estimates in Panel B, Column 1 show that province pairs with greater fiber connectivity experience a statistically significant reduction in the share of trade flows intermediated by wholesalers. This provides evidence that improvements in digital infrastructure lead to **disintermediation in production network**. Indirect trade share is measured as the share of total bilateral manufacturing trade in which the downstream buyer is a wholesaler.

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 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.[35]

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 outcomes.

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

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,[36] indicating an increase in concentration. Column 3 confirms that the result holds when using a firm-specific fiber intensity measure, 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

[35] 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.

[36] 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.

	Concentration			Markup		
	Log Wholesaler Number	Market Share	Market Share	Agg Wholesale Markup	Markup	Markup
Panel A: OLS						
Std Fiber Intensity	-0.048	0.0004		0.017	0.0189	
	(0.019)	(0.0002)		(0.049)	(0.0054)	
Std Fiber Intensity (Firm-specific)			0.0001			0.1243
			(0.0001)			(0.0528)
Panel B: 2SLS						
Std Fiber Intensity	-0.337	0.0010		0.617	0.0227	
	(0.208)	(0.0006)		(0.354)	(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 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. Standard errors (in parentheses) are clustered at the province level.

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 are economically meaningful and consistent across specifications.[37]

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.

[37] 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.

6 Quantitative Exercise

This section presents two key quantitative exercises using the calibrated model: (i) a decomposition of the welfare costs of inefficiencies, and (ii) an evaluation of the impact of fiber internet expansion. Before presenting the results, I first detail the model's calibration.

6.1 Calibration

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	Product adoption rate	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.

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. The wage w is chosen as the numeraire, and L, F_I, F_F, f_m, and f_v are normalized to 1. 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) A final-goods producer's number of direct suppliers is proportional to its total direct purchases raised to the power $1/\beta$. 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 $1/\beta_W$, 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

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

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.[38] 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.

Dep. Variable:	Total Trade	Direct Trade	Indirect Trade	Direct Ext. Margin	Indirect Ext. Margin
Tariff	-4.875 (1.433)	-4.422 (1.669)	-5.773 (2.263)	-2.509 (1.052)	-1.276 (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. 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 di-

[38] 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.

rect 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 product adoption cost f_W is calibrated to match the observed rate of product adoption by wholesalers. Recall that in the model, the product adoption rate S captures the share of upstream products adopted and 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 product adoption rate as the average province-pair product adoption rate 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)[39] *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

[39]*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.

a calibrated aggregate wholesale markup of 1.13.[40]

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 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.17 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.[41] 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.

[40] While the data contain thousands of wholesalers, they do not all compete head-to-head. Manufacturing firms' sourcing from wholesalers is highly localized (Fact 1 and Figure **??**), 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.

[41] 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.

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 direct matching and product adoption — measured as the welfare gain from eliminating them. Recall Proposition 1: a wholesale subsidy equal to the inverse of the wholesale markup, together with a set of taxes/subsidies on direct matching, product adoption, and 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, while Table C.1 in Appendix C.1 reports the results for different alternative parameterizations to highlight the key mechanisms generating inefficiency and to illustrate the sensitivity of the welfare decomposition to perturbations of parameter values. Within each parameterization, I report the level of selected variables and welfare relative to the decentralized equilibrium — for three cases: (i) the decentralized equilibrium, (ii) the equilibrium with only the wholesale subsidy, and (iii) the efficient allocation with the full set of optimal policies. I organize the discussion of the results around each parameterization.

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											
Efficient	0.2200	1.0000	0.0000	0.0435	0.3740	0.3556	0.0230	0.0270	0.2818	0.6682	1.1000
Wholesale subsidy	0.1800	1.0000	0.0000	2.2730	0.2539	0.3914	0.0261	0.0194	0.3895	0.5650	1.0237
Decentralized	0.4800	1.1299	0.0037	4.6653	0.2630	0.3852	0.0799	0.0129	0.3465	0.5608	1.0000

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.

Baseline parameterization. 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.[42] Recall Propositions 2 and 3: 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

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

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 product adoption (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 product adoption (Column 8) also increases, despite the decrease in S, due to the larger number of intermediate goods producers (recall that the product adoption cost rises with both adoption 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 0.024 log points.

If we further implement the matching and entry taxes/subsidies described in Proposition 1, welfare increases by an additional 0.072 log points.[43] These additional instruments address congestion externalities in direct matching and product adoption, as well as double marginalization by manufacturers, by reallocating labor away from match and firm formation toward production. The 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.

Low Ω parameterization. I also 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 2.1 and 3.1: 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

[43] 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.

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.

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 2.1 and 3.1: the degree of misallocation induced by wholesale markups is increasing in their level.

High S parameterization. The next parameterization halves the product adoption cost shifter f_W, raising the equilibrium rate of product adoption 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 3, which implies that the amount of excessive labor allocated to the creation of direct matches increases with the misallocation wedge, itself increasing in the product adoption rate 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 0.091 log points — 0.019 log points 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 0.002 log point. 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 wholesale product adoption — 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 0.007 log points relative to the baseline, reflecting a stronger congestion externality in wholesale product adoption. This effect, however, occurs only when $N_W = 1$, as in the efficient allocation. Recall from Proposition 1 that the optimal tax/subsidy on product adoption 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 adopt 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 product adoption. 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 product adoption. Consequently, the welfare gain from introducing the wholesale subsidy increases slightly, by about 0.0007 log points. Intuitively, the wholesale markup interacts with product adoption inefficiency by lowering the share of indirect trade, thereby dampening the impact of this inefficiency. When λ_W is lower, the overall inefficiency in product adoption is smaller, so this dampening effect is weaker, leading to a larger welfare gain from removing the wholesale markup.

6.3 Evaluating the Welfare Impact of Fiber Internet Expansion

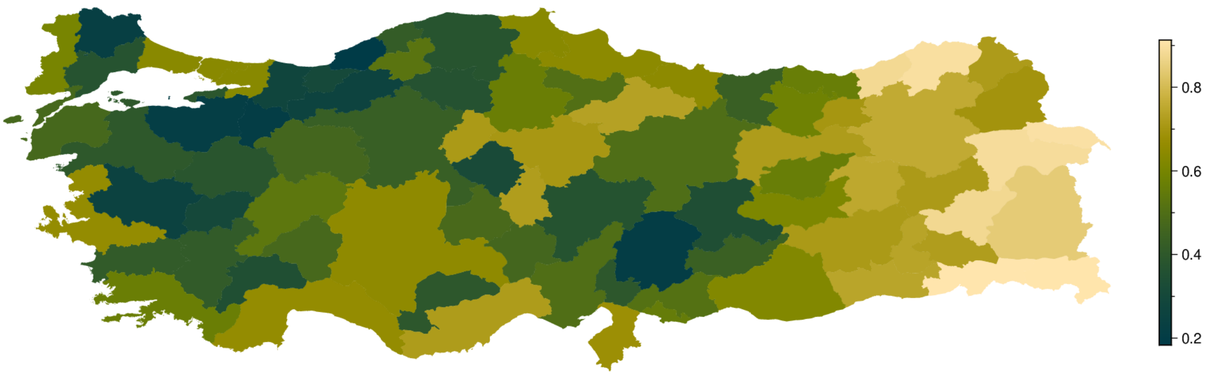


Figure 6: Aggregate indirect trade share across provinces in 2012

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.

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 roll-out and the importance of wholesale trade — and thus provide a richer welfare evaluation. As shown in Figure 6, the importance of indirect trade varies significantly across regions, with an interquartile range of 32 percentage points across Turkish provinces in 2012.[44]

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.19. 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 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 wholesale product adoption cost shifter, f_{Wud}, of the form

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

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 product adoption 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

[44]Figure 6 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.

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

Moment (relative change)	**Data**	**Model**
Targeted moments		
Indirect trade share	-0.517	-0.517
Direct trade flow	3.099	3.076
Indirect trade flow	0.232	0.234
Untargeted moments		
Direct trade (extensive margin)	1.654	2.292
Direct trade (intensive margin)	1.444	0.784
Number of wholesalers	-0.337	-0.696
Number of manufacturing firms	-0.724	-0.234
Wholesale markup	0.617	0.453

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.

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.

Why do we shock only customization productivity? The customization productivity shock, $\hat{\phi}_{cud}$, lowers the relative unit cost of direct sourcing. 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 roll-out.

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, 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 proportional 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 proportional 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. 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.

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

	$\widehat{\Omega}$	$\widehat{\mu^W}$	$\widehat{\sigma^2(\mu)}$	$\widehat{\widetilde{M}}$	$\widehat{S}$	$\widehat{A}$	$\widehat{L_A}$	$\widehat{L_S}$	$\widehat{L_E}$	$\widehat{L_P}$	$\widehat{Welfare}$
Pre Shock											
Wholesale subsidy	0.5804	0.9057	0.0231	0.5640	0.9775	0.9592	0.5263	1.3048	1.1315	1.0605	1.0156
Post Shock											
Wholesale subsidy	0.9505	0.9461	0.3052	0.9191	1.1743	1.0431	0.9251	1.6399	1.0071	1.0339	1.0777
Decentralized	1.5318	1.0895	2.6486	1.4849	1.1503	1.0376	1.6731	0.875	0.8516	1.0089	1.0467

Note: This table reports proportional 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): Proportional Changes Relative to the Pre-Shock Decentralized Equilibrium

	$\widehat{\Omega}$	$\widehat{\mu^W}$	$\widehat{\sigma^2(\mu)}$	$\widehat{M}$	$\widehat{S}$	$\widehat{A}$	$\widehat{L_A}$	$\widehat{L_S}$	$\widehat{L_E}$	$\widehat{L_P}$	$\widehat{Welfare}$
Pre Shock											
Efficient	0.4600	0.8851	0.0000	0.0093	1.4221	0.9231	0.2874	2.1040	0.8132	1.1916	1.1000
Wholesale subsidy	0.3700	0.8851	0.0000	0.4872	0.9653	1.0161	0.3269	1.5104	1.1242	1.0074	1.0237
Post Shock											
Efficient	0.8800	0.8851	0.0000	0.0159	1.4608	0.9790	0.6123	1.5349	0.7679	1.1864	1.1615
Wholesale subsidy	0.6600	0.8851	0.0000	0.7299	1.0909	1.0650	0.6023	1.2584	1.0645	1.0109	1.0765
Decentralized	1.6000	1.1493	2.9731	1.2950	1.1700	1.0122	1.7748	0.8017	0.7958	1.0204	1.0328

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

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:

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

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 6 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 2.1 and 3.1, 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 3.76% (Column 6). Overall, fiber internet expansion increases welfare by 0.0467 log points 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.16). 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 roll-out across provinces imply that those provinces with the fastest internet roll-out 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 0.014 log points. 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:

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

It turns out that allocative efficiency also worsened due to congestion externalities and manufacturer double marginalization, but only by 0.004 log points. 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 5, there is a significant dispersion in the speed of fiber internet roll-out 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[45], 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 roll-out across provinces (Figure 5) against their indirect trade share (Figure 6), 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

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

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 (-0.014 log points in the spatial model, relative to -0.018 log points in the single location model).

This negative correlation between the speed of fiber internet roll-out 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 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 roll-out have seen disproportionately greater degree of disintermediation, compared to eastern provinces that have received similar shocks. Lastly, these dampening and amplification 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 examines how wholesale market power affects the efficiency of production network formation and evaluates the welfare implications of digital infrastructure-induced disintermediation. By developing a model of production network formation featuring endogenous indirect trade shares, wholesale market structure, and markups, I identify significant inefficiencies arising from wholesale market power and congestion externalities due to excessive match formation.

Empirically, I leverage Turkey's fiber internet roll-out as a natural experiment to provide causal evidence supporting the model's predictions. While fiber internet expansion promotes disintermediation and enhances direct trade efficiency, it simultaneously drives increased wholesale market concentration and elevated wholesale markups. These outcomes highlight a nuanced trade-off: technology-driven reductions in direct trade costs may unintentionally strengthen wholesale market power, thereby partially offsetting potential welfare gains.

Quantitative analysis calibrated to the observed empirical patterns reveals that endogenous increases in wholesale markups substantially dampen the welfare benefits from fiber internet investments. Specifically, higher wholesale markups reduce welfare gains by approximately 1.4 percentage points, offsetting roughly 30% of baseline welfare improvements. Taken together, these results underscore the importance of evaluating technology-induced disintermediation in general equilibrium: analyses that abstract from intermediary entry/exit, rising concentration, and upstream reallocation risk overstating the welfare gains inferred from partial-equilibrium evidence. They also highlight the importance of implementing complementary competition policies, such as a wholesale subsidy, to mitigate markup-induced distortions and fully capture the benefits of digital infrastructure investments.

Future research could examine how production network reorganization — triggered by domestic trade cost reductions such as improved digital infrastructure — affects markups and competitive behavior more broadly, not only for wholesalers but also for manufacturing firms. This would parallel the literature on the pro-competitive effects of international trade liberalization, but applied to domestic trade integration. Additionally, while my quantification highlights a substantial welfare cost from congestion externalities, addressing these distortions with size-dependent subsidies or planner-based instruments is often infeasible in practice. An important avenue for future work is to explore alternative, implementable policy tools that could target these inefficiencies indirectly.

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

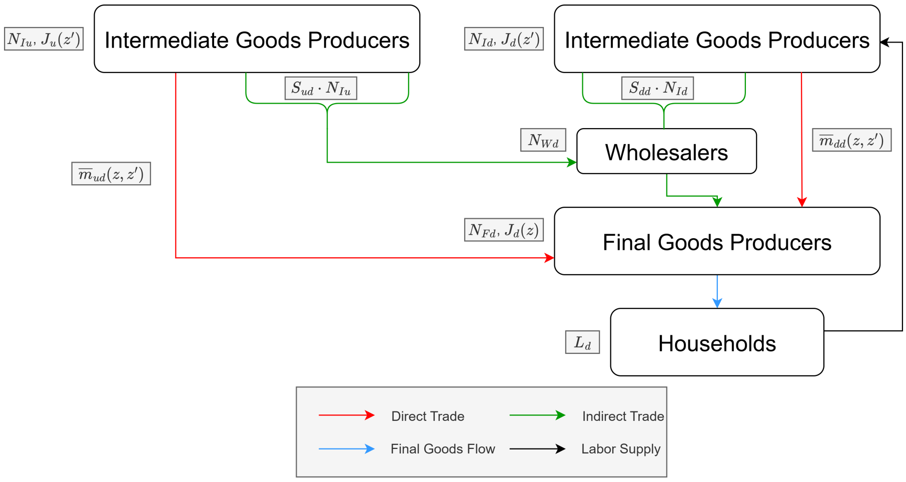


Figure A.1: Graphical Illustration of the Production Network

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 adopt varieties produced in any upstream location u and resell them only to location d final producers. This assumption is motivated by the Motivational Fact 1 that indirect sourcing is mostly intermediated by local wholesalers.

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

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

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 wholesale product adoption are now indexed by location pairs (u, i).

The corresponding unit cost function and its derivative become

$$c_i(z) = \left\{ \sum_{u \in \mathcal{N}} \left[m_{ui}(z) c_{mui}^{1-\sigma} + S_{ui} c_{Wui}^{1-\sigma} - m_{ui}(z) S_{ui} c_{Wmui}^{1-\sigma} \right] \right\}^{\frac{1}{1-\sigma}}$$

$$\frac{dc_i(z)}{dm_{ui}(z)} = \frac{1}{1-\sigma} c_i(z)^\sigma \big(c_{mui}^{1-\sigma} - S_{ui} c_{Wmui}^{1-\sigma} \big).$$

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

$$m_{ui}(z) = \Pi_{H1ui}\, z^{\frac{\gamma}{\beta}}, \qquad \Pi_{H1ui} \equiv \left[\frac{1}{w_i f_{mui} \sigma} \left(\frac{\sigma}{\sigma-1} \right)^{1-\sigma} D_{Hi} \big(c_{mui}^{1-\sigma} - S_{ui} c_{Wmui}^{1-\sigma} \big) \right]^{\frac{1}{\beta-1}}.$$

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

$$\max_{\{p_{Iid}, v_{id}\}} \sum_{d \in \mathcal{N}} p_{Iid}^{1-\sigma} (v_{id} D_{mid} + S_{id} D_{Wid} - v_{id} S_{id} D_{Wmid}) \\ - \sum_{d \in \mathcal{N}} \frac{1}{z} w_i \tau_{id} \Big[p_{Iid}^{-\sigma} (v_{id} D_{mid} + S_{id} D_{Wid} - v_{id} S_{id} D_{Wmid}) \Big] - \sum_{d \in \mathcal{N}} w_i f_{vid} \frac{v_{id}^\beta}{\beta}. \tag{24}$$

The ad-posting FOC becomes

$$v_{id}(z) = \left[\frac{x_{mid}(z)}{w_i f_{vid} \sigma} - \frac{x_{Wmid}(z)}{w_i f_{vid} \sigma} \right]^{\frac{1}{\beta}},$$

with $x_{mid}(z) = p_{Iid}^{1-\sigma} v_{id} D_{mid}$ and $x_{Wmid}(z) = p_{Iid}^{1-\sigma} v_{id} S_{id} D_{Wmid}$.

