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

Matthew Grant[†] and Meredith Startz[‡]

October 2020 [Please click for most recent version]

Abstract

Finished goods may pass along a whole chain of intermediaries on their way from producers to consumers. Using original survey data from Nigeria, we document that there are at least three separate intermediaries between an international manufacturer and a Nigerian consumer on average, and that the characteristics of these intermediaries and their transactions are systematically related to their position along the distribution chain. We build a general framework for understanding why chains with multiple intermediaries form and illustrate their implications for consumer welfare and measuring trade costs. Contrary to the common intuition, consumers can benefit from being at the end of longer chains of intermediation. Taking chains into account also suggests that existing estimates of distance costs in developing countries are biased upward, and may contribute less to consumer-producer price gaps than typically thought. We then build a quantifiable version of the general model, which relates the endogenous chain structure through which goods actually reach consumers in a particular market to fundamentals of geography and demand in many locations, and calibrate it in the context of Nigerian wholesale trade.

*We thank Treb Allen, David Atkin, Kyle Bagwell, Andy Bernard, Lorenzo Casaburi, Arun Chandrasekhar, Cecile Gaubert, 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 simply sells directly to consumers in many locations, within or across countries. The reality is that finished goods may be bought and sold many times, passing through the hands of multiple intermediaries. These chains of intermediation are particularly visible in developing countries, where small-scale traders play a very active economic role. A shopkeeper selling mobile phones in a small town in Nigeria, for instance, is unlikely to have sourced them directly from a producer in China. Instead, she might buy from a trader in a regional market town, who in turn sources from a larger wholesaler, who relies on an importer, and so on. At each step along the route toward consumers, costs are incurred and markups may be charged. Understanding distribution chains is therefore key to understanding the prices and product availability faced by consumers in different locations around the world, and their potential to gain from globalization and trade.

Wholesale and retail firms account for a large share of employment and a large fraction of imports and exports across a diverse set of countries, including China (Ahn, Khandelwal and Wei (2011)) and the Unied States (Bernard et al. (2010)). The literature on agriculture in developing countries has long recognized that crops often pass through domestic supply chains involving multiple intermediaries. However, attempts to better understand the role of intermediation in inter- and intranational trade have hampered by data limitations. Even the most detailed firm data is generally designed to capture a snapshot of transactions at a single point, and so it is difficult to know whether this is in fact one link along a longer distribution chain.

Our first contribution is to empirically document the prevalence of these chains in the context of manufactured consumer goods imported into Nigeria, using original survey data. The data comes from a survey of wholesale and retail traders sampled from a census of shops in the commercial areas of Lagos, Nigeria. Lagos importers source goods like apparel and electronics from suppliers all over the world, and sell to customers throughout Nigeria. Over two-thirds of their suppliers are upstream wholesalers rather than manufacturers, and the majority of sales are to downstream firms rather than final consumers. Conditional on passing through the hands of a trader in Lagos, we find that the typical chain involves a minimum of three intermediaries between manufacturers and consumers. We also observe the details of individual transactions, and show that both firm and transaction features vary systematically with a trader's position along the chain. Traders who are further upstream – i.e., more steps removed from final consumers – have larger businesses, pay lower unit costs to their suppliers, and charge lower markups on their own sales.

Existing empirical work typically takes the observed structure of intermediation as given, and distribution chains with multiple intermediaries do not arise in standard trade models. In order to allow for and endogenize chain formation, we start by building a general framework that departs from these models in a simple way: A good produced in one location may be bought and then *resold* in other locations without further transformation. In a world with fixed as well as variable trade costs, it is not necessarily cost-minimizing for an individual buyer 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 we might expect a rural shopkeeper in in a developing country to source goods from a wholesaler in a local market town rather than directly from China, even when the price is higher. Iterating this logic can lead to a whole chain of intermediaries who source from other intermediaries in response to fixed cost - variable cost trade-offs.

Although this mechanism is simple, it yields meaningful predictions about the structure of distribution chains, how policies will impact consumer welfare across locations, and the empirical measurement of trade costs. Contrary to conventional wisdom, we show that in theory consumers can gain from being at the end of a longer chain, even when intermediaries have market power and simply serve an aggregation role but do not add value to the good. This relates to a longstanding debate – one that is particularly heated in developing countries – about whether middlemen play a positive or negative role. We show that the net impact depends on the trade-off between higher costs incurred along longer chains versus the potential benefits of increased entry of sellers in local markets. Consumers in entrepôts (which are locations that handle large amounts of goods bound for destinations downstream) generally gain from increased downstream demand flowing through their location, while consumers in end destinations may either gain or lose from longer chains.

Accounting for intermediation chains also suggests that conventional estimation strategies overstate distance-related costs, and more so in developing countries and remote areas to the extent that they are served by longer chains. This can shed light on empirical puzzles in the literature, including why trade costs rise more steeply with distance in developing countries than can be accounted for with observables (e.g. Allen (2014); Atkin and Donaldson (2015)) and how large price wedges between consumers and producers coexist with high measured passthrough rates in many contexts (Dillon and Dambro (2017)). Broadly, our framework suggests that more thought should be put into the relative emphasis placed on policies that are intended to be procompetitive (such as regulating intermediation) versus those likely to impact variable trade costs (such as road building) versus fixed trade costs (such as visa policies, access to financial tools, or trade logistics).

