Cutting Out the Middleman: The Structure of Chains of Intermediation*

Matthew Grant[†] and Meredith Startz[‡]

November 2021 [Please click for most recent version]

Abstract

Distribution of goods often involves chains of multiple intermediaries engaged in sequential buying and reselling. Common wisdom holds that connecting producers and consumers more directly will benefit them – we show that this can be, but is not necessarily the case, even when middlemen hold market power. Multi-intermediary chains arise when there are internal economies of scale in trade costs, suggesting consumers in developing countries are more likely to be served via long chains. Equilibrium chain structures are not generally efficient. There is a fundamental welfare trade-off between costs and entry in destination markets, as shorter chains involve lower marginal cost, which benefits consumers, but also lower entry, which can lead to lower product availability, fewer sellers, and higher markups. The proposed mechanism is simple, but can account for empirical patterns in wholesale firm size, prices and markups that we document using original survey data on imported consumer goods in Nigeria. We estimate a structural version of the model for distribution of Chinese-made apparel in Nigeria, and describe endogenous restructuring of chains and the resulting impacts on consumer welfare in response to counterfactual changes in regulation, e-commerce technologies, and transport infrastructure. We find that cutting out middlemen has heterogenous impacts across locations, but often harms more remote consumers.

*We thank Treb Allen, David Atkin, Kyle Bagwell, Andy Bernard, Lorenzo Casaburi, Arun Chandrasekhar, Davin Chor, Swati Dhingra, Cecile Gaubert, Sam Kortum, Nina Pavcnik, Andres Rodriguez-Clare, Ina Simonovksa, and Bob Staiger for their helpful comments. This work would not be possible without the excellent research assistance of Hakeem Bishi, Yemi Ogundare, Samuel Adesanya, Nicole Wilson, Sheah Deilami-Nugent, and the Lagos Trader Project enumerator team.

†Dartmouth College. Email: matthew.w.grant@dartmouth.edu

‡Dartmouth College. Email: meredith.l.startz@dartmouth.edu

1 Introduction

How do goods made in one place reach consumers in another? Models of trade typically abstract from the details, assuming that a producer sells directly to consumers. The reality is that goods may pass through the hands of multiple intermediaries. A shopkeeper selling mobile phones in a small town in Nigeria, for instance, is unlikely to source them directly from the manufacturer in southern China. Instead, she might buy from a wholesaler in a regional market town, who relies on an importer, and so on. At each step toward consumers, costs are incurred and markups may be charged. Understanding intermediation chains is therefore key to understanding the prices and product availability faced by consumers in different locations, and their potential to gain from globalization and trade.

Wholesale and retail firms account for a large share of economic activity all over the world, and there is a great deal of interest in the extent to which they hold market power and mediate price passthrough in both rich and poor countries. Political discourse has often taken a dim view of the trading sector, seeing it as merely driving a price wedge between producers and consumers. This is reflected in policies that restrict the role of middlemen, from limits on where agents can buy and sell (such as India's regulated agricultural marketplaces) to outright prohibition (such as Bangladesh's 2011 ban on delivery order traders). Recently, the idea that changes in trade costs might connect producers and consumers more directly even without heavy-handed policy intervention has gained traction – effectively, that middlemen will be cut out naturally as a result of falling transport costs, improving communication, and new matching technologies such as e-commerce platforms or programs that link farmers with buyers.

Are policies and technologies that cut out middlemen likely to help or harm consumers? Answering this question requires an understanding of why multiple layers of intermediation might arise in the first place. We show that internal economies of scale in the trade costs incurred by an intermediary who buys and resells a good — whether these costs arise from transport, search, or any of the many other functions that intermediaries serve — offer a general microfoundation for the existence of multi-intermediary chains. This insight allows us to consider how chain structure will respond endogenously to changes in policies and costs, and how this in turn will influence consumer welfare. Shorter chains *can* increase welfare but will not necessarily do so, regardless of whether they are induced by policy mandates or technology improvements. Instead, shorter chains imply both lower marginal costs and lower entry of intermediaries in downstream markets, and there is a fundamental welfare tradeoff between these two forces.

We begin by introducing data from an original survey of wholesale and retail traders in Lagos, Nigeria that highlights the ubiquity of multi-intermediary chains and their relevance for consumer prices. These traders source manufactured consumer goods like apparel and electronics from all over the world, and sell to customers throughout Nigeria. A novel feature of the data is that it allows us to observe multiple links along the distribution chain. Over two-thirds of respondents' international suppliers are upstream wholesalers rather than manufacturers, and the majority of their sales are to downstream traders rather than final consumers. It is well-known that agricultural collection chains in developing countries often include multiple levels of intermediation. We document that this is also the case for the distribution of manufactured goods, and that firm size, sourcing costs, and markups vary systematically with chain position.

We next show that a simple mechanism gives rise to such multi-intermediary chains: internal economies of scale in sourcing from a particular market. If a good can be resold, and individual buyers face fixed costs to access different sellers or markets, then it is not necessarily cost-minimizing to source directly from the producer. It may be preferable to purchase the good in a resale market, depending on the quantity being purchased and the trade-off between fixed and variable costs involved. Intermediaries serve a range of functions, and fixed costs (or internal economies of scale more generally) are a feature many of them. Our framework describes how any task or trade cost that has this feature – regardless of the underlying source of the costs – leads to multi-intermediary chains. Examples are everywhere in wholesale and retail trade, ranging from time (e.g. spent searching for products), to transport (e.g. operating a truck or filling a shipping container), to financing (e.g. fees for wire transfers and letters of credit), to regulatory costs (e.g. licenses and port inspection fees).

Intuitively, this is why a rural shopkeeper in Nigeria may prefer to source goods from a local market town rather than directly from China. For a small firm, saving on port fees and the cost of a trip to China may be worth paying a higher unit price to a local wholesaler. Iterating this logic can lead to a whole chain of intermediaries who source from other intermediaries: the rural shopkeeper maximizes profits by sourcing from the market town wholesaler, who optimally chooses to buy from an importer in Lagos, and so on. This intuition holds under general forms of demand and competition. A distribution chain is, essentially, an endogenous set of points at which downstream demand is aggregated up in response to economies of scale in sourcing. The structure of the chain, including the number of links and the degree of competition at each link, will depend on the specifics of demand and trade costs. We show that places with features characteristic of developing country markets, such as small equilibrium firm size and high barriers to accessing production locations, are more likely to be served via long chains. Our framework may therefore be particularly relevant to policy in developing countries, and helpful in understanding cross-country differences in distribution and collection structures.

This framework is also useful for understanding the welfare implications of intermediation chains. The availability of multiple resale markets where the same good can be sourced at different trade costs can be understood as analagous to a choice over production technologies. This technology choice constitutes a new dimension of a classic question about optimal levels of entry when firms have market power. We consider the problem of a planner who cares about welfare in a destination market where there are consumers who demand a good produced elsewhere. The planner takes entry decisions and imperfectly competitive behavior of sellers in wholesale and retail markets as given, but manipulates the menu of sourcing options to achieve second-best outcomes. The planner can, for instance, prohibit sourcing from particular markets, or set policy that affects trade costs. Policies that shorten chains implicitly move the destination market toward sourcing strategies with lower variable and higher fixed costs. Shorter chains benefit consumers by reducing marginal costs, but harm them by reducing entry. Because of these second best considerations, even pure cost reductions – for instance, the removal of bureaucratic barriers to accessing certain markets, or introduction of platforms that reduce communication costs – can shift the market toward sourcing strategies that lower welfare. Broadly, cutting out middlemen can either help or harm consumers.[1]

To pin down these theoretically ambiguous welfare effects, we build a quantifiable model that relates the equilibrium chain structure serving consumers in many locations to fundamentals of geography and demand. The model considers distribution of a single good produced in a single origin market. Heterogenous wholesale and retail traders can enter in each location to serve local consumers, and also to resell onward to downstream traders. Traders at each link charge markups, and so chains with multiple intermediaries feature double marginalization. We provide a tractable approach to modeling wholesale pricing: elasticities of demand pass back along the chain in a way that fully incorporates the endogenous decisions of downstream sellers but also maintains the form of CES markups at each step.

We quantify the model for distribution of Chinese-made apparel throughout Nigeria, combining our original survey data with national administrative data on prices and road travel times. Our estimates are able to match a nuanced set of observed empirical patterns. First, sellers source the same or similar goods from different places. Second, Nigerian consumers are served by long chains. Third, firm size and costs are related to chain position: sellers that are further downstream from producers are smaller and face higher unit costs, while those that are further upstream from consumers are larger and face lower unit costs even conditional on their downstreamness. Finally, more upstream sellers face more elastic demand and charge lower markups.

Using these estimates, we consider several counterfactual policy experiments. We first consider an extreme scenario – what would happen if all wholesaling were prohibited and retailers had to source directly from China? At baseline, most retailers in Nigeria source from a local wholesaler, and so this implies an almost 20-fold increase in their equilibrium fixed cost of sourcing. As a consequence, the number of retail outlets carrying Chinese apparel plummets, consumer spending shifts almost entirely to other goods, and welfare throughout Nigeria falls. Second, we consider the role of investments that target fixed versus variable costs of trade – how much would a business-to-business e-commerce platform need to reduce fixed costs of sourcing from China in order to benefit consumers in Nigeria, and how does this compare to a road improvement program that reduces domestic variable trade costs? Both types of cost reductions actually decrease welfare in some parts of the country, where the extensive margin shift to higher fixed cost sourcing strategies outweighs the gain from lower variable costs. Lagos always benefits from improved international e-commerce, because it is already sourcing from China and so sees gains on the inframargin, but fixed costs of sourcing from China have to fall by at least twenty percent before other parts of the country start to see a net benefit.

[1] Manipulation of resale sourcing technologies is never first-best. Our conclusion is not that policymakers *ought* to regulate intermediation in this way – that depends on the set of policy instruments available and the underlying sources of trade costs in a specific setting. Rather, we offer a way of thinking about how distribution structures will change in response to policies that are undertaken in the real world, as well as actual changes in trade costs and technologies.

We join a small set of papers focused on how wholesale and retail distribution influence consumer welfare in developing countries (Atkin and Donaldson (2015); Lagakos (2016); Atkin, Faber and Gonzalez-Navarro (2018); Emran et al. (2020)). This relates more broadly to a literature on traders in agricultural value chains, and their role in determining price gaps between farmers and consumers (Fafchamps and Hill (2008); Dillon and Dambro (2017); Krishna and Sheveleva (2017); Bergquist and Dinerstein (2019); Casaburi and Reed (2019); Chatterjee (2019); Barrett et al. (2020); Dhingra and Tenreyro (2020)). Although our empirical application is to manufactured goods, our framework is equally applicable to collection chains, such as those in agriculture. These literatures have generally studied prices and market power of intermediaries taking the structure of the chain as given. We offer a way of microfounding chains to understand how the structure itself might respond endogenously to policy or technology changes.

Customs microdata has revealed that wholesaling accounts for a substantial share of international trade, and that selling via an intermediary is more common in smaller transactions. This has motivated models of exporting via a wholesaler as a way for manufacturers to reach consumers at lower fixed cost (Blum, Claro and Horstmann (2009); Bernard et al. (2010); Ahn, Khandelwal and Wei (2011); Crozet, Lalanne and Poncet (2013); Akerman (2018); Ganapati (2020)). More or less explicit in this literature is the idea that wholesalers serve an aggregation function, pooling exports from multiple firms or products to cover fixed costs. Aggregation in response to economies of scale is also at the heart of our view of intermediaries. We extend this logic to yield an endogenous structure of aggregation points and show that chains with more than one intermediary may naturally arise, and consider how this matters for consumer welfare.

Conceptually, we build on a classic literature exploring the efficiency of free entry equilibria in the presence of fixed costs (Spence (1976); Dixit and Stiglitz (1977); Mankiw and Whinston (1986)), and work considering these issues in new trade models (Dhingra and Morrow (2019)). In order to capture the nature of intermediation and resale, we extend this problem to consider the choice over a menu of resale "technologies" (i.e. source markets) featuring different combinations of fixed and variable costs. This introduces another margin along which social and private incentives can diverge, and there is room to achieve second-best outcomes by using policy to constrain these sourcing options. We show that the choice over alternative technologies along the chain is related to the efficiency of free entry within a market conditional on a given technology.

The existence of intermediaries in trade has been attributed to a variety of specific mechanisms, often focused on information problems (Rubinstein and Wolinsky (1987); Biglaiser (1993); Antras and Costinot (2011)) or the provision of financing or transport services. Economies of scale are a feature of a whole class of these intermediation tasks, regardless of the exact mechanism – the ways intermediaries actually solve these problems and the types of costs they incur in doing so often feature internal economies of scale. Our framework is meant to describe how any intermediation task that has this feature leads to multi-intermediary chains and maps into endogenous distribution structures. This can encompass information, regulatory, financial, bureaucratic, and transportation costs of trade. Earlier work shows that travel to overcome search and contracting problems is an important source of fixed costs of trade in this particular Nigerian context (Startz (2021)), but emprical evidence on the substance of trade costs and the existence of economies of scale comes from a wide range of work, including in trucking (Teravaninthorn and Raballand (2009)), container shipping (Cosar and Demir (2018)), trade finance (Niepmann and Schmidt-Eisenlohr (2017)), and licensing and non-tariff barriers.

Our work is complementary to emerging literatures on endogenous production networks, global value chains, and transportation routing. We share the language of chains and some technical considerations, but examine a distinct empirical phenomenon driven by a different underlying mechanism. Intermediation involves buying and reselling without transformation – there is no fixed number of production tasks that need to be completed, and links describe transaction points, rather than routing through physical locations. We attribute the existence of chains with multiple links to internal economies of scale, rather than task-specific productivities that vary across firms or locations (as in the production or value chains literatures) or external economies of scale (as in the transportation literature).

We offer a framework for understanding the endogenous determination of distribution and collection structures, to enable both positive and normative predictions about changes in these structures. We begin in Section 2 by describing wholesale and retail trade in consumer goods in Nigeria, and introducing data that highlights the importance of multi-intermediary chains and empirical patterns associated with chain position. In Section 3, we introduce the main theoretical insights about economies of scale and chain formation in a simple geography. In Section 4, we show how this theoretical insight can be implemented in a quantitative equilbrium model in a more realistic geography, and in Section 5 we estimate that model in the Nigerian setting and consider policy counterfactuals.

2 Distribution chains in Nigeria

2.1 Wholesale and retail trade in Nigeria

The trading sector in Nigeria is large and economically important. As of 2013, wholesale and retail trade accounted for 17% of GDP and 25% of total employment, and was the largest contributor to recent GDP growth (Nigerian National Bureau of Statistics). The sector is highly decentralized, and composed mostly of small-scale informal enterprises—some estimates suggest that as much as 98% of spending on consumer packaged goods in Nigeria goes through small traditional outlets (Nielsen (2015)). Nigeria's trading sector is not exceptional in its economic importance or form relative to other developing countries; for instance, the comparable fraction of sales through traditional outlets is 96% in Ghana and 92% in India. In spite of interest in the large and growing Nigerian consumer goods market among international firms, the presence of large-scale modern retailers is extremely limited. Even genuine branded goods tend to make their way to consumers via an "informal and fragmented" (Leke et al. (2014)) distribution system in which manufacturers exercise "little control over the rest of the distribution chain" (Nielsen (2015)).

This stands in contrast to the modern trading sector that predominates in rich countries, characterized by larger formal firms and higher labor productivity. We are not aware of data that allows for systematic comparisons of the structure of intermediation across countries. However, casual empiricism suggests that the Walmart model – large-scale, centrally managed distribution systems that connect retail operations directly to producers via corporations with national or global reach – is more common in the developed world. Nonetheless, wholesaling is still economically important in rich countries as well – Ganapati (2020) shows that the share of manufactured goods transactions in the United States that were intermediated by wholesalers actually increased from 32% to 50% between 1992 and 2012, although the sector is increasingly consolidated as a result of large fixed cost investments in sourcing and distribution infrastructure.

2.2 Lagos Trader Survey data

We make use of an original data set we collected, which captures part of the distribution chain for manufactured consumer goods imported into Nigeria. The Lagos Trader Survey (LTS) is a panel survey of wholesale and retail traders in Lagos, Nigeria, and includes information about their international and domestic transactions during a five year period, from 2013 to 2017.[2]

LTS participants were identified through a census of over 50,000 shops in commercial areas[3] of Lagos conducted in 2014 and 2015. The survey includes 1,179 traders whose shops were randomly sampled from the census, of whom 620 had imported goods in 2013-14. The sample includes any trader dealing in manufactured consumer goods (excluding food), which we group into six product categories: apparel (including shoes, bags, and textiles), electronics, toiletries and beauty products, hardware, home goods, and miscellanous other products. These goods account for roughly 17 percent of consumer spending in Nigeria.

Consistent with the general picture of Nigeria's trading sector as fragmented and informal, we find that traders' businesses are small, owner-operated wholesale and retail firms. The median firm has one shop, and employs one worker in addition to the owner. On average, each imports US $59,000 in goods annually.

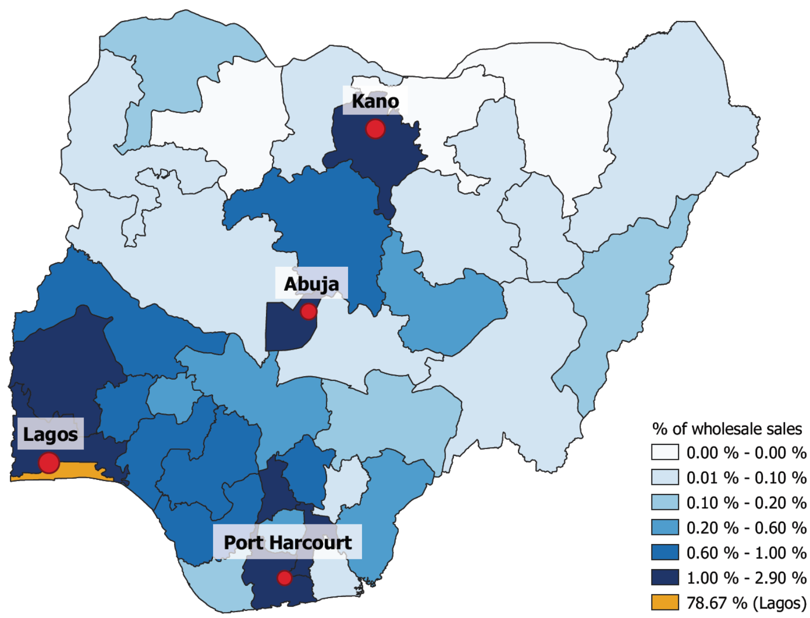
The data captures traders' purchases from international suppliers at the transaction level, and identifies where suppliers are located and whether they are manufacturers or wholesalers.[4] Over thirty different source countries are represented in the data; the largest is China, but the United Arab Emirates, Turkey, Hong Kong, Benin, India, and the United States are also common. Overall, 68% of suppliers are wholesalers rather than manufacturers. For each transaction, we observe the product type and quantity, the cost paid to the supplier, the cost paid to bring the good back to Lagos (including transportation and clearing the port), and the average price the trader in turn charges to buyers for that particular good. This enables us to construct measures of unit costs,

[2] More details about the Lagos Trader Survey can be found in Startz (2021).

[3] The listing focuses on commercial and wholesaling areas of the city, and does not include most residential or manufacturing areas or traditional food markets.

[4] All information is reported by the interviewees, not through direct contact with suppliers or customers. There is likely to be some measurement error, as traders may not know all the details about firms they interact with. We suspect that on net this leads to overestimates of the extent to which traders buy directly from producers, as they may assume any distributor of a known brand is the "manufacturer" of that brand.

Figure 1: Downstream sales in Nigeria



trade costs, and markups, and to relate these to supplier type.

We observe the fraction of traders' sales that are retail versus wholesale, and, starting from 2016, whether they have sales to locations outside of Lagos. In total, 86% of traders in the LTS do some wholesaling, and 52% of observed sales are wholesale.[5] Approximately three-quarters of traders' sales are within Lagos state, but the remainder go to other destinations, consistent with Lagos' role as the main port for all of Nigeria. Figure 1 shows the fraction of wholesale sales that go to each state in Nigeria. Many sell to downstream traders in locations that serve as commercial hubs for other parts of the country such as Abuja, which is the centrally located national capital, and Kano in the north, which has historically served as a hub for trans-Saharan trade routes.