A.4 Direct Matching Across Locations

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

$$V_{ud} = N_{Iu} \int_Z v_{ud}(z) j_u(z) dz, \tag{25}$$

$$M_{ud} = N_{Fd} \int_Z m_{ud}(z) j_d(z) dz, \tag{26}$$

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

$$\theta_{ud}^{v} = \kappa_{ud} V_{ud}^{\lambda_V - 1} M_{ud}^{\lambda_M} \tag{27}$$

$$\theta_{ud}^{m} = \kappa_{ud} V_{ud}^{\lambda_V} M_{ud}^{\lambda_M - 1}. \tag{28}$$

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 Cost Minimization and Demand Shifters

We can similarly solve for the cost minimization problem of a type z' final goods producers in location d and obtain the unit cost of a bundle of intermediate goods:

$$c_d(z') = \left\{ \sum_{u \in \mathcal{N}} \left[m_{ud}(z') c_{mud}^{1-\sigma} + S_{ud} c_{Wud}^{1-\sigma} - m_{ud}(z') S_{ud} c_{Wmud}^{1-\sigma} \right] \right\}^{\frac{1}{1-\sigma}} \tag{29}$$

where

$$c_{mud} = \left[\int_Z p_{Iud}(z)^{1-\sigma} \phi_{cud} \theta_{ud}^{m} \frac{N_{Iu} v_{ud}(z) j_u(z)}{V_{ud}} dz \right]^{\frac{1}{1-\sigma}}$$

$$c_{Wud} = \left[\int_Z p_{ud}^{W}(z)^{1-\sigma} N_{Iu} j_u(z) dz \right]^{\frac{1}{1-\sigma}}$$

$$c_{Wmud} = \left[\int_Z p_{ud}^{W}(z)^{1-\sigma} \theta_{ud}^{m} \frac{N_{Iu} v_{ud}(z) j_u(z)}{V_{ud}} dz \right]^{\frac{1}{1-\sigma}}$$

D_m and $c_m^{1-\sigma} - Sc_{Wm}^{1-\sigma}$ can now be generalized as:

$$D_{mud} = \phi_{cud} \frac{\sigma - 1}{\sigma} \frac{\theta_{ud}^v}{M_{ud}} N_{Fd} \int_Z \frac{x_{Hd}(z)}{c_d(z)^{1-\sigma}} m_{ud}(z) j_d(z) dz, \tag{30}$$

$$c_{mud}^{1-\sigma} - S_{ud} c_{Wmud}^{1-\sigma} = \theta_{ud}^m C_{2ud} \left(\frac{\sigma}{\sigma - 1} w_u \tau_{ud} \right)^{1-\sigma} \frac{\mathbb{E}_u[z^\gamma]}{\mathbb{E}_u[z^{\gamma/\beta}]}. \tag{31}$$

Substituting optimal m_{ud} and v_{ud} gives

$$D_{mud} = \phi_{cud}^{\frac{\beta-1}{\beta}} \theta_{ud}^m \left(\frac{N_{Fd}}{N_{Iu}} \right)^{\frac{\beta-1}{\beta}} \left(\frac{\sigma - 1}{\sigma} \right)^{\frac{\beta-1}{\beta} + \sigma - 1} \left(\frac{f_{vud}}{f_{mud}} \right)^{\frac{1}{\beta}} \left(\frac{C_{2ud}}{C_{1ud}} \right)^{\frac{1}{\beta}} \left(\frac{w_u}{w_d} \right)^{\frac{1}{\beta}} D_{Hd} \left(\frac{\mathbb{E}_d[z^\gamma]}{\mathbb{E}_u[z^\gamma]} \right)^{\frac{\beta-1}{\beta}} \frac{\mathbb{E}_u[z^\gamma]}{\mathbb{E}_u[z^{\gamma/\beta}]}. \tag{32}$$

A.6 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 adoption 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 FOC for adoption is

$$S_{ud} = \left(\frac{X_{ud}^W - X_{mud}^W}{\sigma w_d f_{Wud}} \right)^{\frac{\lambda_W}{\beta_W}} N_{Iu}^{-\lambda_W} N_{Wd}^{\lambda_W \left(2 - \frac{2}{\beta_W} - \frac{1}{\lambda_W} \right)}, \tag{34}$$

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

Demand shifters generalize to

$$D_{ud}^W = \frac{\sigma - 1}{\sigma} N_{Fd} \int_Z \frac{x_{Hd}(z')}{c_d(z')^{1-\sigma}} j_d(z') dz', \qquad D_{mud}^W = \phi_{cud}^{-1} D_{mud}, \tag{35}$$

and the demand for intermediate producers is

$$D_{Wud} = \left(\frac{N_{Wd}\sigma}{N_{Wd}\sigma - 1} \right)^{-\sigma} D_{ud}^W, \qquad D_{Wmud} = \left(\frac{N_{Wd}\sigma}{N_{Wd}\sigma - 1} \right)^{-\sigma} \phi_{cud}^{-1} D_{mud}. \tag{36}$$

A.7 Households

Local household i earns

$$I_i = w_i L_i + \Pi_i^W, \tag{37}$$

faces local price index

$$P_i^H = \left[N_{Fi} \int_Z p_{Fi}(z)^{1-\sigma} j_i(z) dz\right]^{\frac{1}{1-\sigma}}, \tag{38}$$

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

A.8 Free Entry, Labor Market Clearing, and Trade Balance

Location-specific free-entry conditions (compare to (69)-(72)) become

$$w_i F_{Ii} N_{Ii} = \sum_d \left(\frac{1}{\sigma} X_{mid} + \frac{1}{\sigma} X_{Wid} - \frac{1}{\sigma} X_{Wmid}\right) - \sum_d N_{Ii} \int_Z w_i f_{vid} \frac{v_{id}(z)^\beta}{\beta} j_i(z) dz, \tag{39}$$

$$w_i F_{Fi} N_{Fi} = \frac{1}{\sigma} X_{Hi} - \sum_u N_{Fi} \int_Z w_i f_{mui} \frac{m_{ui}(z)^\beta}{\beta} j_i(z) dz, \tag{40}$$

$$w_i F_{Wi} N_{Wi} \leq N_{Wi} \sum_u \left[s_{ui} \theta_{ui}^W N_{Iu} \Pi_{Wui} - w_i f_{Wui} \frac{(s_{ui} N_{Iu}/N_{Wi})^{\beta_W}}{\beta_W}\right]. \tag{41}$$

Labor market clearing in each i:

$$\begin{aligned} L_i = & \sum_d \frac{1}{w_i} \left(\frac{\sigma-1}{\sigma} X_{mid} + \frac{\sigma-1}{\sigma} X_{Wid} - \frac{\sigma-1}{\sigma} X_{Wmid}\right) + \sum_d N_{Ii} \int_Z f_{vid} \frac{v_{id}(z)^\beta}{\beta} j_i(z) dz \\ & + \sum_u N_{Fi} \int_Z f_{mui} \frac{m_{ui}(z)^\beta}{\beta} j_i(z) dz + \sum_u f_{Wui} \frac{(s_{ui} N_{Iu}/N_{Wi})^{\beta_W}}{\beta_W} N_{Wi} + F_{Ii} N_{Fi} + F_{Fi} N_{Ii} + F_{Wi} N_{Wi}. \end{aligned} \tag{42}$$

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

$$\sum_u (X_{mui} + X_{Wui} - X_{Wmui}) = \sum_d (X_{mid} + X_{Wid} - X_{Wmid}). \tag{43}$$

A.9 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 (42), (37), (38), (27), (28), (34), (39), (40), and (41). 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

This appendix presents detailed derivations for the spatial extension of the model. When the number of locations is reduced to one, the analysis collapses to the single-location case discussed in Section 3.

B.1 Detailed Derivation for the Problem of Final Goods Producers

This section shows the detailed derivation of final goods producers' optimal ad posting decision. Recall final goods producers' profit maximization problem:

$$\max_{p_{Fi},\{m_{ui}\}} \quad p_{Fi}^{1-\sigma} D_{Hi} - \frac{1}{z} c_i(z) \left(p_{Fi}^{-\sigma} D_{Hi}\right) - \sum_{u\in\mathcal{N}} w_i f_{mui} \frac{m_{ui}^{\beta}}{\beta}$$

FOC w.r.t. p_F yields the usual monopolistic pricing condition:

$$p_{Fi}(z) = \frac{\sigma}{\sigma-1}\,\frac{1}{z}\,c_i(z)$$

which can be substituted back to the objective function:

$$\max_{\{m_{ui}\}} \quad \frac{1}{\sigma}\left[\frac{\sigma}{\sigma-1}\,\frac{1}{z}\,c_i(z)\right]^{1-\sigma} D_{Hi} - \sum_{u\in\mathcal{N}} w_i f_{mui} \frac{m_{ui}^{\beta}}{\beta}$$

FOC w.r.t. m_{ui}:

$$\begin{aligned}
w_i f_{mui} m_{ui}^{\beta-1} &= \frac{1-\sigma}{\sigma}\left[\frac{\sigma}{\sigma-1}\,\frac{1}{z}\,c_i(z)\right]^{-\sigma} D_{Hi}\,\frac{\sigma}{\sigma-1}\,\frac{1}{z}\,\frac{dc_i(z)}{dm_{ui}} \\
w_i f_{mui} m_{ui}^{\beta-1} &= \frac{1-\sigma}{\sigma}\left[\frac{\sigma}{\sigma-1}\,\frac{1}{z}\,c_i(z)\right]^{-\sigma} D_{Hi}\,\frac{\sigma}{\sigma-1}\,\frac{1}{z}\,\frac{1}{1-\sigma} c_i(z)^{\sigma}(c_{mui}^{1-\sigma} - S_{ui} c_{Wmui}^{1-\sigma}) \\
w_i f_{mui} m_{ui}^{\beta-1} &= \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\,\frac{1}{z}\right)^{1-\sigma} D_{Hi}\,(c_{mui}^{1-\sigma} - S_{ui} c_{Wmui}^{1-\sigma}) \\
m_{ui}(z) &= \left[\frac{1}{w_i\, f_{mui}\,\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} D_{Hi}\,(c_{mui}^{1-\sigma} - S_{ui} c_{Wmui}^{1-\sigma})\right]^{\frac{1}{\beta-1}} z^{\frac{\sigma-1}{\beta-1}} \\
m_{ui}(z) &= \Pi_{H1ui}\, z^{\frac{\gamma}{\beta}} \quad , \quad \Pi_{H1ui} \equiv \left[\frac{1}{w_i\, f_{mui}\,\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} D_{Hi}\,(c_{mui}^{1-\sigma} - S_{ui} c_{Wmui}^{1-\sigma})\right]^{\frac{1}{\beta-1}}
\end{aligned}$$

Therefore, revenue of the final goods producer z can be written as:

$$\begin{aligned}
x_{Hi}(z) \equiv & p_{Fi}(z)^{1-\sigma} D_{Hi} \\
= & \left[\frac{\sigma}{\sigma-1}\,\frac{1}{z}\,c_i(z)\right]^{1-\sigma} D_{Hi} \\
= & \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} D_{Hi}\, z^{1-\sigma} \left\{\sum_{u\in\mathcal{N}} \left[m_{ui}(z)\, c_{mui}^{1-\sigma} + S_{ui} c_{Wui}^{1-\sigma} - m_{ui}(z)\, S_{ui} c_{Wmui}^{1-\sigma}\right]\right\} \\
= & \sum_{u\in\mathcal{N}} \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} S_{ui} c_{Wui}^{1-\sigma}\, D_{Hi}\, z^{\sigma-1} + \sum_{u\in\mathcal{N}} \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} D_{Hi}\,(c_{mui}^{1-\sigma} - S_{ui} c_{Wmui}^{1-\sigma})\, z^{\sigma-1}\, \Pi_{H1ui}\, z^{\frac{\gamma}{\beta}} \\
= & \Pi_{H2i} z^{\sigma-1} + \Pi_{H3i} z^{\gamma}
\end{aligned} \tag{44}$$

where

$$\Pi_{H2i} \equiv \sum_{u\in\mathcal{N}} \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} S_{ui} c_{Wui}^{1-\sigma} D_{Hi} \quad , \quad \Pi_{H3i} \equiv \sum_{u\in\mathcal{N}} \left(\frac{1}{w_i\, f_{mui}\, \sigma}\right)^{\frac{1}{\beta-1}} \left[\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} D_{Hi}\, (c_{mui}^{1-\sigma} - S_{ui} c_{Wmui}^{1-\sigma})\right]^{\frac{\beta}{\beta-1}}$$

The following relationships are used for derivation in other sections. From the definition of $x_{Hi}(z)$:

$$\frac{x_{Hi}(z)}{c_i(z)^{1-\sigma}} = \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} z^{\sigma-1} D_{Hi} \tag{45}$$

Using the FOC w.r.t. m_{ui}:

$$\frac{m_{ui}(z)}{c_i(z)^{1-\sigma}} = \frac{\sigma w_i f_{mui} m_{ui}(z)^{\beta}}{x_{Hi}(z)(c_{mui}^{1-\sigma} - S_{ui} c_{Wmui}^{1-\sigma})} \tag{46}$$

B.2 Detailed Derivation for the Problem of Intermediate Goods Producers

This section shows the detailed derivation of intermediate goods producers' optimal ad posting decision. Recall intermediate goods producers' profit maximization problem:

$$\begin{aligned} \max_{\{p_{Iid}\},\{v_{id}\}} \quad & \sum_{d\in\mathcal{N}} p_{Iid}^{1-\sigma} (v_{id} D_{mid} + S_{id} D_{Wid} - v_{id} S_{id} D_{Wmid}) \\ & - \sum_{d\in\mathcal{N}} \frac{1}{z}\, w_i\, \tau_{id} \left[p_{Iid}^{-\sigma} (v_{id} D_{mid} + S_{id} D_{Wid} - v_{id} S_{id} D_{Wmid})\right] \\ & - \sum_{d\in\mathcal{N}} w_i f_{vid} \frac{v_{id}^{\beta}}{\beta} \end{aligned}$$

where the first line is the total revenue received by the intermediate goods producer, the second line captures the total cost of production, adjusted for the iceberg trade cost of shipping products abroad, while the third line is the sum of search costs to form direct matches.

FOC w.r.t. p_{Iid} yields the usual monopolistic pricing condition:

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

which can be substituted back to the objective function:

$$\max_{\{v_{id}\}} \quad \sum_{d\in\mathcal{N}} \frac{1}{\sigma} \left(\frac{\sigma}{\sigma-1} \frac{w_i}{z} \tau_{id}\right)^{1-\sigma} (v_{id} D_{mid} + S_{id} D_{Wid} - v S_{id} D_{Wmid}) - \sum_{d\in\mathcal{N}} w_i f_{vid} \frac{v_{id}^{\beta}}{\beta}$$

FOC w.r.t. v_{id} yields:

$$w_i\, f_{vid}\, v_{id}^{\beta-1} = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\,\frac{w_i}{z}\tau_{id}\right)^{1-\sigma}(D_{mid} - S_{id}D_{Wmid})$$
$$v_{id}(z) = \left[\frac{x_{mid}(z)}{w_i\, f_{vid}\,\sigma} - \frac{x_{Wmid}(z)}{w_i\, f_{vid}\,\sigma}\right]^{\frac{1}{\beta}}$$

where $x_{mid}(z) \equiv p_{Iid}(z)^{1-\sigma}v_{id}(z)D_{mid}$ is the revenue from direct sales to location d and $x_{Wmid}(z) \equiv p_{Iid}(z)^{1-\sigma}v_{id}(z)S_{id}D_{Wmid}$ is the lost revenue from wholesale trade due to cannibalization by direct sales. Also define $x_{Wid}(z) \equiv p_{Iid}(z)^{1-\sigma}S_{id}D_{Wid}$ as the total potential revenue from wholesale trade before cannibalization. Now, according to the wholesaler's optimization problem in section 3.3.3, we know that $D_{Wmid} = \left(\frac{N_{Wd}\,\sigma}{N_{Wd}\,\sigma-1}\right)^{-\sigma}\phi_{cid}^{-1}D_{mid} < D_{mid}$ (where $\phi_{cid} > 1$ is the customization productivity gain accrued to directly traded intermediate good), which allows us to rewrite:

$$v_{id}(z) = \left[\frac{C_{1id}\,x_{mid}(z)}{w_i\, f_{vid}\,\sigma}\right]^{\frac{1}{\beta}} \quad , \quad C_{1id} \equiv \left[1 - \left(\frac{N_{Wd}\,\sigma}{N_{Wd}\,\sigma-1}\right)^{-\sigma} S_{id}\,\phi_{cid}^{-1}\right]$$

Intuitively, intermediate goods producers post more ads if direct trade is more attractive, which is the case if wholesalers charge a higher markup (resulting in a lower indirect trade demand), if wholesalers adopt less products from location i such that the chance that the intermediate goods producers' products are traded by wholesalers becomes lower, and if the customization productivity gain is larger. These benefits of direct trade from the intermediate goods producer's perspective are summarized by C_{1id}. We can then write the revenue terms as:

$$x_{mid}(z) = \left(\frac{\sigma}{\sigma-1}\,\frac{w_i}{z}\tau_{id}\right)^{1-\sigma}\left[\frac{C_{1id}\,x_{mid}(z)}{w_i\, f_{vid}\,\sigma}\right]^{\frac{1}{\beta}} D_{mid}$$
$$x_{mid}(z)^{\beta-1} = \left(\frac{\sigma}{\sigma-1}\,w_i\tau_{id}\right)^{\beta(1-\sigma)}\left(\frac{C_{1id}}{w_i\, f_{vid}\,\sigma}\right) D_{mid}^{\beta}\, z^{\beta(\sigma-1)}$$
$$x_{mid}(z) = \left[\left(\frac{\sigma}{\sigma-1}\,w_i\tau_{id}\right)^{\beta(1-\sigma)}\left(\frac{C_{1id}}{w_i\, f_{vid}\,\sigma}\right) D_{mid}^{\beta}\right]^{\frac{1}{\beta-1}} z^{\frac{\beta(\sigma-1)}{\beta-1}}$$
$$x_{mid}(z) = \Pi_{mid}z^{\gamma} \tag{47}$$
$$x_{Wmid}(z) = (1 - C_{1id})\,\Pi_{mid}\,z^{\gamma} \tag{48}$$
$$x_{Wid}(z) = \Pi_W\, z^{\sigma-1} \tag{49}$$

where

$$\Pi_{mid} \equiv \left[\left(\frac{\sigma}{\sigma-1}\,w_i\tau_{id}\right)^{\beta(1-\sigma)}\left(\frac{C_{1id}}{w_i\, f_{vid}\,\sigma}\right) D_{mid}^{\beta}\right]^{\frac{1}{\beta-1}} \quad , \quad \gamma \equiv \frac{\beta(\sigma-1)}{\beta-1} \quad , \quad \Pi_W \equiv \left(\frac{\sigma}{\sigma-1}\,w_i\tau_{id}\right)^{1-\sigma} S_{id}D_{Wid}$$

And $v_{id}(z)$ is:

$$v_{id}(z) = \left[\frac{C_{1id}\,\Pi_{mid}}{w_i\, f_{vid}\,\sigma}\right]^{\frac{1}{\beta}} z^{\frac{\gamma}{\beta}} \tag{50}$$

B.3 Detailed Derivation for Final Goods Producers' Cost Minimization

Recall the cost minimization problem of a type z' buyer in location d:

$$\max_{\{y_{Iud}(z)\},\{y^W_{ud}(z)\}} -\left\{\sum_{u\in\mathcal{N}}\int_Z p_{Iud}(z)y_{Iud}(z)\overline{m}_{ud}(z',z)+p^W_{ud}(z)y^W_{ud}(z)S_{ud}[N_{Iu}j_u(z)-\overline{m}_{ud}(z',z)]dz\right\}$$

$$\text{s.t.}\quad Y_{Id}\geq Y$$

$$Y_{Id} = \left\{\sum_{u\in\mathcal{N}}\int_Z y_{Iud}(z)^{\frac{\sigma-1}{\sigma}}\phi_{cud}^{\frac{1}{\sigma}}\overline{m}_{ud}(z',z)+y^W_{ud}(z)^{\frac{\sigma-1}{\sigma}}S_{ud}[N_{Iu}j_u(z)-\overline{m}_{ud}(z',z)]dz\right\}^{\frac{\sigma}{\sigma-1}}$$

Recall $y_{Iud}(z)$, the quantity of a unique intermediate goods variety produced by a firm of type z from location u, that is purchased directly by the type z' buyer in location d for use in its production. Similarly, let $y^W_{ud}(z)$ denote the quantity of a unique intermediate goods variety produced by a firm of type z from location u and used by the type z' buyer in location d in its production, but purchased through a wholesaler. The latter case occurs only when the two firms are not directly matched with each other, and the variety has been adopted by wholesalers in location d. The prices for directly and indirectly sourced varieties are denoted by p_{Iud} and p^W_{ud}, respectively.