In order to quantify the implications of distribution chains, we add specific functional form assumptions to our general model, and extend it to an arbitrary number of locations. This quantitative model relates the endogenous chain structure serving each location in equilibrium to fundamentals of global geography and demand. Trade involves both a fixed and a variable cost between each pair of locations. Intermediaries can enter in each location to sell to both consumers and to any downstream intermediaries who arrive to source the good from them. Final consumers make a discrete choice across intermediaries serving their location, and intermediaries in turn make discrete choices about both which location to source from and which upstream seller to buy from within that location. Indirect sourcing arises when the fixed cost advantages of buying from a resale location are sufficiently high relative to the variable cost disadvantages, which include both the additional trade costs incurred to move the good via a third location and the markups charged by intermediaries in that location. This can lead to a chain in which consumers in a given location may be served by traders who source from other traders, who source from other traders, and so on. Global equilibria need not be unique, and feature two sources of inefficiency. First, traders at each link charge markups, and so chains with multiple intermediaries feature double marginalization. Elasticities of demand compound along the chain in a way that fully incorporates the sequence of profit-maximizing decisions by downstream sellers but also maintains the tractable form of CES markups at each step. Second, an increase in market size induces entry by traders, which increases a location's appeal as a resale market, which further increases market size. These forces create agglomeration economies in trade logistics.

We simulate the model for apparel distribution chains throughout Nigeria under a variety of scenarios, using the Lagos Trader Survey data and publicly available data on consumer spending, population, and travel distances. Although we do not observe the full distribution chain outside of Lagos, a combination of data-based and calibrated parameters allow us to match a nuanced a nuanced set of empirical patterns both within and across markets. First, sellers may source the same good from different places. Second, sellers that are further downstream from the producer will have higher unit costs and lower revenue. Third, all else equal, more upstream sellers will face more elastic demand and will charge lower markups. In our calibrated model, consumer welfare in cities in Nigeria generally falls when middlemen are cut out, even though trade costs and passed through intermediary markups are lower.

We build on a small literature that shows that wholesaling and re-exporting account for a much larger share of trade value than is commonly understood. An empirical regularity that emerges from trade statistics is that buying or selling via an intermediary is more common in smaller transactions. This has motivated a number of studies that model exporting via a wholesaler as a way for manufacturers to reach a given set of consumers at lower fixed costs.[1] 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 is also at the heart of our view of intermediaries, but we allow all buyers to simply choose which market they want to source a good from. This yields an endogenous structure of aggregation points, and lets us focus on how their number, size, and market structure matters.

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 logistics, which may encompass but do not require any of the particular information mechanisms described by earlier work. In addition to information, the flexible combination of fixed and variable trade costs can also capture regulatory, financial, bureaucratic, and physical transportation costs. There is evidence that economies of scale in the latter are important across a variety of contexts – for instance, in trucking in East Africa (Teravaninthorn and Raballand (2009)) and global container shipping.

We share common ground with the emerging literatures on global value chains and production

[1] Blum, Claro and Horstmann (2009); Bernard et al. (2010); Ahn, Khandelwal and Wei (2011); Crozet, Lalanne and Poncet (2013); Akerman (2018)

networks. The value chains literature (Fally and Hillberry (2015); AntrÃ s and de Gortari (2017); Johnson and Moxnes (2019)) takes the number and sequence of required production tasks as fixed and focuses on how the location of each task is determined. In contrast, there is no fixed number of links that need to be involved in distribution – in fact, the baseline is that there need not be any intermediary links. Thus, our focus is on endogenizing the number and structure of links in the chain. Modeling in the value chains literature also generally assumes perfect competition and constant returns to scale. One of our contributions is to build fixed costs and double marginalization into a model of chains in a general geography in a tractable way. Some recent work on production networks has also made progress on these issues (Bernard, Moxnes and Ulltveit-Moe (2018); Lim (2018)), and we build on these models by allowing elasticities of demand to depend on the full sequence of downstream choices along the chain. We also share with a literature on the industrial organization of trade logistics (Brancaccio, Kalouptsidi and Papageorgiou (2017); Ganapati (2018); Wong (2018)) an interest in how actual costs of traded goods are determined in part by endogenous routing responses.