2.3 Characteristics of distribution chains

A unique feature of the LTS data is that it allows us to track goods across multiple links in the distribution chain, conditional on passing through the hands of a Lagos trader. Data on purchases and sales allows us to construct a measure of chain length. We define purchases from manufacturers as one end point on the chain, and retail sales to final consumers as the other. Purchases from or sales to other traders are additional points of intermediation. The data allows us to observe chains with a minimum length of three and maximum length of five. When all of a trader's suppliers are

[5] The heavy weighting toward wholesale in the sample is unsurprising due to the focus on commercial areas of the city and restriction to businesses in permanent physical premises. The tens of thousands of small shops in residential areas and mobile hawkers throughout the city are excluded from the sample, and are likely to purchase their supplies from the wholesalers captured in the LTS data.

Table 1: Chain Length

	Chain length (mean)	Steps upstream of trader	Steps downstream of trader
All	4.23	1.63	1.58
Apparel	4.21	1.65	1.55
Electronics	4.29	1.77	1.48
Beauty	4.26	1.44	1.79
Hardware	4.19	1.60	1.57
Homewares	4.21	1.63	1.57

Note: "Steps upstream" are the average number of agents on the chain above the Lagos trader, and "steps downstream" are the number of agents on the chain below.

manufacturers and all of her sales are retail, this a chain of length three with one intermediary, the trader herself. A chain in which all suppliers are wholesalers and all sales are wholesale has a length of at least five, with three intermediaries – the trader and intermediaries she buys from and sells to.[6]

In this section, we make use of the fact that the data allow us to track multiple vertical links to document a set of novel facts about the intermediaries and distribution chains that bring internationally manufactured goods to Nigerian consumers.

Fact #1: Nigerian consumers are served by long chains with multiple intermediaries

The first and most important fact to come out of the data is that Nigeria is served by multi-intermediary chains. Table 1 shows average chain length for importers overall, and separately for each product category.[7] Chains have, on average, at least two or three independent intermediaries between a foreign producer and a domestic Nigerian consumer. This reflects an approximately even split between middlemen upstream and downstream of the Lagos-based trader. Although there is some variation across products – traders dealing in electronics are somewhat further downstream from manufacturers than the average, while those in beauty and cosmetics are less so – there are multiple layers of intermediation in all product categories.

Fact #2: Smaller traders are more likely to buy in resale markets, and pay higher prices and lower fixed costs

Disaggregating the details Lagos traders' sourcing decisions, we observe patterns that are strongly suggestive of scale economies that vary systematically across source markets. Overall, 68% of purchases are from resellers rather than manufacturers, but this total reflects substantial source market-level variation. For instance, while Lagos traders report large trade flows of clothing purchased from

[6] Our measures of chain length are truncated at length four and five, becase we do not observe the supplier's supplier or the buyer's buyer. The long chains that we observe are therefore actually a lower-bound on the true length of intermediation chains passing through Lagos.

[7] We assume that purchases from all sources are distributed proportionally across all types of buyers. This lets us construct a value-weighted average chain length at the source country - trader level.

Table 2: Costs and firm size when buying from a wholesaler

	(1) Log unit cost ($US)	(2) Fixed trade costs ($US)	(3) Revenue ($US)	(4) Number of employees	(5) Number of shops
% of purchases from wholesaler	0.23	-976.16*	-72187.58**	-0.47**	-0.16*
	(0.17)	(506.04)	(31746.07)	(0.20)	(0.09)
Obs.	403	403	220	403	403
Mean of dep. var.	2.26	2360.27	67,686.77	1.45	0.57
Product FEs	x	x	x	x	x

Note: Observations with measured markups greater than 500% are trimmed in all specifications.

Benin and electronics from Dubai, it seems unlikely that those goods were originally produced in those location, where manufacturing is not a substantial part of the economy. Consistent with a broad characterization of these markets as entrepots, suppliers in those places are heavily reported as wholesalers.[8]

Table 2 shows the relationship between buying from a wholesaler (versus a manufacturer) and traders' firm size and trade costs incurred in sourcing. Column (1) shows that, within a product category, traders pay higher unit costs when buying from a wholesaler. Column (2) shows that they also pay lower fixed costs, measured here as reported costs of travelling to the source market and fees paid to hire agents and clear the port.[9] If buying from a wholesaler allows for a tradeoff between lower fixed and higher variable costs, we should expect larger firms to source more directly and smaller firms to source indirectly.[10] Consistent with this, we see in Columns (3), (4) and (5) that traders who source from wholesalers have smaller businesses, whether measured in terms of total sales, number of workers, or number of shops.

Fact #3: Markups increase along the chain

Traders vary not only in their upstream sourcing strategies, but also in the extent to which they directly serve retail customers versus selling onward to other resellers. Column (1) of Table 3 shows that, in fact, traders' upstream and downstream positions are negatively correlated. Traders who are further upstream from final consumers – i.e. those who sell more wholesale relative to retail –

[8]Even in major manufacturing locations such as China and Germany, a smaller but still substantial fraction of suppliers are wholesalers. This is consistent with the evidence from customs data that a large fraction of exports in many countries, including China, are via wholesalers.

[9]There is some subjectivity in which reported trade costs should be considered fixed. However, the overall pattern that fixed costs are lower when buying from a reseller is robust to the inclusion or exclusion of different cost categories, or to simply allowing the type of seller to modify the relationship between value of goods purchased and total trade costs.

[10]Note that the causality driving this relationship should run both ways – a firm that is smaller for reasons unrelated to trade costs should select into a lower fixed cost sourcing strategy, and that strategy, because it also involves higher variable costs, should cause the firm to sell less and be smaller.

Table 3: Relationship between traders' chain position and firm characteristics

	(1) % of purchases from wholesaler	(2) Log unit cost ($US)	(3) Log markup
% of sales that are wholesale	-0.18**	-0.74**	-0.32***
	(0.09)	(0.31)	(0.09)
% of purchases from wholesaler		0.19	-0.05
		(0.17)	(0.05)
Log unit cost ($US)			-0.13***
			(0.01)
Obs.	403	403	403
Mean of dep. var.	1.52	2.26	.46
Product FEs	x	x	x

Note: Observations with measured markups greater than 500% are trimmed in all specifications.

are more likely to source directly from manufacturers themselves.[11]

Table 3 shows that traders pay different prices to their suppliers and charge different markups to their buyers based on their position in the chain. Column (2) shows that traders who are further upstream pay lower prices to their suppliers. They also charge lower markups to their customers, even when controlling for the fact that they face lower unit costs on average. Column (3) implies that going from entirely retail sales to entirely wholesale sales is associated with a one-third reduction in average markups. This is large, but the magnitude is roughly consistent with traders' self-reports that their wholesale prices are on average 10% lower than their retail prices.

In sum, chains serving Nigerian consumers are long, with on average at least two or three independent middlemen. The characteristics of these middlemen and their transactions vary systematically with their position on the chain, and are suggestive of both source-specific economies of scale and demand becoming less elastic as you move downstream along the chain. We turn next to a theoretical framework that can account for these empirical patterns.

3 A Simple Model of Intermediation

To understand why and how chains with multiple intermediaries might form, we build a framework in which consumers in a destination market demand a good that is produced elsewhere. That good can either be sourced directly from a producer, or indirectly through a resale market. We begin by showing that economies of scale in trade costs can lead to equilibria in which intermediaries maximize profits by sourcing indirectly. We then ask how these indirect sourcing equilibria affect

[11] Note that a negative correlation between upstreamness and downstreamness is not mechanical. If the dominant pattern were that some traders are involved in longer chains than others, we could observe a positive correlation, as Antras and Chor (2018) show for value chains.

consumers in the destination market, and whether consumers benefit from policies or technologies that induce the market to shift to a direct sourcing equilibrium. We work in a stylized setting with a general demand specification in order to highlight the main forces involved. In the following section, we complicate the setting and impose more structure on the model in order to bring it to data.

3.1 Model setup

3.1.1 Consumer demand

We consider the distribution of a single homogeneous good. This good is produced in one location, the "origin", and is demanded by consumers in a destination market. Consumers purchase the good from an endogenous number of local traders, N.[12] Each trader v sells a quantity q_v. We assume a partial equilibrium approach is justified, and normalize the marginal utility of income to one. The consumer payout from this good is given by

$$U = G\left(\sum_{v=1}^{N} f\left(q_v\right)\right) \tag{1}$$

where $G'\left(\cdot\right) > 0$, $G''\left(\cdot\right) < 0$, $f\left(0\right) = 0$, $f'\left(\cdot\right) > 0$ and $f''\left(\cdot\right) \leq 0$. These preferences nest commonly used demand specifications, including linear and CES, and constant passthrough demands more broadly. They also allow for sellers to be either small or large with respect to the market, and for consumers to perceive those sellers as either perfectly or imperfectly substitutable.[13] Imperfect substitutability across sellers implies that consumers may experience gains from *seller* variety even though the good itself is homogeneous. These gains may arise from sellers' different locations within a market, or other factors that affect individual consumers' preferences over purchasing from one seller versus another.[14]

3.1.2 Sourcing technology

Traders are intermediaries who source the good and resell it without transformation. They can purchase the good in the origin, where it is produced. However, the same good can also be sourced from other markets where it is being resold. The costs of sourcing the good in a market i and selling it to consumers in the destination are (c_i, F_i), where c_i represents a constant variable cost per unit

[12] The assumption that consumers buy only in their home market is simplifying but not necessary, and is equivalent to assuming parameters taking values such that it is not cost-effective for a single consumer to source from other markets.

[13] This representation follows Spence (1976) and Mankiw and Whinston (1986). In the Appendix, we generalize it further to encompass the Benassy (1996) extension of CES demands.

[14] For example, consider an address model in which consumers have ideal variety preferences across sellers that cause them to prefer their nearest retailer, but with some willingness to trade off price against time to reach other sellers. This is consistent with the findings of Atkin, Faber and Gonzalez-Navarro (2018) and Lagakos (2016), both of which suggest that developing country consumers consider retail outlets to be imperfect substitutes, and the latter of which connects this explicitly to tradeoffs between price and time/distance.

of the good and F_i is a fixed cost that does not vary with the number of units purchased.[15]

The variable cost includes the price paid to a supplier in the source market, as well as any variable trade costs incurred to get the good to the destination market and sell it there, such as transportation costs, insurance, tariffs, and so on. The fixed cost captures economies of scale that empirical evidence (reviewed in the introduction) suggests are often present in trade costs. For instance, this may include the cost of traveling to a source market to search for a supplier, fees paid at borders or to access letters of credit, time spent gathering information, the cost of operating a truck, and so on. For the time being, we assume that traders take all costs as exogenous. Implicitly, this assumes that the destination market is small and does not affect upstream prices, but this is only to focus attention on the key forces; we show in the following section of the paper that the endogeneity of upstream pricing is both important and interesting.[16]

This setting implies that traders face alternative technologies for serving the destination market with the same good, which feature different combinations of variable and fixed costs. Therefore, we can think of a technology frontier of source markets indexed by j, consisting of the origin, o, and the set of resale markets that have non-dominated cost combinations. So long as markups in resale markets are non-negative (i.e. the good is not being resold for less than cost) and a triangle inequality holds on any variable trade costs,[17] then variable cost must be higher in all resale markets than in the origin, $c_{j\neq o} \geq c_o$. This in turn implies that the fixed costs for all resale markets on the frontier are lower than the origin, $F_{j\neq o} < F_o$. If this were not the case for any resale market, it would be cost-dominated by the origin, and would not be on the frontier. Buying from the producer is therefore a higher fixed cost, lower variable cost sourcing technology compared to buying through a resale market.

3.1.3 Traders optimization and destination market equilibrium

The number of traders serving the destination market is endogenous. There is a large pool of identical potential traders, who can pay a fixed cost $f^e \geq 0$ to enter the market. After entry, each trader v chooses a source market j and a quantity q to buy and resell, in order to maximize their profits:

$$\pi_v\left(q_v, q_{-v}, j, N\right) = \left(p_v - c_j\right) q_v - F_j - f^e \tag{2}$$

[15] We assume this form of economies of scale for simplicity of exposition, but the results extend readily to more general decreasing marginal costs of sourcing.

[16] Although we are not yet explicitly modeling what happens in upstream resale markets, this also rules out traders selling in multiple locations and making joint sourcing decisions across those locations. In the quantitative model in Section 4, there is double marginalization along the chain and therefore potentially a motive to integrate, albeit one that is slightly different from the standard vertical integration problem in a production setting. Empirically, we observe very few firms that enter in multiple markets in the setting we focus on. However, endogenizing integration decisions is an interesting avenue for future work applying this framework to other settings.

[17] The triangle inequality says that it is not less expensive to trade via a third location than to trade directly, e.g. $\tau_{jk} \leq \tau_{jl}\tau_{lk}$ if variable trade costs are denoted by a multiplicative τ_{jk} between locations j and k. While the triangle inequality is not essential – if it does not hold, then there is simply an additional reason that goods may be routed through resale markets – it is an especially innocuous assumption in our setting since we do not consider physically indirect routing that does not involve a transaction to be a chain link. Therefore, most cases that might appear to be violations of the triangle inequality need not be as long as τ_{ij} is defined as the cost over the least-cost actual trade route between i and j.

where p_v is a function of q_v, q_{-v} and N.

Traders enter up to zero profits. We focus on symmetric equilibria in which $q_v = q$ for all v, and ignore integer constraints on the number of firms. An equilibrium in entry, sourcing, and quantity choices is described by the following conditions:

$$\pi\left(q, j, N\right) = 0 \tag{3}$$

$$\pi_v\left(q_v, q_{-v}, j, N\right) \geq \pi_v\left(q_v', q_{-v}, j', N\right) \;\; \forall\, j' \tag{4}$$

$$\frac{\partial}{\partial q_v}\pi_v\left(q_v, q_{-v}, j, N\right) = 0 \tag{5}$$

where q_v' is the quantity that would be chosen by trader v if it deviated to sourcing from market j', holding all other traders' choice fixed.

There are two types of symmetric equilibria: "direct sourcing", in which all traders in the destination source from the origin, and "indirect sourcing", in which they source from a resale market. In a direct sourcing equilibrium, the destination is served via a chain with one intermediary – the trader in the destination – who sources from the producer. In an indirect sourcing equilibrium, the chain has at least two intermediaries – the destination trader and a seller in the resale market – and may have more depending on where the resale market sources the good. We have not yet modeled upstream markets explicitly, but one can imagine that there are traders in the resale market who face an analogous choice between sourcing from the origin and some other set of resale markets. If the equilibrium there also features indirect sourcing, then chains can grow to include more than two intermediaries. In Section 4, we follow this logic through to model upstream wholesale markets and measure the full length of the distribution chain.

3.2 Indirect sourcing

When will the destination market be served by indirect sourcing? The choice across sources is fundamentally a trade-off between fixed and variable costs. In order for there to be an indirect sourcing equilibrium, it must be that traders find the increase in fixed cost associated with direct sourcing to outweigh the gains from the lower variable cost.

A necessary condition for an indirect sourcing equilibrium to exist is that no trader wants to deviate to direct sourcing.[18] To build intuition, we compare the profits from indirect sourcing to those from direct sourcing (holding other firms' choices constant) using a second-order approximation to profits with respect to variable costs. An indirect sourcing equilibrium can exist when:

$$\frac{F_o - F_{j\neq o}}{c_{j\neq o} - c_o} \geq q_j\left[1 + \frac{1}{2}\left(\frac{c_{j\neq o} - c_o}{c_{j\neq o}}\right)\left(\varepsilon_p^q - 1\right)\rho\right] \tag{6}$$

where q_j and c_j are per-firm quantity and unit cost under a symmetric indirect sourcing equilibrium

[18]This condition is necessary but not sufficient. In order for there to be an equilibrium sourcing from any given resale market j, it must also be that no firm wants to deviate to any other resale market, but these sufficient conditions do not provide any additional insight into the question of direct versus (any) indirect sourcing.

from location j, ε_p^q is the price elasticity of demand, and ρ is the passthrough rate.

If the destination is served via indirect sourcing, then Equation 6 must hold for the resale market in question. The first order forces driving sourcing are the relative cost advantages of the origin versus resale markets, and the quantity being sourced. Indirect sourcing is more attractive when the fixed cost advantages of the resale market are large $(F_o - F_{j\neq o})$ relative to the additional variable costs incurred $(c_{j\neq o} - c_o)$ – or, conversely, when the fixed cost barriers to accessing production locations are relatively high. Indirect sourcing is also more attractive when equilibrium firm size is small (i.e. q_j is low), because the total value of the variable cost disadvantage is small relative to the quantity-invariant fixed cost advantage. The second-order forces driving indirect sourcing are the price elasticity of demand facing individual traders, ε_p^q, and the passthrough rate, ρ. When the price elasticity of demand or the passthrough rate are lower, individual firms get less of an advantage from deviating to a lower variable cost strategy.

Equation 6 is a statement about equilibrium relationships,[19] but has substantive empirical content. Small firm size and high barriers to accessing production locations are characteristics of many markets in developing countries. This suggests that indirect sourcing is more common in these places. This result is not driven by any implicit assumption that there are market failures specific to developing countries – only that the characteristics of demand and costs are such that firms are small, and that this is consistent with the existence of indirect sourcing equilibria. However, if there *are* other market failures that further limit the size of firms operating in developing countries – such as span of control problems or credit constraints – these will also push toward more indirect sourcing and longer chains.

3.3 Welfare implications of cutting out the middleman

How are consumers affected by the possibility of sourcing goods via a resale market? We consider the problem of a planner who values welfare in the destination market, and who manipulates sourcing options but takes traders' entry and quantity decisions as given. The sourcing strategy chosen by the market equilibrium may be different from the (second-best) planner's solution, and policy interventions that induce more direct sourcing can either benefit or harm consumers at the end of the chain.

3.3.1 Continuous sourcing technology

Sourcing technologies represent markets, which are inherently discrete objects. However, it is useful to start by imagining that there is a continuous technology frontier, so that we can use derivatives to characterize the margins on which private incentives diverge from social ones. We denote the frontier in terms of variable cost as a function of fixed cost, $c(F)$, and assume that $c'(F) < 0$ and $c''(F) > 0$. The market equilibrium is defined as in Section 3.1.3, except that due to

[19] We keep the form of demand and competition intentionally general in this section, but under narrower assumptions, similar statements can be made in terms of comparative statics with respect to parameters, rather than equilibrium relationships.

the continuity of the cost function, Equation 4 is now:

$$\frac{\partial \pi \left(q_v, q_{-v}, F, N\right)}{\partial F} = 0 = c'\left(F\right) q_v + 1 \tag{7}$$

The planner can select the sourcing technology used by traders in the destination market. What happens if the planner moves the market along the frontier to a slightly higher level of fixed cost, starting from the unconstrained market equilibrium? Marginal cost falls, since $c'(F) < 0$. Under any technology, traders will continue to enter up to zero profits, and to choose quantity to maximize profits. Therefore, quantity per trader increases, $\frac{\partial q}{\partial F} > 0$, and the number of traders in the market falls, $\frac{\partial N}{\partial F} < 0$.

The increase in quantity and decrease in entry follow non-trivially from the form of demand described in Equation 1, and we leave a complete derivation to the Appendix. The intuition, in brief, is that under free entry the overall level of competition must be unchanged due to an envelope condition on the technology choice under the market equilibrium.[20] If competition is unchanged, but marginal cost falls, the profit-maximizing quantity must increase in order for marginal revenue to continue to equal marginal cost. Finally, if quantity per firm increases and competition is unchanged, then entry must fall.