FOC w.r.t. $y_{Iud}(z)$ is:

$$p_{Iud}(z)\overline{m}_{ud}(z',z) = \lambda_d(z')\,Y^{\frac{1}{\sigma}}\,y_{Iud}(z)^{-\frac{1}{\sigma}}\phi_{cud}^{\frac{1}{\sigma}}\overline{m}_{ud}(z',z) \tag{51}$$

FOC w.r.t. $y^W(z)$ is:

$$p^W_{ud}(z)S_{ud}S_{ud}[N_{Iu}j_u(z)-\overline{m}_{ud}(z',z)] = \lambda_d(z')\,Y^{\frac{1}{\sigma}}y^W_{ud}(z)^{-\frac{1}{\sigma}}S_{ud}[N_{Iu}j_u(z)-\overline{m}_{ud}(z',z)] \tag{52}$$

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

$$C_d(z') \equiv \sum_{u\in\mathcal{N}} \int_Z p_{Iud}(z) y_{Iud}(z) \overline{m}_{ud}(z',z) + p^W_{ud}(z) y^W_{ud}(z) S_{ud}[N_{Iu} j_u(z) - \overline{m}_{ud}(z',z)] dz = \lambda_d(z')\, Y \tag{53}$$

Also, define $c_d(z') \equiv \frac{C_d(z')}{Y}$. Now, substitute λ into (51) using (53):

$$p_{Iud}(z) y_{Iud}(z) = \phi_{cud} \left[\frac{p_{Iud}(z)}{c_d(z')}\right]^{1-\sigma} C_d(z')$$

Now, $C_d(z')$ is the total cost of production of a type z' final goods producer in location d, i.e.:

$$C_d(z') = \frac{\sigma-1}{\sigma} x_{Hd}(z')$$

Therefore, the sales of a supplier z from location u to a matched buyer z' in location d is:

$$p_{Iud}(z) y_{Iud}(z) = \phi_{cud} \left[\frac{p_{Iud}(z)}{c_d(z')}\right]^{1-\sigma} \frac{\sigma-1}{\sigma} x_{Hd}(z') \tag{54}$$

Similarly, the sales of a type z intermediate goods variety in location u to a buyer z' in location d through the wholesaler is:

$$p^W_{ud}(z) y^W_{ud}(z) = \left[\frac{p^W_{ud}(z)}{c_d(z')}\right]^{1-\sigma} \frac{\sigma-1}{\sigma} x_{Hd}(z') \tag{55}$$

Lastly, to derive $c_d(z')$, raise both sides of (51) and (52) to power $1-\sigma$, adding the two and using the constraint:

$$\begin{aligned}
\lambda_d(z')^{1-\sigma} =& \sum_{u\in\mathcal{N}} \int_Z p_{Iud}(z)^{1-\sigma} \phi_{cud} \overline{m}_{ud}(z',z) + p^W_{ud}(z)^{1-\sigma} S_{ud}[N_{Iu} j_u(z) - \overline{m}_{ud}(z',z)] dz \\
c_d(z') = \lambda_d(z') =& \left\{ \sum_{u\in\mathcal{N}} \int_Z p_{Iud}(z)^{1-\sigma} \phi_{cud} \overline{m}_{ud}(z',z) + p^W_{ud}(z)^{1-\sigma} S_{ud}[N_{Iu} j_u(z) - \overline{m}_{ud}(z',z)] dz \right\}^{\frac{1}{1-\sigma}} \\
=& \left\{ \sum_{u\in\mathcal{N}} \int_Z p_{Iud}(z)^{1-\sigma} \phi_{cud} m_{ud}(z') \theta^m_{ud} \frac{N_{Iu} v_{ud}(z) j_u(z)}{V_{ud}} \right. \\
& \left. + p^W_{ud}(z)^{1-\sigma} S_{ud} \left[N_{Iu} j_u(z) - m_{ud}(z') \theta^m_{ud} \frac{N_{Iu} v_{ud}(z) j_u(z)}{V_{ud}} \right] dz \right\}^{\frac{1}{1-\sigma}} \\
c_d(z') =& \left\{ \sum_{u\in\mathcal{N}} \left[m_{ud}(z') c^{1-\sigma}_{mud} + S_{ud} c^{1-\sigma}_{Wud} - m_{ud}(z') S_{ud} c^{1-\sigma}_{Wmud} \right] \right\}^{\frac{1}{1-\sigma}}
\end{aligned} \tag{56}$$

where

$$
\begin{aligned}
c_{mud} &= \left[\int_Z p_{Iud}(z)^{1-\sigma} \phi_{cud} \theta^m_{ud} \frac{N_{Iu} v_{ud}(z) j_u(z)}{V_{ud}} dz\right]^{\frac{1}{1-\sigma}} \\
c_{Wud} &= \left[\int_Z p^W_{ud}(z)^{1-\sigma} N_{Iu} j_u(z) dz\right]^{\frac{1}{1-\sigma}} \\
c_{Wmud} &= \left[\int_Z p^W_{ud}(z)^{1-\sigma} \theta^m_{ud} \frac{N_{Iu} v_{ud}(z) j_u(z)}{V_{ud}} dz\right]^{\frac{1}{1-\sigma}}
\end{aligned}
$$

The total direct sales of an intermediate goods producer of type z from location u to matched final goods producers across all locations is therefore:

$$
\begin{aligned}
&= \sum_{d \in \mathcal{N}} \int_Z \phi_{cud} \left[\frac{p_{Iud}(z)}{c_d(z')}\right]^{1-\sigma} \frac{\sigma - 1}{\sigma} x_{Hd}(z')\, v_{ud}(z) \theta^v_{ud} \frac{N_{Fd} m_{ud}(z') j_d(z')}{M_{ud}} dz \\
&= \sum_{d \in \mathcal{N}} p_{Iud}(z)^{1-\sigma} v_{ud}(z) D_{mud}
\end{aligned}
$$

where

$$
D_{mud} = \phi_{cud} \frac{\sigma - 1}{\sigma} \frac{\theta^v_{ud}}{M_{ud}} N_{Fd} \int_Z \frac{x_{Hd}(z)}{c_d(z)^{1-\sigma}} m_{ud}(z) j_d(z) dz \tag{57}
$$

B.4 Derivation for Ad Success Rate

Now, we can begin to solve for D_m and $c_m^{1-\sigma} - c_{Wm}^{1-\sigma}$ which the success rates of ads depend on. First, substitute out $\frac{m_{ud}(z)}{c_d(z)^{1-\sigma}}$ from (30) using (46):

$$
\begin{aligned}
D_{mud} &= \phi_{cud} \frac{\sigma - 1}{\sigma} \frac{\theta^v_{ud}}{M_{ud}} N_{Fd} \int_Z x_{Hd}(z) \frac{\sigma\, w_d\, f_{mud}\, m_{ud}(z)^\beta}{x_{Hd}(z)\, (c^{1-\sigma}_{mud} - S_{ud}\, c^{1-\sigma}_{Wmud})} j_d(z) dz\, j(z) dz \\
D_m &= \phi_{cud} \frac{\sigma - 1}{\sigma} \frac{\theta^v_{ud}}{M_{ud}} N_{Fd} \frac{\sigma\, w_d\, f_{mud}}{c^{1-\sigma}_{mud} - c^{1-\sigma}_{Wmud}} \\
&\qquad \int_Z \left[\frac{1}{w_d\, f_{mud}\, \sigma} \left(\frac{\sigma}{\sigma - 1}\right)^{1-\sigma} D_{Hd}\, (c^{1-\sigma}_{mud} - Sud\, c^{1-\sigma}_{Wmud})\right]^{\frac{\beta}{\beta - 1}} z^\gamma j_d(z) dz \\
D_{mud} &= \phi_{cud} \left(\frac{\sigma - 1}{\sigma}\right)^{1+\gamma} (\sigma\, w_d\, f_{mud})^{\frac{1}{1-\beta}} D_{Hd}^{\frac{\beta}{\beta - 1}} (c^{1-\sigma}_{mud} - S_{ud} \dot{c}^{1-\sigma}_{Wmud})^{\frac{1}{\beta - 1}} \mathbb{E}_d [z^\gamma] \frac{\theta^v_{ud}}{M_{ud}} N_{fd}
\end{aligned} \tag{58}
$$

On the other hand,

$$
\begin{aligned}
c_{mud}^{1-\sigma} - S_{ud} c_{Wmud}^{1-\sigma} = & \left[\int_Z p_{Iud}(z)^{1-\sigma} \phi_{cud} \theta_{ud}^m \frac{N_{Iu} v_{ud}(z) j_u(z)}{V_{ud}} dz\right] - \left[S_{ud} \int_Z p_{ud}^W(z)^{1-\sigma} \theta_{ud}^m \frac{N_{Iu} v_{ud}(z) j_u(z)}{V_{ud}} dz\right] \\
= & \frac{\theta_{ud}^m}{V_{ud}} N_{Iu} \int_Z \left\{ \left(\frac{\sigma}{\sigma-1} \frac{w_u}{z} \tau_{ud}\right)^{1-\sigma} \phi_{cud} - \left[\left(\frac{\sigma}{\sigma-1}\right)^2 \frac{w_u}{z} \tau_{ud}\right]^{1-\sigma} S_{ud} \right\} \\
& \left(\frac{C_{1ud}\, \Pi_{mud}}{w_u\, f_{vud}\, \sigma}\right)^{\frac{1}{\beta}} z^{\frac{\gamma}{\beta}}\, j_u(z) dz \\
= & \frac{\theta_{ud}^m}{V_{ud}}\, N_{Iu}\, C_{2ud} \left(\frac{\sigma}{\sigma-1}\right)^{-\gamma} w_u^{-\gamma}\, \tau_{ud}^{-\gamma} \left(\frac{C_{1ud}}{w_u\, f_{vud}\, \sigma}\right)^{\frac{1}{\beta-1}} \mathbb{E}_u\left[z^\gamma\right] D_{mud}^{\frac{1}{\beta-1}}
\end{aligned} \tag{59}
$$

where I made use of wholesalers' optimal pricing condition (equation 33) to substitute out $p_{ud}^W(z)$, and $C_{2ud} \equiv \phi_{cud} - \left(\frac{N_{Wd}\,\sigma}{N_{Wd}\,\sigma-1}\right)^{1-\sigma} S_{ud}$ captures the cost advantage of direct sourcing for a location d buyer of intermediate goods produced in location u. Substitute out V from (59) using (25):

$$
c_{mud}^{1-\sigma} - S_{ud}\, c_{Wmud}^{1-\sigma} = \theta_{ud}^m\, C_{2ud} \left(\frac{\sigma}{\sigma-1} w_u\, \tau_{ud}\right)^{1-\sigma} \frac{\mathbb{E}_u\left[z^\gamma\right]}{\mathbb{E}_u\left[z^{\frac{\gamma}{\beta}}\right]} \tag{60}
$$

Intuitively, $c_{mud}^{1-\sigma} - S_{ud}\, c_{Wmud}^{1-\sigma}$ captures the benefit for buyer in location d to post an additional ad to search for direct suppliers in location u. An additional ad matches the buyer to direct suppliers with some success rate (θ_{ud}^m), which reduces the input cost conditional on the type of supplier being matched (C_{2ud}), and the ratio $\frac{\mathbb{E}_u[z^\gamma]}{\mathbb{E}_u\left[z^{\frac{\gamma}{\beta}}\right]}$ captures the fact that more productive suppliers charge lower price (proportional to $z^{\sigma-1}$), adjusted for the fact that they also post more ads (proportional to $z^{\frac{\gamma}{\beta}}$, and notice that $\gamma = \sigma - 1 + \frac{\gamma}{\beta}$).

Next, I solve for D_{mud}. Divide (58) by (59) yields:

$$
\frac{D_{mud}}{c_{mud}^{1-\sigma} - S_{ud} c_{Wmud}^{1-\sigma}} = \left(\frac{\sigma-1}{\sigma}\right)^{\frac{\beta-1}{\beta}} \left(\frac{N_{Fd}}{N_{Iu}}\right)^{\frac{\beta-1}{\beta}} C_{2ud}^{\frac{1-\beta}{\beta}}\, C_{1ud}^{-\frac{1}{\beta}}\, w_u^{\sigma-1}\, \tau_{ud}^{\sigma-1} \left(\frac{w_u}{w_d}\right)^{\frac{1}{\beta}} \left(\frac{f_{vud}}{f_{mud}}\right)^{\frac{1}{\beta}} D_{Hd} \left(\frac{\mathbb{E}_d\left[z^\gamma\right]}{\mathbb{E}_u\left[z^\gamma\right]}\right)^{\frac{\beta-1}{\beta}} \phi_{cud}^{\frac{\beta-1}{\beta}} \tag{61}
$$

We can then substitute (60) into (61) to obtain D_{mud}:

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

This allows us to derive the success rates of ads θ^m_{ud} by substituting out $c^{1-\sigma}_{mud} - S_{ud} c^{1-\sigma}_{Wmud}$ and D_{mud}:

$$\theta^m_{ud} \tag{62}$$

$$=\kappa_{ud} V_{ud}^{\lambda_V} M_{ud}^{\lambda_M - 1}$$

$$=\Bigg(\kappa_{ud}\Bigg\{N_{Iu}\left[\left(\frac{\sigma}{\sigma-1}w_u\tau_{ud}\right)^{1-\sigma}\left(\frac{C_{1ud}}{w_u f_{vud}\sigma}\right)\right]^{\frac{1}{\beta-1}}\phi_{cud}^{\frac{1}{\beta}}\left(\frac{N_{Fd}}{N_{Iu}}\right)^{\frac{1}{\beta}}\left(\frac{\sigma-1}{\sigma}\right)^{\frac{1}{\beta}+\frac{\gamma}{\beta}}$$

$$\left(\frac{f_{vud}}{f_{mud}}\right)^{\frac{1}{\beta(\beta-1)}}\left(\frac{C_{2ud}}{C_{1ud}}\right)^{\frac{1}{\beta(\beta-1)}}\left(\frac{w_u}{w_d}\right)^{\frac{1}{\beta(\beta-1)}} D_{Hd}^{\frac{1}{\beta-1}}\left(\frac{\mathbb{E}_d\left[z^\gamma\right]}{\mathbb{E}_u\left[z^\gamma\right]}\right)^{\frac{1}{\beta}}\left(\frac{\mathbb{E}_u\left[z^\gamma\right]}{\mathbb{E}_u\left[z^{\frac{\gamma}{\beta}}\right]}\right)^{\frac{1}{\beta-1}}\mathbb{E}_u\left[z^{\frac{\gamma}{\beta}}\right]\Bigg\}^{\lambda_V}$$

$$\Bigg\{N_{Fd}\left[\frac{1}{w_d\, f_{mud}\,\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}D_{Hd}\,C_{2ud}\left(\frac{\sigma}{\sigma-1}\,w_u\,\tau_{ud}\right)^{1-\sigma}\frac{\mathbb{E}_u\left[z^\gamma\right]}{\mathbb{E}_u\left[z^{\frac{\gamma}{\beta}}\right]}\right]^{\frac{1}{\beta-1}}\mathbb{E}_d\left[z^{\frac{\gamma}{\beta}}\right]\Bigg\}^{\lambda_M-1}\Bigg)^{\frac{\beta-1}{\beta-\lambda_V-\lambda_M}} \tag{63}$$

similarly for θ^v_{ud}.

B.5 Detailed Derivation for Wholesalers' Optimal Price

Recall the demand curve faced by the wholesale sector in reselling a type z variety from location u to a type z' final goods producer from location d (when the two firms are not directly matched with each other):

$$y^W_{ud}(z) = \frac{p^W_{ud}(z)^{-\sigma}}{c_d(z')^{1-\sigma}}\,\frac{\sigma-1}{\sigma}\,x_{Hd}(z')$$

with an inverse demand curve:

$$p^W_{ud}(z) = y^W_{ud}(z)^{-\frac{1}{\sigma}}\left[\frac{\sigma-1}{\sigma}\,\frac{x_{Hd}(z')}{c_d(z')^{1-\sigma}}\right]^{\frac{1}{\sigma}}$$

where $y^W_{ud}(z)$ is the total quantity resold by the wholesalers, and is given by the sum of the quantity resold by the N_{Wd} wholesalers:

$$y^W_{ud}(z) = \sum_{i=1}^{N_{Wd}} y^{Wi}_{ud}(z)$$

where y^{Wi}_{ud} denotes the quantity resold by a cartel i. Conditional on adopting the variety, a wholesaler i would engage in Cournot competition and maximize profit by choosing the quantity to be resold:

$$\max_{y^{Wi}_{ud}(z)} p^W_{ud}(z)\,y^{Wi}_{ud}(z) - p_{Iud}(z)\,y^{Wi}_{ud}(z)$$

FOC:

$$p^W_{ud}(z) - p_{Iud}(z) + y^{Wi}_{ud}(z)\,\frac{\partial p^W_{ud}(z)}{\partial y^{Wi}_{ud}(z)} = 0$$

$$p^W_{ud}(z) - p_{Iud}(z) + y^{Wi}_{ud}(z)\left[\frac{\sigma-1}{\sigma}\,\frac{x_{Hd}(z')}{c_d(z')^{1-\sigma}}\right]^{\frac{1}{\sigma}}\left(-\frac{1}{\sigma}\right)y^W_{ud}(z)^{-\frac{1}{\sigma}-1} = 0$$

$$y_{ud}^{Wi}(z) = \sigma\, y_{ud}^{W}(z) \left(1 - \frac{p_{Iud}(z)}{p_{ud}^{W}(z)}\right)$$

Note that $y_{ud}^{W}(z) = N_{Wd}\, y_{ud}^{Wi}(z)$ in equilibrium due to symmetry, therefore:

$$p_{ud}^{W}(z) = \frac{N_{Wd}\,\sigma}{N_{Wd}\,\sigma - 1}\, p_{Iud(z)}$$

B.6 Detailed Derivation for Household's Problem

Household's cost minimization problem:

$$\max_{c_i(z)} - N_{Fi} \int_Z p_{Fi}(z) c_i(z) j_i(z) dz$$

$$\text{s.t.} \left[N_{Fi} \int_Z c_i(z)^{\frac{\sigma-1}{\sigma}} j_i(z) dz\right]^{\frac{\sigma}{\sigma-1}} \geq C_i$$

FOC w.r.t. $c_i(z)$

$$p_{Fi}(z) j_i(z) = \lambda_i\, C_i^{\frac{1}{\sigma}}\, c_i(z)^{-\frac{1}{\sigma}}\, j_i(z) \tag{64}$$

Using the constraint:

$$I_i = \int_Z p_{Fi}(z) c_i(z) j_i(z) dz = \lambda_i\, C_i \quad , \quad \lambda_i = \frac{I_i}{C_i} \tag{65}$$

Raise both sides of (64) to power $1 - \sigma$, integrating, yields:

$$P_i^H \equiv \frac{I_i}{C_i} = \lambda_i = \left[N_{Fi} \int_Z p_{Fi}(z)^{1-\sigma} j_i(z) dz\right]^{\frac{1}{1-\sigma}} \tag{66}$$

Now, continue to solve for P_i^H:

$$\begin{aligned}
(P_i^H)^{1-\sigma} =& N_{Fi} \int_Z \left[\frac{\sigma}{\sigma - 1}\, \frac{1}{z}\, c_i(z)\right]^{1-\sigma} j_i(z) dz \\
=& N_{Fi} \int_Z \left(\frac{\sigma}{\sigma - 1}\right)^{1-\sigma} \left\{\sum_{u \in \mathcal{N}} \left[m_{ui}(z) c_{mui}^{1-\sigma} + S_{ui}\, c_{Wui}^{1-\sigma} - m_{ui}(z)\, S_{ui}\, c_{Wmui}^{1-\sigma}\right]\right\} z^{\sigma-1} j_i(z) dz
\end{aligned}$$

Substitute out $m_{ui}(z)$, c_{mui}, c_{Wui} and c_{Wmui} using (29), (60) and (A.2):

$$\begin{aligned}
P_i^H =& N_{Fi}^{\frac{1}{1-\sigma}} \left(\frac{\sigma}{\sigma-1}\right) \left(\sum_{u\in\mathcal{N}} \left\{ (w_i\, f_{mui}\, \sigma)^{\frac{1}{1-\beta}} \left(\frac{\sigma}{\sigma-1}\right)^{-\gamma\left(\frac{1}{\beta}+1\right)} D_{Hi}^{\frac{1}{\beta-1}}\, w_u^{-\gamma}\, \tau_{ui}^{-\gamma} \right.\right. \\
& \theta_{ui}^{m\,\frac{\beta}{\beta-1}}\, C_{2ui}^{\frac{\beta}{\beta-1}} \left(\frac{\mathbb{E}_u\left[z^\gamma\right]^{\frac{1}{\beta}}}{\mathbb{E}_u\left[z^{\frac{\gamma}{\beta}}\right]}\right)^{\frac{\beta}{\beta-1}} \mathbb{E}_u\left[z^\gamma\right]\, \mathbb{E}_i\left[z^\gamma\right] \\
& \left.\left. + S_{ui}\, N_{Iu} \left(\frac{\sigma}{\sigma-1}\right)^{2(1-\sigma)} w_u^{1-\sigma}\, \tau_{ui}^{1-\sigma}\, \mathbb{E}_u\left[z^{\sigma-1}\right]\, \mathbb{E}_i\left[z^{\sigma-1}\right] \right\}\right)^{\frac{1}{1-\sigma}}
\end{aligned} \tag{67}$$

Finally, substitute $\lambda_i = P_i^H$ into (64):

$$\begin{aligned}
p_{Fi}(z) =& P_i^H\, C_i^{\frac{1}{\sigma}}\, c_i(z)^{-\frac{1}{\sigma}} \\
p_{Fi}(z)c_i(z) =& \left[\frac{p_{Fi}(z)}{P_i^H}\right]^{1-\sigma} I_i \\
p_{Fi}(z)c_i(z) =& p_{Fi}(z)^{1-\sigma}\, D_{Hi}
\end{aligned}$$

We get the sales of each type z final goods producer to the household in location i, where

$$D_{Hi} \equiv \frac{I_i}{P_i^{H\,1-\sigma}} \tag{68}$$

B.7 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

$$w\, F_I\, N_I = \frac{1}{\sigma} X_m + \frac{1}{\sigma} X_W - \frac{1}{\sigma} X_{Wm} - N_I \int_Z w\, f_v\, \frac{v(z)^\beta}{\beta}\, j(z)\, dz \tag{69}$$

Free entry condition for final goods producers

$$w\, F_F\, N_F = \frac{1}{\sigma} X_H - N_F \int_Z w\, f_m\, \frac{m(z)^\beta}{\beta}\, j(z)\, dz \tag{70}$$

where

$$X_H \equiv N_F \int_Z x_H(z)\, j(z)\, dz = N_F \int_Z p_F(z)^{1-\sigma}\, D_H\, j(z)\, dz = P^{H\,1-\sigma}\, D_H = I = w\,L + \Pi^W \tag{71}$$

is total sales of final goods producers.