Price gaps between producers and consumers have long been a subject of interest to both policymakers and economists, particularly in developing countries. The debate is still open on exactly how large these price gaps are, and to what extent they reflect high trade costs, lack of competition, or both. Indirect estimates of total trade costs are consistently much higher than the tangible components we know how to measure (Anderson and van Wincoop (2004); Allen (2014)), and distance costs in developing countries are higher relative to those in rich countries than we are able to account for via differences in transport costs (Atkin and Donaldson (2015)). An obvious explanation for this gap is trader markups (Emran et al. (2017); Chatterjee (2018)). However, studies that focus on passthrough yield mixed results (Fafchamps and Hill (2008); Bergquist and Dinerstein (2019); Casaburi and Reed (2019)), and on balance may point toward high passthrough and competitive intermediary markets (Dillon and Dambro (2017)). We don't attempt to close this debate, but offer a framework for thinking about how both markups and realized trade costs arise from the endogenous structure of intermediation, and can compound along a chain. We show that even frontier methods for estimating trade costs may be biased upward if they fail to take the structure of the chain into account. This can help reconcile the apparent conflicts within the body of existing evidence – particularly the paucity of evidence for sufficiently high market power at any one link to explain the difference between directly and indirectly measured trade costs.

2 Distribution chains in Nigeria

Economists' empirical understanding of intermediaries is generally hampered by data limitations.[2] Although data on distribution chains are rare, we make use of an unusual data set that

[2] As more VAT data becomes available to researchers, especially from developing countries where we expect long intermediation chains to be more common, this may be a promising avenue for future exploration. Most currently available VAT data we are aware of have some combination of limitations that make them non-ideal for investigating the questions we pursue in this paper, such as lacking prices or product identifiers, challenges in distinuishing production from intermediation, missing small firms or small transactions, and not tracing flows across country borders.

captures part of the chain for consumer goods imported into Nigeria. The Lagos Trader Survey (LTS) is a panel survey of wholesale and retail traders dealing in consumer goods 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 of Lagos conducted in 2014 and 2015. The listing focuses on commercial and wholesaling areas of the city, and does not include most residential or manufacturing areas or traditional food markets. The initial round of the survey includes 1,179 traders whose shops were randomly sampled from the census, of whom 620 had imported goods in the previous two years. The sample includes any trader dealing in manufactured consumer goods (excluding food), which we group products into six categories: apparel (which includes shoes, bags, and textiles), electronics, toiletries and beauty products, hardware, home goods, and miscellanous other products.

The LTS data captures Lagos traders' purchases from suppliers at the transaction level. Crucially for the purposes of analyzing intermediation chains, the data identifies where suppliers are located and whether those suppliers are manufacturers or wholesalers.[4] 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 charged to buyers. This enables us to contruct measures of unit costs and markups, and to relate these to supplier type. We also observe the fraction of traders' sales that are retail versus wholesale, and, starting from 2016, whether they have sales to locations outside of Lagos.

2.1 Purchases in the LTS data

Imports in the LTS data cover a wide variety of products and are sourced from more than thirty different countries. Figure 1 shows the most common source countries. The largest source is China, but the United Arab Emirates (specifically, Dubai), Turkey, Hong Kong, Benin, India, and perhaps surprisingly, the United States and United Kingdom are also common.

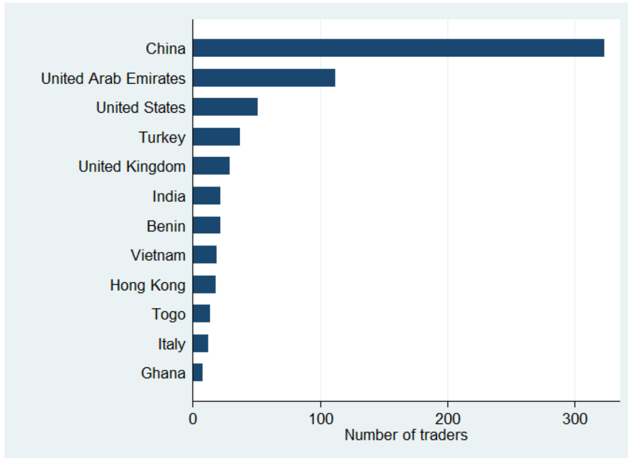
A number of these source countries appear to serve as entrepôts—clothing from Benin and electronics from Dubai are unlikely to have been originally manufactured in those locations. Indeed, overall 68% of suppliers are wholesalers rather than manufacturers. This reflects substantial variation across countries, as shown in Table 1. Nearly all West African suppliers are other wholesalers, while buying directly from a manufacturer is much more common in Europe. 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 trade statistics that a large fraction of exports in many countries are sold by wholesalers.

Table 3 shows that traders who make a larger fraction of their purchases from wholesalers (rather

[3]More details about the Lagos Trader Survey can be found in Startz (2018).

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

Figure 1: Top source countries



than directly from manufacturers) are on average smaller, in terms of both their annual revenue, and the number of workers they employ. They also pay higher average prices to their suppliers, although the coefficient in column (3) is not statistically significant.