With these comparative statics in hand, we can turn to how welfare changes in response to this change in technology. Destination market welfare is consumer surplus, as profits are zero in equilibrium:

$$W = U - Npq$$

The effect of the increase in F on welfare in the destination can be decomposed as:

$$\frac{\partial W}{\partial F} = \left(\frac{\partial U}{\partial N} - pq\right)\frac{\partial N}{\partial F} + N\left(p - c\right)\frac{\partial q}{\partial F} - N\left(c'\left(F\right) q + 1\right) \tag{8}$$

The first two terms on the right side of Equation 8 have opposite signs. The first term captures the gain or loss from a change in the number of sellers, where $\frac{\partial U}{\partial N} - pq$ is the difference between the social benefit and social cost from an additional seller. We have already shown that $\frac{\partial N}{\partial F} < 0$. Under our assumptions on the form of utility, welfare is weakly increasing in seller variety, so that $\frac{\partial U}{\partial N} - pq > 0$.[21] Therefore, the first term is negative, and an increase in fixed cost reduces welfare due to a reduction in the number of sellers. The second term corresponds to the gain or loss from a change in quantity per-seller. Markups, $p - c$, capture the difference between the social benefit and social cost of an additional unit. Since markups are weakly positive and $\frac{\partial q}{\partial F} > 0$, the entire second

[20] In many similar-looking problems, we are accustomed to thinking about a comparative static with respect to F while holding c constant. In that case, the result is different and more trivial – an increase in F decreases profits, and so there must be exit under free entry, and the level of competition falls. In contrast, in this setting, F and c move together, and traders have already optimized with respect to the technology choice. The combination of the envelope condition on technology and the zero profit condition imply that competition does not change in response to local technology changes around the market equilibrium.

[21] $\frac{\partial U}{\partial N} - pq = G'\left(Nf\left(q\right)\right)\left[f\left(q\right) - qf'\left(q\right)\right]$ where $f\left(q\right) - qf'\left(q\right) \geq 0$ due to the concavity of $f\left(\cdot\right)$ and $G'\left(\cdot\right) > 0$ by assumption.

term is positive. Therefore, welfare increases due to a reduction in the quantity distortion arising market power.

The third term on the right side of Equation 8 is zero when starting from the market equilibrium, because of the cost-minimization condition on technology shown in Equation 7. From this, we can infer that the planner's technology choice and the market technology choice will not generally coincide. The first-order condition of the planner's problem will set the entirety of Equation 8 to zero, while the third term alone is set to zero in the market equilibrium. Therefore, the market equilibrium sourcing technology will not be efficient, except in cases where the first two terms are zero or cancel one another out. The former occurs under perfect competition, and the latter under standard CES demand with monopolistic competition, which is a knife-edge case in which variety and quantity distortions are perfectly offsetting. Otherwise, the planner is willing to accept some distortion of cost minimization in order to counterbalance net variety and quantity distortions that arise in the market equilibrium.

Will the planner's preferred technology lie above or below the market equilibrium along the frontier?

Proposition 1. If the assumptions on the form of demand from Section 3.1.1 hold, and there is a continuous sourcing cost frontier, the planner's preferred sourcing technology may have either a higher or lower fixed cost than the market equilibrium technology.

Proof. See Appendix.

As previously shown, the total welfare effect of a shift along the technology frontier is the net of the first two terms on the right side of Equation 8 when starting from the market equilibrium. Increasing the fixed cost reduces seller variety and mitigates the quantity distortion. The planner wishes to move to a higher fixed cost sourcing technology if the quantity distortion dominates, or a lower one if the variety distortion dominates. The concavity of welfare with respect to fixed costs means that if the planner wishes to lower (or raise) the fixed cost locally, then the planner's solution must lie below (or above) the market equilibrium.

The fact that a higher fixed cost technology mitigates the quantity distortion in the area around the market equilibrium does not imply that market power itself is decreasing. Defining the level of competition faced by trader v as $z_v = \sum_{v' \neq v} f(q_{v'})$, it must be that $\frac{\partial z}{\partial F} = 0$ at the market equilibrium so that the free entry condition is satisfied. When F increases locally, entry is lower and per firm quantity is higher, but these two forces exactly offset so that the overall level of competition is unchanged. Price can be expressed as a function of marginal cost and the level of competition, $p(c, z)$, and so even though the level of competition is fixed, markups may rise or fall, depending on whether passthrough of the lower marginal cost is greater than or less than one.

Finally, what are the implications of changes in technology away from the market equilibrium, as opposed to small changes around the point of tangency? The envelope condition on entry no longer holds, and so changes in fixed cost induce changes in the level of competition. In particular, starting from technologies that have a higher fixed cost than the market equilibrium, a further increase in fixed cost must decrease the level of competition, so that $\frac{\partial z}{\partial F} < 0$. In brief, technology changes that move further from the market equilibrium reduce each trader's profits, holding quantity and the level of competition constant. In order for the free entry condition to hold, profit must stay at zero. Equation 5 implies that quantity has already been optimized, and so competition must decrease in order to bring profits back to zero. (See a full proof in the Appendix.) This result will be key to understanding the implications of discrete changes in technology.

3.3.2 Discrete sourcing technology

In reality, the technology frontier is made up of a finite number of potential source markets, and the relevant changes in sourcing are discrete, rather than continuous. The planner can dictate sourcing from a particular market, which may involve a large change in fixed and variable costs relative to the source chosen by the market equilibrium. We are particularly interest in whether the planner should cut out middlemen by requiring direct sourcing, or conversely, by prohibiting sourcing from resale markets. Does this type of policy benefit consumers in the destination?

Under a discrete technology frontier, the change in welfare due to a shift from an equilibrium with sourcing from j to one with sourcing from j' is:

$$\Delta_j W = \left[\Delta_{N_j} U\left(N_j, q_{j'}\right) - p_{j'} q_{j'} \Delta_j N_j\right] + N_j \left[\Delta_{q_j} \frac{U\left(N_j, q_j\right)}{N_j} - c_j \Delta_j q_j\right] - N_j\left(q_{j'} \Delta_j c_j + \Delta_j F_j\right) \tag{9}$$

where Δ denotes a partial difference with respect to the subscripted variable, evaluated at j versus j' and holding other arguments fixed. For instance, $\Delta_{N_j} U\left(N_j, q_{j'}\right) \equiv U\left(N_j, q_{j'}\right) - U\left(N_{j'}, q_{j'}\right)$, and $\Delta_{q_j} U$ is defined analogously with respect to q rather than N. This is equivalent to the decomposition in equation (8), but accounts for the fact that a change in sourcing technology will involve a discrete jump in fixed and variable costs. The three terms on the right side of the equation – describing changes in welfare due to entry, quantity per firm, and total cost, respectively – have the same interpretation as those in the continuous version.

As noted in Section 3.1.2, all resale markets on the frontier have lower fixed costs than the origin market. Requiring direct sourcing therefore involves an increase in fixed cost and decrease in variable cost, so that $\Delta_j F_j > 0$ and $\Delta_j c_j < 0$. The sign of the first term on the right side of Equation 9 follows from the same reasoning as the continuous case. A higher fixed cost source implies a fall in entry, $\Delta_j N_j < 0$. Consumers always weakly gain from seller variety, and so welfare unambiguously falls due to the reduction in the number of sellers.

The logic of the second and third terms do not follow from the continuous case. For small changes around the market equilibrium, we showed that quantity per firm unambiguously increased. For discrete changes, however, quantity may either increase or decrease, $\Delta_j q_j \lesseqgtr 0$. As described in the previous section, an increase in F above the market equilibrium level leads to a reduction in overall competition. When marginal cost and competition both fall, optimal quantity per firm may either increase or decrease, depending on the passthrough rate. Therefore, the sign of the second term is ambiguous – quantity per firm may rise or fall and this may increase or decrease welfare.

Finally, the third term is unambiguously negative. In contrast to the continuous case, where small changes in technology around the market equilibrium had zero effect on total cost due to an envelope condition on the technology choice, large changes in technology will affect total cost. By definition, the technology chosen by the market minimized total cost conditional on quantity. Therefore, by revealed preference, any change in the technology will increase total cost and reduce welfare.

In sum, when the planner dictates a shift from indirect to direct sourcing, consumers in the destination are harmed by the reduction in entry and the increase in total cost. They may either gain or lose from the change in the quantity distortion, depending on whether the reduction in competition or the passed through reduction in marginal cost wins out. On net, therefore, consumers may either gain or lose from policy that requires direct sourcing.

3.4 Discussion of welfare considerations

We have just shown that cutting out middlemen can either help or hurt consumers at the end of the chain. Welfare gains or losses are driven by a fundamental trade-off between costs and entry. Direct sourcing involves lower variable cost and higher fixed cost compared to sourcing indirectly through an upstream resale market. The lower variable cost benefits consumers, but the higher fixed cost leads to lower entry of sellers in the destination, which harms consumers through loss of seller variety and reduced competition. In this section, we draw out a few additional considerations that are useful for understanding how this fundamental trade-off relates to real policies and empirical contexts.

3.4.1 Product variety

In Section 3.3.2, we showed how a policy-induced direct sourcing equilibrium may be better or worse for consumers than indirect sourcing under a market equilibrium. It is worth noting, however, than a direct sourcing equilibrium may not exist. This is the case when the potential profits for a monopolist are insufficient to cover the fixed costs of entry and origin sourcing. If indirect sourcing is prohibited and a monpolist cannot earn positive profits under direct sourcing, then the destination market will not be served at all and consumer surplus goes to zero.

While we have thus far highlighted the implications of chain structure for welfare due to seller variety, the fact that the destination market may not be served at all in our single good set-up means there are also implications for product variety. Several recent papers find that small, remote markets in developing countries have less product variety (Atkin and Donaldson (2015); Gunning, Krishnan and Mengistu (2018)). This empirical pattern is predicted by any trade model that has product-level scale economies. Our framework implies that the existence of intermediation chains can ameliorate the tendency for small, remote markets to have less product variety. Goods for which local demand is not large enough to support procurement from the production location may still be available when they can be bought in resale markets that feature lower fixed costs of sourcing. Therefore, policies that cut out middlemen can harm remote consumers by eliminating this alternative channel for accessing products.

3.4.2 Technology improvements

The welfare analysis of Section 3.3 considered policy interventions that mandate movements along a technology frontier. However, the second-best considerations driving those results also imply that changes that strictly reduce the costs of some sourcing options – shifting the technology frontier inward – also have ambiguous effects on welfare.

Consider a market that is initially in an indirect sourcing equilibrium, and then experiences an exogenous reduction in the fixed cost of direct sourcing. This cost reduction could arise from removal of bureaucratic barriers, like reducing the red tape involved in getting a visa to visit the source market. Or, it could arise from actual technology improvements, like the introduction of a platform that facilitates communication between destination traders and origin sellers. Either way, suppose that this cost reduction is large enough that the market equilibrium switches from indirect to direct sourcing. Even though direct sourcing involves a lower fixed cost than it would have before the change, it may still involve a higher fixed cost than indirect sourcing. If this happens, the new equilibrium may feature lower entry and lower welfare. Of course, it may also increase welfare, and the same ambiguity holds for a change in the costs of indirect sourcing as well as direct sourcing. Importantly, these perverse effects of technology improvements can only arise through changes on the extensive margin of source market choice – if the costs of accessing a market decreases and traders continue to use that source, those inframarginal gains can only improve welfare.

3.4.3 Mechanisms driving gain and loss

Conventional wisdom often assumes implicitly that shortening distribution and collection chains will help consumers and small producers in developing countries. Our analysis suggests the need for more caution. Even under what might appear to be the circumstances least favorable to the need for middlemen – when the destination market is always served, when consumers do not care about seller variety, and when middlemen have market power and do not add value to the good itself – cutting out middlemen can reduce welfare by decreasing competition in the destination market.

Of course, consumers will not *necessarily* be hurt by such a policy, and can in fact be helped. It may be equally surprising to some readers that the sourcing structure selected by the market equilibrium is not generally efficient, and that this offers a rationale for intervention. However, even when there are gains from cutting out middlemen, the mechanism is not the one highlighted by the conventional wisdom, which often focuses on reducing the market power of intermediaries. Instead, in this framework, there are net gains to consumers when variable cost savings outweigh *increases* in market power.[22]

[22] Direct sourcing may reduce variable costs in the destination in part through elimination of markups charged by intermediaries in upstream resale markets, as we show in Section 4. A more complete argument, therefore, is that shortening chains can reduce the role of market power of held by intermediaries who are cut out of the chain, while

Saying more about the conditions under which policy is likely to help requires imposing more structure on demand and competition. Fundamentally, this is an empirical question, and so in the next section we turn to building these theoretical insights into a quantitative framework that can be taken to data.

4 A Many Location Model

The core insights of our framework are quite general: chains may arise in response to internal economies of scale in trade costs, and shorter chains can either help or harm consumers depending on a trade-off between lower costs and lower entry in downstream markets. In order to quantify these forces and consider how chain structure endogenously responds to changes in policy or technology, we now build the mechanism from Section 3 into a more narrowly specified model in a more realistic geography. Doing so requires three steps. First, we provide specific functional forms for demand and competition. Second, we model wholesaling explicitly, endogenizing prices in hub markets. Third, we make traders small with respect to the market, but allow them to be heterogenous. This both makes the model more tractable and allows for richer equilibrium chain structures, including layers of wholesaling and retailing within a location.

4.1 Environment

We still model the distribution of a single good, which is produced in the "origin", o. The good is manufactured by perfectly competitive firms using a constant returns to scale production technology, and is sold at a price p_o. In contrast to the simple geography in Section 3, we now allow the good to demanded by consumers in an arbitrary set of locations $i \in \{1, \ldots, J\}$.

4.1.1 Consumer demand

In every location i, a measure of identical consumers of length Z_i demand two goods: good 0 which is freely and costlessly traded, and good 1 (which we also refer to as the intermediated good) which consumers buy from a measure of retailers, $\Omega_{r,i}$, indexed by ω. Utility for a consumer who consumes X_0 units of good 0 and X_1 units of good 1 from trader ω is given by

$$U\left(X_0, X_1, \omega\right) = X_0 + A\left(X_1 \varepsilon\left(\omega\right)\right)^{\alpha}$$

where $\varepsilon\left(\omega\right)$ is an idiosyncratic iid Frechet draw of match value for a given consumer with seller ω. This Frechet draw has shape parameter $\frac{1}{\mu_c}$ with $\mu_c < 1$[23], and scale parameter $\frac{\Omega_{r,i}^{\gamma-\mu_c}}{\Gamma(1-\mu_c)}$ where $\Gamma\left(\cdot\right)$ denotes the gamma function and $\gamma \in [0, \frac{1-\alpha}{\alpha} \cdot \frac{\sigma_c-2}{\sigma_c-1})$ parameterizes a distinction between the

increasing that of those who remain, and that the latter effect may be larger than the former. However, there can be gains to consumers in our framework even if there are no upstream markups.

[23]As will become clear shortly, this condition is necessary to deliver a consumer elasticity of demand greater than 1.

private and social valuation of seller variety, as in Benassy (1996).[24] The parameter $\alpha \in \left(0, \frac{1}{1-\mu_c}\right)$[25] captures the price elasticity of demand for good 1, while $A > 0$ acts as a demand shifter for good 1.

Consumers are endowed with Y_i units of good 0. They first choose $X_{0,i}$, after observing all prices but before observing their match values with particular sellers. Then they observe their match draws and choose to buy from the trader who delivers the largest value of $X_1\varepsilon(\omega) = (Y_i - X_{0,i})\,\varepsilon(\omega)$. We assume that consumers in all locations are endowed with a sufficiently large initial quantity of good 0 so that they always consume a positive amount of good 0 in equilibrium. This leads to CES-form demands across sellers with elasticity of substitution $\sigma_c = 1 + \frac{1}{\mu_c}$.

The maximized consumer payout from the intermediated sector is

$$\mathbb{E}\left[\left(\max_{\omega} \frac{Y - X_0}{p(\omega)}\varepsilon(\omega)\right)^{\alpha}\right] = \left(\frac{\alpha A}{\Omega_{r,i}^{-\frac{1}{1-\sigma_c}-\gamma} P_{r,i}}\right)^{\frac{\alpha}{1-\alpha}}$$

where $p(\omega)$ is the price charged by trader ω , $P_{r,i} = \left(\int_{\omega\in\Omega_{r,i}} p(\omega)^{1-\sigma_c}\,d\omega\right)^{\frac{1}{1-\sigma_c}}$ is the conventional CES price index, and $\tilde{P_{r,i}} \equiv \Omega_{r,i}^{\frac{1}{1-\sigma_c}-\gamma} P_{r,i}$ is the true consumer price index given the social valuation of variety implied by γ.

Therefore, overall expected utility is

$$\mathbb{E}\left[\max_{X_0}\left(X_0 + A\left(\max_{\omega}\frac{Y-X_0}{p(\omega)}\varepsilon(\omega)\right)^{\alpha}\right)\right] = Y - \left[\alpha A\left((\Omega_i^r)^{\frac{1}{1-\sigma_c}-\gamma} P_{r,i}\right)^{-\alpha}\right]^{\frac{1}{1-\alpha}} + \left(\frac{\alpha A}{(\Omega_i^r)^{-\frac{1}{1-\sigma_c}-\gamma} P_{r,i}}\right)^{\frac{\alpha}{1-\alpha}}$$

Total expenditure on the intermediated good will be $Z_i\,(Y_i - X_{0,i})$. We will denote this aggregate expenditure by E_i.

4.1.2 Geography and trade costs

Although the good must be purchased at the origin location at some point, it can be re-sold at other locations. As in the framework from Section 3, it can be traded from j to i with payment of both fixed ($F_{ij} > 0$) and multiplicative variable ($\tau_{ij} > 1$) trade costs, which are both denominated in units of the numeraire and conform to triangle inequalities $F_{ij} \leq F_{ik} + F_{kj}$ and $\tau_{ij} \leq \tau_{ik}\tau_{kj}$.

In equilibrium, the good will be sourced from multiple locations by different traders selling in the same home market. It will therefore be useful to define a "chain": a sequence of links (i.e. transactions at particular locations) through which the good moves from the origin to final consumers in a given location. A chain is denoted z and is defined by a vector of length N_z whose elements are the ordered locations at which the good is bought and sold. The location $z(N_z)$ is always the origin, $z(1)$ is the location where the good is consumed, and $z(n)$ is the location of the n^{th} transaction in the chain.

[24]The lower bound on γ is so that consumers (weakly) gain from variety. The upper bound delivers uniqueness of the equilibrium, as the proof of Proposition 2 shows.

[25]The upper bound on α ensures that firm profits will decline in the price index.

4.1.3 Intermediation

There are two types of intermediaries, wholesalers (w) and retailers (r). There is an infinite set of potential traders that can enter into either type of intermediation in every location. Firms can enter and sell only in a single location,[26] and can only sell to other intermediaries if wholesaling or to consumers if retailing. Firms must pay a fixed cost $f_{u,e}$ (which is sunk) to enter as type $u \in \{w, r\}$.

After entering, traders observe the distribution of wholesale prices in all source markets and draw idiosyncratic additive shocks to their fixed cost of sourcing from each, before choosing a source location. After choosing a source, each trader chooses a chain to serve and a supplier within their chosen source market, subject to multiplicative Frechet shocks to their variable profits on each chain ($\zeta_u(z)$) and for each supplier ($\varepsilon_z(\omega)$). They then set their own sale price, and buy the relevant quantity. Thus, the overall payout for a trader in location i, of type u, sourcing from location j, and sourcing from seller ω to serve chain z will have the form

$$\Pi_{iujz\omega} = \arg\max_p \zeta_u(z)\, \varepsilon_z(\omega)\, \pi(p, c(\omega), z) - F_{ij} - \xi_u(j) - f_{u,e}$$

where $\pi(p, c(\omega), z)$ denotes variable profits given own price p, unit cost $c(\omega)$, and choice to serve chain z. The individual profit shocks $\zeta_u(z)$ and $\varepsilon_z(\omega)$ are iid and distributed Frechet, the former with shape parameter $\frac{1}{\beta_u}$, and the latter with shape parameter $\frac{1}{\mu_t}$. For each, the scale of the distribution is normalized so that the expected value of the maximized payout is equal to book profits, for reasons discussed below and as described in detail in the Appendix. The idiosyncratic component of fixed cost, $\xi_u(j)$, is an iid draw from a Gumbel distribution with scale parameter s.

