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

Matthew Grant† and Meredith Startz‡

September 2021 [Please click for most recent version]

Abstract

Distribution of goods often involves chains of multiple intermediaries engaged in sequential buying and reselling. Why do such chains arise, and how will changes in their structure due to changes in policy or trade costs affect consumers? We show that internal economies of scale in trade costs lead to chains with multiple intermediaries, and that this suggests developing countries markets are more likely to be served via long chains. Contrary to common wisdom, cutting middlemen out can, but does not necessarily, benefit consumers, even when intermediaries hold market power. Instead, there is a fundamental tradeoff between costs and entry that means even pure reductions in trade costs can have perverse effects. 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, and Ina Simonovksa 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 a producer 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, viewing it at best as a necessary evil that drives a price wedge between producers and consumers. This is reflected in policies intended to restrict the role of middlemen, ranging from prohibition (such as Bangladesh's 2011 ban on delivery order traders) to limits on where agents can buy and sell (such as India's regulated agricultural marketplaces). 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 economies of scale in trade costs — whether they occur in transport, search, or any of the many other functions that middlemen serve — offer a general microfoundation for the existence of multi-intermediary chains. This allows us to consider how chain structure will respond endogenously to changes in policies and costs, and how this in turn will influence 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 in local markets, and there is a fundamental tradeoff between these two forces. Policy makers may prefer either more or less intermediation than what is provided by the market, and chains are subject to second best considerations that mean even pure reductions in trade costs can have perverse effects.

We begin by showing that a simple mechanism gives rise to chains of intermediation: internal economies of scale in sourcing from or selling to a particular market. If a good produced in one location may be bought and resold without further transformation, and individual buyers face scale economies in trade costs to access different 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. 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 the cost of a trip to China and port fees 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. 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. This suggests that our framework may be particularly relevant to policy in developing countries, and helpful in understanding cross-country differences in the structure of wholesale and retail distribution.

The existence of intermediaries in trade has been attributed to a variety of specific mechanisms, often focused on information problems such as search or quality control, or the provision of financing or transport services. Which of these explanations are at play surely varies across contexts. We do not argue that economies of scale are an alternative, but rather a feature of a whole class of these more specific mechanisms – the ways intermediaries actually solve these problems and the types of costs they incur in doing so often feature internal economies of scale. Examples are everywhere in wholesale and retail trade, ranging from time (e.g. spent traveling to source markets to search for products or doing research to gather information), 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). 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.[1]

This mechanism provides a framework for thinking about 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 adds a new dimension to classic questions about optimal entry when firms have market power. We consider the problem of a planner who cannot control prices or the number of firms in wholesale or retail markets, but can manipulate the menu of sourcing options to achieve second-best outcomes. The planner can, for instance, prohibit sourcing from or selling to particular markets, or tax or subsidize trade costs. Policies that shorten chains implicitly move the destination market toward sourcing strategies with lower variable and higher fixed costs, and therefore reduce the number of firms under free entry. Shorter chains benefit consumers by reducing marginal costs, but harm them by reducing entry. Depending on the set of available source markets and the features of demand and competition, the planner may prefer either a more or less direct sourcing strategy than the one selected by the market equilibrium. Because of these second best considerations, even unambiguous technology improvements – for instance, the removal of bureaucratic barriers to accessing certain

[1] Our goal is to illustrate the implications of scale economies for equilibrium chain structure and welfare, since it seems clear that these scale economies do, in fact, exist in many settings. It is possible that there are additional explanations for the existence of multi-intermediary chains, although we are not aware of any other systematic accounts of them.

markets, or introduction of platforms that reduce communication costs – can have perverse effects by shifting the market toward sourcing strategies that lower welfare. Broadly, cutting out middlemen can either help or harm consumers.[2]

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 compound 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 using data from an original survey of wholesale and retail traders in Lagos, Nigeria. These traders source goods manufactured consumer goods like apparel and electronics from suppliers all over the world, and sell to customers throughout Nigeria. They are part of long chains with multiple intermediaries: Over two-thirds of their international suppliers are upstream wholesalers rather than manufacturers, and the majority of their sales are to downstream traders rather than final consumers. We estimate the model for distribution of Chinese-made apparel throughout Nigeria. Our estimates are able to match a nuanced set of observed empirical patterns from the survey data. First, sellers source the same or similar goods from different places. Second, Nigerian consumers are served by long chains, which the model predicts are longer on average in poorer, smaller, more remote markets. 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 fixed cost of sourcing. As a consequence, the number of retail outlets carrying Chinese apparel plummets, and consumer spending shifts almost entirely to other goods. 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 marginal shift to higher

[2]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 faced by intermediaries 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 changes in trade costs and technologies.

fixed cost sourcing strategies outweighs the inframarginal gain in variable costs. While Lagos always benefits from improved e-commerce as the main port of entry, 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.