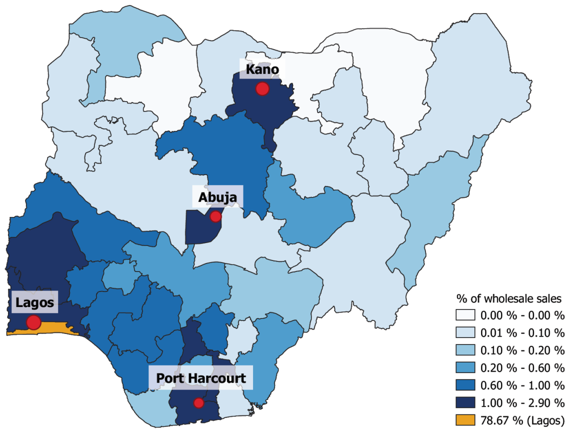
2.2 Sales in the LTS data

In total, 86% of traders in the LTS are wholesalers (and 94% of importers), and 52% of observed sales are wholesale. Unsurprisingly, traders who are more heavily involved in wholesaling have larger businesses in terms of both revenue and number of employees, as shown in Table 4, below. There is also clear evidence of price discrimination: 95% of wholesalers report that they set different prices for wholesale buyers than retail buyers, and that on average wholesale prices are 9% lower.

Starting in the second round of the LTS, we are also able to observe the destination of wholesale sales. Approximately three-quarters are within Lagos State, but the remainder go to other destinations, consistent with Lagos' role as the main port and commercial hub for all of Nigeria.[5] Figure 2 shows the fraction of traders who report sales to each state in Nigeria. Many Lagos traders sell in geographically nearby states in southern Nigeria. However, a large fraction also sell to downstream traders in much further locations that plausibly serve as commercial hubs for the central and northern parts of the country. Aside from the Lagos hinterland in the southwest, the largest flows go to states that contain Abuja, which is the centrally located national capital; Kano in the north, which has historically served as a hub for trans-Saharan trade routes; and Port Harcourt in the southeast, which contains the country's other major port and is the center of the oil business.

[5]Note also that this is a lower bound on the fraction of imports that come through Lagos that are bound for another destination within the country, since many wholesalers who only buy domestically (presumably many of them from these importers) also sell to other states.

Figure 2: Downstream sales in Nigeria



2.3 Defining chain length

The data on purchases and sales allows us to construct a truncated measure of the length of distribution chains, conditional on goods passing through the hands of a Lagos trader. We define purchases from manufacturers as one end point on the chain, and retail sales to final consumers as the other. Purchases from or sales to other traders are additional points of intermediation. 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 the wholesalers 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 we consider this to be a measure of chain length that 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.

Supplier identity is observed at the level of the individual transaction for imports, but buyer information is only observed at the trader level, reported as fractions of total sales. For a clothing trader who buys both from a producer in Turkey and a wholesaler in Dubai, and sells both to consumers in Lagos and to wholesalers in Ogun State in Nigeria, we cannot observe whether the products imported from Dubai are mainly sold to Ogun wholesalers and those from Turkey to Lagos consumers. Instead, we assume that purchases from all sources are distributed proportionally across all types of buyers. This lets us construct a value-weighted average chain length at the source country - trader level. Table 2 shows average chain length for importers overall, and separately for each product category.

In addition to chain length, we also construct measures of "upstreamness" and "downstreamness". "Steps upstream" are the average number of agents on the chain above the Lagos trader (e.g. a manufacturer and a wholesaler is two steps), and "steps downstream" are the number of agents on the chain below (e.g. a retailer and a final consumer is two steps). Although these measures are truncated due to the same data limitations as the chain length measure, they are useful for distinguishing whether a trader with average chain length four is buying directly but selling to wholesalers, buying from wholesalers but selling directly to consumers, or some of each.

2.4 Position on the chain

The features of traders' businesses and transactions vary systematically with their position on the chain, both upstream and downstream. Column (1) of Table 4 shows that traders' upstream and downstream positions are correlated. Traders who serve more downstream demand – i.e. those who sell more wholesale relative to retail – are more likely to source directly from manufacturers themselves. Traders who sell more wholesale 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 (which is also associated with larger size, in Table 3).

Table 4 also shows that traders pay different costs to their suppliers and charge different markups on those costs to their buyers based on their position in the chain. Traders who sell more wholesale pay lower unit costs to their suppliers, as do traders who source more directly. They also charge lower markups, a fact which is magnified in column (6) when controlling for the fact that they pay lower unit costs on average, since lower costs are otherwise associated with higher markups. Column (6) implies that going from entirely retail sales to entirely wholesale sales is associated with a roughly one-third reduction in average markups. This is large, but the magnitude is consistent with traders' self-reports that they charge 10% lower prices in wholesale sales, given that the average markup is just under 50%. The theoretical framework that follows will attempt to capture these patterns associated with chain position.

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 resale markets. We work in a simple geography with a relatively general setup, which is intended to highlight the main forces involved. This setup is not necessarily tractable or quantifiable in a general geography, and so in the following section we impose more structure on the model in order to bring it to data.

3.1 Model

Geography

We consider the distribution of a single good, in a simple geography with only three locations. This good is produced in one location, which we refer to as the "origin", o, and is available for purchase there at a fixed price p_o. The good is demanded by consumers in two other locations, h and d.