This formulation of intermediaries' problem has several important implications for the equilibrium behavior of distribution chains. First, requiring intermediaries to choose one chain to serve means that pricing is separable across chains. This ensures that there is a pricing equilibrium in pure strategies.[27] The single chain assumption is stylized, but generates realistic aggregate behavior while yielding standard markup rules and well-behaved demands.[28]

Second, normalizing the idiosyncratic components of variable profits implies that the expected value of sourcing from a location is not increasing in the number of sellers there. In the absence of this normalization, an increase in the number of wholesalers in a market increases the expected maximized match value for a trader sourcing there. This generates an agglomeration force – larger markets attract more wholesale sourcing, which further increases their size as more downstream

[26] This rules out integration across locations. This is a reasonable simplication in the context of Nigerian consumer goods, as discussed in Section 2. However, endogenizing the extent of cross-location integration in this framework may be important for understanding differences in distribution structure across countries with different income levels, and is an interesting subject for future work.

[27] If this were not the case, sellers would face demand from multiple types of traders who arrive to make wholesale purchases in a given location, who may represent different types of downstream demand. When a given seller changes her price, she would expect the composition of traders who choose to buy from to her to shift. In contrast to the typical principle that consumers will be more elastic at higher prices, we would expect a seller to serve relatively more of the inelastic types of wholesale buyers when her price increases, which encourages further price increases.

[28] An assumption that traders serve all chains that go from their chosen source j to their home location i and can perfectly price discriminate across chains yields similar aggregate behavior.

demand flows through them and induces entry. We shut down this force in order to focus attention on the consumer welfare tradeoffs we highlighted in Section 3, which arise regardless of agglomeration in wholesaling.[29] This also ensures a unique equilibrium and increases the computational tractability of the model.

Third, the combination of a measure of small traders, idiosyncratic shocks to sourcing decisions, and the lack of choke prices ensures that there will be non-zero flows along every chain in equilibrium. This allows the equilibrium to be characterized using first-order conditions, and avoids the intractable combinatorial optimization problem that is a familiar roadbloack in the international trade literature.

Finally, our assumptions on the form of demand and that traders are small eliminate the pro-competitive pricing effects that we allow for in Section 3. Because the setting is (modified) CES with monopolistically competitive traders, markups do not vary in response to entry. The welfare conclusions remain the same as those in Section 3, but allow variety effects to stand in for both variety and procompetitive forces, which move in the same direction, following a long tradition in the international trade literature.

4.2 Intermediary choices

Traders make sequential optimization decisions to enter, choose a source market, choose a chain and then a supplier, and finally to set sale prices in their home market. In this section we solve for each step, working backward from pricing to entry.

4.2.1 Price setting, wholesale demand elasticity, and supplier choice

One of the issues highlighted by our model is the need for a microfoundation for the elasticities of demand faced by wholesalers selling to other intermediaries that capture downstream behavior. In a typical discrete choice framework for final consumer demand, the elasticity of substitution across sellers comes from the distribution of an idiosyncratic shock capturing a buyers' match value with each seller. As a seller raises its price, the share of buyers who still find that seller the most rewarding falls in a way that is related to the distribution of the shock. Retail pricing in our model follows this standard form, with a markup over marginal cost as a function of the own-price elasticity of demand that follows directly from the consumer utility function, $\sigma_c = 1 + \frac{1}{\mu_c}$. For wholesalers, however, the problem is more complicated. Wholesalers' buyers are also resellers, and so their elasticity of demand will be governed not only by their own match values, but also the elasticity of demand they in turn face as they pass changes in their purchase price through to their sale price. The problem at any stage of the chain, therefore, depends on the full sequence of decisions by downstream agents. [30] We therefore solve simultaneously for pricing and sourcing rules that account for this downstream dependence at all points on the chain.

[29] Both seller variety gains and pro-competitive effects of entry into wholesaling would create agglomeration forces, which may be empirically relevant in some settings, and a subject for future work.

[30] In principle, models of production networks could face this same challenge. In practice, the literature has modeled production in ways that make the elasticity of demand for each input separable from the elasticity of demand for the output.

We start by considering the relationship between elasticities at steps n and $n+1$ along this chain, counting up from sales to final consumers at step 1 (so that agents at step n buy from agents at step $n+1$). For simplicity, we suppress $iujz$ subscripts since all transactions are by definition within a specific chain. Suppose that the agent at step n is an intermediary who faces CES demands (in terms of own price) of the form

$$\mathbb{E}\left[q_n\left(p_n\right)\right] = b_n p_n{}^{-\sigma_n}$$

The firm's marginal cost will be $c_n\left(\omega\right) = \tau_{n+1} p_{n+1}\left(\omega\right)$ where $p_{n+1}\left(\omega\right)$ is the price charged by the selected supplier ω at step $n+1$ and $\tau_{n+1}$is the relevant variable trade cost given the location of the supplier at step $n+1$. The firm has optimal markup rule:

$$p_n = \frac{\sigma_n}{\sigma_n - 1}\tau_{n+1} p_{n+1}\left(\omega\right)$$

so that profits when buying from supplier ω with match value $\varepsilon\left(\omega\right)$ can be expressed as

$$\pi_n\left(\omega\right) = \varepsilon\left(\omega\right) b_n \left(\frac{1}{\sigma_n}\right)\left(\frac{\sigma_n}{\sigma_n - 1}\tau_{n+1} p_{n+1}\left(\omega\right)\right)^{1-\sigma_n}$$

The firm will choose the upstream seller to maximize these profits, and since $\varepsilon\left(\omega\right)$ is distributed Frechet with shape parameter $\frac{1}{\mu_t}$, the probability of choosing seller ω from the set of potential sellers on chain z, Ω_z, is

$$\Pr\left(\omega\right) = \frac{p_{n+1}\left(\omega\right)^{\frac{1-\sigma_n}{\mu_t}}}{\int_{\Omega_z} p_{n+1}\left(\omega'\right)^{\frac{1-\sigma_n}{\mu_t}} d\omega'}$$

This implies that the expected sales of upstream sellers at step $n+1$ can be expressed as

$$\mathbb{E}\left[q_{n+1}\left(p_{n+1}\left(\omega\right)\right)\right] = b_{n+1}\left(p_{n+1}\left(\omega\right)\right)^{\frac{1-\sigma_n}{\mu_t} - \sigma_n}$$

where

$$b_{n+1} \equiv b_n \frac{\left(\frac{\sigma_n}{\sigma_n - 1}\tau_{n+1}\right)^{-\sigma_n}}{\int_{\Omega_z} p_{n+1}\left(\omega'\right)^{\frac{1-\sigma_n}{\mu_t}} d\omega'}$$

Note that if $\sigma_{n+1} \equiv \frac{\sigma_n - 1}{\mu_t} + \sigma_n$ and demand at step n has the hypothesized form $\mathbb{E}\left[q_n\left(p_n\right)\right] = b_n p_n{}^{-\sigma_n}$, then we have shown that demand at step $n+1$ also has this form. Since consumer demand at the final step of the chain is CES, by induction, demand will fit this CES form at every stage of the chain. Furthermore, if the shape of wholesale buyer-seller match values is the same μ_t in all locations, then this implies that at every step n of all chains the own-price elasticity of demand faced by wholesalers is:

$$\sigma_{n+1} - 1 = \left(\sigma_c - 1\right)\left(\frac{\mu_t + 1}{\mu_t}\right)^n$$

This formulation implies that elasticities of demand are increasing – and therefore, multiplicative markups are decreasing – in upstreamness from final consumers.

4.2.2 Chain choice

A trader sourcing in location j from market i chooses which chain z to serve.[31] The pricing rule shown in the previous section implies that all suppliers at step $n+1$ of chain z are symmetric and so we can think of a single price p_z for chain z. Since $\zeta_u(z)$ is distributed Frechet with shape parameter $\frac{1}{\beta_u}$, the probability of choosing chain z is

$$\Pr(z) = \frac{\pi_z^{\frac{1}{\beta_u}}}{\sum_{z'} \pi_{z'}^{\frac{1}{\beta_u}}}$$

which, for retailers, can be simplified to

$$\Pr(z) = \frac{p_z^{\frac{1}{\beta_r}(1-\sigma_c)}}{\sum_{z'} p_{z'}^{\frac{1}{\beta_r}(1-\sigma_c)}}$$

4.2.3 Source market choice

The Gumbel distributed shock $\xi_u(j)$ to fixed costs yields standard logit shares for sourcing decisions. Thus, the share of traders in i choosing to source from j of type u (which we denote χ_{ij}^u) is

$$\chi_{ij}^u = \frac{\exp\left(\frac{\pi_{ij}^u}{s}\right)}{\sum_{j'} \exp\left(\frac{\pi_{ij'}^u}{s}\right)}$$

where π_{ij}^u are the expected profits of trader type u when serving route ij – taking into account the probability of choosing each type of chain and the expected profits from serving that chain. These terms also take into account the share component of fixed cost from serving the ij route, F_{ij}. A positive measure of traders will source from every location in equilibrium.

4.2.4 Trader entry

Traders of both types will enter up to zero expected profits. Given the distribution of the fixed cost shock, this means for every location i

$$f_{i,e} = s \ln\left(\sum_j \exp\left(\frac{\pi_{ij}^u}{s}\right)\right)$$

This pins down the number of traders in every location.

[31] This choice is from within the set of feasible chains: either wholesale or retail (depending on the trader's type) with steps from j to i. Technically, this is actually a choice of step and chain in the case of chains with multiple wholesale links with the same source and destination, e.g. a chain like (o, i, i, i).

4.3 Equilibrium

4.3.1 Definition

The prior subsections can be summarized in the following definition of equilibrium. For each of retail and wholesale firms and for every location:

1. The measure of firms is pinned down by entry up to zero expected profits.
2. Firms choose source locations to maximize expected profits given their idiosyncratic shocks to the fixed cost of sourcing from every location.
3. Firms choose to participate in the chain which yields them the highest expected profits given their idiosyncratic draws draws across all chains.
4. Firms pick the seller on that chain which yields them the highest expected profits given their idiosyncratic draws across all relevant sellers.
5. Firms set prices that maximize their expected profits given their choice of source location, chain, and seller.
6. Consumers allocate expenditure across the numeraire and intermediated good, and pick a retail seller to maximize their utility.

4.3.2 Consumer price index

Our goals in this subsection are twofold. First, we provide a definition of the consumer price index. Perhaps more importantly, we will also show that each location presents a single expected sourcing cost to intermediaries from all downstream locations, up to variable trade costs. Thus, the model we have described can partly be understood as a microfoundation for a shared wholesale price at the market level. This also provides some initial intuition for why the equilibrium we define is unique (foreshadowing Proposition 2 in the next subsection), as these prices are independent of traders' sourcing decisions.

We start by defining the consumer price index. Consumer welfare in each destination i is a function of the true consumer price index, $\tilde{P}_{r,i} \equiv \Omega_{r,i}^{-\frac{1}{1-\sigma_c}-\gamma} P_{r,i}$. This is composed of a variety adjustment term $\Omega_{r,i}^{-\frac{1}{1-\sigma_c}-\gamma}$ and the conventional CES retail price index in location i, which we define as

$$P_{r,i}^{1-\sigma_c} = \Omega_{r,i} \sum_j \chi_{ij}^r P_{ij}^{1-\sigma_c}$$

where P_{ij} is what the conventional price index would be if a measure one of retailers sourced from location j.

We now turn to the $P_{ij}^{1-\sigma_c}$ term. This term reflects the average price charged by retailers who source for i from location j. In the appendix, we show that the only i-specific component of P_{ij} is the variable trade cost:

$$P_{ij} = m_c \tau_{ij} \tilde{P}_j$$

where $\tilde{P}_j$ is wholesale price index in market j which is paid by all retailers (on average) regardless of where they arrive from. This term is defined (depending on whether $j = o$ or not) by

$$\tilde{P}_{j\neq o}^{1-\sigma_c} \equiv \frac{1}{\phi_j^{1-\sigma_c}} \left[p_o^{(1-\sigma_c)\left(1+\frac{1}{\beta_r}\right)} + \sum_{j'} \left(\frac{M_2}{m_c} \tau_{jj'} \tau_{j'o} p_o \right)^{(1-\sigma_c)\left(1+\frac{1}{\beta_r}\right)} + \sum_{j''} \sum_{j'} \left(\frac{M_3}{m_c} \tau_{oj'} \tau_{j'j''} \tau_{j''o} p_o \right)^{(1-\sigma_c)\left(1+\frac{1}{\beta_r}\right)} + \dots \right]$$
$$\tilde{P}_{o}^{1-\sigma_c} \equiv \frac{1}{\phi_o^{1-\sigma_c}} \left[p_o^{(1-\sigma_c)\left(1+\frac{1}{\beta_r}\right)} + \sum_{j'} \left(\frac{M_2}{m_c} \tau_{ijj} \tau_{j'o} p_o \right)^{(1-\sigma_c)\left(1+\frac{1}{\beta_r}\right)} + \sum_{j''} \sum_{j'} \left(\frac{M_3}{m_c} \tau_{oj'} \tau_{j'j''} \tau_{j''o} p_o \right)^{(1-\sigma_c)\left(1+\frac{1}{\beta_r}\right)} + \dots \right]$$

The expressions weight the price of chains both based on their share in consumer expenditure (related to their cost and $1 - \sigma_c$) and the propensity of retailers to carry them (related to their cost and $\frac{1}{\beta_r}$). Consumer markups and transportation cost from j to i are shared across all chains and thus can be factored out. The final piece is ϕ_j – this is an index of profits across all chains from location j and influences the propensity of retailers to carry a particular chain. Importantly, this expression is also independent of i.

Formally, $\phi_j^{1-\sigma}$ is defined (depending on whether $j = o$ or not) by

$$\phi_{j\neq o}^{1-\sigma} \equiv \left(\frac{M_2}{m_c} \tau_{jo} p_o \right)^{\frac{1}{\beta_r}(1-\sigma_c)} + \sum_{j'} \left(\frac{M_3}{m_c} \tau_{jj'} \tau_{j'o} p_o \right)^{\frac{1}{\beta_r}(1-\sigma_c)} + \sum_{j''} \sum_{j'} \left(\frac{M_4}{m_c} \tau_{jj'} \tau_{j'j''} \tau_{j''o} p_o \right)^{\frac{1}{\beta_r}(1-\sigma_c)} + \dots$$
$$\phi_{j\neq o}^{1-\sigma} \equiv p_o^{\frac{1}{\beta_r}(1-\sigma_c)} + \sum_{j'} \left(\frac{M_2}{m_c} \tau_{oj'} \tau_{j'o} p_o \right)^{\frac{1}{\beta_r}(1-\sigma_c)} + \sum_{j''} \sum_{j'} \left(\frac{M_3}{m_c} \tau_{oj'} \tau_{j'j''} \tau_{j''o} p_o \right)^{\frac{1}{\beta_r}(1-\sigma_c)} + \dots$$

and where M_N denotes the markup at the end of a chain length N.

To summarize, despite the chain-specific pricing and chain history, each source market has a single price index that it presents to all downstream retail locations. This price index is a function only of fundamentals, which as we will show in Section 5 gives our model substantial empirical tractability.

4.3.3 Uniqueness

Proposition 2. If the assumptions on the setting and consumer and trader payouts of Section 4.1 hold, then the equilibrium is unique.

Proof. See Appendix.

The intuition for the uniqueness of the equilibrium is as follows. First, the consumer price at the end of each chain is a function only of fundamental parameters governing variable trade costs and buyer-seller match distributions. However, firm revenues are affected both by expenditure on the intermediated good and the conventional price index – both of which are influenced by the number of firms and their sourcing choices. It is straightforward to show that given a measure of firms, the sourcing choices of those firms are unique. Showing that there is a unique number of firms requires that the profits are strictly decreasing in the number of firms. This is more complicated, as an increase in the number of firms increases aggregate expenditure, decreases the conventional price index, and changes firm routing decisions. The provided conditions on γ (in particular, that γ is not too large) mean that the increase in expenditure is not so large as to outweigh the decrease in the conventional price index, so that holding sourcing fixed, firm revenue falls. A simple revealed preference argument suffices to show that changes in sourcing cannot increase firm profits, as otherwise the changed sourcing pattern would have been adopted in the first place. Therefore, there is a unique measure of retail traders that satisfies the zero profit condition, and thus a unique set of sourcing shares and consumer price indices in all locations.

Notably, the choices of retail traders and the consumer price index are independent of wholesalers' sourcing decisions: it suffices that some wholesalers serve every route (and make optimal pricing decisions); the measure of wholesalers serving a particular link doesn't matter for retailers. Given the retail trader decisions, it is possible to describe the ultimate consumer expenditure on each route. The cross-chain, sourcing, and entry decisions of wholesalers are unique for the same reasons as for retail traders.

4.3.4 Weighted average chain length

One additional feature of the equilibrium with empirical content is the average chain length serving consumers in a particular location. The formula for this outcome, which we denote by L_i for average chain length of chains ending in location i is

$$L_i = \sum_j \Omega_{r,i} \chi_{ij}^r \left(\frac{P_{ij}}{P_{i,r}} \right)^{1-\sigma_c} L_{ij}$$

where as before, $\Omega_{r,i}\chi_{ij}^r$ is the measure of retail traders sourcing from j, $\left(\frac{P_{ij}}{P_{i,r}} \right)^{1-\sigma_c}$ is the share of expenditure on these chains up to seller effects, and L_{ij} is the weighted average chain length on routes arriving from j.

In turn, for locations $j \neq o$, we show in the Appendix that

$$L_{ij} = \frac{\tilde{L}_j}{\left(\phi_j \tilde{P}_j \right)^{1-\sigma_c}}$$

where, depending on whether or not j is the origin,

$$\tilde{L}_{j \neq o} \equiv 2 \left(\frac{M_2}{m_c} \tau_{jo} p_o \right)^{\frac{1}{\beta_r}(1-\sigma_c)} + 3 \sum_{j'} \left(\frac{M_3}{m_c} \tau_{jj'} \tau_{j'o} p_o \right)^{\frac{1}{\beta_r}(1-\sigma_c)} + 4 \sum_{j''} \sum_{j'} \left(\frac{M_4}{m_c} \tau_{jj'} \tau_{j'j''} \tau_{j''o} p_o \right)^{\frac{1}{\beta_r}(1-\sigma_c)} + \dots$$

$$\tilde{L}_o \equiv \left(\frac{M_1}{m_c} p_o \right)^{(1-\sigma_c)\left(1+\frac{1}{\beta_r}\right)} + 2 \sum_{j'} \left(\frac{M_2}{m_c} \tau_{joj} \tau_{j'o} p_o \right)^{(1-\sigma_c)\left(1+\frac{1}{\beta_r}\right)} + 3 \sum_{j''} \sum_{j'} \left(\frac{M_3}{m_c} \tau_{j'j''} \tau_{j''o} p_o \right)^{(1-\sigma_c)\left(1+\frac{1}{\beta_r}\right)} + \dots$$

5 Quantification for Nigeria

We now return to the Nigerian context, and apply our many-location model to quantify the welfare implications of distribution chains. We will calibrate the model to capture the empirical features of the distribution of Chinese-made apparel to locations across Nigeria. Our goals are, first, to show that the model is able to capture nuanced empirical patterns in the part of the trade network that we observe, at realistic parameter values. Second, we aim to provide a sense of the magnitude of implications for consumer welfare and trade cost measurement.

In our baseline calibration, we will consider the distribution of Chinese-made apparel to consumers in Dubai, Lagos, and the other 36 states in Nigeria.[32] We treat apparel as a single good, and take each of the 39 locations to be a single market.[33]

5.1 Data

Our calibration relies mainly on the LTS survey data, supplemented by similarly structured original survey data on apparel traders in the capital city of Oyo state, and secondary data on the Nigerian economy.

From the LTS data, we observe the retail fraction of each Lagos trader's sales, and their total annual revenue. At the supply transaction level, we observe the source location, manufacturer or intermediary status of the supplier, per unit purchase and average sale prices, and reported variable trade costs. From the survey data on apparel trade in Oyo state, we use information on the retail fraction of sales, total annual revenue, the location of suppliers, and source location-specific variable trade costs. These data allow us to calculate the share of traders in Lagos and Oyo who source from each location.

For each state in Nigeria, we use administrative data from the National Bureau of Statistics on population, GDP per capita, and the average unit price of apparel goods taken from CPI microdata. From the World Bank Global Consumption Database we take the overall fraction of Nigerian GDP per capita spent on apparel. We use estimates of the number of wholesale and retail traders in each state from the Small and Medium Enterprises Development Agency of Nigeria (SMEDAN). Finally, we take average travel times by road between the capital cities of each state from Google Maps queries.