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); 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 taking the structure of the chain itself 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 buying or 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 extending these insights into new trade models (Dhingra and Morrow (2019)). In order to capture the nature of intermediation and resale, we add to this problem the choice over a menu of resale "technologies" (i.e. source markets) featuring different combinations of fixed and variable costs. We show that this introduces another margin along which social and private incentives can diverge, and that there is room to achieve second-best outcomes by using policy to constrain these sourcing options.

A theoretical literature on intermediation typically casts middlemen as solvers of information problems who mitigate information asymmetries (Biglaiser (1993)) or facilitate matching (Rubinstein and Wolinsky (1987); Antras and Costinot (2011)). We allow the existence of intermediaries to be driven by general economies of scale in trade costs, which may encompass but do not require any of these particular mechanisms. In addition to information, the combination of fixed and variable trade costs can also capture regulatory, financial, bureaucratic, and transportation costs. 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)), travel to find or inspect goods (Startz (2021)), and licensing and non-tariff barriers. A wave of recent papers study the introduction or expansion of new e-commerce or matching platforms intended to reduce information costs of trade, some of which are explicitly motivated by the desire to cut out or curb the market power of middlemen.

[TREB: Can we cut this paragraph? Feels defensive and maybe not necessary now that we've explained what we're doing better?] Our work is complementary to emerging literatures on the geography of global value chains and transportation networks. We share the language of chains and technical considerations related to modeling networks, but examine a distinct empirical phenomenon and offer a different theoretical mechanism. One branch of the value chains literature (summarized in Antras and Chor (2021)) asks why steps in a production process might take place in different locations. It generally assumes a fixed sequence of tasks and attributes chains to location- and task-specific productivity. The recent spatial transport networks literature (Allen and Arkolakis (2019); Fajgelbaum and Schaal (2020); Ganapati, Wong and Ziv (2020)) models transport as constant returns to scale from the perspective of each agent, and attributes indirectness to external economies or diseconomies of scale, through which aggregate choices may influence trade costs. In contrast, we consider intermediation chains, where links involve transactions but no transformation of goods, and there is no fixed set of tasks. We attribute chains to trade costs rather than productivity differences, and indirectness to *internal* economies of scale, so that goods do not necessarily reach consumers via the lowest cost route.

Questions about the role played by intermediaries are old, but have gained new life with the introduction of technologies and platforms for connecting consumers more directly to producers. 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, and 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, higher labor productivity, and sometimes, centrally managed supply chains that connect retail operations directly to producers. These are epitomized by "big box" stores that are part of corporate chains with national or global reach, such as Walmart. 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 increased from 32% to 50% between 1992 and 2012, but that the sector is increasingly consolidated as a result of large fixed cost investments in sourcing and distribution infrastructure.

We are not aware of data that allows for systematic comparisons of the structure of intermediation across countries. However, this qualitative contrast between rich and poor countries is broadly consistent with historical accounts of the development of agricultural markets and the structure of commerce in the United States, which describe collection and distribution chains as having first expanded, then fragmented, and finally shortened with the arrival of improved transportation and communication technologies, starting with railroads (Chandler (1977); Cronon (1991)).

2.2 Lagos Trader Survey data

We make use of an unusual data set that 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.[3]

LTS participants were identified through a census of over 50,000 shops in commercial areas[4] 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

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

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

$59,000 in goods annually. 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.

2.2.1 Purchases

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.[5] 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, trade costs, and markups, and to relate these to supplier type.

Overall, 68% of suppliers are wholesalers rather than manufacturers. This reflects substantial variation across source countries. Clothing from Benin and electronics from Dubai are unlikely to have been originally manufactured in those locations, and indeed, suppliers in those places are heavily reported as wholesalers. Even in major manufacturing locations such as China and Germany, however, a 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.

2.2.2 Sales

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.[6]

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.

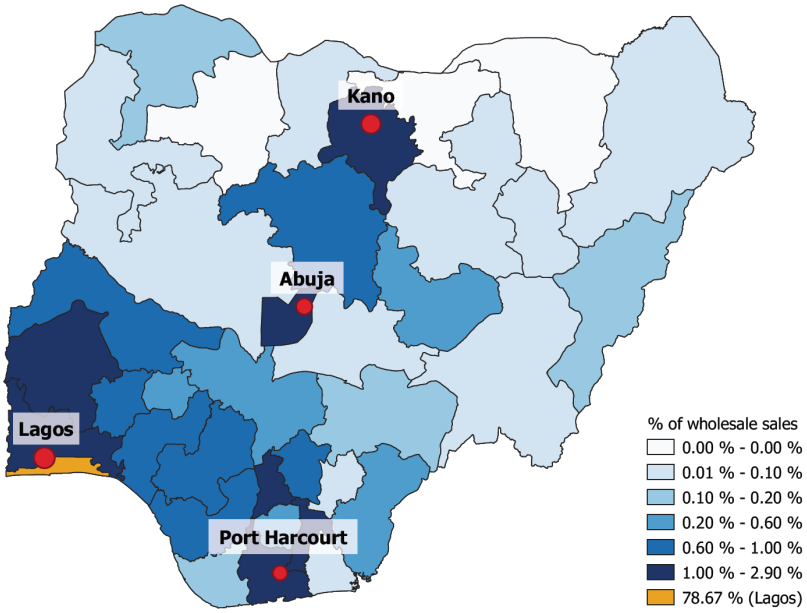
2.3 Characteristics of distribution chains

The data on purchases and sales allows us to construct a measure of the length of distribution chains, conditional on goods passing through the hands of a Lagos trader. We define purchases

[5] 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.