Trade costs

It 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 modeled as a multiplicative cost, τ_{ij}. The fixed cost to source from location j for traders in location i is denoted F_{ij}. Each type of trade cost conforms to a triangle inequality, such that $F_{ij} \leq F_{ik} + F_{kj}$ and $\tau_{ij} \leq \tau_{ik}\tau_{kj}$ (i.e. it is not more costly to trade the good directly between two locations than indirectly through a third location).[6] These trade costs can include physical transportation costs but also regulatory, financial, bureaucratic, information, marketing, and other types of costs associated with buying a good in one location, potentially taking it to another location, and selling it.

Resale markets

In each location, the good may be sold by a set of traders who number Ω_i, and who can individually be either large or small with respect to the market. We refer interchangeably to traders and intermediaries, who are agents who buy and sell the good but do transform it.[7] Traders can only sell in a single location[8], but may source from any location where the good is sold. We assume that if consumers buy the good, they must buy it from traders serving their location.[9]

Traders potentially face two types of demand – from local consumers in their home location, and from other traders who arrive there to purchase the good for resale. We allow traders to price discriminate, so that they can set a retail price for local consumers and a wholesale price for downstream traders. We assume that traders are identical, and focus on a symmetric equilibrium so that we can express the retail and wholesale prices charged by each trader in location i as p_i^r and p_i^w.

[6] This assumption is not necessary, but simplifies the exposition. If the triangle inequality does not always hold, then there is an additional reason that goods may be routed through third locations. Our point is that indirect trade can be explained even under this more restrictive assumption. Note, however, that most cases that might initially appear to be violations of the triangle inequality (e.g. road networks that pass through other locations, or shipping routes that involve unlading in a third port en route) 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 an additional transaction point to be an intermediate link on a chain. One can also imagine true violations, however, such as avoidance of policy barriers that requires a party in another country to take custody, or legal or linguistic advantages in a third location that cannot be provided at arms length.

[7] In reality, brokers may play an important intermediation role in many places, earning a commission for matching a buyer and seller or facilitating a transaction without ever taking ownership of the goods. We allow broker fees to fall under the fixed and variable trade costs that may be incurred by traders, but do not consider them as a separate step in the chain.

[8] This assumption rules out integration to serve consumers in multiple locations, which may be relevant for sourcing patterns. This is a reasonable simplication in the context of Nigerian consumer goods, as the vast majority of Lagos traders have only one shop, and even multiple shops owned by the same trader tend to be located in the same city (or even the same market). However, this may not be equally innocuous in all contexts, and is a subject for future work.

[9] We make this assumption for simplicity, but note that this is equivalent to assuming that parameters take values such that it will never be cost-effective for a single consumer to pay the fixed costs of sourcing. Alternatively, we could allow consumers to source directly if they prefer that to purchasing from an available local trader; this complicates the set of potential equilibria but does not otherwise change our conclusions.

(We use p^r_{iv} and p^w_{iv} to denote prices chosen by trader v when taking derivatives, but $p^r_{iv} = p^r_i \; \forall v$ in equilibrium.) Each trader will therefore earn profits when selling in i and sourcing from j of $\pi_{ijv} \equiv (p^r_{iv} - c_{ij}) \, q^r_{iv} + (p^w_{iv} - c_{ij}) \, q^w_{iv} - F_{ij}$, where $c_{ij} = p^w_j \tau_{ij}$ is the marginal cost of sourcing from j for location i and q^r_{iv} and q^w_{iv} are the quantities sold to consumers and other traders, respectively, by trader v (again, $q^r_{iv} = q^r_i \; \forall v$ in equilibrium).

Because traders are identical, there will be two potential types of symmetric equilibria: ones in which all traders in both downstream location source directly from the origin, and ones in which traders in one downstream location source from traders in the other, who in turn source from the origin. Without loss of generality, we will denote the location that sources directly in the second type of equilibrium as h (for the "hub") and the one that sources indirectly as d (for the "destination").

3.1.1 Retail demand

Consumers in location $i \in \{h, d\}$ have indirect utility over consumption of the good $V_i(\mathbf{p}^r_i, Y_i)$, where $\mathbf{p}^r_i$ denotes a vector of retail prices charged by traders in location i. This yields consumer demand $q_{iv}(\mathbf{p}^r_i, Y_i)$ faced by each trader. Note that this allows for consumers to experience gains from variety even though the good itself is homogeneous – this may incorporate location, information or other access factors that affect individual consumer's costs or preferences over other aspects of the purchasing experience.

The retail price will be chosen to maximize traders' variable profits from retail sales and will conform to the standard markup rule, so that $p^r_i = \frac{\eta^r_i}{\eta^r_i - 1} c_{ij}$ where $\eta^r_{iv}(\mathbf{p}^r_i, Y_i)$ is the elasticity of q^r_{iv} with respect to the retail price charged by that trader, p^r_{iv}. Symmetric equilibrium implies (again) $p^r_{iv} = p^r_i \; \forall v$, $\eta^r_{iv} = \eta^r_i \; \forall v$, and $q^r_{iv} = q^r_i \; \forall v$.