Free entry condition for wholesalers

$$w\, F_W\, N_W \leq N_W \left[s\, \theta^W\, N_I\, \Pi_W - w\, f_W\, \frac{(s\, N_I/N_W)^{\beta_W}}{\beta_W} \right] \tag{72}$$

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' product adoption, and for financing entry of firms and wholesalers.

Labor market clearing condition

$$\begin{aligned}
\text{Labor Supply} =& \text{Production Labor} + \text{Ads Labor} + \text{Product Adoption Labor} + \text{Entry Cost Labor} \\
L =& \frac{1}{w} \left(\frac{\sigma-1}{\sigma} X_m + \frac{\sigma-1}{\sigma} X_W - \frac{\sigma-1}{\sigma} X_{Wm} \right) \\
&+ N_I \int_Z f_v\, \frac{v(z)^\beta}{\beta}\, j(z)\, dz + N_F \int_Z f_m\, \frac{m(z)^\beta}{\beta}\, j(z)\, dz \\
&+ f_W\, \frac{(s\, N_I/N_W)^{\beta_W}}{\beta_W}\, N_W + F_I\, N_F + F_F\, N_I + F_W\, N_W
\end{aligned} \tag{73}$$

B.8 Endogenous Entry

Using intermediate goods producer's FOC, we can rewrite the sum of their search cost as:

$$\begin{aligned}
& \sum_{d \in \mathcal{N}} N_{Ii} \int_Z w_i\, f_{vid}\, \frac{v_{id}(z)^\beta}{\beta}\, j_i(z) dz \\
=& \sum_{d \in \mathcal{N}} \frac{1}{\beta\sigma} N_{Ii} \int_Z [x_{mid}(z) - x_{Wid}(z)] j_i(z) dz \\
=& \sum_{d \in \mathcal{N}} \frac{1}{\beta\sigma} (X_{mid} - X_{Wmid}) \\
=& \frac{1}{\beta\sigma}\, \overline{\psi}_i^R \sum_{d \in \mathcal{N}} (X_{mid} + X_{Wid} - X_{Wmid}) \\
=& \frac{1}{\beta\sigma}\, \overline{\psi}_i^R \sum_{u \in \mathcal{N}} (X_{mui} + X_{Wui} - X_{Wmui}) \\
=& \frac{1}{\beta\sigma}\, \overline{\psi}_i^R\, \xi_i \sum_{u \in \mathcal{N}} (X_{mui} + X_{ui}^W - X_{mui}^W) \\
=& \frac{\sigma-1}{\beta\sigma^2}\, \overline{\psi}_i^R\, \xi_i\, w_i L_i
\end{aligned}$$

where

$$\overline{\psi}_i^R \equiv \frac{\sum_{d \in \mathcal{N}} (X_{mid} - X_{Wmid})}{\sum_{d \in \mathcal{N}} (X_{mid} + X_{Wid} - X_{Wmid})} \;\; , \;\; \xi_i \equiv \frac{\sum_{u \in \mathcal{N}} (X_{mui} + X_{Wui} - X_{Wmui})}{\sum_{u \in \mathcal{N}} (X_{mui} + X_{ui}^W - X_{mui}^W)}$$

The fifth line makes use of the trade balance condition (43). The seventh line makes use of the accounting relationship that the total cost of final goods production is equal to the total purchase of intermediate goods, and also household's budget constraint. Substitute it back to the free entry condition (39) yields:

$$
\begin{aligned}
w_i F_{Ii} N_{Ii} =& \sum_{d\in\mathcal{N}} \left(\frac{1}{\sigma} X_{mid} + \frac{1}{\sigma} X_{Wid} - \frac{1}{\sigma} X_{Wmid}\right) - \frac{\sigma-1}{\beta\sigma^2} \overline{\psi}_i^R \, \xi_i \, w_i L_i \\
w_i F_{Ii} N_{Ii} =& \frac{\sigma-1}{\sigma^2} \, \xi_i \, w_i L_i - \frac{\sigma-1}{\beta\sigma^2} \overline{\psi}_i^R \, \xi_i \, w_i L_i \\
N_{Ii} =& \frac{\sigma-1}{\beta\sigma^2} \xi_i (\beta - \overline{\psi}_i^R) \frac{L_i}{F_{Ii}} \\
d\log N_{Ii} =& d\log \xi_i - \frac{\overline{\psi}_i^R}{\beta - \overline{\psi}_i^R} d\log \overline{\psi}_i^R
\end{aligned} \tag{74}
$$

Consider an exogenous shock to iceberg trade costs. The first term increases with location i's direct trade share in sourcing, because there would then be more labor allocated away from financing the entry of wholesalers to the production of intermediate goods, which raises intermediate goods producer's share of total output and therefore the number of intermediate goods producers N_{Ii}. The second term decreases with location i's direct trade share in sales. The intuition is that a higher direct sales share raises the average fixed cost of operation (to search for customers), which effectively raises the entry barrier and lowers the number of intermediate goods producers.

Using final goods producer's FOC, we can rewrite the sum of their search cost as:

$$
\begin{aligned}
& \sum_{u\in\mathcal{N}} N_{Fi} \int_Z w_i \, f_{mui} \, \frac{m_{ui}(z)^\beta}{\beta} \, j_i(z) dz \\
=& \sum_{u\in\mathcal{N}} N_{Fi} \int_Z \frac{1}{\beta\sigma} \left(\frac{\sigma}{\sigma-1} \frac{1}{z}\right)^{1-\sigma} D_{Hi} \, \left(m_{ui}(z) \, c_{mui}^{1-\sigma} - m_{ui}(z) \, S_{ui} \, c_{Wmui}^{1-\sigma}\right) \, j_i(z) dz \\
=& \sum_{u\in\mathcal{N}} \frac{1}{\beta\sigma} \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} D_{Hi} \, \Lambda_{ui} \, \psi_{ui} \, N_{Fi} \int_Z \left(\frac{1}{z} \, c_i(z)\right)^{1-\sigma} j_i(z) dz \\
=& \sum_{u\in\mathcal{N}} \frac{1}{\beta\sigma} \Lambda_{ui} \, \psi_{ui} \, X_{Hi} \\
=& \frac{1}{\beta\sigma} \, \psi_i \, X_{Hi}
\end{aligned}
$$

where

$$
\begin{aligned}
\Lambda_{ui} \equiv & \frac{\int_Z \left[m_{ui}(z) c_{mui}^{1-\sigma} + S_{ui}\, c_{Wui}^{1-\sigma} - m_{ui}(z)\, S_{ui}\, c_{Wmui}^{1-\sigma}\right]\, z^{\sigma-1}\, j_i(z) dz}{\sum\limits_{l\in\mathcal{N}} \left\{\int_Z \left[m_{li}(z) c_{mli}^{1-\sigma} + S_{li}\, c_{Wli}^{1-\sigma} - m_{li}(z)\, S_{li}\, c_{Wmli}^{1-\sigma}\right]\, z^{\sigma-1}\, j_i(z) dz\right\}} \\
\psi_{ui} \equiv & \frac{\int_Z \left[m_{ui}(z) c_{mui}^{1-\sigma} - m_{ui}(z)\, S_{ui}\, c_{Wmui}^{1-\sigma}\right]\, z^{\sigma-1}\, j_i(z) dz}{\int_Z \left[m_{ui}(z) c_{mui}^{1-\sigma} + S_{ui}\, c_{Wui}^{1-\sigma} - m_{ui}(z)\, S_{ui}\, c_{Wmui}^{1-\sigma}\right]\, z^{\sigma-1}\, j_i(z) dz} \\
\psi_i \equiv & \frac{\sum\limits_{l\in\mathcal{N}} \left\{\int_Z \left[m_{li}(z) c_{mli}^{1-\sigma} - m_{li}(z)\, S_{li}\, c_{Wmli}^{1-\sigma}\right]\, z^{\sigma-1}\, j_i(z) dz\right\}}{\sum\limits_{l\in\mathcal{N}} \left\{\int_Z \left[m_{li}(z) c_{mli}^{1-\sigma} + S_{li}\, c_{Wli}^{1-\sigma} - m_{li}(z)\, S_{li}\, c_{Wmli}^{1-\sigma}\right]\, z^{\sigma-1}\, j_i(z) dz\right\}}
\end{aligned}
$$

Substitute it back to (40) yields:

$$
\begin{aligned}
w_i F_{Fi} N_{Fi} = & \frac{1}{\sigma} X_{Hi} - \frac{1}{\beta\sigma}\, \psi_i\, X_{Hi} \\
N_{Fi} = & \frac{1}{\beta\sigma}\, (\beta - \psi_i)\, \frac{L_i}{F_{Fi}} \\
d\log N_{Fi} = & -\frac{\psi_i}{\beta - \psi_i} d\log \psi_i
\end{aligned}
$$

Again, I consider an exogenous shock to iceberg trade costs. An increase in direct sourcing share lowers the mass of final goods producers. The intuition is that a higher direct sourcing share raises the average fixed cost of operation, which effectively raises the entry barrier.

Using wholesalers' FOC, we can rewrite the sum of their cost of product adoption as:

$$
\begin{aligned}
& w_i\, f_{Wui}\, \frac{(s_{ui}\, N_{Iu}/N_{Wi})^{\beta_W}}{\beta_W} \\
= & \frac{1}{\beta_W}\, S_{ui}\, N_{Iu}\, \Pi_{Wui}
\end{aligned}
$$

Substitute it back to (41) yields:

$$
\begin{aligned}
w_i F_{Wi} N_{Wi} = & N_{Wi} \sum_{u\in\mathcal{N}} \frac{\beta_W - 1}{\beta_W} S_{ui}\, N_{Iu}\, \Pi_{Wui} \\
w_i F_{Wi} = & \frac{\beta_W - 1}{\beta_W \sigma}\, N_{Wi}^{-2} \sum_{u\in\mathcal{N}} (X_{ui}^W - X_{mui}^W) \\
w_i F_{Wi}\, {N_{Wi}}^2 = & \frac{\beta_W - 1}{\beta_W \sigma}\, (1 - \Omega_i) \sum_{u\in\mathcal{N}} (X_{mui} + X_{ui}^W - X_{mui}^W) \\
w_i F_{Wi}\, {N_{Wi}}^2 = & \frac{\beta_W - 1}{\beta_W \sigma}\, (1 - \Omega_i)\, \frac{\sigma - 1}{\sigma}\, w_i L_i \\
N_{Wi} = & \left[\frac{(\beta_W - 1)(\sigma - 1)}{\beta_W \sigma^2}\, (1 - \Omega_i)\, \frac{L_i}{F_{Wi}}\right]^{\frac{1}{2}}
\end{aligned}
$$

where

$$\Omega_i \equiv \frac{\sum\limits_{u \in \mathcal{N}} X_{mui}}{\sum\limits_{u \in \mathcal{N}} (X_{mui} + X_{ui}^W - X_{mui}^W)}$$

is the total share of direct trade in location i's purchases of intermediate goods, and the fourth line makes use of equation (77).

B.9 Detailed Derivation for Social Planner's Problem

In this section, I present the detailed setup of the social planner's problem. The social planner chooses the quantities of intermediate goods produced and traded directly, $y_I(z, z')$, and indirectly, $y^W(z, z')$; the number of ads posted by firms, $m(z)$ and $v(z)$; the effort of product adoption chosen by wholesalers, s; and the levels of firm and wholesaler entry, 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 sourced directly and indirectly:

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

This is subject to a labor resource constraint:

$$L = L_p + L_A + L_S + L_E,$$

where

L_p = labor used in intermediate goods production,

L_A = labor used for posting search ads,

L_S = labor used for product adoption by wholesalers,

L_E = fixed labor costs for firm and wholesaler entry.

First, conditional on the network and without markups, the optimal quantity of directly and indirectly traded inputs chosen by the social planner implies the following final consumption price index:

$$
\begin{aligned}
P^H &= \left[N_F \int_Z p_F(z)^{1-\sigma} j(z) dz \right]^{\frac{1}{1-\sigma}} \\
&= \left\{ N_F \int_Z \left[\frac{1}{z} c(z) \right]^{1-\sigma} j(z) dz \right\}^{\frac{1}{1-\sigma}} \\
&= \left\{ N_F \int_Z \left[m(z)\, c_m^{1-\sigma} + S\, c_W^{1-\sigma} - m(z)\, S\, c_{Wm}^{1-\sigma} \right] z^{\sigma-1}\, j(z) dz \right\}^{\frac{1}{1-\sigma}}
\end{aligned}
$$

where

$$
\begin{aligned}
c_W^{1-\sigma} &= \int_Z p^W(z)^{1-\sigma}\, N_I\, j(z) = \int_Z \left(\frac{w}{z} \right)^{1-\sigma} N_I\, j(z) dz = N_I\, \mathbb{E}\left[z^{\sigma-1} \right] \\
c_m^{1-\sigma} &= \int_Z p_I(z)^{1-\sigma}\, \phi_c\, \theta^m\, \frac{N_I\, v(z)\, j(z)}{V} dz = \phi_c\, \frac{\widetilde{M}}{M\, V}\, N_I \int_Z z^{\sigma-1}\, v(z)\, j(z) dz \\
c_{Wm}^{1-\sigma} &= \phi_c^{-1}\, c_m^{1-\sigma}
\end{aligned}
$$

Therefore:

$$
\begin{aligned}
P^H =& \Bigg\{ (\phi_c - S)\, \frac{\widetilde{M}}{M\, V} \left[N_F \int_Z z^{\sigma-1}\, m(z)\, j(z) dz \right] \left[N_I \int_Z z^{\sigma-1}\, v(z)\, j(z) dz \right] \\
&+ N_F\, N_I\, S\, \left(\mathbb{E}\left[z^{\sigma-1} \right] \right)^2 \Bigg\}^{\frac{1}{1-\sigma}}
\end{aligned}
$$

We can then reformulate the social planner's problem into the following:

$$
\max_{\{m(z), v(z), s, N_I, N_F, N_W\}} \quad \frac{I}{P^H} = \frac{L - L_A - L_S - L_E}{P^H}
$$

s.t.

$$
\begin{aligned}
L =& L_p + L_A + L_S + L_E \\
P^H =& \Bigg\{ (\phi_c - S)\, \frac{\widetilde{M}}{M\, V} \left[N_F \int_Z z^{\sigma-1}\, m(z)\, j(z) dz \right] \left[N_I \int_Z z^{\sigma-1}\, v(z)\, j(z) dz \right] \\
&+ N_F\, N_I\, S\, \left(\mathbb{E}\left[z^{\sigma-1} \right] \right)^2 \Bigg\}^{\frac{1}{1-\sigma}}
\end{aligned}
$$

where

$$
\begin{aligned}
L_A =& \int_Z \left[f_m \frac{m(z)^\beta}{\beta} + f_v \frac{v(z)^\beta}{\beta} \right] j(z) dz \\
L_S =& f_W \, \frac{(s \, N_I / N_W)^{\beta_W}}{\beta_W} \\
L_E =& F_I \, N_I + F_F \, N_F + F_W \, N_W \\
S =& s \, \theta^W = s^{\lambda_W} \, {N_W}^{\lambda_W - 1} \\
M =& N_F \int_Z m(z) j(z) dz \\
V =& N_I \int_Z v(z) j(z) dz \\
\widetilde{M} =& \kappa V^{\lambda_V} M^{\lambda_M}
\end{aligned}
$$

FOC w.r.t. $m(z)$:

$$
\begin{aligned}
& f_m \, m(z)^{\beta-1} \\
=& (L - L_A - L_S - L_E) \, \frac{1}{\sigma - 1} \left\{ (\phi_c - S) \, \frac{\widetilde{M}}{M \, V} \, \left[N_F \, \int_Z z^{\sigma-1} \, m(z) \, j(z) dz \right] \, \left[N_I \, \int_Z z^{\sigma-1} \, v(z) \, j(z) dz \right] \right. \\
& \left. + N_F \, N_I \, S \, \left(\mathbb{E} \left[z^{\sigma-1} \right] \right)^2 \right\}^{-1} (\phi_c - S) \, \left[N_I \, \int_Z z^{\sigma-1} \, v(z) \, j(z) dz \right] \, \frac{\widetilde{M}}{M \, V} \, z^{\sigma-1} \\
& + (L - L_A - L_S - L_E) \, \frac{1}{\sigma - 1} \, \left(P^H \right)^{\sigma-1} (\phi_c - S) \, \left[N_F \, \int_Z z^{\sigma-1} \, m(z) \, j(z) dz \right] \, \left[N_I \, \int_Z z^{\sigma-1} \, v(z) \, j(z) dz \right] \\
& \kappa V^{\lambda_V - 1} \, (\lambda_M - 1) \, M^{\lambda_M - 2}
\end{aligned}
$$