[6] 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.

Figure 1: Downstream sales in Nigeria



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.

We observe chains with a minimum length of three and maximum length of five. When all of a trader's suppliers are 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. We emphasize that it is *at least* five, because we are not able to observe the supplier's suppliers or the buyer's buyers. It is possible that there are more steps of intermediation on either end of the chain, and so this measure of chain length is truncated at one step away from the Lagos respondent on both ends. All results that follow therefore reflect a lower-bound measure of the length of intermediation 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 layters of intermediation in all product categories.

[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 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.

Table 2: Relationship between traders' chain position and firm characteristics

	(1) % of purchases from wholesaler	(2) Revenue ($US)	(3) Number of workers	(4) Log unit cost ($US)	(6) Log markup
% of sales that are wholesale	-0.18**	104,424.74*	0.53	-0.74**	-0.32***
	(0.09)	(56,127.23)	(0.35)	(0.31)	(0.09)
% of purchases from wholesaler		-64,470.80**	-0.44**	0.19	-0.05
		(31,835.56)	(0.20)	(0.17)	(0.05)
Log unit cost ($US)					-0.13***
					(0.01)
Obs	403	220	403	403	403
Mean	1.52	67,687	1.45	5.2	.46
Product FEs	x	x	x	x	

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

The features of traders' businesses and transactions vary systematically with their position on the chain. Column (1) of Table 2 shows that traders' upstream and downstream positions are correlated. Traders who are further upstream from final consumers – i.e. those who sell more wholesale relative to retail – are more likely to source directly from manufacturers themselves.[8] Trading firms that are more upstream from consumers are also larger in terms of both revenue and number of workers, as shown in columns (2) and (3), even controlling for the fact that they also source more directly. Those who are further downstream from manufacturers are smaller.

Table 2 also shows that traders pay different prices to their suppliers and charge different markups to their buyers based on their position in the chain. Column (4) shows that traders who are further upstream pay lower prices to their suppliers. They also charge lower markups, even when controlling

[8]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.

for the fact that they pay lower unit costs on average. Column (5) 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 they charge on average 10% lower prices in wholesale sales, given that the average markup is a little over 50%.

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. 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 a single good can be sourced either directly from a production location, or indirectly through a resale market. We work in a simple geography with a general demand specification in order to highlight the main forces involved. In the following section, we work in a more realistic geography and impose more structure on the model in order to bring it to data.

3.1 Model setup

3.1.1 Geography

We consider the distribution of a single good. This good is produced in one location, the "origin", o, and is available for purchase there at a fixed price p_o. The good is demanded by consumers in a "destination", d. The good is also available in a third location h (the "hub") at a fixed price p_h, where it is resold without transformation after having been produced in the origin.[9]

The good can be traded between any pair of locations with payment of fixed and variable trade costs. The variable cost to trade the good from location j to location i ($i, j \in \{o, h, d\}$) is a multiplicative cost, $\tau_{ij} > 1$. The fixed cost to source from location j for traders in location i is denoted $F_{ij} > 0$. Each conforms to a triangle inequality, such that $F_{ij} \leq F_{ik} + F_{kj}$ and $\tau_{ij} \leq \tau_{ik}\tau_{kj}$.[10] These trade costs can include physical transportation but also regulatory, financial, bureaucratic, information, and other types of costs associated with buying a good in one location, potentially taking it to another location, and selling it.

Let $c_{ij} = p_j\tau_{ij}$ represent the total per-unit cost of sourcing the good from location j to sell it in location i. The triangle inequality implies that the unit cost of sourcing from the hub (to the destination) is strictly greater than the unit cost of sourcing from the origin, $c_{hd} > c_{od}$, because

[9] In effect, we are assuming that the destination market is small and does not affect prices in the origin or hub. We do so only to focus attention on the main forces of interest, and endogenize wholesale pricing in the following section.

[10] This assumption is not necessary. If the triangle inequality does not always hold, then there is an additional reason that goods may be routed through third locations. Note, however, that 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, since we do not count physically indirect routing that does not involve a transaction to be a chain link.