5.2 Estimation

Our model relies on three families of fundamental parameters:

[32] Including the Federal Capital Territory (FCT).

[33] The level of aggregation chosen to define a market is not neutral, since it has implications for the ability of buyers arriving at a location to benefit from variety. We choose an aggregation that corresponds to the level at which we have the data described below for each market, but there is no conceptual barrier to further disaggregation, and the quantification is sufficiently computationally tractable to allow for a larger number of locations.

Consumer utility These parameters govern the utility function of final consumers. A shifts expenditure on the intermediated good relative to the numeraire. The elasticity of expenditure on the intermediated good with respect to the true price index is α. In order to avoid building in the efficiency of sourcing and entry by construction, we allow γ to parameterize a difference between the private and social value of variety. Finally, μ_c describes the dispersion of consumers' idiosyncratic match values with individual retailers.

Intermediary matching These parameters decribe the willingness of intermediaries to substitute across upstream sellers. The dispersion of match values across chains is governed by β_r, while the dispersion of the idiosyncratic match value with individual wholesalers is described by μ_t.

Intermediary costs Last are parameters describing the cost fundamentals faced by intermediaries. Costs in all locations will depend on the manufacturer price at the origin, p_o. There is a fixed cost of entry f_{ei}, which we allow to take two values, one for Lagos and one for other locations in Nigeria; we assume the entry cost is the same between retail and wholesale. Last, all traders face shared fixed trade costs $\{F_{ij}\}_{\forall\ i,j}$ and multiplicative variable trade costs $\{\tau_{ij}\}_{\forall\ i,j}$ that are specific to trade from market j to market i. The fixed costs are subject to an additive idiosyncratic shock from a Gumbel distribution with scale parameter s.

5.2.1 Approximation to wholesale markups

In order to make use of matrix equations that improve the computational tractability of the model, we use an approximation to the aggregate markup on a chain across all steps (i.e. the wedge between the physical cost of the goods accounting for the origin price and aggregate trade costs and prices charged to consumers). This wedge will only depend on the total number of traders in the chain, which we will denote N, and we will preserve the notation of Section 4.2.1 in denoting by σ_n the elasticity faced by a trader $n-1$ transactions above final consumers (who buy at transaction 1). Using these definitions, we can write

$$M_N = \prod_{n=1}^{N} \frac{\sigma_n}{\sigma_n - 1}$$

where σ_n is defined recursively as in Section 4.2.1.

We take a second order approximation to $\ln M_N$ around $N = 2$ and $\frac{\mu_t+1}{\mu_t} = 1$ – an approximation that will be appropriate for relatively short chains and for large values of μ_t. Both assumptions are consistent with our estimates. This approximation yields

$$M_N \approx e^{\frac{1}{\mu_t \sigma_c}} \left(m_c e^{-\frac{1}{\mu_t \sigma_c}} \right)^N$$

where $m_c = \frac{\sigma_c}{\sigma_c - 1}$ is the markup charged to consumers. This expression captures consumer markups exactly, and the tendency of wholesale markups to be lower than retail markups (note that $m_n =$

$m_c e^{-\frac{1}{\mu_t \sigma_c}}$ for $n \geq 2$ which is strictly less than m_c).[34] For parsimony, we denote $B = e^{-\frac{1}{\mu_t \sigma_c}}$ so that

$$M_N \approx m_c \left(B m_c\right)^{N-1}$$

This approximation to markups permits straightforward computation of source-specific price indexes. Returning to our expressions in Section 4.3.2, we show in the Appendix that it is possible to approximate the price index when buying in location j, $\tilde{P}_j$, as

$$(j \neq o)\ \tilde{P}_j = \frac{p_o^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)}}{\phi_j^{1-\sigma_c}} \left[\left(B m_c \tau_{jo}\right)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} + \mathbf{T}_{jo} \left(\mathbf{I} - \mathbf{T}\right)^{-1} \mathbf{T}_{jj'}^T \right]$$

$$\tilde{P}_o = \frac{p_o^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)}}{\phi_o^{1-\sigma_c}} \left[1 + \mathbf{T}_{jo} \left(\mathbf{I} - \mathbf{T}\right)^{-1} \mathbf{T}_{oj'}^T \right]$$

where

$$\mathbf{T} \equiv \begin{pmatrix} \left(B m_c \tau_{11}\right)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} & \cdots & \left(B m_c \tau_{L1}\right)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} \\ \vdots & \ddots & \vdots \\ \left(B m_c \tau_{1L}\right)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} & \cdots & \left(B m_c \tau_{LL}\right)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} \end{pmatrix}$$

where $\mathbf{T}_{jj'}^T$ is the j-th column and $\mathbf{T}_{j'o}$ is the first row of $\mathbf{T}$.

A similar expression is possible for the profit index when sourcing from location j, ϕ_j, is

$$(j \neq o)\ \phi_j^{1-\sigma_c} = p_o^{\frac{1}{\beta_r}(1-\sigma_c)} \left[\left(B m_c \tau_{jo}\right)^{\frac{1}{\beta_r}(1-\sigma_c)} + \hat{\mathbf{T}}_{j'o} \left(\mathbf{I} - \hat{\mathbf{T}}\right)^{-1} \hat{\mathbf{T}}_{jj'}^T \right]$$

$$\phi_o^{1-\sigma_c} = p_o^{\frac{1}{\beta_r}(1-\sigma_c)} \left[1 + \hat{\mathbf{T}}_{j'o} \left(\mathbf{I} - \hat{\mathbf{T}}\right)^{-1} \hat{\mathbf{T}}_{oj'}^T \right]$$

where $\mathbf{I}$ is the identify matrix,

$$\hat{\mathbf{T}} \equiv \begin{pmatrix} \left(B m_c \tau_{11}\right)^{\frac{1}{\beta_r}(1-\sigma_c)} & \cdots & \left(B m_c \tau_{L1}\right)^{\frac{1}{\beta_r}(1-\sigma_c)} \\ \vdots & \ddots & \vdots \\ \left(B m_c \tau_{1L}\right)^{\frac{1}{\beta_r}(1-\sigma_c)} & \cdots & \left(B m_c \tau_{LL}\right)^{\frac{1}{\beta_r}(1-\sigma_c)} \end{pmatrix}$$

and $\hat{\mathbf{T}}_{jj'}^T$ is the j-th column and $\hat{\mathbf{T}}_{j'o}$ is the first row of $\hat{\mathbf{T}}$.

Similarly, the expressions for expenditure-weighted chain length in Section 4.3.4 can be approx-

[34] Note that we assume that $\sigma_c > 1$ and $\mu_t > 1$, $m_n > 1$.

Table 4: Baseline parameter values

Parameter	Meaning	Estimate	Source
Utility parameters			
μ_c	Dispersion of consumer match to retailers	0.80	Calculated from LTS
A	Shifts expenditure relative to numeraire	7.45	Estimated from NBS
α	Elasticity of expenditure to price index	0.19	Estimated from NBS
γ	Private versus social value of variety	1.41	Estimated from NBS
Intermediary matching			
μ_t	Dispersion of wholesale match	8.93	Calculated from LTS
β_r	Dispersion of chain match	0.22	Calibrated to match LTS
Intermediary costs			
p_o	Price in the origin	\$8.70	Calculated from LTS
τ_{ij}	Variable trade costs	various	Estimated from LTS and Google Maps
F_{ij}	Fixed trade costs	various	Estimated from LTS and Google Maps
s	Dispersion of fixed cost shocks	\$129	Calibrated from Startz (2021)
$f_{e(\text{Lagos})}$	Fixed cost of entry	\$3,866	Estimated from LTS
$f_{e(\text{other})}$		\$1,101	

Note: LTS is Lagos Trader Survey data and NBS is National Bureau of Statistics. The full matrix of fixed and variable cost estimates is available on request.

imated as

$$(j \neq o)\ \tilde{L}_j = \frac{p_o^{\frac{1}{\beta_r}(1-\sigma_c)}}{\left(\tilde{P}_j\phi_j\right)^{1-\sigma_c}}\left[2\left(Bm_c\tau_{jo}\right)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} + \mathbf{T}_{jo}\left(\sum_{n=0}^{\infty}(n+3)\,\mathbf{T}^n\right)\mathbf{T}_{jj'}^T\right]$$

$$\tilde{L}_o = \frac{p_o^{\frac{1}{\beta_r}(1-\sigma_c)}}{\left(\tilde{P}_o\phi_o\right)^{1-\sigma_c}}\left[1 + \mathbf{T}_{jo}\left(\sum_{n=0}^{\infty}(n+2)\,\mathbf{T}^n\right)\mathbf{T}_{oj'}^T\right]$$

where $\mathbf{T}$, $\mathbf{T}_{jo}$, $\mathbf{T}_{jj'}^T$, $\tilde{P}_j$ and ϕ_j are as previously defined.

When the relevant parameters are known, the price indexes are easy to compute, and the model counterfactuals can be solved quickly.

5.2.2 Estimation strategy

In this section, we discuss some additional assumptions made to take the model to the available data, and describe how each parameter is calibrated or estimated. Table 4 shows the estimated parameters.

We set μ_c to match markups charged by retailers in the LTS data,

$$\hat{\mu}_c = \hat{m}_c - 1$$

where $\hat{m}_c$ is the average observed retail markup.

Next, we estimate the demand parameters A, α, and γ via the following regression of per capita expenditure on the intermediated good on prices and numbers of traders across states within Nigeria:[35]

$$\ln\left(\frac{E_i}{Z_i}\right) = \delta + \beta_1 \ln \bar{p}_{ir} + \beta_2 \ln \Omega_{ri} + \varepsilon_i$$

where $\bar{p}_{ir}$ is the average retail price and Ω_{ri} is the number of retail traders. Given estimates of δ, β_1, and β_2, we can then calculate the demand parameters $\alpha = \frac{\beta_1}{\beta_1 - 1}$, $\gamma = -\frac{\beta_2}{\beta_1}$, and $A = \frac{\beta_1 - 1}{\beta_1} \exp\left(\frac{\delta}{1-\beta_1}\right)$. Our estimate of $\gamma = 1.41$ implies that the social value of seller variety exceeds the private value, and so the market equilibrium will feature insufficient entry. In general, this will make interventions that shift traders toward higher fixed cost sourcing strategies less valuable from a consumer welfare perspective.

The cross-chain elasticity parameter, β_r, is calibrated to approximately match the average chain length in Lagos that we observe in the LTS data. We estimate μ_t to match average observed markups charged by wholesalers in the LTS data, using the approximation to wholesale markups in all locations described in Section 5.2.1, which implies that:

$$\mu_t = \left(\frac{\mu_c}{\mu_c + 1}\right) \frac{1}{\ln m_c - \ln m_t}$$

The final set of parameters needed describe trade and entry costs faced by intermediaries. We take variable trade costs from China and Dubai to Lagos directly from the average values reported for apparel in the LTS data.[36] We assume that the following functional form describes variable trade costs between locations within Nigeria:

$$\tau_{ij} = \phi_\tau + \upsilon_\tau d_{ij}$$

where d_{ij} is the travel time by road between locations. Using the reported values for within Lagos and from Lagos to Oyo state from the supplementary survey data, we exactly fit ϕ_τ and υ_τ given $d_{i(\text{Lagos})}$. With estimates of ϕ_τ and υ_τ in hand, we calculate the implied τ_{ij} between all other locations in Nigeria.[37]

The scale parameter s governing the distribution of idiosyncratic components of fixed costs is set to match the variance of the distribution of fixed costs of travel for West Africa from Startz (2021). Given s, we can estimate fixed costs of sourcing from all locations j to i, relative to any fixed cost of sourcing from the home market via:

[35] We assume these parameters are the same across locations, but note that the features of demand and retail equilibrium in China and Dubai have no influence on our estimates or counterfactuals for locations within Nigeria.

[36] Note that it is not necessary to use reported trade costs. Instead, what is required is any two of the following three data points: retail and wholesale prices, variable trade costs, and markups (or the elasticity of substitution across retail outlets).

[37] Rather than assuming this same relationship holds internationally, we set $\tau_{(\text{China})(\text{Dubai})}$ to make the triangle inequality hold with equality from Lagos, based on the average estimated trade costs in Startz (2021).

$$F_{ij} - F_{ii} = \pi^v_{rij} - \pi^v_{rii} - s \ln \frac{\chi_{rij}}{\chi_{rii}}$$

In principle, this allows for estimation of fixed costs for any locations where sourcing shares and variable profits are observed. In our case, this means fixed trade costs from China and Dubai to Lagos, and from Lagos to Oyo states. We set F_{ii} based on the non-transport costs of sourcing reported by traders in Lagos who source from within Lagos.[38] Because we do not observe sourcing shares for other locations, we assume that fixed costs within Nigeria follow:

$$F_{ij} = F_{ii} + \upsilon_F d_{ij}$$

where we can fit υ_F exactly based on our Lagos and Oyo estimates, and extrapolate to calculate the implied F_{ij} between all other locations in Nigeria. For both fixed and variable trade costs, we assume that costs are symmetric from j to i and i to j, and that the triangle inequality holds with equality for importing from China or Dubai directly to locations within Nigeria. This implies, for instance, that the trade cost to move goods from China to Kano is exactly the cost to move them from China to Lagos and then Lagos to Kano.[39]

Finally, we can estimate the fixed cost of entry, inclusive of any fixed cost of sourcing from the home market via:

$$f_{ei} + F_{ii} = \pi^v_{rij} - F_{ij} - s \ln \chi_{ij}$$

which we can calculate for Lagos and Oyo. We assume that the estimate for Oyo holds for all other non-Lagos locations within Nigeria, and normalize entry costs in China and Dubai to one since they do not influence outcomes in Nigeria.

5.3 Baseline outcomes

With these parameter estimates in hand, we simulate the full model for all 39 locations. This yields equilibrium sourcing shares, retail price indexes, and weighted average chain lengths, allowing us to describe the distribution chains serving consumers in each location.

The average length of chains serving consumers in Nigeria is 4.6, implying that there are on average two or three intermediaries between the manufacturer and the final consumer. Table 5 shows that chain length is decreasing in population and income, and increasing in remoteness. A one standard deviation increase in log GDP per capita or state population corresponds to a 0.32 standard deviation or 0.39 standard deviation decrease in average retail chain length, respectively. A one standard deviation increase in travel time from Lagos – corresponding to approximately 9

[38] This includes both financial costs incurred by the trader, and an estimate of the average time cost of the reported one day per week spent on purhcasing.

[39] This assumption is consistent with the fact that Lagos is the main port of entry for goods from overseas which would then be transported by road within Nigeria. Air travel from anywhere in Nigeria to China or Dubai is also likely to go through Lagos.

Table 5: Relationship between chain length and location features

	(1) Chain length (z-score)
Log GDP/capita	-0.32**
(z-score)	(0.14)
State population	-0.39***
(z-score)	(0.11)
Travel time to Lagos	0.39***
(z-score)	(0.14)
Obs	37

Note: All variables are de-meaned and standardized.

hours of driving time by road – is associated with a 0.39 standard deviation increase in retail chain length.

5.4 Counterfactuals

5.4.1 Direct sourcing from the origin

We next turn to comparing outcomes under our baseline calibration to outcomes under a variety of alternative scenarios. We begin with the most basic – what happens if indirect sourcing is prohibited, so that retailers in all locations must source directly from the origin? This exercise compares a world with intermediation chains (our baseline) to the one implicit in most trade models, in which goods must be sourced directly from their production location. Although complete prohibition of wholesaling isn't a common policy, this counterfactual also has the flavor of more realistic policies that prohibit or discourage wholesaling in particular places along the chain. For instance, India's policy on Agricultural Produce Market Committee markets, or mandis, makes it difficult for traders to buy at farmgates or in more local, unregulated marketplaces. There has been a great deal of debate recently over whether deregulation of these markets would benefit farmers or consumers, or indeed, whether it would lead to more or less intermediation.

In practice, we arrive at a direct-sourcing-only equilibrium by setting trade costs between consumer locations and all sources other than China to high levels so that the share of firms sourcing from China approaches one hundred percent.[40] By construction, the result is that all traders in all downstream locations source from the origin, and all consumers are served by chains of length three with only one intermediary, their local retailer. The equilibrium fixed sourcing cost paid per retailer increases from approximately US$1,000 to over US$16,000. As a consequence, the number of retail outlets carrying Chinese apparel crashes throughout Nigeria. In response to the lack of seller variety, consumers reduce their spending on the intermediated good by 99% on average across all

[40] Due to the idiosyncratic Gumbel shock to fixed costs, the share is never actually zero to other locations, but becomes extremely small.

states, shifting toward the outside sector. Consumer surplus from Chinese apparel falls to 0.09% of its baseline level. While this is a stylized example, it reflects realistic forces in response to mandatory disintermediation: small retail shops cannot aggregate enough demand to cover the fixed costs of direct sourcing, and consumers will not be willing to travel to the couple of larger stores in the state capital that do carry the imported good.

5.4.2 E-commerce and domestic transportation costs

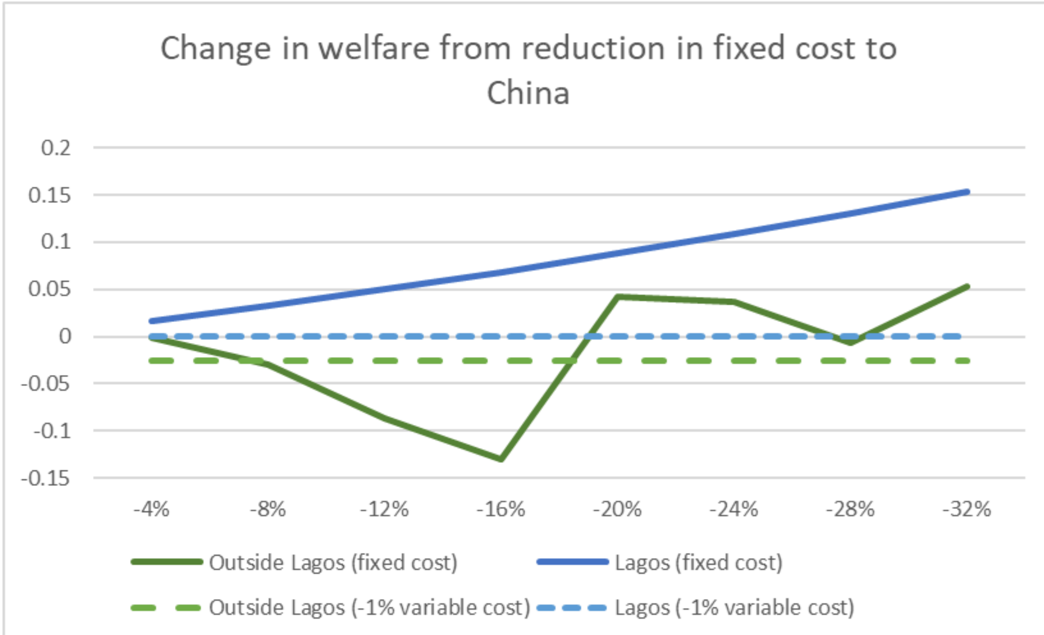
Even in circles in which regulation of intermediation is not popular, the idea that falling trade costs might connect producers and consumers more directly is often viewed favorably. One approach that has gained particular traction is the development of platforms to improve buyer-seller matching and reduce search costs or information asymmetries. These take a variety of forms, ranging from business-to-business e-commerce platforms—for instance, the 2018 expansion of Taobao into rural markets in China, studied in Faber et al. 2020, or ConnectAmericas, studied in Peru in Carballo et al. 2018—to public exchanges, such as the Ethiopian commodity exchange established in 2008 and since emulated in many other developing countries. In agricultural markets, the last five years has seen a wave of both public and private investment in platforms to link farmers with buyers in agricultural markets, in places like India, Ghana, and Uganda.

It is implicitly assumed that such technologies will benefit small producers and consumers, and in a framework with direct trade and efficient levels of firm entry, that would indeed be the case. However, in a world with intermediation chains and potentially inefficient sourcing equilibria, it is important to take endogenous restructuring of distribution into account. We model the introduction of such a platform as a reduction in the fixed cost of sourcing from China, to any location in Nigeria. We compare this to the impact of a more traditional investment to increase market access, such as a road improvement program that reduces variable trade costs domestically.