FOC w.r.t. $v(z)$:

$$
\begin{aligned}
& f_v \, v(z)^{\beta-1} \\
=& (L - L_A - L_S - L_E) \, \frac{1}{\sigma - 1} \left\{ (\phi_c - S) \, \frac{\widetilde{M}}{M \, V} \, \left[N_F \, \int_Z z^{\sigma-1} \, m(z) \, j(z) dz \right] \, \left[N_I \, \int_Z z^{\sigma-1} \, v(z) \, j(z) dz \right] \right. \\
& \left. + N_F \, N_I \, S \, \left(\mathbb{E} \left[z^{\sigma-1} \right] \right)^2 \right\}^{-1} (\phi_c - S) \, \left[N_F \, \int_Z z^{\sigma-1} \, m(z) \, j(z) dz \right] \, \frac{\widetilde{M}}{M \, V} \, z^{\sigma-1} \\
& + (L - L_A - L_S - L_E) \, \frac{1}{\sigma - 1} \, \left(P^H \right)^{\sigma-1} (\phi_c - S) \, \left[N_F \, \int_Z z^{\sigma-1} \, m(z) \, j(z) dz \right] \, \left[N_I \, \int_Z z^{\sigma-1} \, v(z) \, j(z) dz \right] \\
& \kappa M^{\lambda_M - 1} \, (\lambda_V - 1) \, V^{\lambda_V - 2}
\end{aligned}
$$

FOC w.r.t. s:

$$
\begin{aligned}
& N_W\, f_W\, \left(\frac{s\, N_I}{N_W}\right)^{\beta_W - 1}\, \frac{N_I}{N_W} \\
=& (L - L_A - L_S - L_E)\, \frac{1}{\sigma - 1} \Bigg\{ (\phi_c - S)\, \frac{\widetilde{M}}{M\, V}\, \left[N_F \int_Z z^{\sigma-1}\, m(z)\, j(z) dz\right] \left[N_I \int_Z z^{\sigma-1}\, v(z)\, j(z) dz\right] \\
& + N_F\, N_I\, S\, \left(\mathbb{E}\left[z^{\sigma-1}\right]\right)^2 \Bigg\}^{-1} \Bigg\{ N_F\, N_I\, \left(\mathbb{E}\left[z^{\sigma-1}\right]\right)^2 \\
& - \frac{\widetilde{M}}{M\, V}\, \left[N_F \int_Z z^{\sigma-1}\, m(z)\, j(z) dz\right] \left[N_I \int_Z z^{\sigma-1}\, v(z)\, j(z) dz\right] \Bigg\}\, s^{\lambda_W - 1}\, N_W^{\lambda_W - 1}\, \lambda_W
\end{aligned}
$$

FOC w.r.t. N_W:

$$
\begin{aligned}
F_W =& (\beta_W - 1)\, f_W\, \frac{(s\, N_I / N_W)^{\beta_W}}{\beta_W} - (L - L_A - L_S - L_E)\, \frac{1}{\sigma - 1}\, \left(P^H\right)^{\sigma-1} \\
& \left\{ N_F\, N_I\, \left(\mathbb{E}\left[z^{\sigma-1}\right]\right)^2 - \frac{\widetilde{M}}{M\, V}\, \left[N_F \int_Z z^{\sigma-1}\, m(z)\, j(z) dz\right] \left[N_I \int_Z z^{\sigma-1}\, v(z)\, j(z) dz\right] \right\} \\
& (1 - \lambda_W)\, s^{\lambda_W}\, N_W^{\lambda_W - 2}
\end{aligned}
$$

FOC w.r.t. N_F:

$$
\begin{aligned}
& F_F\, N_F + N_F \int_Z f_m \frac{m(z)^\beta}{\beta}\, j(z) dz \\
=& \frac{1}{\sigma - 1}\, (L - L_A - L_S - L_E) + (L - L_A - L_S - L_E) \frac{1}{\sigma - 1}\, \left(P^H\right)^{\sigma-1} (\phi_c - S) \\
& \left[N_F \int_Z z^{\sigma-1}\, m(z)\, j(z) dz\right] \left[N_I \int_Z z^{\sigma-1}\, v(z)\, j(z) dz\right] (\lambda_M - 1)\, \frac{\widetilde{M}}{M\, V}
\end{aligned}
$$

FOC w.r.t. N_I:

$$
\begin{aligned}
& F_I\, N_I + N_I \int_Z f_v \frac{v(z)^\beta}{\beta}\, j(z) dz \\
=& \frac{1}{\sigma - 1}\, (L - L_A - L_S - L_E) + (L - L_A - L_S - L_E) \frac{1}{\sigma - 1}\, \left(P^H\right)^{\sigma-1} (\phi_c - S) \\
& \left[N_F \int_Z z^{\sigma-1}\, m(z)\, j(z) dz\right] \left[N_I \int_Z z^{\sigma-1}\, v(z)\, j(z) dz\right] (\lambda_V - 1)\, \frac{\widetilde{M}}{M\, V} \\
& - N_W\, f_W\, \left(\frac{s\, N_I}{N_W}\right)^{\beta_W}
\end{aligned}
$$

Comparing these FOC against those in the decentralized equilibrium gives us the following wedges discussed in Section 4.1:

$$
\begin{aligned}
&\Delta_{markup}(m(z)) \\
=&\frac{\left\{(\phi_c - S)\,\frac{\widetilde{M}}{M\,V}\,\left[N_F\,\int_Z z^{\sigma-1}\,m(z)\,j(z)dz\right]\,\left[N_I\,\int_Z z^{\sigma-1}\,v(z)\,j(z)dz\right] + N_F\,N_I\,S\,\left(\mathbb{E}\left[z^{\sigma-1}\right]\right)^2\right\}}{\left\{(\phi_c - (\mu^W)^{-\sigma} S)\,\frac{\widetilde{M}}{M\,V}\,\left[N_F\,\int_Z z^{\sigma-1}\,m(z)\,j(z)dz\right]\,\left[N_I\,\int_Z z^{\sigma-1}\,v(z)\,j(z)dz\right] + (\mu^W)^{-\sigma}\,N_F\,N_I\,S\,(\mathbb{E}\left[z^{\sigma-1}\right])^2\right\}} \\
&\frac{\phi_c - S\,(\mu^W)^{1-\sigma}}{\phi_c - S}\,\frac{\sigma}{\sigma-1} > 1
\end{aligned}
$$

$$
\begin{aligned}
&\Delta_{congestion}(m(z)) \\
=&(L - L_A - L_S - L_E)\,\frac{1}{\sigma-1}\,(P^H)^{\sigma-1}\,(\phi_c - S)\,\left[N_F\,\int_Z z^{\sigma-1}\,m(z)\,j(z)dz\right]\,\left[N_I\,\int_Z z^{\sigma-1}\,v(z)\,j(z)dz\right] \\
&\kappa V^{\lambda_V - 1}\,(1-\lambda_M)\,M^{\lambda_M - 2} \geq 0
\end{aligned}
$$

$$
\begin{aligned}
&\Delta_{markup}(v(z)) \\
=&\frac{\left\{(\phi_c - S)\,\frac{\widetilde{M}}{M\,V}\,\left[N_F\,\int_Z z^{\sigma-1}\,m(z)\,j(z)dz\right]\,\left[N_I\,\int_Z z^{\sigma-1}\,v(z)\,j(z)dz\right] + N_F\,N_I\,S\,\left(\mathbb{E}\left[z^{\sigma-1}\right]\right)^2\right\}}{\left\{(\phi_c - (\mu^W)^{-\sigma} S)\,\frac{\widetilde{M}}{M\,V}\,\left[N_F\,\int_Z z^{\sigma-1}\,m(z)\,j(z)dz\right]\,\left[N_I\,\int_Z z^{\sigma-1}\,v(z)\,j(z)dz\right] + (\mu^W)^{-\sigma}\,N_F\,N_I\,S\,(\mathbb{E}\left[z^{\sigma-1}\right])^2\right\}} \\
&\frac{\phi_c - S\,(\mu^W)^{-\sigma}}{\phi_c - S} > 1
\end{aligned}
$$

$$
\begin{aligned}
&\Delta_{congestion}(v(z)) \\
=&(L - L_A - L_S - L_E)\,\frac{1}{\sigma-1}\,(P^H)^{\sigma-1}\,(\phi_c - S)\,\left[N_F\,\int_Z z^{\sigma-1}\,m(z)\,j(z)dz\right]\,\left[N_I\,\int_Z z^{\sigma-1}\,v(z)\,j(z)dz\right] \\
&\kappa M^{\lambda_M - 1}\,(1-\lambda_V)\,V^{\lambda_V - 2} \geq 0
\end{aligned}
$$

$$
\begin{aligned}
&\Delta_{markup}(s) \\
=&\frac{\left\{(\phi_c - S)\,\frac{\widetilde{M}}{M\,V}\,\left[N_F\,\int_Z z^{\sigma-1}\,m(z)\,j(z)dz\right]\,\left[N_I\,\int_Z z^{\sigma-1}\,v(z)\,j(z)dz\right] + N_F\,N_I\,S\,\left(\mathbb{E}\left[z^{\sigma-1}\right]\right)^2\right\}}{\left\{(\phi_c - (\mu^W)^{-\sigma} S)\,\frac{\widetilde{M}}{M\,V}\,\left[N_F\,\int_Z z^{\sigma-1}\,m(z)\,j(z)dz\right]\,\left[N_I\,\int_Z z^{\sigma-1}\,v(z)\,j(z)dz\right] + (\mu^W)^{-\sigma}\,N_F\,N_I\,S\,(\mathbb{E}\left[z^{\sigma-1}\right])^2\right\}} \\
&\left(\mu^W\right)^{1-\sigma}\,\frac{\mu^W - 1}{\mu^W}\,\sigma
\end{aligned}
$$

$$
\Delta_{congestion}(s) = \frac{1}{\lambda_W} \geq 1
$$

$$
\begin{aligned}
&\Delta_{congestion}(N_W) = (L - L_A - L_S - L_E)\,\frac{1}{\sigma - 1}\,\left(P^H\right)^{\sigma-1}\,(1 - \lambda_W)\,s^{\lambda_W}\,N_W^{\lambda_W - 2} \\
&\left\{N_F\,N_I\,\left(\mathbb{E}\left[z^{\sigma-1}\right]\right)^2 - \frac{\widetilde{M}}{M\,V}\,\left[N_F\,\int_Z z^{\sigma-1}\,m(z)\,j(z)dz\right]\,\left[N_I\,\int_Z z^{\sigma-1}\,v(z)\,j(z)dz\right]\right\} \geq 0
\end{aligned}
$$

$$
\begin{aligned}
&\Delta_{congestion}(N_I) \\
=&(L - L_A - L_S - L_E)\,\frac{1}{\sigma - 1}\,\left(P^H\right)^{\sigma-1}\,(\phi_c - S)\,\left[N_F\,\int_Z z^{\sigma-1}\,m(z)\,j(z)dz\right]\,\left[N_I\,\int_Z z^{\sigma-1}\,v(z)\,j(z)dz\right] \\
&(1 - \lambda_V)\,\frac{\widetilde{M}}{M\,V} \geq 0
\end{aligned}
$$

$$
\Delta_{AdoptionCost}(N_I) = -N_W\,f_W\,\left(\frac{s\,N_I}{N_W}\right)^{\beta_W} < 0
$$

$$
\begin{aligned}
&\Delta_{markup}(N_F) \\
=&\frac{\left\{(\phi_c - \left(\mu^W\right)^{1-\sigma} S)\,\frac{\widetilde{M}}{M\,V}\,\left[N_F\,\int_Z z^{\sigma-1}\,m(z)\,j(z)dz\right]\,\left[N_I\,\int_Z z^{\sigma-1}\,v(z)\,j(z)dz\right] + \left(\mu^W\right)^{1-\sigma}\,N_F\,N_I\,S\,\left(\mathbb{E}\left[z^{\sigma-1}\right]\right)^2\right\}}{\left\{(\phi_c - (\mu^W)^{-\sigma} S)\,\frac{\widetilde{M}}{M\,V}\,\left[N_F\,\int_Z z^{\sigma-1}\,m(z)\,j(z)dz\right]\,\left[N_I\,\int_Z z^{\sigma-1}\,v(z)\,j(z)dz\right] + (\mu^W)^{-\sigma}\,N_F\,N_I\,S\,(\mathbb{E}\left[z^{\sigma-1}\right])^2\right\}} \\
&\frac{\sigma}{\sigma - 1} > 1
\end{aligned}
$$

$$
\begin{aligned}
&\Delta_{congestion}(N_F) \\
=&(L - L_A - L_S - L_E)\,\frac{1}{\sigma - 1}\,\left(P^H\right)^{\sigma-1}\,(\phi_c - S)\,\left[N_F\,\int_Z z^{\sigma-1}\,m(z)\,j(z)dz\right]\,\left[N_I\,\int_Z z^{\sigma-1}\,v(z)\,j(z)dz\right] \\
&(1 - \lambda_M)\,\frac{\widetilde{M}}{M\,V} \geq 0
\end{aligned}
$$

B.10 Proof of Proposition 1

Proof. Define the following terms only for this proof:

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

$$P_Q^H \equiv \left\{ (\phi_c - (\tau^W\, \mu^W)^{-\sigma}\, S)\, \frac{\widetilde{M}}{M\,V} \left[N_F \int_Z z^{\sigma-1}\, m(z)\, j(z) dz \right] \left[N_I \int_Z z^{\sigma-1}\, v(z)\, j(z) dz \right] + (\tau^W\, \mu^W)^{-\sigma}\, N_F\, N_I\, S\, \left(\mathbb{E}\left[z^{\sigma-1} \right] \right)^2 \right\}^{\frac{1}{1-\sigma}}$$

$$P_X^H \equiv \left\{ (\phi_c - (\tau^W\, \mu^W)^{1-\sigma}\, S)\, \frac{\widetilde{M}}{M\,V} \left[N_F \int_Z z^{\sigma-1}\, m(z)\, j(z) dz \right] \left[N_I \int_Z z^{\sigma-1}\, v(z)\, j(z) dz \right] + (\tau^W\, \mu^W)^{1-\sigma}\, N_F\, N_I\, S\, \left(\mathbb{E}\left[z^{\sigma-1} \right] \right)^2 \right\}^{\frac{1}{1-\sigma}}$$

We can rewrite the FOC in the decentralized equilibrium with taxes and subsidies as follows:
FOC w.r.t. $m(z)$:

$$\begin{aligned} \tau_m^M\, f_m\, m(z)^{\beta-1} =& (L - L_A - L_S - L_E)\, \frac{1}{\sigma-1}\, (P^H)^{\sigma-1}\, (\phi_c - S) \left[N_I \int_Z z^{\sigma-1}\, v(z)\, j(z) dz \right] \frac{\widetilde{M}}{M\,V}\, (z^{\sigma-1} + \lambda_M - 1) \\ &+ \left(\frac{P^H}{P_Q^H} \right)^{1-\sigma} \frac{\phi_c - S\,(\tau^W\, \mu^W)^{1-\sigma}}{\phi_c - S} \cdot \frac{z^{\sigma-1}}{z^{\sigma-1} + \lambda_M - 1} \cdot \frac{\sigma}{\sigma-1} \end{aligned}$$

FOC w.r.t. $v(z)$:

$$\begin{aligned} \tau_v^M\, f_v\, v(z)^{\beta-1} =& (L - L_A - L_S - L_E)\, \frac{1}{\sigma-1}\, (P^H)^{\sigma-1}\, (\phi_c - S) \left[N_F \int_Z z^{\sigma-1}\, m(z)\, j(z) dz \right] \frac{\widetilde{M}}{M\,V}\, (z^{\sigma-1} + \lambda_V - 1) \\ &+ \left(\frac{P^H}{P_Q^H} \right)^{1-\sigma} \frac{\phi_c - S\,(\tau^W\, \mu^W)^{1-\sigma}}{\phi_c - S} \cdot \frac{z^{\sigma-1}}{z^{\sigma-1} + \lambda_V - 1} \end{aligned}$$

FOC w.r.t. s:

$$\begin{aligned} & \tau_s^M\, N_W\, f_W \left(\frac{s\, N_I}{N_W} \right)^{\beta_W - 1} \frac{N_I}{N_W} \\ =& (L - L_A - L_S - L_E)\, \frac{1}{\sigma-1}\, (P^H)^{\sigma-1} \\ & \cdot \left\{ N_F\, N_I\, \left(\mathbb{E}\left[z^{\sigma-1} \right] \right)^2 - \frac{\widetilde{M}}{M\,V} \left[N_F \int_Z z^{\sigma-1}\, m(z)\, j(z) dz \right] \left[N_I \int_Z z^{\sigma-1}\, v(z)\, j(z) dz \right] \right\} s^{\lambda_W - 1}\, N_W^{\lambda_W - 1}\, \lambda_W \\ & \cdot (\tau^W\, \mu^W)^{1-\sigma} \cdot \frac{\mu^W - 1}{\mu^W} \cdot \sigma \cdot \frac{1}{\lambda_W} \end{aligned}$$

FOC w.r.t. N_W:

$$\tau_W^E\, F_W \leq \left[\beta_W \left(\frac{2\,\lambda_W - 1}{\lambda_W} \right) - 1 \right] f_W\, \frac{(s\, N_I / N_W)^{\beta_W}}{\beta_W} \cdot \frac{(\beta_W - 1)\, \tau_s^M}{\beta_W \left(\frac{2\,\lambda_W - 1}{\lambda_W} \right) - 1}$$

FOC w.r.t. N_F:

$$\begin{aligned} \tau_F^E\, F_F\, N_F =& \frac{1}{\sigma-1}\, (L - L_A - L_S - L_E) \left\{ (1 - \frac{\psi^*}{\beta}) + \left[\psi^*\, (\lambda_M - 1) - \frac{\psi^*}{\beta}\, (\lambda_M - 1)\, \frac{M^*}{M_{Z^*}} \right] \right\} \\ & \cdot \left(\frac{P_X^H}{P_Q^H} \right)^{1-\sigma} \cdot \frac{\sigma}{\sigma-1} \cdot \frac{1 - \frac{\psi}{\beta}}{(1 - \frac{\psi^*}{\beta}) + \left[\psi^*\, (\lambda_M - 1) - \frac{\psi^*}{\beta}\, (\lambda_M - 1)\, \frac{M^*}{M_{Z^*}} \right]} \end{aligned}$$

FOC w.r.t. N_I:

$$\tau_I^E\, F_I\, N_I = \frac{1}{\sigma-1}\left(L - L_A - L_S - L_E\right)\left\{(1-\frac{\psi^*}{\beta}) + \left[\psi^*\left(\lambda_V - 1\right) - \frac{\psi^*}{\beta}\left(\lambda_V - 1\right)\frac{V^*}{V_Z{}^*}\right] - \left(1-\Omega^*\right)\left(\frac{S^*}{\theta^{W^*}}\right)^{1-\lambda_W}\lambda_W\right\}$$
$$\cdot \frac{1 - \left(\frac{P_X^H}{P_Q^H}\right)^{1-\sigma}\cdot\frac{\psi}{\beta}}{(1-\frac{\psi^*}{\beta}) + \left[\psi^*\left(\lambda_V - 1\right) - \frac{\psi^*}{\beta}\left(\lambda_V - 1\right)\frac{V^*}{V_Z{}^*}\right] - \left(1-\Omega^*\right)\left(\frac{S^*}{\theta^{W^*}}\right)^{1-\lambda_W}\lambda_W}$$