3.1.2 Wholesale demand

Traders may also serve downstream demand from traders in the other non-origin location. In symmetric equilibria with identical traders, it will never be the case that traders who source indirectly will themselves also face positive wholesale demand, and so as noted above we can think of this as traders in h facing wholesale demand from d but not vice versa. We model a general discrete choice framework to pair selling and buying traders within the market, so that each downstream buyer from d will each match with a particular seller in h with probability $\phi_{hd}(\mathbf{p}^w_h)$, where $\mathbf{p}^w_h$ is a vector of wholesale prices in location h.[10] A trader in h will sell downstream quantity $q^w_h = \phi_{hd} \Omega_{hd} q^r_d$, where Ω_{hd} is the number of traders in d sourcing from h. In a direct sourcing equilibrium Ω_{hd} will be zero, and therefore wholesale demand in h will also be zero.

This permits the price elasticity of sales to downstream sellers to be decomposed into elasticities

[10] We are agnostic here about what form this matching function takes and whether it enters traders' objective function or not. This is irrelevant for welfare considerations (taking the set of traders in equilibrium as given) since traders enter up to zero expected profits.

of matching and pass-through of downstream elasticity:

$$\frac{\partial q^w_{hv}}{\partial p^w_{hv}} = \left(\frac{\partial \phi_{hd}}{\partial p^w_{hv}} \Omega_{hd} q^r_d + \phi_{hd} \frac{\partial \Omega_{hd}}{\partial p^w_{hv}} q^r_d + \phi_{hd} \Omega_{hd} \frac{\partial q^r_d}{\partial p^w_{hv}} \right)$$
$$= \frac{q^w_{hv}}{p^w_{hv}} \left(\varepsilon^{\phi}_{p^w_{hv}} + \varepsilon^{\Omega}_{p^w_{hv}} + \eta^r_d \chi^r_d \right)$$

where $\varepsilon^{\phi}_{p^w_{hv}}$ is the elasticity of the matching rate with respect to wholesale price charged by a given trader, $\varepsilon^{\Omega}_{p^w_{hv}}$ is the elasticity of the number of traders in d who source from h with respect to the wholesale price charged by a given trader, and χ^r_d is the (symmetric) pass-through rate to the retail price in d. The elasticity of downstream sales for agents in h is therefore:

$$\eta^w_h = \varepsilon^{\phi}_{p^w_{hv}} + \varepsilon^{\Omega}_{p^w_{hv}} + \eta^r_d \chi^r_d \tag{1}$$

This expression says that the elasticity perceived by traders selling wholesale is the sum of the matching elasticity, the elasticity of the number of downstream traders who arrive in the market, and the downstream elasticity multiplied by the pass-through ratio.

3.1.3 Equilibrium

Traders must pay a fixed cost to enter in location i, f_i, and will enter up to zero expected profits. They then choose which location to source from to maximize their profits. We assume that the good is available from the origin in unlimited supply at a constant price, and that consumers in i are endowed with exogenous income, Y_i. In equilibrium, consumers maximize utility, traders maximize profits (in their entry, sourcing and pricing decisions), and goods markets clear. Given our h and d notation, functionally traders in h always source from the origin in either type of equilibrium. In direct sourcing equilibrium, traders in d also source from the origin, while in an indirect sourcing equilibrium, they source from h.

3.2 Implications

The choice of traders in d to source directly from o or indirectly via h is fundamentally a trade-off between fixed and variable costs. If the triangle inequality holds, then the variable cost of purchasing the good in h and bringing it home is always weakly higher than the variable cost of sourcing directly.[11] The only reason to source indirectly is if the fixed costs involved are sufficiently lower to make it worth paying the higher variable cost.[12]

There are a few points worth noting about an indirect sourcing equilibrium. First, and most obviously, there is more "trade" between the origin and destination d than there appears to be – d consumes a good produced in the origin but there are no trade flows between the two locations. The

[11] The triangle inequality implies that $\tau_{od} \leq \tau_{oh}\tau_{hd}$, and traders in h won't sell the good below cost, so $p^w_h \geq \tau_{oh} p_o$. Therefore, $\tau_{hd} p^w_h \geq \tau_{od} p_o$.

[12] Of course, it could be that the fixed costs of sourcing from h are also higher. In that case, we should not expect to see an indirect sourcing equilibrium.

flip side of this is that there is trade between h and d, even though h does not produce anything. Because of this, the gap in prices between the origin and d cannot be interpreted as reflecting trade costs between those locations, or even a combination of trade costs and markups. We will return to this point later.

Second, as Equation 1 shows, wholesale demand and pricing will reflect downstream retail demand. Traders who serve more elastic consumers will themselves be more elastic in purchasing, and changes in elasticity get "passed up" the distribution chain into upstream pricing. If traders are more price sensitive in choosing sellers than the consumers they serve (i.e. if $\varepsilon^{\phi}_{p^w_{hv}} + \varepsilon^{\Omega}_{p^w_{hv}} > \eta^r_d$), then trivially Equation 1 says that the total wholesale elasticity will be higher than the downstream retail elasticity. If passthrough is complete (i.e. $\chi^r_d = 1$), then wholesale sales will also be weakly more elastic than retail. If, however, passthrough is incomplete and traders' are imperfectly price elastic in matching to sellers, then the relative elasticity can be ambiguous.