Figure 2 shows the welfare impacts of increasingly large reductions in the fixed cost of sourcing from China, and compares this to the impact of a roads program that reduces domestic variable trade costs by 1 percent. A reduction in domestic variable trade costs has a very small positive effect on consumer welfare in Lagos – close to zero net impact – and a small negative impact on average welfare in locations outside Lagos. In contrast, reductions in the fixed cost of sourcing from China always benefit Lagos, and increasingly so as the cost reduction gets larger.

The pattern outside Lagos is more nuanced. Reductions in fixed costs of sourcing from China have very little impact at low levels because they are not large enough to generate much of a shift in sourcing shares. This is consistent with one of the patterns that has emerged so far from the new agricultural matching platforms – they seem to have little impact, in part because few of the intended "direct" transactions end up taking place. If such platforms don't reduce costs by enough to make it worth a discrete switch in strategy for a substantial number of agents, then we should not expect to see much in the way of gains or losses. In our counterfactual scenario, as the reductions get larger, more traders do shift toward direct sourcing. This initially has a net negative welfare impact, as the marginal shifts in sourcing and resulting decreases in entry outweigh the gains from lower

Figure 2: Change in welfare due to fixed versus variable trade cost reductions



passed through costs. Consumers in most of Nigeria don't start to see gains until cost decreases are around 20% of the baseline level.[41]

6 Conclusion

We began by asking what might be lost in the implicit assumption common to the trade literature that goods go directly from producers to consumers. Everyday experience makes clear that this is almost never the case, and that wholesale and retail firms play a major role in the distribution of goods all over the world. Both the literature on agriculture in developing countries and our own survey data on manufactured goods in Nigeria document that there may in fact be long chains in which goods pass sequentially through the hands of more than one intermediary. We show that economies of scale in trade costs faced by individual firms are sufficient to give rise to such chains when goods can be resold by agents other than the original producer. While extremely simple, this conceptual insight yields a rich microfoundation for the endogenous structure of distribution chains and generates substantive empirical predictions about what chains will look like across locations and what firms and transactions will look like along chains.

Thinking about intermediaries as facing a menu of source markets with different fixed and variable costs is also the key to understanding how policies or technologies that cut out middlemen and shorten chains will impact consumers. We show that the market equilibrium does not generally select

[41] The construction of the model rules out downstream gains from increased wholesale entry in upstream markets. In reality, intermediaries elsewhere in Nigeria might gain from wholesale entry in Lagos, through gains from seller variety or procompetitive effects on wholesale prices. To the extent that this happens, it would mitigate the losses we see from small cost reductions.

efficient distribution, and that the optimal second best distribution equilibrium may involve either longer or shorter chains. These second best considerations arise from a trade off between minimizing the variable costs of serving a particular set of consumers and offsetting entry distortions in settings with market power and fixed costs. Quantifying these forces in the context of Chinese-made apparel sold in Nigeria shows that shortening chains may indeed reduce consumer welfare, whether it is due to regulatory intervention or technology improvements.

Price gaps between producers and consumers seem to be large on average in developing countries, and policymakers are extremely interested in reducing them. International organizations frequently fixate on reduction of "marketing costs" as a win-win solution to the "classic food price dilemma" (World Bank 2009): how to raise prices for poor producers without raising them for poor consumers. Accounting for the endogenous structure of intermediation chains is necessary for understanding which policy levers matter; for instance, whether the key is to reduce physical transport costs (e.g. through road-building or other infrastructure improvements), encourage entry into intermediation (e.g. by removing artificial regulatory barriers or reducing capital constraints), or to decrease fixed costs of sourcing from particular locations (e.g. through personal travel costs or restrictions, information frictions, or red tape or banking barriers). It is common for policy makers to assume that cutting out intermediaries will reduce their influence on prices and eliminate extra costs, but we demonstrate that it cannot be assumed that more direct connections are good for small farmers or consumers.

It may be tempting for economists to take the opposite stance, assuming that the market equilibrium delivers efficient distribution structures, and that therefore heavy-handed regulation of intermediation can only cause harm and reductions in trade costs can only do good. This is also not the case – policy can potentially increase welfare by restricting intermediation, and even pure technology improvements can have perverse effects. This highlights the importance of not building efficiency into models of intermediation or agricultural trade by construction, for instance through the common reliance on standard CES demands with monopolistic competition. It also implies that allowing for and documenting the extent of endogenous chain restructuring should be important for empirical work on intermediation, rather than simply studying changes in prices and passthrough holding chains fixed.

Although we provide rich empirical evidence on specific manufactured consumer goods in Nigeria, this paper highlights the need for much more systematic empirical documentation of distribution structures across settings, products, and countries. Casual empiricism suggests that consumers in rich countries may, on average, be served by shorter chains than those in poor ones. This would be consistent with the predictions of our model. New data collection may be needed, especially in developing countries, but there is also potential to use linked customs and VAT microdata as these become available across an increasingly wide range of countries.

Explaining cross-country patterns that we suspect will arise from such data may also require additions to our modeling framework. We considered distribution of a single good, by firms that only serve a single location. While these simplifications are reasonable and realistic in the context we study in Nigeria, we have abstracted from several forces that are likely to be important for understanding differences in the structure of distribution chains across rich and poor countries. The first is economies of scope that could be achieved if one intermediary can source and sell multiple goods at a total cost that is less than the cost of dealing in each separately. Scope decisions also introduce an additional "one stop shopping" role for entrepôt locations. For instance, a trader might be able to go to Dubai to buy goods from both China and India, rather than having to pay the costs to source from each separately.

A second force that will be key to understanding, for instance, the supermarket revolution in some parts of the developing world, is the motive to integrate by serving consumers in multiple locations. Integration allows intermediaries to both take advantage of greater economies of scale in sourcing, and to eliminate double marginalization at at least one step of the chain. If intermediaries in the developing world face more constraints on scale or the ability to have multiple outlets than those in the rich world – for instance due to differences in credit constraints or span of control – then allowing for scope and integration decisions becomes particularly important for explaining differences in chain structure across countries. We think this is likely to be a fruitful direction for future research, particularly if improved data also makes empirical comparisons across developing and developed countries possible.

References

Ahn, JaeBin, Amit K. Khandelwal, and Shang-Jin Wei. 2011. "The role of intermediaries in facilitating trade." *Journal of International Economics*, 84(1): 73–85.

Akerman, Anders. 2018. "A theory on the role of wholesalers in international trade based on economies of scope." *Canadian Journal of Economics*, 51(1): 156–185.

Antras, Pol, and Arnaud Costinot. 2011. "Intermediated Trade." *The Quarterly Journal of Economics*, 126(3): 1319–1374.

Antras, Pol, and Davin Chor. 2018. "On the Measurement of Upstreamness and Downstreamness in Global Value Chains." National Bureau of Economic Research Working Paper 24185. Series: Working Paper Series.

Atkin, David, and Dave Donaldson. 2015. "Who's Getting Globalized? The Size and Implications of Intra-National Trade Costs." Social Science Research Network SSRN Scholarly Paper ID 2642224.

Atkin, David, Benjamin Faber, and Marco Gonzalez-Navarro. 2018. "Retail Globalization and Household Welfare: Evidence from Mexico." *Journal of Political Economy*, 126(1): 1–73. Publisher: The University of Chicago Press.

Barrett, Christopher B., Thomas Reardon, Johan Swinnen, and David Zilberman. 2020. "Agri-food Value Chain Revolutions in Low- and Middle-Income Countries." *Journal of Economic Literature*.

Benassy, Jean-Pascal. 1996. "Taste for variety and optimum production patterns in monopolistic competition." *Economics Letters*, 52(1): 41–47.

Bergquist, Lauren Falcao, and Michael Dinerstein. 2019. "Competition and Entry in Agricultural Markets: Experimental Evidence from Kenya."

Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott. 2010. "Wholesalers and Retailers in US Trade." *American Economic Review*, 100(2): 408–413.

Biglaiser, Gary. 1993. "Middlemen as Experts." *The RAND Journal of Economics*, 24(2): 212–223.

Blum, Bernardo S., Sebastian Claro, and Ignatius J. Horstmann. 2009. "Intermediation and the Nature of Trade Costs: Theory and Evidence."

Casaburi, Lorenzo, and Tristan Reed. 2019. "Interlinked Transactions and Competition: Experimental Evidence from Cocoa Markets."

Chatterjee, Shoumitro. 2019. "Market Power and Spatial Competition in Rural India." Faculty of Economics, University of Cambridge Cambridge Working Papers in Economics.

Cosar, A. Kerem, and Banu Demir. 2018. "Shipping inside the box: Containerization and trade." *Journal of International Economics*, 114: 331–345.

Crozet, Matthieu, Guy Lalanne, and Sandra Poncet. 2013. "Wholesalers in international trade." *European Economic Review*, 58: 1–17.

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

Dhingra, Swati, and Silvana Tenreyro. 2020. "The Rise of Agribusiness and the Distributional Consequences of Policies on Intermediated Trade." Social Science Research Network SSRN Scholarly Paper ID 3535470, Rochester, NY.

Dillon, Brian, and Chelsey Dambro. 2017. "How Competitive Are Crop Markets in Sub-Saharan Africa?" *American Journal of Agricultural Economics*, 99(5): 1344–1361.

Dixit, Avinash K., and Joseph E. Stiglitz. 1977. "Monopolistic Competition and Optimum Product Diversity." *The American Economic Review*, 67(3): 297–308. Publisher: American Economic Association.

Emran, M. Shahe, Dilip Mookherjee, Forhad Shilpi, and M. Helal Uddin. 2020. "Credit Rationing and Pass-Through in Supply Chains: Theory and Evidence from Bangladesh." National Bureau of Economic Research Working Paper 26615. Series: Working Paper Series.

Fafchamps, Marcel, and Ruth Vargas Hill. 2008. "Price Transmission and Trader Entry in Domestic Commodity Markets." *Economic Development and Cultural Change*, 56(4): 729–766.

Ganapati, Sharat. 2020. "The Modern Wholesaler: Global Sourcing, Domestic Distribution, and Scale Economies."

Gunning, Jan Willem, Pramila Krishnan, and Andualem T. Mengistu. 2018. "Fading Choice: Transport Costs and Variety in Consumer Goods." Centre for the Study of African Economies, University of Oxford 2018-05-2. Publication Title: CSAE Working Paper Series.

Krishna, Kala, and Yelena Sheveleva. 2017. "Wheat or Strawberries? Intermediated Trade with Limited Contracting." *American Economic Journal: Microeconomics*, 9(3): 28–62.

Lagakos, David. 2016. "Explaining Cross-Country Productivity Differences in Retail Trade." *Journal of Political Economy*, 124(2): 579–620. Publisher: The University of Chicago Press.

Leke, Acha, Reinaldo Fiorini, Richard Dobbs, Fraser Thompson, Aliyu Suleiman, and David Wright. 2014. "Nigeria's renewal: Delivering inclusive growth." McKinsey & Company Global Institute.

Mankiw, N. Gregory, and Michael D. Whinston. 1986. "Free Entry and Social Inefficiency." *The RAND Journal of Economics*, 17(1): 48–58. Publisher: [RAND Corporation, Wiley].

Nielsen. 2015. "Africa: How To Navigate the Retail Distribution Labyrinth." Nielsen.

Niepmann, Friederike, and Tim Schmidt-Eisenlohr. 2017. "International trade, risk and the role of banks." *Journal of International Economics*, 107: 111–126.

Rubinstein, Ariel, and Asher Wolinsky. 1987. "Middlemen." *The Quarterly Journal of Economics*, 102(3): 581–593.

Spence, Michael. 1976. "Product Selection, Fixed Costs, and Monopolistic Competition." *The Review of Economic Studies*, 43(2): 217–235. Publisher: [Oxford University Press, Review of Economic Studies, Ltd.].

Startz, Meredith. 2021. "The value of face-to-face: Search and contracting problems in Nigerian trade."

Teravaninthorn, Supee, and Gael Raballand. 2009. *Transport Prices and Costs in Africa: A Review of the Main International Corridors.* World Bank Publications.

I. Material from Section 3

I.i. Derivation of equation 6

To derive Equation 6, we take a second order approximation to profits around the equilibrium level of variable costs, c, and fixed costs, holding all other firms' behavior fixed

$$\pi\left(c', F'\right) \approx \pi\left(c, F\right) + \frac{\partial \pi}{\partial c}\left(c' - c\right) + \frac{1}{2}\frac{\partial^2 \pi}{\partial c^2}\left(c' - c\right)^2 + F' - F$$

It follows immediately from the envelope theorem that $\frac{\partial \pi}{\partial c} = -q$, and thus

$$\begin{aligned}\frac{\partial^2 \pi}{\partial c^2} &= \frac{\partial q}{\partial p}\frac{\partial p}{\partial c} \\ &= \frac{q}{c}\left(\varepsilon_p^q - 1\right)\rho\end{aligned}$$

where ρ is the equilibrium passthrough rate and ε_p^q is the price elasticity of demand. The second line follows from firms' profit-maximizing quantity choices. Re-arranging the original second order approximation

$$\pi\left(c'\right) - \pi\left(c\right) = -q\left(c' - c\right)\left[1 + \frac{1}{2}\frac{\left(c' - c\right)}{c}\left(\varepsilon_p^q - 1\right)\rho\right] + F' - F$$

Thus, for an indirect sourcing equilibrium to hold, it must be that a change from indirect to direct sourcing yields (weakly) decreasing profits, giving the condition presented in the text.

I.ii. Proposition 1

Proposition 1. If the assumptions on the form of demand from Section 3.1.2 hold, and there is a continuous sourcing cost frontier in terms of fixed and variable costs, the planner's preferred sourcing technology may lie either above or below the market equilibrium along the frontier.

The first part of the proof is to derive expressions for $\frac{\partial N}{\partial F}$ and $\frac{\partial q}{\partial F}$ at the free market equilibrium.

The jumping off point are the free entry and optimal quantity setting conditions presented in the text. By differentiating the free entry condition with respect to F, we obtain

$$\begin{aligned}0 &= \left(\frac{\partial \pi_v}{\partial q_v} + \frac{\partial \pi_v}{\partial q_{-v}}\right)\frac{\partial q}{\partial F} + \frac{\partial \pi}{\partial N}\frac{\partial N}{\partial F} - c'\left(F\right)q - 1 \\ &= \frac{\partial p_v}{\partial q_{-v}}\frac{\partial q}{\partial F} + \frac{\partial p_v}{\partial N}\frac{\partial N}{\partial F}\end{aligned}$$

where the second line follows from optimal choices of quantity and technology by firms at the free market equilibrium. Similarly, from the firm quantity setting condition, we obtain

$$
\begin{aligned}
0 &= \left(\frac{\partial^2 \pi_v}{\partial q_v^2} + \frac{\partial^2 \pi_v}{\partial q_v \partial q_{-v}}\right)\frac{\partial q}{\partial F} + \frac{\partial^2 \pi_v}{\partial q_v \partial N}\frac{\partial N}{\partial F} - c'(F) \\
c'(F) &= \frac{\partial^2 \pi_v}{\partial q_v^2}\frac{\partial q}{\partial F} + q_v\left(\frac{\partial^2 p_v}{\partial q_v \partial q_{-v}}\frac{\partial q}{\partial F} + \frac{\partial^2 p_v}{\partial q_v \partial N}\frac{\partial N}{\partial F}\right)
\end{aligned}
$$

where to obtain the second line we use the derivative of the free entry condition with respect to F.

The prior results have been fully general; to simplify further we turn to the form of preferences specified in the text. We use N_{-v} to denote the number (or measure) of other firms in the economy. We use this notation to encompass both settings in which firms are large and $-N_v = N - 1$ and settings in which firms are small so that $N_{-v} = N$. We take appropriate derivatives and obtain (note that we suppress arguments of functions to simplify the expressions) to obtain

$$
\begin{aligned}
p_v &= G' f' \\
\frac{\partial p_v}{\partial N} &= \frac{f}{N_{-v} f'}\frac{\partial p_v}{\partial q_{-v}} \\
\frac{\partial^2 p_v}{\partial q_v \partial N} &= \frac{f}{N_{-v} f'}\frac{\partial^2 p_v}{\partial q_v \partial q_{-v}}
\end{aligned}
$$

so that

$$
\begin{aligned}
\frac{\partial^2 p_v}{\partial q_v \partial q_{-v}}\frac{\partial q}{\partial F} + \frac{\partial^2 p_v}{\partial q_v \partial N}\frac{\partial N}{\partial F} &= \frac{\partial p_v}{\partial q_{-v}}\frac{\partial q}{\partial F} + \frac{\partial p_v}{\partial N}\frac{\partial N}{\partial F} \\
&= 0
\end{aligned}
$$

We finish this section of the proof by solving for $\frac{\partial q}{\partial F}$ and $\frac{\partial N}{\partial F}$

$$
\begin{aligned}
\frac{\partial q}{\partial F} &= \frac{c'(F)}{\frac{\partial^2 \pi_v}{\partial q_v^2}} < 0 \\
\frac{\partial N}{\partial F} &= -\left(\frac{f}{N_{-v} f'}\right)^{-1}\frac{c'(F)}{\frac{\partial^2 \pi_v}{\partial q_v^2}} > 0
\end{aligned}
$$

where the signs in the second and third lines follow from the second order condition of firm profits with respect to own quantity.

We now turn to the proof of the main proposition. Equation (7) follows immediately from the first order condition of consumer surplus with respect to fixed cost and free entry. Using the general form of utility it can be re-written as (again, dropping the arguments of functions to simplify the expressions)

$$\begin{aligned}\frac{\partial W}{\partial F} &= G' \cdot \left(f - qf'\right) \frac{\partial N}{\partial F} + N\left(G'f' - c\right)\frac{\partial q}{\partial F} \\ &= -G' \cdot \left(f - qf'\right)\left(\frac{f}{N_{-v}f'}\right)^{-1}\frac{\partial q}{\partial F} + N\left(G'f' - c\right)\frac{\partial q}{\partial F} \\ &= N\left[\left(G'f' - c\right) - \frac{N_{-v}}{N}G'f' \cdot \left(1 - q\frac{f'}{f}\right)\right]\frac{\partial q}{\partial F}\end{aligned}$$

Inside the braces, the first term is strictly positive (since the marginal costs are constant and there are fixed costs of entry, price must be above marginal cost), while the second term is negative. We can increase the relative magnitude of the second term by assuming firms are small and f is very concave, while in the event that f is linear the second term disappears entirely. To show the ambiguous sign, consider the family of examples where firms are small, G is linear with slope $S > 0$, $f = aq - \frac{b}{2}q^2$ (where we assume a is sufficiently large) for positive constants a, b, and $c\left(F\right) = e^{-F}$.[42] Firm optimal quantity choices imply $q = \frac{a-c}{2b}$ so that $c = a + \sqrt{\left(\frac{a}{2}\right)^2 + 2b}$ and

$$\frac{\partial W}{\partial F} = NS\left[-\frac{c}{S} + \frac{\left(a+c\right)^2}{3a+c}\right]\frac{\partial q}{\partial F}$$

So, we can always pick an S such that this equation takes on a positive or negative sign.

I.iii. Derivation of equation 8 and signing of $\frac{\partial p}{\partial F}$

The derivation comes from differentiating welfare with respect to the fixed cost

$$\begin{aligned}W &= U - Npq \\ \frac{\partial W}{\partial F} &= \left(\frac{\partial U}{\partial N} - pq\right)\frac{\partial N}{\partial F} - Nq\frac{\partial p}{\partial F}\end{aligned}$$

and in turn, we can express p as a function of c and N, so that

$$\frac{\partial p}{\partial F} = \frac{\partial p}{\partial N}\frac{\partial N}{\partial F} + \frac{\partial p}{\partial c}c'\left(F\right)$$

combining these expressions yields equation (8).