Comparing these against the social planner FOCs in Appendix B.9 confirms that the two sets of FOCs coincide when the taxes and subsidies satisfy the expressions in Proposition 1. □

B.11 Proof of Proposition 2

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:

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

Step 3. Income in terms of revenues The household budget constraint (71) and the identical markup used by final goods firms imply

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

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

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

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

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

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 2. □

B.12 Corollary 2.1

Corollary 2.1 (Productivity gap and dispersion, first-order approximation with varying Ω^*)**.** *In the single-location economy of Proposition 2, define*

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

where $A(\mu^W, \Omega^*)$ and $A_{\textit{efficient}}(\Omega^*)$ are given in Proposition 2, $\mu_D = (\sigma/(\sigma-1))^2$, $\mu_I = \mu_D\, \mu^W$, and $\Omega^* \equiv A_D/(A_D + A_I)$ 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 2,

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

Step 2 (compact ratio). Define

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

so that

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

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

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

For $\mu^W > 1$, $-\mu^W V + U = A_D(1-\mu^W) < 0$, 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 $\Omega^* \equiv A_D/(A_D+A_I)$) 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.13 Proof of Proposition 3

Proof. Under the simplifying assumptions of proposition 3, the social planner's FOC for m and v, derived in Appendix B.9, 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 FOC 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:

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

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 2 for the decentralized equilibrium to arrive at $A_{\text{decentralized}}$. □

B.14 Corollary 3.1

Corollary 3.1 (Welfare gap and dispersion, first-order approximation with varying Ω^*)**.** *Under the assumptions of Proposition 3, define*

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

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 3). Consequently, for small changes, the welfare wedge* $\mathcal{W}_{\text{efficient}} - \mathcal{W}$ *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. From (**??**), $\mathcal{W} = L_p(\mu^W)\; X(\mu^W)^{1/(\sigma-1)}$ with

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

and $L_p(\mu^W)$ and $\widetilde{M}(\mu^W)$ given in Proposition 3. 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

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

where $\mu_D = (\sigma/(\sigma-1))^2$ and $\mu_I = \mu_D\,\mu^W$. Following Proposition 3, 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 3). 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. $\square$

B.15 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:*

$$\widehat{\frac{I_i}{P_i^H}} = \widehat{T}_i \left(\frac{\widehat{\Omega_{ii}}}{\widehat{N_{Fi}}\,\widehat{\overline{M}_{ii}}} \, \widehat{\Lambda_{ii}} \right)^{\frac{1}{1-\sigma}} \tag{75}$$

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

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

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:

$$\Lambda_{ud} \equiv \frac{\int_Z \left[m_{ud}(z) c_{mud}^{1-\sigma} + S_{ud}\, c_{Wud}^{1-\sigma} - m_{ud}(z)\, S_{ud}\, c_{Wmud}^{1-\sigma}\right] z^{\sigma-1}\, j_d(z) dz}{\sum\limits_{l \in \mathcal{N}} \left\{\int_Z \left[m_{ld}(z) c_{mld}^{1-\sigma} + S_{ld}\, c_{Wld}^{1-\sigma} - m_{ld}(z)\, S_{ld}\, c_{Wmld}^{1-\sigma}\right] z^{\sigma-1}\, j_d(z) dz\right\}} = \frac{X_{mud} + X_{ud}^W - X_{mud}^W}{\sum\limits_{l \in \mathcal{N}} \left(X_{mld} + X_{ld}^W - X_{mld}^W\right)}$$

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

$$\Omega_{ud} \equiv \frac{\int_Z m_{ud}(z) c_{mud}^{1-\sigma}\, z^{\sigma-1}\, j_d(z) dz}{\int_Z \left[m_{ud}(z) c_{mud}^{1-\sigma} + S_{ud}\, c_{Wud}^{1-\sigma} - m_{ud}(z)\, S_{ud}\, c_{Wmud}^{1-\sigma}\right] z^{\sigma-1}\, j_d(z) dz} = \frac{X_{mud}}{X_{mud} + X_{mud}^W - X_{mud}^W}$$

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:

$$\begin{aligned}
(P_i^H)^{1-\sigma} =& N_{Fi} \int_Z \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \left\{\sum_{u \in \mathbb{N}} \left[m_{ui}(z)\, c_{mui}^{1-\sigma} + S_{ui}\, c_{Wui}^{1-\sigma} - m_{ui}(z)\, S_{ui}\, c_{Wmui}^{1-\sigma}\right]\right\} z^{\sigma-1}\, j_i(z) dz \\
(P_i^H)^{1-\sigma} =& N_{Fi} \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \left\{\int_Z \left[m_{ii}(z)\, c_{mii}^{1-\sigma} + S_{ii}\, c_{Wii}^{1-\sigma} - m_{ii}(z)\, S_{ii}\, c_{Wmii}^{1-\sigma}\right] z^{\sigma-1}\, j_i(z) dz\right\} \Lambda_{ii}^{-1} \\
(P_i^H)^{1-\sigma} =& N_{Fi} \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \int_Z m_{ii}(z)\, c_{mii}^{1-\sigma}\, z^{\sigma-1}\, j_i(z) dz\, \Omega_{ii}^{-1}\, \Lambda_{ii}^{-1} \\
(P_i^H)^{1-\sigma} =& N_{Fi} \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \Pi_{H1ii}\, c_{mii}^{1-\sigma}\, \mathbb{E}_i\left[z^{\gamma}\right]\, \Omega_{ii}^{-1}\, \Lambda_{ii}^{-1} \\
(\widehat{P_i^H})^{1-\sigma} =& \widehat{N_{Fi}}\, \widehat{\Pi_{H1ii}}\, \widehat{c_{mii}}^{1-\sigma}\, \widehat{\Omega_{ii}}^{-1}\, \widehat{\Lambda_{ii}}^{-1}
\end{aligned}$$

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

$$\begin{aligned}
\widetilde{M}_{ii} =& \theta_{ii}^m\, M_{ii} \\
\widehat{\widetilde{M}}_{ii} =& \widehat{\theta_{ii}^m}\, \widehat{M_{ii}} = \widehat{\theta_{ii}^m}\, \widehat{N_{Fi}}\, \widehat{\Pi_{H1ii}} = \widehat{N_{Fi}}\, \widehat{\overline{M}}_{ii} \\
\widehat{\overline{M}}_{ii} \equiv& \widehat{\theta_{ii}^m}\, \widehat{\Pi_{H1ii}}
\end{aligned}$$

Also, from (29) we know that

$$\widehat{c_{mii}}^{1-\sigma} = \widehat{\theta_{ii}^m}\, \widehat{w_i}^{1-\sigma}$$

Therefore,

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

□

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[46]:

$$\widehat{N_{Fi}} = \frac{\beta}{\beta - \psi_i} - \frac{\psi_i}{\beta - \psi_i}\widehat{\psi_i} \tag{76}$$

where

$$\psi_i \equiv \frac{\sum\limits_{l \in \mathcal{N}} \left\{ \int_Z \left[m_{li}(z) c_{mli}^{1-\sigma} - m_{li}(z)\, S_{li}\, c_{Wmli}^{1-\sigma} \right] z^{\sigma-1}\, j_i(z) dz \right\}}{\sum\limits_{l \in \mathcal{N}} \left\{ \int_Z \left[m_{li}(z) c_{mli}^{1-\sigma} + S_{li}\, c_{Wli}^{1-\sigma} - m_{li}(z)\, S_{li}\, c_{Wmli}^{1-\sigma} \right] z^{\sigma-1}\, j_i(z) dz \right\}}$$

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 (76) 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 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

[46] A detailed derivation is provided in Appendix B.8

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 product adoption 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 (75) 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.17.

B.16 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:

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

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 (71) 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:

$$I_i = \frac{\sigma}{\sigma-1}\sum_{u\in\mathcal{N}}\left(X_{mui} + X_{ui}^W - X_{mui}^W\right) \tag{77}$$

We can then log-linearize equation (77) 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:*

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

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 $X_{ui}^W - X_{mui}^W$ from equation (77) using equations (79) and (81), and using equation (68) to substitute out D_{Hi} yields the following log-linearized equation:

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

where $\overline{M}_{ud}\equiv\frac{\widetilde{M}_{ud}}{N_{Fd}}$ and $\overline{M}^W_{ud}\equiv S_{ud}\,N_{Iu}$ represent the average number of direct and (potential) indirect matches per final goods producer, $\mu^W_i\equiv\frac{N_{Wi}\,\sigma}{N_{Wi}\,\sigma-1}$ is wholesale markup in location i.

Simplify the above equation yields:

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

which can be rewritten in vector form:

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

where $\boldsymbol{\Omega}$ is a $|\mathcal{N}|\times|\mathcal{N}|$ diagonal matrix whose (i,i)-th element are Ω_i; $\boldsymbol{\chi}$, $\boldsymbol{\chi_m}$, $\boldsymbol{\chi^W}$, $\boldsymbol{\chi_m^W}$, $d\log\boldsymbol{\phi_c}$, $d\log\overline{\boldsymbol{M}}$, 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; $\boldsymbol{1}$ is a $|\mathcal{N}|\times 1$ column vector whose entries are all 1 . Note that $\odot$ denotes element-wise multiplication between matrices. Also:

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

$$\chi_{mud} \equiv \frac{X_{mud}}{\sum\limits_{l\in\mathcal{N}} (X_{mld})}\ ,\ \chi_{ud}^W \equiv \frac{X_{ud}^W}{\sum\limits_{l\in\mathcal{N}} (X_{ld}^W)}\ ,\ \chi_{mud}^W \equiv \frac{X_{mud}^W}{\sum\limits_{l\in\mathcal{N}} (X_{mld}^W)}$$

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:

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

□

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 (77)), 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 rate of product adoption. 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[47]:

$$\widehat{N_{Wi}} = \left[\left(\frac{1}{1-\Omega_i} - \frac{\Omega_i}{1-\Omega_i}\, \widehat{\Omega_i} \right) \frac{\widehat{L_i}}{\widehat{F_{Wi}}} \right]^{\frac{1}{2}} \tag{78}$$

Equation (78) 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 on the same shocks and the **same change in the number of matches**, the current model features

[47] A detailed derivation is provided in Appendix B.8

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.

B.17 Gravity Equation

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

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

where the second line makes use of equation (47). Now we can substitute out D_{mud} using B.4:

$$\begin{aligned}
X_{mud} =& N_{Iu}\, \mathbb{E}_u\left[z^{\gamma}\right] \left[\left(\frac{\sigma}{\sigma-1} w_u \tau_{ud} \right)^{\beta(1-\sigma)} \left(\frac{C_{1ud}}{w_u\, f_{vud}\, \sigma} \right) \right]^{\frac{1}{\beta-1}} \theta_{ud}^{m\,\frac{\beta}{\beta-1}} \phi_{cud} \\
& \left(\frac{N_{Fd}}{N_{Iu}} \right) \left(\frac{\sigma-1}{\sigma} \right)^{1+\gamma} \left(\frac{f_{vud}}{f_{mud}} \right)^{\frac{1}{\beta-1}} \left(\frac{C_{2ud}}{C_{1ud}} \right)^{\frac{1}{\beta-1}} \left(\frac{w_u}{w_d} \right)^{\frac{1}{\beta-1}} D_{Hd}^{\frac{\beta}{\beta-1}} \frac{\mathbb{E}_d\left[z^{\gamma}\right]}{\mathbb{E}_u\left[z^{\gamma}\right]} \left(\frac{\mathbb{E}_u\left[z^{\gamma}\right]}{\mathbb{E}_u\left[z^{\frac{\gamma}{\beta}}\right]} \right)^{\frac{\beta}{\beta-1}}
\end{aligned} \tag{79}$$

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:

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

where the second line makes use of equation (26), and the third line makes use of equation (60). It can then be shown that:

$$
X_{mud} = \widetilde{M}_{ud} \left(\frac{\sigma}{\sigma - 1} \right)^{1-2\sigma} w_u^{1-\sigma} \tau_{ud}^{1-\sigma} \phi_{cud} w_d L_d P_d^{H\sigma-1} \frac{\mathbb{E}_d \left[z^{\gamma} \right]}{\mathbb{E}_d \left[z^{\frac{\gamma}{\beta}} \right]} \frac{\mathbb{E}_u \left[z^{\gamma} \right]}{\mathbb{E}_u \left[z^{\frac{\gamma}{\beta}} \right]} \tag{80}
$$

Location d wholesalers' total sales of location u intermediate goods is:

$$
\begin{aligned}
& X_{ud}^W - X_{mud}^W \\
=& \mu_d^W X_{Wud} - \mu_d^W X_{Wmud} \\
=& \mu_d^W \left(\frac{\sigma}{\sigma - 1} \right)^{1-\sigma} \left(w_u \tau_{ud} \right)^{1-\sigma} S_{ud} D_{Wud} N_{Iu} \mathbb{E}_u \left[z^{\sigma-1} \right] - \mu_d^{W1-\sigma} S_{ud} \phi_{cud}{}^{-1} X_{mud}
\end{aligned}
$$

Now,

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

where the second and third lines make use of equations (35) and (45). Substitute D_{Wud} out then yields:

$$\begin{aligned} X_{ud}^W - X_{mud}^W =& \mu_d^{W\,1-\sigma} \left(\frac{\sigma}{\sigma-1}\right)^{1-2\sigma} (w_u\,\tau_{ud})^{1-\sigma}\; S_{ud}\, D_{Hd}\, N_{Fd}\, N_{Iu}\, \mathbb{E}_d\left[z^{\sigma-1}\right]\, \mathbb{E}_u\left[z^{\sigma-1}\right] \\ & - \mu_d^{W\,1-\sigma}\, S_{ud}\, {\phi_{cud}}^{-1}\, X_{mud} \end{aligned} \tag{81}$$

B.18 Gravity Equation and Distance Puzzle

In this section, I show that the model continues to admit a tractable characterization of gravity equation after incorporating indirect trade through wholesalers in the production network. As shown in the appendix B.17, the total direct sales of location u intermediate goods producers to location d final goods producers X_{mud} is given by:

$$X_{mud} = \rho^D\, \chi_{ud}^D\, \xi_u^D\, \xi_d^D$$

where ρ^D is some constant, while ξ_u^D and ξ_d^D are the origin and destination shifters. χ_{ud}^D is the multilateral resistance term given by:

$$\chi_{ud}^D = \kappa_{ud}^{\frac{\beta}{\beta-1}}\, f_{vud}^{-\frac{\lambda}{\beta-1}}\, f_{mud}^{-\frac{1-\lambda}{\beta-1}}\, (\tau_{ud}^{1-\sigma})^{\frac{\beta}{\beta-1}}\, C_{1ud}^{\frac{\lambda}{\beta-1}}\, C_{2ud}^{\frac{1-\lambda}{\beta-1}}\, \phi_{cud}^{1+\frac{\lambda}{\beta-1}}$$

where κ_{ud} governs the matching efficiency, f_{vud} and f_{mud} control the levels of search cost, τ_{ud} is the iceberg trade cost, C_{1ud} and C_{2ud} summarize the benefit of direct trade versus indirect trade, while ϕ_{cud} is the customization productivity gain incurred in direct trade.

On the other hand, the indirect sales of location u intermediate goods to location d final goods producers through wholesalers $X_{ud}^W - X_{mud}^W$ is given by:

$$\begin{aligned} X_{ud}^W - X_{mud}^W =& \left(\frac{\sigma}{\sigma-1}\right)^{2-3\sigma} (w_u\,\tau_{ud})^{1-\sigma}\; S_{ud}\, D_{Hd}\, N_{Fd}\, N_{Iu}\, \mathbb{E}_d\left[z^{\sigma-1}\right]\, \mathbb{E}_u\left[z^{\sigma-1}\right] \\ & - \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \phi_{cud}^{-1}\, S_{ud}\, X_{mud} \end{aligned}$$

where the first term is the total potential indirect trade through wholesalers were there no direct trade at all, and the second term captures the part of that potential indirect trade that is cannibalized by direct trade. As the expression shows, the total indirect trade flow no longer admits a form that is multiplicative in the multilateral resistance terms and the origin and destination shifters, and so as the total trade flow ($X_{ud} \equiv X_{mud} + X_{ud}^W - X_{mud}^W$). However, that does not prevent us from deriving the elasticity of total trade flow with respect to the multilateral resistance terms. In the following, I am going to derive the trade elasticity (elasticity of total trade flow with respect to the iceberg trade cost) as an illustration.

First, to simplify the exposition, I am going to assume that manufacturing firms and wholesalers do not internalize the fact that their trading cannibalize each other's trade. This tends to overstate the magnitude of the trade elasticity. For example, an increase in the iceberg trade cost tends to lower product adoption, which reduces cannibalization of direct trade and dampens the extent that firms cut back on forming direct matches. But reassuringly, these cannibalization forces do not affect the ranking of direct and indirect trade, as confirmed in simulation exercise. Under this assumption, the direct trade elasticity is simply given by:

$$\frac{d\log X_{mud}}{d\log \tau_{ud}} = \frac{\beta(1-\sigma)}{\beta-1} = -\gamma$$

To derive the indirect trade elasticity, note that without internalizing cannibalization, $\frac{d\log S_{ud}}{d\log \tau_{ud}} = \frac{1-\sigma}{\beta_W-1}$. Thus,

$$\begin{aligned}\frac{d\log(X^W_{ud}-X^W_{mud})}{d\log\tau_{ud}} =& \frac{X^W_{ud}}{X^W_{ud}-X^W_{mud}}\left[\frac{d\log S_{ud}}{d\log\tau_{ud}}+(1-\sigma)\right] - \frac{X^W_{mud}}{X^W_{ud}-X^W_{mud}}\left(\frac{d\log S_{ud}}{d\log\tau_{ud}}+\frac{d\log X_{mud}}{d\log\tau_{ud}}\right)\\ =& -\gamma_W + \frac{X^W_{mud}}{X^W_{ud}-X^W_{mud}}\frac{\gamma}{\beta} > -\gamma_W > -\gamma\,,\ \gamma_W \equiv \frac{\beta_W(\sigma-1)}{\beta_W-1}\end{aligned}$$

The indirect trade elasticity is smaller in absolute value than the direct trade elasticity because of two reasons. First, the elasticity of the number of suppliers with respect to sales measured in the data suggests that $\beta_W > \beta$, and thus $\gamma_W < \gamma$. Second, the indirect trade elasticity is further dampened by the fact that firms that form direct matches with each other do not trade through the wholesalers. An increase in iceberg trade cost tends to reduce the occasion of such cannibalization and therefore dampens the reduction in indirect trade flow. This effect is captured by the positive component $\frac{X^W_{mud}}{X^W_{ud}-X^W_{mud}}\frac{\gamma}{\beta}$ appearing in the indirect trade elasticity. This is true even when wholesalers and manufacturing firms do not internalize such effect when making their product adoption and ad posting decisions.

The overall trade elasticity is simply the weighted average of the direct and indirect trade elasticity:

$$\begin{aligned}\frac{d\log X_{ud}}{d\log\tau_{ud}} =& \Omega_{ud}(-\gamma) + (1-\Omega_{ud})\left(-\gamma_W + \frac{X^W_{mud}}{X^W_{ud}-X^W_{mud}}\frac{\gamma}{\beta}\right)\\ =& -\left(\gamma_W - \frac{X^W_{mud}}{X^W_{ud}-X^W_{mud}}\frac{\gamma}{\beta}\right) - \Omega_{ud}\left(\gamma-\gamma_W+\frac{X^W_{mud}}{X^W_{ud}-X^W_{mud}}\frac{\gamma}{\beta}\right)\end{aligned}$$

where $\Omega_{ud} \equiv \frac{X_{mud}}{X_{ud}}$ is the share of direct trade.