To get at the welfare implications of indirect sourcing, we can compare what happens in both h and d under an indirect sourcing equilibrium to what would happen if we exogenously disallowed indirect sourcing (i.e. only allow traders to source from the origin, but keep all other features of the economy the same). Consumers in the hub h unambiguously lose when indirect demand stops flowing through their location, because the decrease in total demand lowers expected profits and therefore must reduce the number of active traders, Ω_h, under a zero profit condition. This may lower consumer welfare both through an increase in markups (i.e., if $\frac{\partial \eta^r_h(p^r_h, \Omega_h)}{\partial \Omega_h} > 0$, where in a slight abuse of notation where we denote retail elasticity as a function of the symmetric retail price and measure of traders $\eta^r_h(p^r_h, \Omega_h)$,) and through a decrease in seller variety, which consumers may value per se (i.e., if $\frac{\partial V_h(p^r_h, \Omega_h, Y_h)}{\partial \Omega_h} > 0$ where we define the indirect utility as a function of the symmetric retail price and measure of traders $V_h(p^r_h, \Omega_h, Y_h)$).

In contrast, indirect sourcing has ambiguous welfare effects for consumers in the destination d. Again comparing an indirect sourcing equilibrium to the equilibrium that arises under the same fundamentals but with indirect sourcing disallowed, there are two potentially countervailing effects. First, the number of traders Ω_d is lower under direct than indirect sourcing, as we show in the Appendix. To the extent that there are pro-competitive effects of entry on prices or consumers value having a variety of sellers per se, then direct sourcing hurts final consumers. However, direct sourcing always carries lower marginal costs, because it involves lower variable trade costs and does not include any markup passed through from the hub under indirect sourcing. The difference in consumer prices in the destination under indirect versus direct sourcing can therefore be decomposed into:

$$\ln p_d - \ln \hat{p}_d = \underbrace{(\ln \tau_{oh} + \ln \tau_{hd} - \ln \tau_{od})}_{\Delta Var.\ Trade\ Cost} + \underbrace{\ln \frac{\eta^w_h}{\eta^w_h - 1}}_{Hub\ Markup} - \underbrace{\left(\ln \frac{\hat{\eta}^r_d}{\hat{\eta}^r_d - 1} - \ln \frac{\eta^r_d}{\eta^r_d - 1} \right)}_{\Delta Home\ Markup} \tag{2}$$

where "hats" denote the equilibrium object under direct sourcing. Depending on whether decreased costs or increased destination markups wins out, consumer prices in d may be higher or lower under direct compared to indirect sourcing. Any additional welfare effects due to differences in seller variety will weigh in favor of indirect sourcing, since these always go in the same direction as the pro-competitive effect on markups.

These basic forces also hold true in a variety of settings – the discrete choice framework is not necessary, and the same forces also arise with additive rather than multiplicative trade costs and markups, or quantity-setting rather than price-setting. To make these points, we've relied on a simple geography, and abstracted from more complicated equilibria and routing patterns. To move toward a fully quantifiable model, we will allow for a general geography and the full set of possible chains in equilibrium (including chains of infinite length and with two-way trade). In order to maintain tractability and make things quantifiable, we will require a more restrictive specification. This will maintain all the intuition of the simple model, but will rely on particular specifications of demand and idiosyncratic match factors, as well as killing off incomplete passthrough and true pro-competitve price effects (as opposed to variety gains, which operate in the same way).

3.3 Measuring trade costs

With the exception of a small number of recent studies that experimentally induce price variation at a specific point in a chain (Bergquist and Dinerstein (2019); Casaburi and Reed (2019)), most empirical studies of passthrough and trade costs rely on comparisons of prices across locations. Even with panel data on prices and exogenous cost shocks, drawing inferences about trade costs from spatial price gaps is plagued by a variety of problems, including, as highlighted in Atkin and Donaldson (2015), heterogeneity in product quality or characteristics across locations, inclusion of locations that are not actually trading pairs, and the potential for markups to vary across locations in ways that are correlated with distance.

To that list, we add a fourth problem: goods that originate in one location and are consumed in another may be traded via multiple intermediary links. When this is the case, trade costs may be over-estimated and the trading sector in remote locations may be more competitive than it appears. If chains with multiple levels of intermediation are more common in remote areas and in developing countries, it can reconcile three empirical regularities that the literature has struggled to bring together: 1) large price wedges between producers and consumers, 2) relatively high passthrough at specific links on a chain, and 3) lower measured trade costs than are implied by indirect estimates. All of these facts make sense and can be reconciled with one another in the presence of intermediation chains.