We show that $\frac{\partial p}{\partial F}<0$

$$\begin{aligned}\frac{\partial p}{\partial F} &= \frac{\partial p}{\partial N}\frac{\partial N}{\partial F} + \frac{\partial p}{\partial q}\frac{\partial q}{\partial F} \\ &= \left(\frac{\partial p_v}{\partial q_v} + \frac{\partial p_v}{\partial q_{-v}}\right)\frac{\partial q}{\partial F} + \frac{\partial p_v}{\partial N}\frac{\partial N}{\partial F}\end{aligned}$$

and as we showed earlier in the appendix, the free entry condition implies $\frac{\partial p_v}{\partial q_{-v}}\frac{\partial q}{\partial F} + \frac{\partial p_v}{\partial N}\frac{\partial N}{\partial F} = 0$, and

[42] Set the entry cost so that there are no net profits given the other parameters.

we showed $\frac{\partial q}{\partial F} > 0$. Consequently

$$\begin{aligned} \frac{\partial p}{\partial F} &= \frac{\partial p_v}{\partial q_v}\frac{\partial q}{\partial F} \\ &< 0 \end{aligned}$$

I.iv. Choice of source between the hub and origin

First, we establish that a switch to the higher fixed cost lower variable cost source increases per-firm quantity and reduces the number of firms. Starting from the free-entry condition

$$\begin{aligned} 0 &= \Delta_j \left[q_j \left(p_j - c_j \right) \right] - \Delta F_j \\ \frac{\Delta_j F_j - \left(p_{j'} - c_{j'} \right) \Delta_j q_j}{q_j} &= \Delta_j \left(p_j - c_j \right) \end{aligned}$$

Second, taking differences from the condition for profit-maximizing quantity (where we substitute in the prior expression at the appropriate points)

$$\begin{aligned} 0 &= \Delta_j \left(p_j - c_j \right) + \Delta_j \left[q_j \frac{\partial p_j \left(q_i, q_{-i}, N_j \right)}{\partial q_i} \right] \\ &= \Delta_j \left(p_j - c_j \right) + q_{j'} \Delta_j \left(\frac{\partial p_j \left(q_i, q_{-i}, N_j \right)}{\partial q_i} \right) + \left(\frac{\partial p_j \left(q_i, q_{-i}, N_j \right)}{\partial q_i} \right) \Delta_j q_j \\ &= \frac{\Delta_j F_j - \left(p_{j'} - c_{j'} \right) \Delta_j q_j}{q_j} + q_{j'} \Delta_j \left(\frac{\partial p_j \left(q_i, q_{-i}, N_j \right)}{\partial q_i} \right) + \left(- \frac{\left(p_{j'} - c_{j'} \right) \Delta_j q_j}{q_j} \right) \Delta_j q_j \\ &= \frac{\Delta_j F_j}{q_j} + q_{j'} \Delta_j \left(\frac{\partial p_j \left(q_i, q_{-i}, N_j \right)}{\partial q_i} \right) + \left(\frac{\partial p_j \left(q_i, q_{-i}, N_j \right)}{\partial q_i} - \frac{\left(p_{j'} - c_{j'} \right)}{q_j} \right) \Delta_j q_j \\ &= \frac{\Delta_j F_j}{q_j} + q_{j'} \Delta_j \left(\frac{\partial p_j \left(q_i, q_{-i}, N_j \right)}{\partial q_i} \right) + \left(\frac{\partial p_j \left(q_i, q_{-i}, N_j \right)}{\partial q_i} - \frac{\left(p_{j'} - c_{j'} \right)}{q_j} \right) \Delta_j q_j \end{aligned}$$

When changing from hub to origin, $\frac{\Delta_j F_j}{q_j} > 0$, and $\frac{\partial p_j (q_i, q_{-i}, N_j)}{\partial q_i} - \frac{(p_{j'} - c_{j'})}{q_j}$ is negative. Therefore, $\Delta_j q_j > 0$ or $\Delta_j \left(\frac{\partial p_j (q_i, q_{-i}, N_j)}{\partial q_i} \right) < 0$. However, due to the second order condition on marginal revenue, $\Delta_j \left(\frac{\partial p_j (q_i, q_{-i}, N_j)}{\partial q_i} \right) < 0$ only if $\Delta_j q_j > 0$ – therefore $\Delta_j q_j > 0$. (Note that the same logic with a different sign on $\Delta_j F_j$ implies that when switching from origin to hub, the per-firm quantity falls). And $\Delta_j N_j$ must move in the opposite direction of $\Delta_j q_j$, as otherwise a firm would find it profitable to make a unilateral deviation. (If the number of firms didn't change, then the fall in markups from only a single firm changing its quantity would be less than when all firms change quantity. Thus if the zero profit condition held with a fixed number of firms when sourcing is changed, a firm would earn positive profits from a unilateral deviation).

We then proceed to the derivation of the welfare impact presented in the text.

$$
\begin{aligned}
\Delta_j CS_j &= \Delta_j \left(U\left(N_j, q_j\right) - N_j q_j p_j\right) \\
&= \Delta_{N_j} U\left(N_j, q_{j'}\right) + \Delta_{q_j} U\left(N_j, q_j\right) - \left(N_j \Delta_j \left(p_j q_j\right) + p_{j'} q_{j'} \Delta_j N_j\right) \\
&= \Delta_{N_j} U\left(N_j, q_{j'}\right) + \Delta_{q_j} U\left(N_j, q_j\right) - \left(N_j \Delta_j \left(c_j q_j + F_j\right) + p_{j'} q_{j'} \Delta_j N_j\right) \\
&= \left(\Delta_{N_j} U\left(N_j, q_{j'}\right) - p_{j'} q_{j'} \Delta_j N_j\right) + \Delta_{q_j} U\left(N_j, q_j\right) - N_j \left(c_j \Delta_j q_j + q_{j'} \Delta_j c_j + \Delta_j F_j\right) \\
&= \left[\Delta_{N_j} U\left(N_j, q_{j'}\right) - p_{j'} q_{j'} \Delta_j N_j\right] + N_j \left[\Delta_{q_j} \frac{U\left(N_j, q_j\right)}{N_j} - c_j \Delta_j q_j\right] - N_j \left(q_{j'} \Delta_j c_j + \Delta_j F_j\right)
\end{aligned}
$$

I.v. Technology improvements

The effect of a small change in technology on the equilibrium per-firm quantity and the number of firms will be ambiguous: any combination of signs for both objects is possible. We demonstrate this by solving for both of these outcomes.

To derive this result, we differentiate all three equations with respect to t (since equilibrium is always governed by the free market) to find

$$
\begin{aligned}
0 &= \frac{\partial \pi}{\partial q_{-i}} \frac{\partial q_{-i}}{\partial t} + \frac{\partial \pi}{\partial N} \frac{\partial N}{\partial t} + \frac{\partial \pi}{\partial F} \frac{\partial F}{\partial t} - c_t \\
0 &= \frac{\partial^2 \pi}{\partial q_i^2} \frac{\partial q}{\partial t} + \frac{\partial^2 \pi}{\partial q_i \partial q_{-i}} \frac{\partial q}{\partial t} + \frac{\partial^2 \pi}{\partial q_i \partial N} \frac{\partial N}{\partial t} + \frac{\partial^2 \pi}{\partial q_i \partial F_i} \frac{\partial F_i}{\partial t} - c_t \\
0 &= \frac{\partial^2 \pi}{\partial F_i^2} \frac{\partial F_i}{\partial t} + \frac{\partial^2 \pi}{\partial F_i \partial q_i} \frac{\partial q_i}{\partial t} + \frac{\partial^2 \pi}{\partial F_i \partial t}
\end{aligned}
$$

which can be simplified to

$$
\begin{aligned}
0 &= \frac{\partial p_i}{\partial q_{-i}} \frac{\partial q}{\partial t} + \frac{\partial p}{\partial N} \frac{\partial N}{\partial t} - c_t \\
0 &= \frac{\partial^2 \pi}{\partial q_i^2} \frac{\partial q}{\partial t} + q \left(\frac{\partial^2 p}{\partial q_i \partial q_{-i}} \frac{\partial q}{\partial t} + \frac{\partial^2 p}{\partial q_i \partial N} \frac{\partial N}{\partial t}\right) - c_F \frac{\partial F}{\partial t} \\
0 &= c_{FF} q \frac{\partial F}{\partial t} + c_F \frac{\partial q}{\partial t} + c_{Ft} q
\end{aligned}
$$

As before, using our assumptions on the general form of the utility implies

$$
\begin{aligned}
\frac{\partial p}{\partial q_{-i}} &= \frac{N_{-i} f'}{f} \frac{\partial p}{\partial N} \\
\frac{\partial^2 p}{\partial q_i \partial q_{-i}} &= \frac{N_{-i} f'}{f} \frac{\partial^2 p}{\partial q_i \partial N}
\end{aligned}
$$

which permits us to simplify the free entry and optimal quantity conditions

$$\frac{c_t}{\frac{\partial p}{\partial N}} = \frac{N_{-i}f'}{f}\frac{\partial q}{\partial t} + \frac{\partial N}{\partial t}$$

$$\frac{\partial F}{\partial t} = \frac{1}{c_F}\left[\frac{\partial^2 \pi}{\partial q_i^2}\frac{\partial q}{\partial t} + q\frac{c_t}{\frac{\partial p}{\partial N}}\frac{\partial^2 p}{\partial q_i \partial N}\right]$$

At this point, we can solve for $\frac{\partial q}{\partial t}$ and $\frac{\partial N}{\partial t}$

$$\frac{\partial q}{\partial t} = -\frac{q^2 \frac{c_{FF}}{(c_F)^2}\frac{c_t}{\frac{\partial p}{\partial N}}\frac{\partial^2 p}{\partial q_i \partial N} + c_{Ft}}{q\frac{c_{FF}}{(c_F)^2}\frac{\partial^2 \pi}{\partial q_i^2} + 1}$$

$$\frac{\partial N}{\partial t} = \frac{c_t}{\frac{\partial p}{\partial N}} + \frac{N_{-i}f'}{f}\frac{q^2 \frac{c_{FF}}{(c_F)^2}\frac{c_t}{\frac{\partial p}{\partial N}}\frac{\partial^2 p}{\partial q_i \partial N} + c_{Ft}}{q\frac{c_{FF}}{(c_F)^2}\frac{\partial^2 \pi}{\partial q_i^2} + 1}$$

and it is clear that the impact on these outcomes is ambiguous.

The expression for the welfare effect of a change in technology derived

$$W = U - N\left(cq + F + f^e\right)$$

$$\frac{\partial W}{\partial t} = \frac{\partial W}{\partial N}\frac{\partial N}{\partial t} + \frac{\partial W}{\partial q}\frac{\partial q}{\partial t} + \frac{\partial W}{\partial F}\frac{\partial F}{\partial t} + \frac{\partial W}{\partial t}$$

$$= \left(\frac{\partial U}{\partial N} - qp\right)\frac{\partial N}{\partial t} + N\left(p - c\right)\frac{\partial q}{\partial t} - Nc_t q$$

since, following firm choices of fixed cost, $\frac{\partial W}{\partial F} = 0$.

I.vi. Generalization to encompass utility like Benassy (1996)

In this section, we consider CES utility with a divergence between the social and private value of variety as in Benassy (1996) which does not fit in the utility form described in Equation (1). We show that our results still hold in this framework. In particular, we will consider utility of the form

$$U = G\left(\sum_{v=1}^{N} h\left(N\right) f\left(q_v\right)\right)$$

In addition to the assumptions we made before, we additionally assume $h > 0$, $N\frac{h'}{h} + 1 \geq 0$ (i.e. variety never makes consumers worse off), and $\frac{h}{h'N} + 1 < -\frac{h}{\varepsilon^{G'}}$ where $\varepsilon^{G'}$ is the elasticity of G' with respect to its argument (which will be negative, so that the right hand side is positive). This condition implies that $\frac{dp}{dN} < 0$.

First, the welfare decompositions presented in the text do not rely on the specific form of the utility and so clearly hold in this framework. However, the the interpretation changes slightly. There is a slightly different expression for the gains from variety, as

$$\begin{aligned}\frac{\partial U}{\partial N} &= NG'h'f + G'hf \\ &= p\left(N\frac{h'}{h}\frac{f}{f'} + \frac{f}{f'}\right)\end{aligned}$$

so that

$$\frac{\partial U}{\partial N} - pq = p\left[N\frac{h'}{h}\frac{f}{f'} + \frac{f}{f'} - q\right]$$

The second two terms $p\left(\frac{f}{f'} - q\right)$ are the same as before. However, the addition of the firs term means this expression can be negative in the event that f is not very concave and $h' < 0$.

Second, we show that under the provided conditions, $\frac{\partial N}{\partial F} < 0$ and $\frac{\partial q}{\partial F} > 0$ as before. It is obvious that the specific Benassy form will yield $\frac{\partial N}{\partial F} < 0$ and $\frac{\partial q}{\partial F} > 0$ – firms still have CES markups and complete passthrough, so that in response to a fall in variable costs, firms will increase output. The increased competitiveness and rise in fixed cost will then force firms to exit to maintain the zero profit condition. But we show that this holds more generally if $h'\left(2G'' - \frac{G'G'''}{G''}\right) \geq 0$ (note that this is a sufficient condition but not a necessary one). Our jumping off point is the simplified system of equations we derived earlier in this appendix before imposing any assumptions on utility

$$\begin{aligned}0 &= \frac{\partial p_v}{\partial q_{-v}}\frac{\partial q}{\partial F} + \frac{\partial p_v}{\partial N}\frac{\partial N}{\partial F} \\ c'(F) &= \frac{\partial^2 \pi_v}{\partial q_v^2}\frac{\partial q}{\partial F} + q_v\left(\frac{\partial^2 p_v}{\partial q_v \partial q_{-v}}\frac{\partial q}{\partial F} + \frac{\partial^2 p_v}{\partial q_v \partial N}\frac{\partial N}{\partial F}\right)\end{aligned}$$

and by substituting from one to the other

$$c'(F) = \frac{\partial^2 \pi_v}{\partial q_v^2}\frac{\partial q}{\partial F} + q_v\left(-\frac{\partial^2 p_v}{\partial q_v \partial q_{-v}}\frac{\frac{\partial p_v}{\partial N}}{\frac{\partial p_v}{\partial q_{-v}}} + \frac{\partial^2 p_v}{\partial q_v \partial N}\right)\frac{\partial N}{\partial F}$$

Working with our new utility form, and we find that

$$\begin{aligned}\frac{\partial^2 p_v}{\partial q_v \partial N} &= 2\left(f'\right)^2 G''hh' + h'G'f'' + \frac{\chi f}{N_{-v}f'}\frac{\partial MR_i}{\partial q_{-i}} \\ -\frac{\frac{\partial p_v}{\partial N}}{\frac{\partial p_v}{\partial q_{-v}}} &= \frac{h'G'f' + h^2\chi G''f'f}{N_{-v}G''\left(hf'\right)^2}\end{aligned}$$

So that

$$-\frac{\partial^2 p_v}{\partial q_v \partial q_{-v}}\frac{\frac{\partial p_v}{\partial N}}{\frac{\partial p_v}{\partial q_{-v}}} + \frac{\partial^2 p_v}{\partial q_v \partial N} = q_i hh'\left(f'\right)^2\left(2G'' - \frac{G'G'''}{G''}\right)$$

From the FOC of the free entry condition with respect to F, we establish that $\frac{\partial q}{\partial F}$ and $\frac{\partial N}{\partial F}$ have opposite signs. Thus, as long as

$$-\frac{\partial^2 p_v}{\partial q_v \partial q_{-v}} \frac{\frac{\partial p_v}{\partial N}}{\frac{\partial p_v}{\partial q_{-v}}} + \frac{\partial^2 p_v}{\partial q_v \partial N} < 0$$

We will obtain $\frac{\partial N}{\partial F} > 0$ and $\frac{\partial q}{\partial F} < 0$. This holds true if

$$h' \left(2G'' - \frac{G' G'''}{G''} \right) \geq 0$$

which is the assumption we make.

II. Material from Section 4

II.i. Consumer utility

We provide details for the statements about consumer utility provided in the paper.

The payout for agent with expenditure E_1 from buying variety ω at price $p(\omega)$ with idiosyncratic match value is

$$\frac{E_1}{p(\omega)} \varepsilon(\omega)$$

where $\varepsilon(\omega)$ is distributed Frechet with shape parameter $\frac{1}{\mu_c}$ and scale parameter $\frac{\Omega_r^{\gamma-\mu_c}}{\Gamma(1-\mu_c)}$ where Ω_r is the number (or measure) of retailers ω and γ is a parameter.

This gives rise to CES demands across varieties ω. Maximizing sub-utility is the same as maximizing the monotone transformation

$$\ln E_1 - \ln p(\omega) + \ln \varepsilon(\omega)$$

where $\ln \varepsilon(\omega)$ is distributed Gumbel with location parameter $(\gamma - \mu_c)\ln \Omega_r - \ln \Gamma(1-\mu_c)$ and scale parameter μ_c. As shown in Anderson, de Palma, and Thisse (1987) maximizing this monotone transformation of utility gives rise to CES demands with elasticity of substitution

$$\sigma_c = 1 + \frac{1}{\mu_c}$$

Next, determining the expenditure will depend on the expected value of the maximum draw. Note that because $\left(\max_\omega \frac{E_1}{p(\omega)} \varepsilon(\omega)\right)^\alpha$ is a monotone transformation of $\max_\omega \frac{E_1}{p(\omega)} \varepsilon(\omega)$, so the parameter α will not affect the choice of the utility maximizing seller. And, due to the properties of the Frechet distribution, $\max_\omega \frac{E_1}{p(\omega)} \varepsilon(\omega)$ will be distributed Frechet with shape parameter $\frac{1}{\mu_c}$ and scale parameter $\left(\frac{E_1 \Omega_r^{\gamma-\mu_c}}{\Gamma\left(1-\frac{1}{\mu_c}\right)}\right) \left(\sum_\omega p(\omega)^{-\frac{1}{\mu_c}}\right)^{\mu_c}$. Thus, $\left(\max_\omega \frac{E_1}{p(\omega)} \varepsilon(\omega)\right)^\alpha$ will be distributed Frechet will scale $\left(\frac{E_1 \Omega_r^{\gamma-\mu_c}}{\Gamma\left(1-\frac{1}{\mu_c}\right)}\right)^\alpha \left(\sum_\omega p(\omega)^{-\frac{1}{\mu_c}}\right)^{\alpha\mu_c}$ and shape parameter $\frac{1}{\alpha\mu_c}$, so that the expected value for the maximum draw will be

$$\mathbb{E}\left[\left(\max_{\omega} V(\omega)\right)^{\alpha}\right] = \frac{E_1^{\alpha}}{\left(\sum_{\omega}\left(\frac{1}{(p(\omega))}\right)^{-\frac{1}{\mu_c}}\right)^{-\alpha\mu_c}}\Omega_r^{\alpha(\gamma-\mu_c)}$$

If we define the price index in the usual way (and where following Anderson, de Palma, and Thisse (1987) we define $1-\sigma_c = -\frac{1}{\mu_c}$), i.e. that

$$P_r \equiv \left(\sum_{\omega}(p(\omega))^{1-\sigma_c}\right)^{\frac{1}{1-\sigma_c}}$$

then

$$\mathbb{E}\left[\max_{\omega} V(\omega)^{\alpha}\right] = \left(\frac{E_1}{P_r\Omega_r^{-\frac{1}{1-\sigma_c}-\gamma}}\right)^{\alpha}$$

Note that this suggests that $\max_{\omega} V(\omega)$ is identical to standard CES utility in the event that $\gamma = \frac{1}{1-\sigma_c}$. If not, following Benassy (1996), γ is a parameter governing gains from variety.

Thus, given a distribution of prices and measure of sellers which lead to the price index $\Omega_r^{-\frac{1}{1-\sigma_c}-\gamma}P_r$, following immediately from the FOC for E_1, the consumer will choose

$$E_1 = \left[\alpha A\left(\Omega_r^{-\frac{1}{1-\sigma_c}-\gamma}P_r\right)^{-\alpha}\right]^{\frac{1}{1-\alpha}}$$

Thus consumers will have expected utility from the differentiated sector of

$$\mathbb{E}\left[\left(\max_{\omega} V(\omega)\right)^{\alpha}\right] = \left(\frac{\alpha A}{\Omega_r^{-\frac{1}{1-\sigma_c}-\gamma}P_r}\right)^{\frac{\alpha}{1-\alpha}}$$

II.ii. Intermediary payouts

In this section, we provide details about the distribution of intermediaries' shocks.