Next, I would like to illustrate formally how this model is able to reconcile the distance puzzle. In particular, I will show how a reduction in information friction leads to an ambiguous effect on the distance elasticity of total trade flow. To begin, I need to take a stance on how various multilateral resistance terms are related to distance $d_{ud} > 1$. Specifically, I assume that only iceberg trade cost increases with distance:

$$\tau_{ud} = {d_{ud}}^{\epsilon},\ \ d_{ud} > 1,\ \ \epsilon > 0$$

As derived above, the iceberg trade cost elasticity of direct trade is larger than that of indirect trade, and so when iceberg trade cost is the only multilateral resistance term that increases with distance, it implies that direct trade would also have a larger distance elasticity. While this is enough to illustrate the key argument to rationalize the distance puzzle, one can also assume other multilateral resistance terms to be increasing with distance, and that would lead to a larger difference between the distance elasticity of direct and indirect trade flows, further reinforcing the argument.

Under the above parameterization, the distance elasticity of total trade flow is given by:

$$\frac{d\log X_{ud}}{d\log d_{ud}} = \epsilon \frac{d\log X_{ud}}{d\log \tau_{ud}} = -\epsilon\left(-\gamma_W - \frac{X^W_{mud}}{X^W_{ud}-X^W_{mud}}\frac{\gamma}{\beta}\right) - \epsilon\,\Omega_{ud}\left(\gamma-\gamma_W + \frac{X^W_{mud}}{X^W_{ud}-X^W_{mud}}\frac{\gamma}{\beta}\right)$$

Suppose a reduction in information friction implies that iceberg trade cost increases with distance at a slower rate, i.e. it can be thought of as a reduction in ϵ. A frictionless world is one in which $\epsilon = 0$, such that distance does not affect the iceberg trade cost. To see the impact on the distance elasticity, I derive the following derivative:

$$\begin{aligned}\frac{d\frac{d\log X_{ud}}{d\log d_{ud}}}{d\epsilon} = &-\left(-\gamma_W - \frac{X^W_{mud}}{X^W_{ud}-X^W_{mud}}\frac{\gamma}{\beta}\right) - \Omega_{ud}\left(\gamma-\gamma_W + \frac{X^W_{mud}}{X^W_{ud}-X^W_{mud}}\frac{\gamma}{\beta}\right)\\ &-\epsilon\left(\gamma-\gamma_W + \frac{X^W_{mud}}{X^W_{ud}-X^W_{mud}}\frac{\gamma}{\beta}\right)\frac{d\Omega_{ud}}{d\epsilon}\\ &+\epsilon\,(1-\Omega_{ud})\frac{\gamma}{\beta}\frac{d}{d\epsilon}\frac{X^W_{mud}}{X^W_{ud}-X^W_{mud}}\end{aligned}$$

where the first line negative, and capture the direct effect of a reduction in information friction (decrease in ϵ) that lowers the distance elasticity (less negative). The second is the compositional effect that captures the change of the direct trade share, which is equal to:

$$\begin{aligned}\frac{d\Omega_{ud}}{d\epsilon} =&\Omega_{ud}\frac{d\log\Omega_{ud}}{d\log\tau_{ud}}\frac{d\log\tau_{ud}}{d\epsilon}\\ =&\Omega_{ud}\left(\frac{d\log X_{mud}}{d\log\tau_{ud}} - \frac{d\log X_{ud}}{d\log\tau_{ud}}\right)\log d_{ud}\\ =&\Omega_{ud}(1-\Omega_{ud})\left(\frac{d\log X_{mud}}{d\log\tau_{ud}} - \frac{d\log(X^W_{ud}-X^W_{mud})}{d\log\tau_{ud}}\right)\log d_{ud}\end{aligned}$$

i.e. a reduction in information friction tends to raise the share of direct trade, which follows directly from the fact that the distance elasticity is larger in absolute value for direct trade. Therefore, the compositional effect tends to raise the distance elasticity, rendering the overall effect of a reduction in information friction on the distance elasticity ambiguous. Notice that the last line is positive and tends to reinforce the direct effect of a reduction in information friction. However, since the share of cannibalized trade measured in the data is in general an order of magnitude smaller than the direct trade share, the last line is numerically insignificant compared to the second line.

The distance puzzle is important in that the non-declining distance elasticity seems to suggest that technological improvement has been ineffective in offsetting distance as a trade barrier. But the analysis here suggests that such an inference is problematic: one has to account for the compositional change of trade flow in order to draw correct inference of the impact of technological improvement in reducing the impact of distance. As a result of biased inference, one might obtain a vastly different conclusion on the welfare impact of technological improvement [48].

B.19 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:

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

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 is wholesalers' FOC for product adoption (34); the fifth equation comes from the free entry condition of final goods producers and their FOC; the sixth equation comes from the free entry condition of intermediate goods producers and their FOC; and the last equation comes from the free entry condition of wholesalers and their FOC. The last three equations are derived in the appendix B.8. These can be

[48] Other theories suggest that technological improvement does not necessarily reduce the impact of distance, see for example Leamer (2007) and Akerman, Leuven and Mogstad (2022). But what I suggest here is that conditional on the assumption that it does, then it is important to account for the compositional effect to infer the magnitude of such reduction of distance's impact.

rearranged in terms of hat variables:

$$
\begin{aligned}
\widehat{I_i} =& \sum_{u\in\mathcal{N}} \left[\frac{\frac{C_{2ui}}{\phi_{cui}}\, X_{mui}}{\sum\limits_{l\in\mathcal{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\mathcal{N}} \left\{ \left[\frac{X_{ui}^{W} - X_{mui}^{W}}{\sum\limits_{l\in\mathcal{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\limits_{l\in\mathcal{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\mathcal{N}} \left[\frac{\frac{C_{2id}}{\phi_{cid}}\, X_{mid}}{\sum\limits_{l\in\mathcal{N}} \left(\frac{C_{2li}}{\phi_{cli}}\, X_{mli} + X_{li}^{W} \right)}\, \frac{\widehat{C_{2id}}}{\widehat{\phi_{cid}}}\, \widehat{X_{mid}} \right] + \sum_{d\in\mathcal{N}} \left\{ \left[\frac{X_{id}^{W} - X_{mid}^{W}}{\sum\limits_{l\in\mathcal{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\limits_{l\in\mathcal{N}} \left(\frac{C_{2li}}{\phi_{cli}}\, X_{mli} + X_{li}^{W} \right)} \right] \widehat{X_{id}^{W}} \right\} \\
&+ \sum_{u\in\mathcal{N}} \left\{ \frac{1}{N_{Wi}\,\sigma} \left[\frac{X_{ui}^{W} - X_{mui}^{W}}{\sum\limits_{l\in\mathcal{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\limits_{l\in\mathcal{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\mathcal{N}} \left[\frac{1}{N_{Wi}\,\sigma}\, \frac{\left(1 - \frac{C_{2ui}}{\phi_{cui}}\right) X_{mui}}{\sum\limits_{l\in\mathcal{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\mathcal{N}} \left\{ \frac{1}{N_{Wd}\,\sigma} \left[\frac{X_{id}^{W} - X_{mid}^{W}}{\sum\limits_{l\in\mathcal{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\limits_{l\in\mathcal{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\mathcal{N}} \left[\frac{1}{N_{Wd}\,\sigma}\, \frac{\left(1 - \frac{C_{2id}}{\phi_{cid}}\right) X_{mid}}{\sum\limits_{l\in\mathcal{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} \Bigg\{ \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} \\
&- \frac{\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)}{\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]} \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}}{\widetilde{N_{Wi}}^{2}}}\, \widehat{N_{Wi}}^{-1} - \frac{\frac{N_{Wi}}{\widetilde{N_{Wi}}^{2}}}{N_{Wi}^{-1} - \frac{N_{Wi}}{\widetilde{N_{Wi}}^{2}}}\, \frac{\widehat{N_{Wi}}}{\widehat{\widetilde{N_{Wi}}^{2}}} \right) \Bigg\}^{-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{L_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{\overline{\psi_i^{R}}} \right) \frac{\widehat{L_i}}{\widehat{F_{Ii}}} \\
\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}}
\end{aligned}
$$

where

$$
\begin{aligned}
\widehat{\psi_i} =& \left\{ \sum_{u\in\mathcal{N}} \left[\frac{\frac{C_{2ui}}{\phi_{cui}} X_{mui}}{\sum\limits_{l\in\mathcal{N}} \left(\frac{C_{2li}}{\phi_{cli}} X_{mli} \right)} \, \widehat{\frac{C_{2ui}}{\phi_{cui}}} \widehat{X_{mui}} \right] \right\} \\
& \left\{ \sum_{u\in\mathcal{N}} \left[\frac{\frac{C_{2ui}}{\phi_{cui}} X_{mui}}{\sum\limits_{l\in\mathcal{N}} \left(\frac{C_{2li}}{\phi_{cli}} X_{mli} + X^W_{li} \right)} \, \widehat{\frac{C_{2ui}}{\phi_{cui}}} \widehat{X_{mui}} \right] + \sum_{u\in\mathcal{N}} \left[\left(\frac{X^W_{ui} - X^W_{mui}}{\sum\limits_{l\in\mathcal{N}} \left(\frac{C_{2li}}{\phi_{cli}} X_{mli} + X^W_{li} \right)} + \frac{\left(1 - \frac{C_{2ui}}{\phi_{cui}}\right) X_{mui}}{\sum\limits_{l\in\mathcal{N}} \left(\frac{C_{2li}}{\phi_{cli}} X_{mli} + X^W_{li} \right)} \right) \widehat{X^W_{ui}} \right] \right\}^{-1} \\
\widehat{\widehat{\psi^R_i}} =& \left\{ \sum_{d\in\mathcal{N}} \left[\frac{C_{1id} X_{mid}}{\sum\limits_{l\in\mathcal{N}} (C_{1il} X_{mil})} \, \widehat{C_{1id}} \widehat{X_{mid}} \right] \right\} \\
& \left\{ \sum_{d\in\mathcal{N}} \left[\frac{C_{1id} X_{mid}}{\sum\limits_{l\in\mathcal{N}} (C_{1il} X_{mil} + X_{Wil})} \, \widehat{C_{1id}} \widehat{X_{mid}} \right] + \sum_{d\in\mathcal{N}} \left[\left(\frac{X_{Wid} - X_{Wmid}}{\sum\limits_{l\in\mathcal{N}} (C_{1il} X_{mil} + X_{Wil})} + \frac{(1 - C_{1id}) X_{mid}}{\sum\limits_{l\in\mathcal{N}} (C_{1il} X_{mil} + X_{Wil})} \right) \widehat{X_{Wid}} \right] \right\}^{-1} \\
\widehat{\xi_i} =& (\mu^W_i \, \widehat{\mu^W_i})^{-1} \, \xi_i^{-1} + \left[\frac{1}{N_{Wi}\, \sigma} \, \frac{\sum\limits_{u\in\mathcal{N}} X_{mui}}{\sum\limits_{u\in\mathcal{N}} (X_{mui} + X^W_{ui} - X^W_{mui})} \, \xi_i^{-1} \, \frac{1}{\widehat{N_{Wi}}} \sum_{u\in\mathcal{N}} \left(\frac{X_{mui}}{\sum\limits_{l\in\mathcal{N}} X_{mli}} \, \widehat{X_{mui}} \right) \right. \\
& \left. \left\{ \sum_{u\in\mathcal{N}} \left[\frac{\frac{C_{2ui}}{\phi_{cui}} X_{mui}}{\sum\limits_{l\in\mathcal{N}} \left(\frac{C_{2li}}{\phi_{cli}} X_{mli} + X^W_{li} \right)} \, \widehat{\frac{C_{2ui}}{\phi_{cui}}} \widehat{X_{mui}} \right] + \sum_{u\in\mathcal{N}} \left[\left(\frac{X^W_{ui} - X^W_{mui}}{\sum\limits_{l\in\mathcal{N}} \left(\frac{C_{2li}}{\phi_{cli}} X_{mli} + X^W_{li} \right)} + \frac{\left(1 - \frac{C_{2ui}}{\phi_{cui}}\right) X_{mui}}{\sum\limits_{l\in\mathcal{N}} \left(\frac{C_{2li}}{\phi_{cli}} X_{mli} + X^W_{li} \right)} \right) \widehat{X^W_{ui}} \right] \right\}^{-1} \right] \\
\widehat{\Omega_i} =& \left\{ \sum_{u\in\mathcal{N}} \left[\frac{X_{mui}}{\sum\limits_{l\in\mathcal{N}} (X_{mli})} \, \widehat{X_{mui}} \right] \right\} \\
& \left\{ \sum_{u\in\mathcal{N}} \left[\frac{\frac{C_{2ui}}{\phi_{cui}} X_{mui}}{\sum\limits_{l\in\mathcal{N}} \left(\frac{C_{2li}}{\phi_{cli}} X_{mli} + X^W_{li} \right)} \, \widehat{\frac{C_{2ui}}{\phi_{cui}}} \widehat{X_{mui}} \right] + \sum_{u\in\mathcal{N}} \left[\left(\frac{X^W_{ui} - X^W_{mui}}{\sum\limits_{l\in\mathcal{N}} \left(\frac{C_{2li}}{\phi_{cli}} X_{mli} + X^W_{li} \right)} + \frac{\left(1 - \frac{C_{2ui}}{\phi_{cui}}\right) X_{mui}}{\sum\limits_{l\in\mathcal{N}} \left(\frac{C_{2li}}{\phi_{cli}} X_{mli} + X^W_{li} \right)} \right) \widehat{X^W_{ui}} \right] \right\}^{-1} \\
\widehat{X^W_{ud} - X^W_{mud}} =& \left[1 + \frac{\left(\mu^W_d\right)^{1-\sigma} S_{ud}\, \phi^{-1}_{cud}\, X_{mud}}{X^W_{ud} - X^W_{mud}} \right] \widehat{X^W_{ud}} - \frac{\left(\mu^W_d\right)^{1-\sigma} S_{ud}\, \phi^{-1}_{cud}\, X_{mud}}{X^W_{ud} - X^W_{mud}} \left(\widehat{\mu^W_d}\right)^{1-\sigma} \widehat{S_{ud}}\, \widehat{\phi_{cud}}^{-1}\, \widehat{X_{mud}} \\
\widehat{X_{mud}} =& \widehat{K_{mud}}\, \widehat{\theta^m_{ud}}^{\frac{\beta}{\beta-1}}\, \widehat{N_{Fd}}\, \widehat{w_d}^{\frac{1}{1-\beta}}\, I_d^{\frac{\beta}{\beta-1}}\, \widehat{P^H_d}^{\gamma}\, \widehat{w_u}^{-\gamma}\, \widehat{C_{2ud}}^{\frac{1}{\beta-1}} \\
\widehat{X^W_{ud}} =& \left(\widehat{\mu^W_d}\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^H_d}\right)^{\sigma-1} \\
\widehat{\theta^m_{ud}} =& \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^H_d}^{\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^{-1}_{1ud} - C^{-1}_{1ud} \left(\mu^W_d\right)^{-\sigma} S_{ud}\, \phi^{-1}_{cud} \left(\widehat{\mu^W_d}\right)^{-\sigma} \widehat{S_{ud}}\, \widehat{\phi_{cud}}^{-1} \\
\widehat{C_{2ud}} =& C^{-1}_{2ud}\, \phi_{cud}\, \widehat{\phi_{cud}} - C^{-1}_{2ud} \left(\mu^W_d\right)^{1-\sigma} S_{ud} \left(\widehat{\mu^W_d}\right)^{1-\sigma} \widehat{S_{ud}} \\
\widehat{\mu^W_i} =& \frac{\widehat{N_{Wi}}\, (N_{Wi}\, \sigma - 1)}{\widehat{N_{Wi}}\, N_{Wi}\, \sigma - 1} \\
\widetilde{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{\widetilde{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}}
\end{aligned}
$$

C Sensitivity Analysis for Counterfactuals

C.1 Decomposing the Welfare Cost of Inefficiencies

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

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%.

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 0.005 log points, 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%, 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%. In fact, the change in the markup level is so small that the *misallocation* markup dispersion, evaluated at the *efficient* direct trade share post-shock, is likely to have decreased. This explains why misallocation actually improves by 0.005 log points 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 misallocation markup dispersion.