$c_{hd} = p_o\tau_{oh}\mu_{oh}\tau_{hd} > p_o\tau_{od} = c_{od}$ (where $\mu_{oh} \geq 1$ is any resale markup in the hub). Finally, we assume that the hub is a place where $F_{hd} < F_{od}$ in order to restrict attention to cases where there is potential for indirect sourcing. Note that there may be many potential resale locations where this is not the case, but agents from the destination will never purchase there if the origin dominates on both fixed and variable costs.

3.1.2 Retail demand

The good is demanded by consumers in the destination, who purchase it from N local traders.[11] 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 common demand specifications (including linear and CES, and constant passthrough demands more broadly) with or without imperfect substitutability across sellers, and allow for individual firms to be either small or large with respect to the market.[12] Note that this allows for the possibility that consumers experience gains from seller variety even though the good itself is homogeneous – this may represent sellers' locations within a market, or other access factors that affect individual consumer's preferences over purchasing from one seller versus another.[13]

3.1.3 Traders and retail market equilibrium

Traders' business involves purchasing the good in a source market, paying trade costs, and reselling it. There is a large pool of identical potential entrants. Firms pay a fixed cost of entry $f^e > 0$ and choose a source location and quantity to maximize their profits:

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

where $j \in \{o, h\}$, p_v is a function of q_v and q_{-v}, and N is the total number of firms in the market.

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

[11] The assumption that consumers do not source directly 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 pay the fixed costs of sourcing.

[12] 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, which we use in Section 4, and all of the results continue to hold.

[13] 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.

choices is described by following conditions:

$$\frac{\partial}{\partial q_v}\pi_v\left(q_v, q_{-v}, j, N\right) = 0 \tag{3}$$

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

$$\pi\left(q, j, N\right) = 0 \tag{5}$$

where q_v' is the quantity that would be chosen by firm v if it deviated to sourcing from market j', holding all other firms' choice fixed. Clearly, there are two potential types of symmetric equilibria: "direct sourcing" equilibria in which all traders in the destination source from the origin, and "indirect sourcing" equilibria in which they source from the hub resale market.

3.2 Indirect sourcing

When will an indirect sourcing equilibrium exist? Firms' choice to source directly or indirectly is fundamentally a trade-off between fixed and variable costs. The only reason to source indirectly is if the fixed costs involved are sufficiently lower to make it worth paying the higher variable cost.

Simply, an indirect sourcing equilibrium exists when no trader wants to deviate to direct sourcing conditional on other traders' choices. 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. Firms will source indirectly when:

$$\frac{F_o - F_h}{c_h - c_o} \geq q\left[1 + \frac{1}{2}\frac{c_h - c_o}{c}\left(\varepsilon_p^q - 1\right)\rho\right] \tag{6}$$

where q and c are symmetric equilibrium per-firm quantity and unit cost, ε_p^q is the price elasticity of demand, and ρ is the passthrough rate.

6 shows that the first-order forces driving indirect sourcing are the relative cost advantage of available resale markets and equilibrium firm size in the destination. When the fixed cost advantages of resale markets are large relative to the additional variable costs incurred – or, conversely, when the fixed cost barriers to accessing production locations are high – then indirect sourcing will be more desirable. Indirect sourcing is also more likely when firm size is small. 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 sourcing strategy. These are statements about equilibrium relationships,[14] but have substantive empirical content. For instance, it predicts that, all else equal, places where we observe smaller firms are more likely to be served indirectly.[15]

[14] 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. For instance, under many specifications, a fall in entry costs reduces firm size and induces more indirect sourcing. We other parametric examples in the Appendix.

[15] In this simple geography, sourcing is either direct or indirect, and so the chain has either one intermediary or two. However, one can imagine that there are traders in the hub who face an analogous choice between sourcing from

Small firm size and high barriers to accessing production locations are characteristic of many markets in developing countries. This theory suggests that chains will be longer on average in these places. Equilibrium firm size is pinned down by fixed costs in this model, but if market failures affecting firm size are present these will also push toward indirect sourcing. For instance, if span of control problems or credit constraints are more likely to keep firms in developing countries from expanding, we should expect to see longer chains.

3.3 Welfare implications of cutting out the middleman

How are consumers affected by the possibility of sourcing goods indirectly via a resale market? Our framework allows for standard inefficiencies in the level of entry in the destination market, conditional on costs, which arise due to firms' pricing power and a potential discrepancy between socially and privately optimal levels of entry. In this section, we show that the choice between source markets can be conceptualized as a choice across "production" (resale) technologies with different levels of fixed and variable costs, and introduces another margin along which private incentives may diverge from social ones.