We assume that (for each intermediation activity u), $\xi_u(j)$ is distributed iid Gumbel with scale parameter s and location parameter $-s\Gamma'(1)$ where $\Gamma'(1)$ is the derivative of the gamma function evaluated at 1 (and is equal to the negative of the Euler-Mascheroni constant).

We assume that (for each intermediation activity u), $\zeta_u(z)$ is distributed iid Frechet with shape parameter $\frac{1}{\beta_u}$ and scale parameter $\frac{1}{\Gamma(1-\beta_u)}\left[\frac{\sum_{z\in z_{iju}}\pi_z\pi_z^{\frac{1}{\beta_r}}}{\sum_{z'\in z_{iju}}\pi_{z'}^{\frac{1}{\beta_r}}} - \left(\sum_{z\in z_{iju}}\pi_z^{\frac{1}{\beta_u}}\right)^{\beta_u}\right]$ where $\Gamma(\cdot)$ is the gamma function, π_z is the expected profits conditional on choosiing chain z, and z_{iju} is the set of chains going from j to i for intermediation activity u.

Finally, we assume that $\varepsilon_z(\omega)$ is distributed iid Frechet with shape parameter $\frac{1}{\mu_t}$ and scale parameter $\frac{\Omega_z^{\frac{1}{1-\sigma_z}}}{\Gamma(1-\mu_t)}$ where $\Gamma(\cdot)$ is the gamma function, Ω_z is the measure of sellers on chain z, and

σ_z is the elasticity of demand at the relevant stage of chain z (note that this elasticity is solved for in the body of the paper). Note that this formulation means that the number of sellers serving a chain does not affect downstream payouts.

II.ii. Consumer price index

In this section, we provide details for the derivation of the expression of the consumer price index presented in the paper, P_{ij}. This is equal to the probability a given retailer carries any particular chain times that chain's contribution to the consumer price index. Note that the probability of carrying chain z for a retailer is (as provided in Section 4.2.2 of the paper)

$$\Pr(z) = \frac{p_z^{\frac{1}{\beta_r}(1-\sigma_c)}}{\sum_{z'} p_{z'}^{\frac{1}{\beta_r}(1-\sigma_c)}}$$

where we can ignore the role of the number of sellers due to the shape parameter of seller match distribution – all that matters for trader payouts are prices. For convenience, we will define ϕ_{ij} by $\left((m_c\tau_{ij})^{\frac{1}{\beta_r}}\phi_{ij}\right)^{1-\sigma_c} = \sum_{z'} p_{z'}^{\frac{1}{\beta_r}(1-\sigma_c)}$ – i.e. it is a (transformation of the) profit index. Using this definition, we can write for location $j \neq o$

$$\begin{aligned} P_{ij}^{1-\sigma_c} &= (M_2\tau_{jo}\tau_{ij}p_o)^{1-\sigma_c}\frac{(M_2\tau_{jo}\tau_{ij}p_o)^{\frac{1}{\beta_r}(1-\sigma_c)}}{\left((m_c\tau_{ij})^{\frac{1}{\beta_r}}\phi_{ij}\right)^{1-\sigma_c}} + \sum_{j'}\left(M_3\tau_{j'o}\tau_{jj'}\tau_{ij}p_o\right)^{1-\sigma_c}\frac{\left(M_3\tau_{j'o}\tau_{jj'}\tau_{ij}p_o\right)^{\frac{1}{\beta_r}(1-\sigma_c)}}{\left((m_c\tau_{ij})^{\frac{1}{\beta_r}}\phi_{ij}\right)^{1-\sigma_c}} + \sum_{j''}\sum_{j'}\left(M_4\tau_{j''o}\tau_{j'j''}\tau_{jj'}\tau_{ij}p_o\right)^{1-\sigma_c}\frac{\left(M_4\tau_{j''o}\cdots\right.}{\left(\cdots\right.} \\ &= \frac{(m_c\tau_{ij})^{1-\sigma_c}}{\phi_{ij}^{1-\sigma}}\left[\left(\frac{M_2}{m_c}\tau_{jo}p_o\right)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} + \sum_{j'}\left(\frac{M_3}{m_c}\tau_{j'o}\tau_{jj'}p_o\right)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} + \sum_{j''}\sum_{j'}\left(\frac{M_4}{m_c}\tau_{j''o}\tau_{j'j''}\tau_{jj'}p_o\right)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} + \ldots\right] \end{aligned}$$

and by close analogy

$$P_{io}^{1-\sigma_c} = \frac{(m_c\tau_{io})^{1-\sigma_c}}{\phi_{io}^{1-\sigma}}\left[p_o^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} + \sum_{j'}\left(\frac{M_2}{m_c}\tau_{j'o}\tau_{oj'}p_o\right)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} + \sum_{j''}\sum_{j'}\left(\frac{M_3}{m_c}\tau_{j''o}\tau_{j'j''}\tau_{oj'}p_o\right)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} + \ldots\right]$$

We next turn to the profit indexes, and we show (for $j \neq i$, we will neglect the origin as the definition is provided in the text and by this point it is clear by analogy)

$$\begin{aligned} \left((m_c\tau_{ij})^{\frac{1}{\beta_r}}\phi_{ij}\right)^{1-\sigma_c} &= (M_2\tau_{jo}\tau_{ij}p_o)^{\frac{1}{\beta_r}(1-\sigma_c)} + \sum_{j'}(M_3\tau_{j'o}\tau_{jj'}\tau_{ij}p_o)^{\frac{1}{\beta_r}(1-\sigma_c)} + \sum_{j''}\sum_{j'}(M_4\tau_{j''o}\tau_{j'j''}\tau_{jj'}\tau_{ij}p_o)^{\frac{1}{\beta_r}(1-\sigma_c)} + \ldots \\ \phi_{ij}^{1-\sigma_c} &= \left(\frac{M_2}{m_c}\tau_{jo}p_o\right)^{\frac{1}{\beta_r}(1-\sigma_c)} + \sum_{j'}\left(\frac{M_3}{m_c}\tau_{j'o}\tau_{jj'}p_o\right)^{\frac{1}{\beta_r}(1-\sigma_c)} + \sum_{j''}\sum_{j'}\left(\frac{M_4}{m_c}\tau_{j''o}\tau_{j'j''}\tau_{jj'}p_o\right)^{\frac{1}{\beta_r}(1-\sigma_c)} + \ldots \end{aligned}$$

where since it is independent of i, we define $\phi_{ij}^{1-\sigma_c} \equiv \phi_j^{1-\sigma_c}$ shared across all i. This then lets us define the $\tilde{P}_j$ as in the text, so that

$$P_{ij}^{1-\sigma_c} = (m_c\tau_{ij})^{1-\sigma_c}\,\tilde{P}_j^{1-\sigma_c}$$

yielding the expression in the text that

$$P_{ij} = m_c \tau_{ij} \tilde{P}_j$$

II.iii. Proposition 2

Proposition 2. If the assumptions on the setting and consumer and trader payouts of Section 4.1 hold then the equilibrium is unique.

As established in the prior subsection of this appendix, the P_{ij} are only a function of model parameters and not the equilibrium choices of intermediaries. Thus uniqueness boils down the sourcing choices of retailers and consumer price indeces, and these choices are separable across locations.

First, we show that, holding the measure of traders in a final location fixed, this implies a unique sourcing pattern. In particular, we show that when we define the function $e\left(P_{r,i}\right)$ (for notational simplicity)

$$e\left(P_{r,i}\right) = P_{r,i}^{1-\sigma_c} - \Omega_{r,i} \frac{\sum_j \exp\left(\frac{\pi_{ij}^r}{s}\right)}{\sum_{j'} \exp\left(\frac{\pi_{ij'}^r}{s}\right)} P_{ij}^{1-\sigma_c}$$

this function is strictly decreasing in P_{ij} ; this montonicity implies a unique value of P_{ij} (which in turn implies a unique sourcing pattern). The intuition for this result is that a rise in the consumer price index will make sourcing from all locations more profitable, but disproportionately so for locations which have the lowest price indeces. Formally, we start from the expression for expenditure in terms of the price index

$$\frac{E_i}{P_{r,i}^{1-\sigma_c}} = C \Omega_{r,i}^{\delta} P_{r,i}^{\psi}$$

where have adopted for notational convenience $C \equiv \frac{Z_i (\alpha A)^{\frac{1}{1-\alpha}}}{s \sigma_c}$,$\delta \equiv \frac{-\alpha}{1-\alpha}\left(-\frac{1}{1-\sigma_c} - \gamma\right)$ $\psi \equiv \sigma_c - 1 - \frac{\alpha}{1-\alpha}$. Then holding $\Omega_{r,i}$ fixed

$$\frac{\partial}{\partial P_{r,i}} \exp\left(\frac{\pi_{ij}^r}{s}\right) = \delta C \Omega_{r,i}^{\delta} P_{ij}^{1-\sigma_c} P_{r,i}^{\psi-1} \exp\left(\frac{\pi_{ij}^r}{s}\right)$$

Taking the derivative with respect to $P_{r,i}$ again holding $\Omega_{r,i}$ fixed:

$$e'\left(P_{c,i}\right) = \left(1-\sigma_c\right) P_{r,i}^{-\sigma_c} - \psi \Omega_{r,i}^{\delta} P_{r,i}^{\psi-1} \Omega_{r,i} \cdot \sum_j P_{ij}^{1-\sigma_c} \chi_{ij} \left(P_{ij}^{1-\sigma_c} - \frac{P_{r,i}^{1-\sigma_c}}{\Omega_{r,i}}\right)$$

where $\chi_{ij} \equiv \frac{\exp\left(\frac{\pi_{ij}^r}{s}\right)}{\sum_{j'} \exp\left(\frac{\pi_{ij'}^r}{s}\right)}$ is the sourcing share from j. Note that $\Omega_{r,i} \sum_j \chi_{ij} P_{ij}^{1-\sigma_c} = P_{r,i}^{1-\sigma_c}$ by definition, so that

$$\sum_j P_{ij}^{1-\sigma_c} \chi_{ij} \left(P_{ij}^{1-\sigma_c} - \frac{P_{r,i}^{1-\sigma_c}}{\Omega_{r,i}}\right) = \sum_j \chi_{ij} \left(P_{ij}^{1-\sigma_c} - \frac{P_{r,i}^{1-\sigma_c}}{\Omega_{r,i}}\right)\left(P_{ij}^{1-\sigma_c} - \frac{P_{r,i}^{1-\sigma_c}}{\Omega_{r,i}}\right)$$

which is simply the variance in the price index across locations and by definition weakly positive. Since restrictions on α imply $\delta > 0$ and $\sigma_c > 1$, it follows that $e'(P_{r,i}) < 0$ for all $P_{r,i}$. Thus, the solution to $e(P_{r,i}) = 0$ is unique holding the measure of firms fixed.

Second, we show that there is a unique measure of traders in equilibrium. Per-trader profits are strictly decreasing in the number of traders, such that there is only one measure of traders such that the expected profits are equal to the fixed cost of entry.

Before addressing this point directly, we develop a few expressions which will simplify future steps. We return to our expression for $\frac{E_i}{P_{r,i}^{1-\sigma_c}}$ from earlier, and now express it as

$$\frac{E_i}{P_{r,i}^{1-\sigma_c}} = C\Omega_{r,i}^{\delta'} \left(\sum_j \chi_{ij} P_{ij}^{1-\sigma_c}\right)^{\frac{\psi}{1-\sigma_c}}$$

where $\delta' = \delta + \frac{\psi}{1-\sigma_c}$. Furthermore, it can be shown that

$$\frac{\partial \chi_{ij}}{\partial \Omega_{r,i}} = \frac{1}{s\sigma_c}\chi_{ij}\left(P_{ij}^{1-\sigma} - \frac{P_{r,i}^{1-\sigma_c}}{\Omega_{r.,i}}\right) \cdot \frac{\partial}{\partial \Omega_{r,i}}\left(\frac{E_i}{P_{r,i}^{1-\sigma_c}}\right)$$

Using both of these expressions, we find

$$\frac{\partial}{\partial \Omega_{r,i}}\left(\frac{E_i}{P_{r,i}^{1-\sigma_c}}\right) = \frac{\delta'}{\Omega_{r,i}}\frac{E_i}{P_{r,i}^{1-\sigma_c}} + \frac{\frac{\psi}{1-\sigma_c}\Omega_{r,i}}{P_{r,i}^{1-\sigma_c}}\frac{E_i}{P_{r,i}^{1-\sigma_c}}\sum_j P_{ij}^{1-\sigma_c}\frac{\partial \chi_{ij}}{\partial \Omega_{r,i}}$$

$$\frac{\partial}{\partial \Omega_{r,i}}\left(\frac{E_i}{P_{r,i}^{1-\sigma_c}}\right) = \frac{\frac{\delta'}{\Omega_{r,i}}\frac{E_i}{P_{r,i}^{1-\sigma_c}}}{\left(\frac{1}{s\sigma_c}\frac{\frac{\psi}{\sigma_c-1}\Omega_{r,i}}{P_{r,i}^{1-\sigma_c}}\frac{E_i}{P_{r,i}^{1-\sigma_c}}V + 1\right)}$$

where we define $V \equiv \sum_j \chi_{ij}\left(P_{ij}^{1-\sigma} - \frac{P_{r,i}^{1-\sigma_c}}{\Omega_{r.,i}}\right)P_{ij}^{1-\sigma_c}$ as the variance in the P_{ij}.Note that $V > 0$. Following our assumption that $\gamma < \frac{1-\alpha}{\alpha}\frac{\sigma_c-2}{\sigma_c-1}$, so that $\delta < -\frac{\alpha}{1-\alpha}\left(\frac{1}{\sigma_c-1}\right) + \frac{\sigma_c-2}{\sigma_c-1}$ and $\delta' < 0$. Furthermore, our assumptions on α imply $\psi > 0$. Consequently, $\frac{\partial}{\partial \Omega_{r,i}}\left(\frac{E_i}{P_{r,i}^{1-\sigma_c}}\right) < 0$.

We now turn to the main result. Following the distribution of trader shocks, retail firm profits conditional on entry are

$$\mathbb{E}[\pi] = s\ln\left(\sum_j \exp\left(\frac{\pi_{ij}^r}{s}\right)\right)$$

Thus,

$$\frac{\partial \mathbb{E}[\pi]}{\partial \Omega_{r,i}} = C\Omega_{r,i}^{\delta'}\sum_j \chi_{ij}\frac{\partial}{\partial \Omega_{r,i}}\left(\frac{E_i}{P_{r,i}^{1-\sigma_c}}\right)$$

This will be strictly negative under the conditions provided, and there is a unique measure of traders which will enter retail.

II.iv. Chain length

The expenditure-weighted length of chains going from location j to location i will reflect the share of expenditure on the route allocated to each chain times the length of that chain. Formally, for a location $j \neq o$,

$$L_{ij} = \frac{1}{\left((m_c\tau_{ij})^{\frac{1}{\beta_r}}\ \phi_{ij}P_{ij}\right)^{1-\sigma_c}} \left[2\left(M_2\tau_{ij}\tau_{jo}p_o\right)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} + 3\sum_{j'}\left(M_3\tau_{ij}\tau_{jj'}\tau_{j'o}p_o\right)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} + 4\sum_{j''}\sum_{j'}\left(M_4\tau_{ij}\tau_{jj'}\tau_{j'j''}\tau_{j''o}p_o\right)^{\left(1+\right.}\right.$$

As with the price indeces, it is possible to factor out the i-specific components

$$L_{ij} = \frac{1}{\left(\phi_j\tilde{P}_j\right)^{1-\sigma_c}}\left[2\left(\frac{M_2}{m_c}\tau_{jo}p_o\right)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} + 3\sum_{j'}\left(\frac{M_3}{m_c}\tau_{jj'}\tau_{j'o}p_o\right)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} + 4\sum_{j''}\sum_{j'}\left(\frac{M_4}{m_c}\tau_{jj'}\tau_{j'j''}\tau_{j''o}p_o\right)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} + .\right.$$

so that we can define the shared part of chain length as

$$\tilde{L}_j \equiv 2\left(\frac{M_2}{m_c}\tau_{jo}p_o\right)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} + 3\sum_{j'}\left(\frac{M_3}{m_c}\tau_{jj'}\tau_{j'o}p_o\right)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} + 4\sum_{j''}\sum_{j'}\left(\frac{M_4}{m_c}\tau_{jj'}\tau_{j'j''}\tau_{j''o}p_o\right)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} + \dots$$

And for chains from the origin

$$L_{io} = \frac{1}{\left((m_c\tau_{io})^{\frac{1}{\beta_r}}\ \phi_{io}P_{io}\right)^{1-\sigma_c}}\left[\left(M_1\tau_{io}p_o\right)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} + 2\sum_{j'}\left(M_2\tau_{ij'}\tau_{j'o}p_o\right)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} + 3\sum_{j''}\sum_{j'}\left(M_3\tau_{ij'}\tau_{j'j''}\tau_{j''o}p_o\right)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)}\right.$$

so that we similarly simplify and define

$$\tilde{L}_o = (p_o)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} + 2\sum_{j'}\left(\frac{M_2}{m_c}\tau_{j'o}p_o\right)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} + 3\sum_{j''}\sum_{j'}\left(\frac{M_3}{m_c}\tau_{j'j''}\tau_{j''o}p_o\right)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} + \dots$$

Using these definitions, for all ij pairs we can write

$$L_{ij} = \frac{\tilde{L}_j}{\left(\phi_j\tilde{P}_j\right)^{1-\sigma_c}}$$

III. Material from Section 5

III.i. Markups approximation

We start with the expression for the markup at step n and find

$$\frac{\sigma_n}{\sigma_n - 1} = m_c\left[1 - \frac{1}{\sigma_c}\left(1 - \left(\frac{\mu_t + 1}{\mu_t}\right)^{1-n}\right)\right]$$

and then substitute into the expression for the log of the aggregate markup

$$\ln M_N = N \ln m_c + \sum_{n=1}^{N} \ln \left(1 - \frac{1}{\sigma_c}\left(1 - \left(\frac{\mu_t + 1}{\mu_t}\right)^{1-n}\right)\right)$$

We next take a second order approximation to $\ln M_N$ around $N = 2$ and $\frac{\mu_t+1}{\mu_t} = 1$ (i.e. as μ_t gets very large) to obtain

$$\ln M_N \approx N \ln m_c - \frac{1}{\sigma_c}\left(\frac{1}{\mu_t}\right)(N-1)$$

By exponentiating both sides we obtain the expression in the text.

III.ii. Price index approximation

Using our markup approximation, we are able to substantially simplify our expressions for the $\tilde{P}_j$. Starting from the expression for $\tilde{P}_j$ in the text for a non-origin source

$$\tilde{P}_j^{1-\sigma_c} \approx \frac{1}{\phi_j^{1-\sigma_c}} \left[(B m_c \tau_{jo} p_o)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} + \sum_{j'} \left((B m_c)^2 \tau_{j'o} \tau_{jj'} p_o \right)^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)} + \sum_{j''} \sum_{j'} \left((B m_c)^3 \tau_{j''o} \tau_{j'j''} \tau_{jj'} p_o \right)^{(}\right.$$

Then, by using the formulas for $\mathbf{T}$ in the text

$$\begin{aligned} \tilde{P}_{j\neq o}^{1-\sigma_c} &\approx \frac{p_o^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)}}{\phi_j^{1-\sigma_c}} \left[(B m_c \tau_{jo})^{\frac{1}{\beta_r}(1-\sigma_c)} + \mathbf{T}_{jo} \left(\sum_{n=0}^{\infty} \mathbf{T}^n \right) \mathbf{T}_{jj'}^T \right] \\ &= \frac{p_o^{\left(1+\frac{1}{\beta_r}\right)(1-\sigma_c)}}{\phi_o^{1-\sigma_c}} \left[(B m_c \tau_{jo})^{\frac{1}{\beta_r}(1-\sigma_c)} + \mathbf{T}_{jo} \left(\mathbf{I} - \mathbf{T} \right) \mathbf{T}_{jj'}^T \right] \end{aligned}$$

Essentially identical steps can be taken to derive $\tilde{P}_o^{1-\sigma_c}$, $\phi_{j\neq o}^{1-\sigma_c}$ and $\phi_o^{1-\sigma_c}$.