High S parameterization. The next parameterization halves the product adoption cost shifter f_W, raising the equilibrium rate of product adoption 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

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

	$\widehat{\Omega}$	$\widehat{\mu^W}$	$\widehat{\sigma^2(\mu)}$	$\widehat{\tilde{M}}$	$\widehat{S}$	$\widehat{A}$	$\widehat{L_A}$	$\widehat{L_S}$	$\widehat{L_E}$	$\widehat{L_P}$	$\widehat{Welfare}$
Baseline (Pre Shock)											
Efficient	0.4600	1.1493	0.0000	0.0093	1.4221	0.9231	0.2874	2.1040	0.8132	1.1916	1.1000
Wholesale subsidy	0.3700	1.0000	0.0000	0.4872	0.9653	1.0161	0.3269	1.5104	1.1242	1.0074	1.0237
Decentralized (level)	0.4800	1.1299	0.0037	0.0005	0.2630	0.3852	0.0799	0.0129	0.3465	0.5608	0.2160
Baseline (Post Shock)											
Efficient	0.8800	1.1493	0.0000	0.0159	1.4608	0.9790	0.6123	1.5349	0.7679	1.1864	1.1615
Wholesale subsidy	0.6600	1.0000	0.0000	0.7299	1.0909	1.0650	0.6023	1.2584	1.0645	1.0109	1.0765
Decentralized	1.6000	1.1493	2.9731	1.2950	1.1700	1.0122	1.7748	0.8017	0.7958	1.0204	1.0328
Low Ω (Pre Shock)											
Efficient	0.2500	1.1493	0.0000	0.0053	1.5374	0.9010	0.1523	1.7307	0.7505	1.2140	1.0939
Wholesale subsidy	0.2400	1.0000	0.0000	0.3284	0.9576	0.9993	0.2082	1.1743	1.0535	1.0179	1.0172
Decentralized (level)	0.2500	1.1299	0.0028	0.0002	0.2642	0.3891	0.0427	0.0189	0.3850	0.5534	0.2153
Low Ω (Post Shock)											
Efficient	0.6400	1.1493	0.0000	0.0113	1.6526	0.9419	0.4269	1.5395	0.7341	1.2107	1.1404
Wholesale subsidy	0.5200	1.0000	0.0000	0.5841	1.0812	1.0396	0.4662	1.0870	1.0266	1.0197	1.0601
Decentralized	1.5000	1.0000	1.2522	1.3142	1.1282	1.0300	1.5684	0.8062	0.9349	1.0081	1.0384
Low μ^W (Pre Shock)											
Efficient	0.6000	1.0649	0.0000	0.0114	1.5854	0.9349	0.3857	3.8512	0.8061	1.1651	1.0892
Wholesale subsidy	0.6600	1.0000	0.0000	0.7509	0.9843	1.0040	0.6250	1.2770	1.0708	1.0052	1.0092
Decentralized (level)	0.4800	1.0610	0.0009	0.0005	0.2374	0.3889	0.0771	0.0064	0.3408	0.5757	0.2239
Low μ^W (Post Shock)											
Efficient	1.0000	1.0649	0.0000	0.0183	1.6120	0.9960	0.7600	2.6234	0.7532	1.1600	1.1553
Wholesale subsidy	0.9900	1.0000	0.0000	0.9745	1.1092	1.0597	0.9690	0.9828	0.9953	1.0071	1.0672
Decentralized	1.3000	1.0207	1.6658	1.1809	1.1532	1.0573	1.3904	0.9220	0.9044	1.0052	1.0628
High S (Pre Shock)											
Efficient	0.4000	1.1493	0.0000	0.0082	1.4380	0.9226	0.2424	2.1653	0.8166	1.1922	1.0999
Wholesale subsidy	0.3500	1.0000	0.0000	0.4601	0.9664	1.0171	0.3027	1.5218	1.1242	1.0073	1.0246
Decentralized (level)	0.4800	1.1299	0.0037	0.0005	0.2960	0.3984	0.0777	0.0129	0.3487	0.5607	0.2234
High S (Post Shock)											
Efficient	0.8500	1.1493	0.0000	0.0154	1.4642	0.9779	0.5816	1.5668	0.7731	1.1860	1.1598
Wholesale subsidy	0.6400	1.0000	0.0000	0.7129	1.0914	1.0651	0.5828	1.2683	1.0656	1.0108	1.0767
Decentralized	1.6000	1.1493	2.9029	1.3132	1.1654	1.0097	1.8209	0.7699	0.7913	1.0213	1.0312
Low λ_V, λ_M (Pre Shock)											
Efficient	0.9300	1.1493	0.0000	0.0087	1.2771	0.9424	0.5716	1.4735	0.7697	1.1920	1.1233
Wholesale subsidy	0.4100	1.0000	0.0000	0.5431	0.9651	1.0171	0.3622	1.4597	1.1141	1.0083	1.0255
Decentralized (level)	0.4800	1.1299	0.0037	0.0005	0.2631	0.3852	0.0788	0.0130	0.3476	0.5606	0.2159
Low λ_V, λ_M (Post Shock)											
Efficient	1.3000	1.1493	0.0000	0.0113	1.3145	1.0070	0.8972	0.9575	0.7129	1.1935	1.2018
Wholesale subsidy	0.6800	1.0000	0.0000	0.7582	1.0904	1.0655	0.6227	1.2300	1.0585	1.0114	1.0777
Decentralized	1.6000	1.1493	3.0803	1.2791	1.1758	1.0096	1.7784	0.8301	0.7982	1.0196	1.0294
Low λ_W (Pre Shock)											
Efficient	0.4500	1.1493	0.0000	0.0095	1.3806	0.9345	0.2840	1.9335	0.8304	1.1853	1.1077
Wholesale subsidy	0.3700	1.0000	0.0000	0.4833	0.9680	1.0168	0.3232	1.5129	1.1248	1.0074	1.0244
Decentralized (level)	0.4800	1.1299	0.0037	0.0005	0.2631	0.3853	0.0798	0.0129	0.3466	0.5608	0.2160
Low λ_W (Post Shock)											
Efficient	0.8600	1.1493	0.0000	0.0160	1.4378	0.9872	0.6001	1.4277	0.7818	1.1820	1.1668
Wholesale subsidy	0.6700	1.0000	0.0000	0.7379	1.0834	1.0633	0.6129	1.2483	1.0621	1.0110	1.0750
Decentralized	1.6000	1.1493	3.1245	1.2885	1.2294	1.0193	1.7412	0.8587	0.8042	1.0188	1.0384

Note: Within each set of parameterization, this table reports the *level* of the pre-shock decentralized equilibrium, as well as *proportional 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%.

more with the increase in the wholesale markup post-shock when S is higher, by 0.019 log points compared with 0.018 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 0.021 log points, compared with 0.018 log points 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 wholesale product adoption. As discussed earlier, the overall inefficiency in adoption 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 product adoption, which then strengthens the markup's dampening effect on adoption inefficiency by reducing the share of indirect trade. The result is a more muted increase in the welfare gain from removing the markup: 0.011 log points post-shock.

D Tables and Figures

D.1 Tables and Figures for Motivational Facts

Fact 6: Manufacturing firms do not face additional markdowns when selling to wholesalers

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

	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

Robust standard errors in parentheses

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. Robust standard errors are reported in parentheses.

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 manufacturing 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. 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. Standard errors are reported in parentheses.

D.2 Tables and Figures for Fiber Internet Expansion Empirics

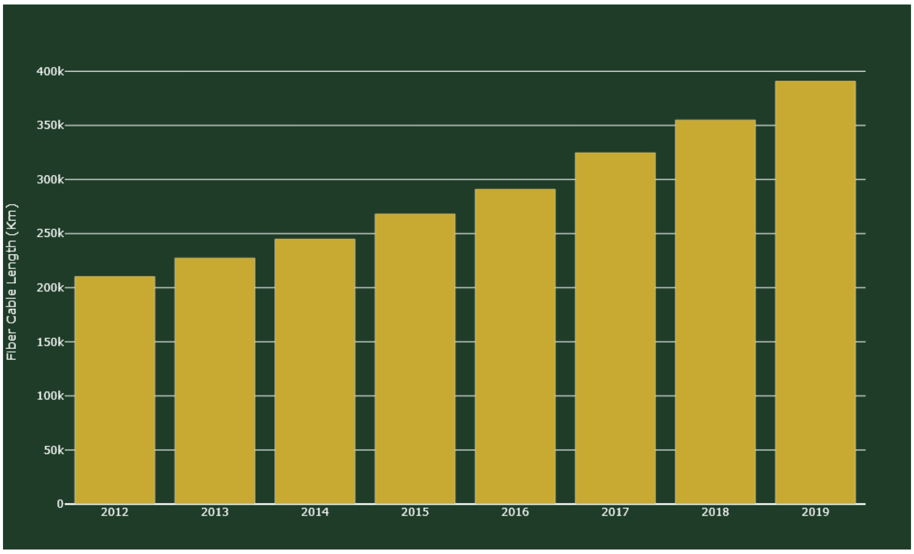


Figure D.1: Evolution of the total length of fiber cable deployed in Turkey between 2012-2019

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).

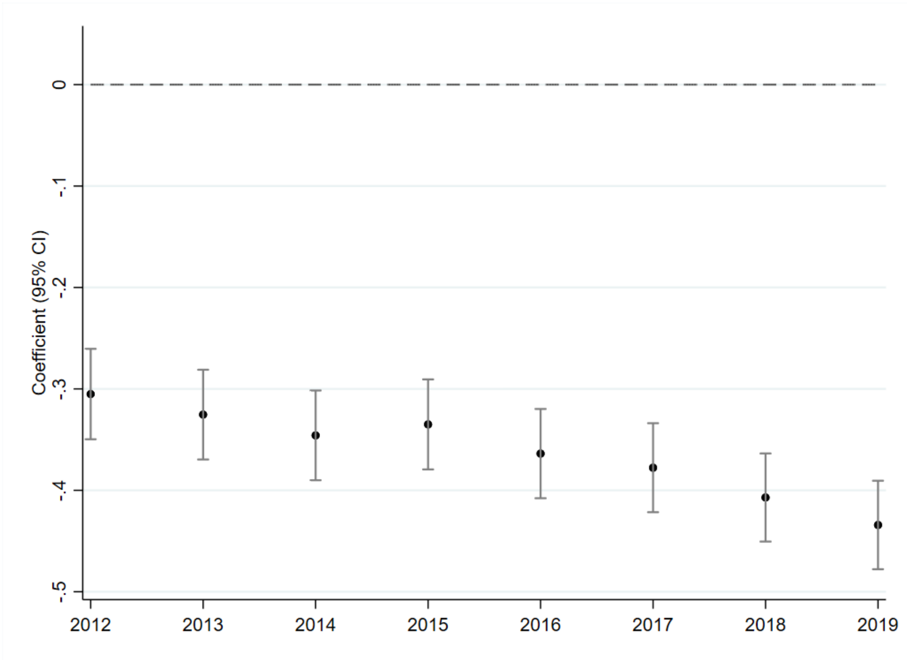


Figure D.2: First-stage coefficient estimates of distance to oil pipeline interacted with year dummies

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

	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. Standard errors clustered at the province-pair level are reported in parentheses.

Table D.4 shows that the pattern of disintermediation also holds at the firm level. Firms operating in provinces with faster fiber roll-out 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

	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.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. 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)
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. Robust standard errors are reported in parentheses.

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 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**	
Dep. Variable: Markup	(1)	(2)	(3)	(4)
Std Fiber Intensity	0.0126	0.0985	0.0081	0.1882
	(0.0042)	(0.0169)	(0.0063)	(0.0680)
Intensity x Indirect Sales Share	-0.0013	0.0009	-0.0031	-0.0009
	(0.0013)	(0.0013)	(0.0026)	(0.0024)
Intensity x 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.

D.4 Tables and Figures for Counterfactuals

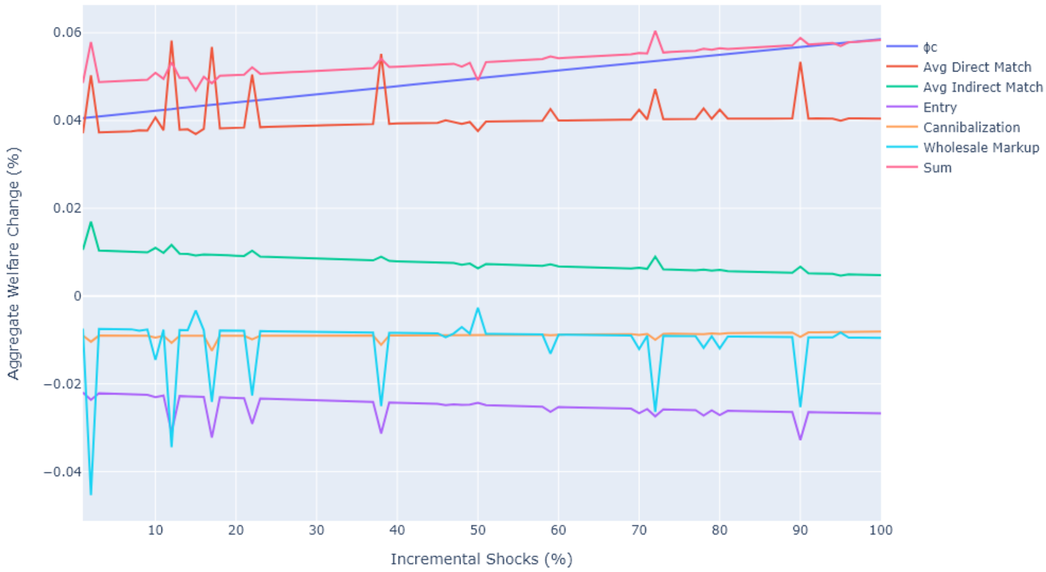


Figure D.4: Stepwise First Order Effects of Internet Shock

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.

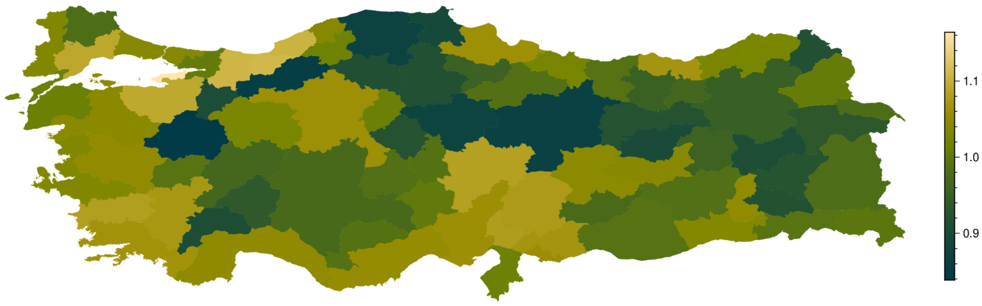


Figure D.5: Proportional Change in Welfare

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

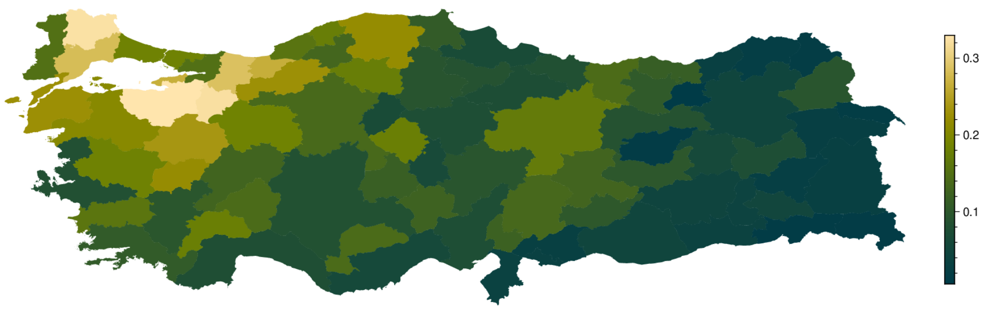


Figure D.6: Ratio of direct purchases from Istanbul to province-level GDP

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.

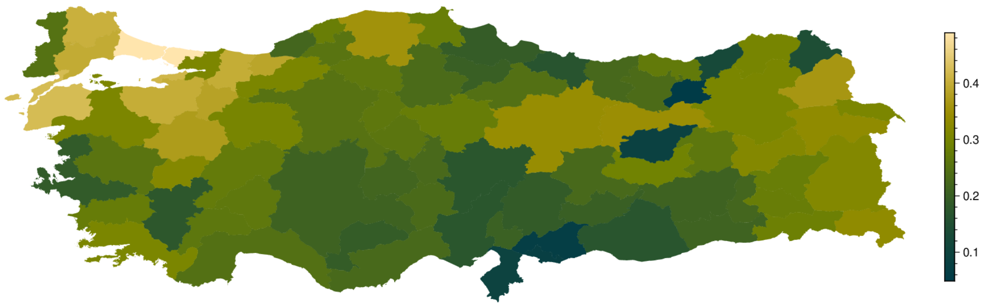


Figure D.7: Ratio of total purchases from Istanbul to province-level GDP

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.

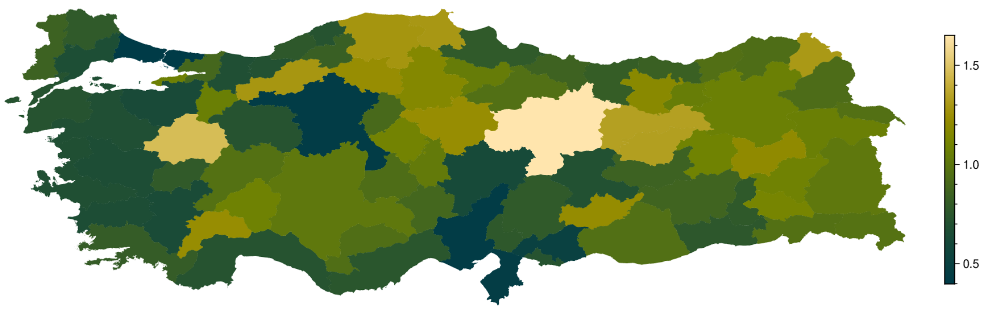


Figure D.8: Proportional Change in Indirect Trade Share

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.

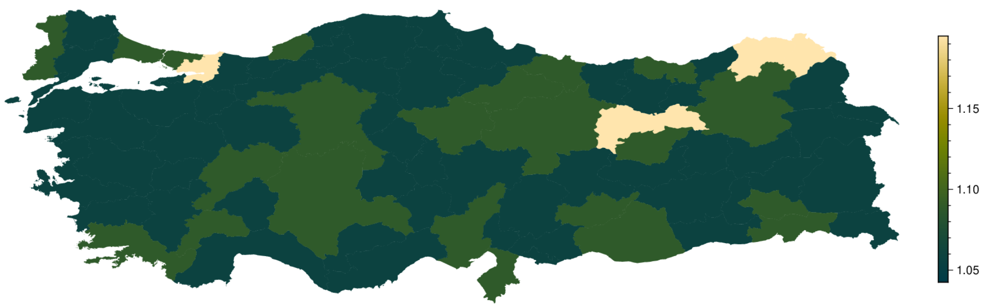


Figure D.9: Wholesale Markup (Pre-Shock)

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

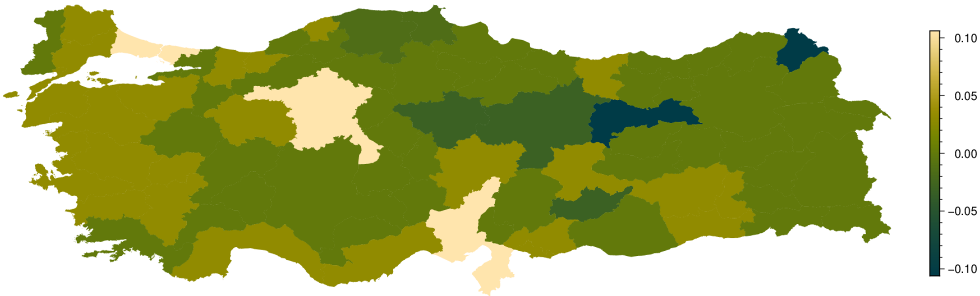


Figure D.10: Proportional Change in Wholesale Markup

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

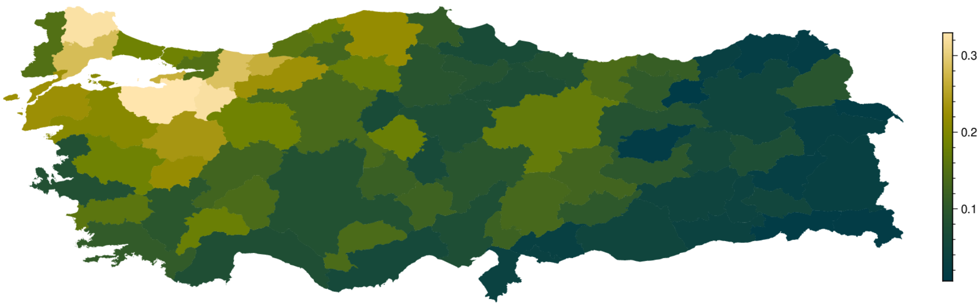


Figure D.11: Ratio of Direct Purchases from Istanbul to Province GDP

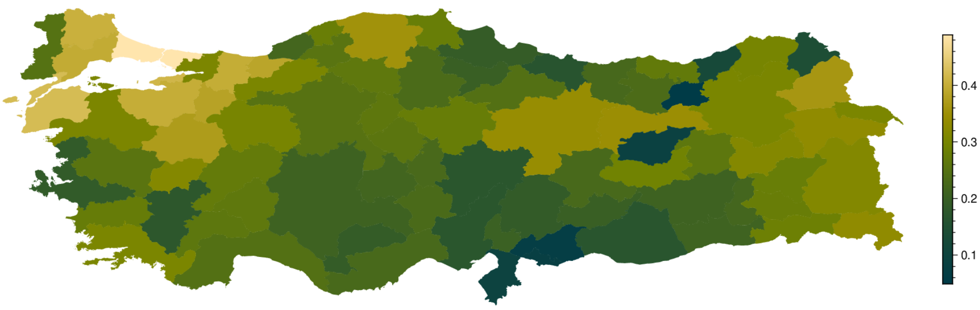


Figure D.12: Ratio of Total Purchases from Istanbul to Province GDP