Destination market welfare under a symmetric equilibrium with sourcing from market j is:[16]

$$W_j = CS_j = U_j - N_j p_j q_j$$

We consider the problem of a policy maker who is concerned with welfare in the destination market, and who cannot control prices or the number of firms directly but can manipulate the menu of sourcing technologies. The policy maker can, for instance, prohibit sourcing from or selling to particular markets, or tax or subsidize fixed or variable trade costs. We show that the sourcing strategy chosen by the market equilibrium may be different from the (second-best) planner's solution, and that 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 specific source markets; it is natural to think of these as forming a discrete menu of options. However, it is expositionally useful to begin by imagining a continuous cost frontier, so that we can use derivatives to characterize the margins on which private incentives diverge from social ones. (Note, however, that this abstracts from some considerations that complicate the results in the following section when we return to the discrete setting.) We denote the variable cost along this frontier as a function of fixed cost, $c(F)$, and assume that $c'(F) < 0$ and $c''(F) > 0$.

We start by considering the difference between the planner's technology choice and that chosen

the origin and some set of resale markets. If the equilibrium in the hub also features indirect sourcing, then chains can grow to include more than two intermediaries. What kind of destinations will be served by these longer chains depends not only on their own characteristics, but on the full geography. Long chains will arise in places that are inclined to source indirectly from markets that are also inclined to source indirectly.

[16] This excludes profits, which are zero in equilibrium.

by the market. Market equilibrium is defined as in Section 3.1.3, except that due to the continuity of the cost function, 4 is now

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

In other words, every firm chooses the level of technology which minimizes its total costs, conditional on its choice of quantity. In contrast, the planner selects the technology that maximizes welfare, conditional on free entry and profit-maximizing quantities chosen by firms. A useful decomposition of the planner's first order condition is:

$$\frac{\partial W}{\partial F} = 0 = \left(\frac{\partial U}{\partial N} - pq\right)\frac{\partial N}{\partial F} + N\left(p - c\right)\frac{\partial q}{\partial F} - N\left(c'\left(F\right)q + 1\right) \tag{7}$$

The first term on the right side captures any direct gain or loss from a change in the number of (seller) varieties provided, where $\frac{\partial U}{\partial N} - pq$ is the difference between the social benefit and social cost from an additional seller. The second term corresponds to how a movement along the cost frontier changes the distortion of per-seller quantity due to market power, where markups, $p - c$, capture the difference between the social benefit and social cost of an additional unit. The final term captures the distortion of the cost minimizing technology choice of firms, $c'\left(F\right)q + 1$.

Clearly, the planner's technology choice will not generally be the same as that selected by the market, as the third term in equation (7) alone is the same as the entirety of the first order condition for firms. There is one particular case in which the planner and market solutions coincide: if the first two terms precisely counterbalance one another and net out to zero, as is the case under CES demand with monopolistic competition. Otherwise, the planner is willing to accept some distortion of cost minimization in order to counterbalance net variety and quantity distortions under the free entry equilibrium.

Will the planner's preferred technology lie above or below the market equilibrium along the cost curve? Assume the market equilibrium is an interior solution, and consider a small shift toward a higher fixed cost starting from the market equilibrium. Importantly, the planner takes as given firms' behavior conditional on the cost of sourcing, and so the free entry condition continues to hold as we move along the cost frontier. In this setting, an increase in the fixed cost level decreases the number of firms and increases per-firm quantity, i.e. $\frac{\partial N}{\partial F} < 0$ and $\frac{\partial q}{\partial F} > 0$.[17] These opposing effects on variable costs and entry are the key trade-off involved in shortening intermediation chains, and the net effect on welfare is ambiguous.

[17] Although a full proof is slightly tedious and is left for the Appendix, the key intuition comes in two parts. First, the free entry condition (and the form of utility) means that a small change in fixed cost will not affect the competitive forces affecting a given firm. Although the number of firms and per firm quantity will change, these two forces will exactly cancel out and prices will not be affected by the net changes to other firms. Second, firms optimally set quantity where marginal revenue equals marginal cost. When the fixed cost increase, marginal costs fall. And since there is no change in competition, the necessary fall in marginal revenue must come from an increase in own quantity (due to the second order condition on profits). Since quantity per firm increases it must be that the number of firms falls for the level of competitiveness to remain unaffected.

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 have either a higher or lower fixed cost than the market equilibrium technology.

Proof. See Appendix.

To gain intuition for Proposition 1, consider the decomposition of the welfare effect due to a small shift toward a higher fixed cost shown in equation (7), starting from the market equilibrium. Profits under free entry are always zero and do not enter into changes in welfare. The net effect is the sum of three terms representing the welfare effects of changes in variety, quantity per firm, and total cost per firm. The sign of the variety term $\left(\left(\frac{\partial U}{\partial N} - pq\right)\frac{\partial N}{\partial F}\right)$ is weakly negative. Under our assumption about the form of utility

$$\frac{\partial U}{\partial N} - pq = G'\left(Nf\left(q\right)\right)\left[f\left(q\right) - qf'\left(q\right)\right]$$

where $f\left(q\right) - qf'\left(q\right) \geq 0$ due to the concavity of $f\left(\cdot\right)$ and is only equal to 0 if $f''\left(q\right) = 0$, and $G'\left(\cdot\right) > 0$ by assumption. In other words, welfare is weakly increasing in seller variety. Since $\frac{\partial N}{\partial F} < 0$, an increase in fixed cost reduces the number of sellers and harms welfare. The sign of the second term $\left(N\left(p - c\right)\frac{\partial q}{\partial F}\right)$, representing the effect of the change in per firm quantity, is positive because firms charge prices that are strictly greater than marginal cost $(p > c)$, and quantity per firm increases. The third, total cost term is zero by definition when starting at the market equilibrium, because firms optimally choose F to minimize their total costs conditional on their choice of quantity.

.

The total welfare effect of a shift along the technology frontier is therefore the net of the first two effects when starting from the market equilibrium. A higher fixed cost technology reduces 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.

It follows from the mitigation of the quantity distortion that prices decrease under the higher fixed cost technology. Note, however, that this does not imply that market power is decreasing. On the contrary, entry is lower and the number of competitors is smaller. Markups may rise or fall, depending on whether passthrough is greater than or less than one.

There may be a particular interest in the case where sellers exercise market power but consumers experience no gains from seller variety per se, as in Cournot oligopoly with linear demand. In this case, the gains from variety term is zero and the planner would always like to push the market toward a higher fixed cost sourcing strategy than that chosen in the market equilibrium. This result is broadly consistent with policymakers' instinct to enact policies that shorten collection and distribution chains, although the mechanism – that cost savings will outweigh increases in markups – does not exactly accord with their reasoning, which often focuses on *reducing* the market power of

intermediaries.[18] However, we show in the following section that this does not imply that sourcing directly from the origin is optimal, and in the absence of the convenient but fictional continuous trade cost frontier, does not hold at all.

3.3.2 Discrete sourcing technology

We now return to the case where there are a finite number of potential source markets, which implies a discrete menu of sourcing technology options. Under a discrete sourcing technology menu, the change in welfare due to a shift from an equilibrium with sourcing from j to one with sourcing from j' is:

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

where the notation $\Delta_{N_j} U\left(N_j, q_{j'}\right) \equiv U\left(N_j, q_{j'}\right) - U\left(N_{j'}, q_{j'}\right)$ – i.e. this denotes a partial difference with respect to N at j versus j' holding other arguments fixed, and $\Delta_{q_j} U$ is defined analogously. This is equivalent to the decomposition in equation (7), but accounts for the fact that a change in sourcing technology will involve a discrete jump in fixed and variable costs. Consider a case in which the market equilibrium involves indirect sourcing in our three location geography. Can a policymaker increase welfare by prohibiting sourcing from the hub, efffectively forcing traders to source using a higher fixed cost and lower variable technology? The answer is "maybe".

First, we note that the destination market may not be served at all. If traders cannot source from the hub, then the only option is to source directly from the origin. However, a direct sourcing equilibrium may or may not exist – if it does not, consumer surplus goes to zero. This is the case when the potential profits for a monopolist are insufficient to cover the fixed costs of entry and origin sourcing. Several recent papers find that small, remote markets in developing countries have less product variety (Atkin and Donaldson (2015); Gunning, Krishnan and Mengistu (2018)). Our framework implies that existence of intermediation chains ameliorates this tendency. 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. Therefore, policies that cut out middlemen can harm remote consumers by eliminating this alternative channel for accessing products.

If a direct sourcing equilibrium exists, then the three margins of distortion discussed in the continuous case come into play. The ambiguity carries over from the continuous case – the sourcing strategy selected by the market equilibrium may involve a higher or lower fixed cost than that preferred by the planner. However, even if the policymaker would prefer a higher fixed cost strategy than the one selected by the market, forcing traders to use a higher fixed cost technology by prohibiting hub sourcing nonetheless may increase or decrease welfare. This is because the direct sourcing

[18] Direct sourcing reduces variable costs in part through elimination of passed through markups charged by hub intermediaries, as we show in Section 4. A more complete argument, therefore, is that shortening chains reduces the role of market power of held by intermediaries who are cut out of the chain, while increasing that of those who remain.

equilibrium may "overshoot", and feature more insufficient entry than the hub sourcing equilibrium did excess entry. The change in welfare due to the change in the quantity distortion, is still unambiguously positive, as it was in the continuous case. The third term, describing the change in the cost distortion, is negative. The net of the last two terms is ambiguous: when jumping to origin sourcing, the increase total costs may outweigh the reduction in the per firm quantity distortion. Thus, even if there are no gains from seller variety, the move to a higher fixed cost technology may help or harm consumers.

3.3.3 Technology improvements

So far, we have considered policy interventions that mandate movements along a cost frontier. However, because of the second-best considerations we show above, even changes that strictly reduce the costs of some sourcing options have ambiguous effects on welfare.

Consider a simple example, in which the market is initially in an indirect sourcing equilibrium. There is an exogenous reduction in the fixed cost of direct sourcing that is not accompanied by any other change, such as a removal of bureaucratic barriers or the introduction of a platform that facilitates communication between destination retailers and origin sellers. This cost reduction is large enough that the market equilibrium switches from indirect to direct sourcing. Even though this change decreases fixed costs *conditional on sourcing strategy*, it may increase equilibrium fixed costs through the switch from indirect to direct 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.)

Although the intuition is simplest in the discrete case, the conclusion carries over to the continuous case as well. We re-parameterize the cost frontier as $c(F,t)$ where t is an exogenous technology shifter that changes the tradeoff between fixed and variable costs faced by firms. As before, $c > 0$, $c_F < 0$, and $c_{FF} > 0$. We additionally assume $c_{Ft} < 0$: at a higher level of technology, marginal cost is lower for any given level of fixed cost, and we will assume $c_t(0,t) = 0$. This implies that $c_t \leq 0$. We now consider the effect of a small improvement in the technology level, starting from the market equilibrium at the initial technology:

$$\frac{\partial W}{\partial t} = \left(\frac{\partial U}{\partial N} - qp\right)\frac{\partial N}{\partial t} + N(p-c)\frac{\partial q}{\partial t} - N(c_F q + 1) - N c_t q$$

As before, the envelope condition implies that the third term is zero around the initial market equilibrium. The fourth term is new, and represents the change in the total resource constraint in the economy due to the technology improvement – this is unambiguously positive. The first two terms represent the "substitution effect" of the technology improvement – the change in the shape of the trade cost frontier causes the market equilibrium to move to a new sourcing strategy. The sign of these terms is both jointly and individually ambiguous (see formal discussion in the Appendix).

3.3.4 Summing up welfare considerations

Conventional wisdom in both policy and academia often assumes implicitly that shortening distribution and collection chains will help consumers and small producers in developing countries. Overall, our welfare 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 increase consumer prices by increasing markups in the destination market. (Or, in a collection chain, it can reduce producer prices by increasing monopsony power.)

Of course, consumers will not *necessarily* be hurt by such a policy, and can in fact be helped. The direction of the effect will depend on the relative strength size of the passed through cost savings versus local market consolidation. 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

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

where $\varepsilon(\omega)$ 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 $\mu_c < 1$[19], and scale parameter $\frac{\Omega_{r,i}^{\gamma-\mu_c}}{\Gamma(1-\mu_c)}$ where $\Gamma(\cdot)$ 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 private and social valuation of seller variety, as in Benassy (1996).[20] The parameter $\alpha \in \left(0, \frac{1}{1-\mu_c}\right)$[21] 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

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

[20] 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.

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

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.

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,[22] 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.[23] The single chain assumption is stylized, but generates realistic aggregate behavior

[22] 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.

[23] 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

while yielding standard markup rules and well-behaved demands.[24]

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 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.[25] 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

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.

[24] 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.

[25] 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.

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. [26] We therefore solve simultaneously for pricing and sourcing rules that account for this downstream dependence at all points on the chain.

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 $c_n(\omega) = \tau_{n+1} p_{n+1}(\omega)$ where $p_{n+1}(\omega)$ 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:

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

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

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

The firm will choose the upstream seller to maximize these profits, and since $\varepsilon(\omega)$ 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

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

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

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

where

[26] 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.

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

Note that if $\sigma_{n+1} \equiv \frac{\sigma_n - 1}{\mu_t} + \sigma_n$ and demand at step n has the hypothesized form $\mathbb{E}[q_n(p_n)] = b_n {p_n}^{-\sigma_n}$, 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:

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

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.[27] 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

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

[27] 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).

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.

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,

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

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

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.[28] We treat apparel as a single good, and take each of the 39 locations to be a single market.[29]

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.

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

[29] 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.

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:

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).[30] 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

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

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

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

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

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

and $\hat{\mathbf{T}}^T_{jj'}$ 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 approximated as

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

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

where $\mathbf{T}$, $\mathbf{T}_{jo}$, $\mathbf{T}^T_{jj'}$, $\tilde{P}_j$ 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 3 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:[31]

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

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 $A =$

[31] 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.

Table 3: 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.

$\frac{\beta_1-1}{\beta_1}\exp\left(\frac{\delta}{1-\beta_1}\right)$. 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.[32] 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

[32]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).

$d_{i(\text{Lagos})}$. With estimates of ϕ_τ and υ_τ in hand, we calculate the implied τ_{ij} between all other locations in Nigeria.[33]

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:

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

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.[34] 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.[35]

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

[33] 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).

[34] 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.

[35] 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 4: 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.

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 4 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 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.[36] 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 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.

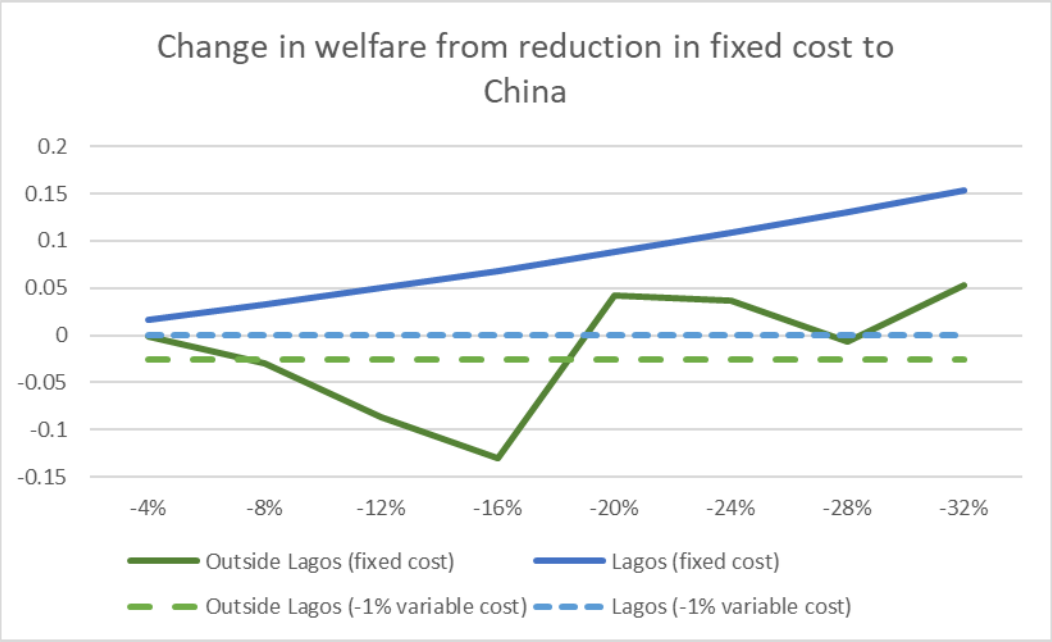
It is implicitly assumed that such technologies will benefit small producers and consumers, and in a framework with direct trade and efficient levels of firm entry, that would indeed be the case. However, in a world with intermediation chains and potentially inefficient sourcing equilibria, it is important 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

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

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



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 passed through costs. Consumers in most of Nigeria don't start to see gains until cost decreases are around 20% of the baseline level.[37]

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

[37]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.

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 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.

Allen, Treb, and Costas Arkolakis. 2019. "The Welfare Effects of Transportation Infrastructure Improvements." National Bureau of Economic Research Working Paper 25487. Series: Working Paper Series.

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.

Antras, Pol, and Davin Chor. 2021. "Global Value Chains." National Bureau of Economic Research Working Paper 28549. 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."

Chandler, Alfred. 1977. *The Visible Hand: the Managerial Revolution in American Business.* Belknap.

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.

Cronon, William. 1991. *Nature's Metropolis: Chicago and the Great West.* Norton.

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.

Fajgelbaum, Pablo D., and Edouard Schaal. 2020. "Optimal Transport Networks in Spatial Equilibrium." *Econometrica*, 88(4): 1411–1452. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA15213.

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

Ganapati, Sharat, Woan Foong Wong, and Oren Ziv. 2020. "EntrepÃŽt: Hubs, Scale, and Trade Costs." Social Science Research Network SSRN Scholarly Paper ID 3571905, Rochester, NY.

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.

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 $c(F) = e^{-F}$.[38] Firm optimal quantity choices imply $q = \frac{a-c}{2b}$ so that $c = a + \sqrt{\left(\frac{a}{2}\right)^2 + 2b}$ and

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

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

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

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

[38] 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 $\frac{\partial p_j(q_i, q_{-i}, N_j)}{\partial q_i} - \frac{\left(p_{j'} - c_{j'}\right)}{q_j}$ 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

$$W = U - N\left(cq + F + f^e\right)$$
$$\begin{aligned} \frac{\partial W}{\partial t} &= \frac{\partial W}{\partial N} \frac{\partial N}{\partial t} + \frac{\partial W}{\partial q} \frac{\partial q}{\partial t} + \frac{\partial W}{\partial F} \frac{\partial F}{\partial t} + \frac{\partial W}{\partial t} \\ &= \left(\frac{\partial U}{\partial N} - qp \right) \frac{\partial N}{\partial t} + N\left(p - c\right) \frac{\partial q}{\partial t} - N c_t q \end{aligned}$$

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

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

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 $j \neq o$

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

and by close analogy

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

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)

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

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)

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

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 $\chi_{ij} \equiv \frac{\exp\left(\frac{\pi_{ij}^r}{s}\right)}{\sum_{j'} \exp\left(\frac{\pi_{ij\prime}^r}{s}\right)}$ 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

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

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

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

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

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

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

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

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