The last part of the puzzle – a gap between measured and estimated trade costs – takes two forms. First, that the variable trade cost implied by the flow of goods between a given pair of locations is higher than observable components of this trade cost, as in Allen (2014). A second, slightly different version is that in Atkin and Donaldson (2015), which finds that the of slope of estimated trade costs with respect to distance is steeper in Ethiopia and Nigeria than it is in the United States. The difference in slope remains even after accounting for differences in road network, travel times, and actual measured costs of trucking across the countries. Intermediation chains offer three possible explanations for these observations.

The first part of the explanation is mechanical: if goods sometimes move from a production location to a consumption location via other locations, then the distance the goods actually travel is greater than the distance between the production and consumption points. Actual distance will be weakly greater than measured distance, even if the measure accounts for potentially indirect travel via e.g. the actual road network. It is important to note that this explanation does not mean that actual trade costs as incurred are lower than implied trade costs, but it does imply that both actual and implied trade costs will be higher than trade costs calculated by relating observables to the shortest distance or route between two locations.

A second reason that estimated trade costs may be higher than both actual and measured trade costs has to do with the way trade flows are typically measured. Imports and exports (or flows between locations within a country) generally capture only the sending and receiving location, and not the point of original production or other upstream or downstream locations along a chain if one exists.[13] To the extent that locations along a chain are on average closer to one another than the origin and final destination points are, standard data will show that there is more trade between nearby places and less between places that are more distant from one another. Any gravity-style estimation framework will therefore imply that, all else equal, the trade cost between the nearby places must be low relative to the trade cost between the more distant places. In the model of chains presented here, this is need not be true – variable trade costs could be the same between the origin and d as they are between h and d (or even lower), and equilibrium routing might still have the vast majority of flows going from the origin to d via h, depending on the fixed costs, market size, and (in the full model below) the geography of the rest of the world. We should expect gravity models that assume direct trade to therefore systematically overestimate trade costs, and more so in places where chains are more common.

Third, even frontier methods that don't rely on gravity should be expected to overestimate distance costs in the presence of chains. Atkin and Donaldson (2015), for instance, measure passthrough from an origin to a final destination using shocks to origin prices. They then use the estimated passthrough rate to purge price gaps of spatially varying markups so that the remainder can be attributed to trade costs. However, if there are intermediation points between the origin and destination locations, the estimated passthrough rate (ρ_{od}) in this framework will actually be the compounded passthrough rate from the origin to the intermediation location and the intermediation location to the destination ($\rho_{oh}\rho_{hd}$). Since some trade costs are incurred on the second leg and only subject to the second part of the compound passthrough rate and not the total, price gaps will be overcorrected and total trade costs overestimated relative to markups.

At the same time, in this type of framework, the competitiveness of the destination market will be underestimated. This is true in spite of the fact that trade costs are overestimated, and that therefore a larger part of the total price gap must be due to markups, for two reasons. First, if the competitiveness of the destination market is increasing in the estimated passthrough rate (as is true

[13] Young (1999) makes a version of this point, writing that if goods reach their destination via intermediary points where their prices change, this "suggests that our existing conceptual and practical criteria for defining origins and destinations, all of which are central to the testing and development of trade theory, are problematic".

in the Atkin and Donaldson (2015) framework), then if passthrough is incomplete the estimated compound passthrough rate will be lower than the actual passthrough rate in the destination, implying that the market there is less competitive. Another way to think about this is that the role of markups in the total price gap is not due to markups in the destination alone, but the sum of the markups at each step on the route. Second, part of the price gap reflects double marginalization, which means that the market power of any particular intermediary along the route is lower than would be implied by those markups under integration. Unobserved intermediation chains can therefore simultaneously explain why estimated trade costs are higher than we are able to account for and at the same time why the measured passthrough rate at a particular point in the chain might be quite high (as many studies have found) even if the total price gap between two locations is large.

4 A Many Location Model

In order to quantify the implications of the intermediation chains for consumers, we want to build the basic insights laid out in Section 3 into a more complete equilibrium model with a general geography. There are two main technical challenges to accomplishing this. The first is that moving from a three-location setting to an arbitrary number of locations makes solving for the global equilibrium a potentially intractable combinatorial optimization problem – a familiar roadbloack in both the global value chains literature and the trade literature more generally. To get around this, we will set up a framework that allows us to guarantee positive trade flows between all locations in equilibrium, by assuming a continuum of traders in each location who make heterogeneous sourcing choices, and choosing a demand specification that does not have choke prices. This allows the equilibrium to be characterized using first-order conditions.

The second technical challenge is that the simple retail versus wholesale price discrimination from Section 3 is not sufficient to guarantee a pure-strategy within-market pricing equilibrium when chains can pass through more than one point of intermediation. In the general geography, there will be 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, therefore, she will expect the composition of traders who choose to buy from to her to shift. In contrast to the typical intuition that retail buyers will be more elastic at higher prices (following Marshall's Second Law of Demand), in general we should expect a seller to serve relatively more of the inelastic types of wholesale buyers when her price increases, which encourages further price increases. To address this problem, we will allow sellers to price discriminate at the chain level, and buyers to choose the optimal seller for each chain they serve. This introduces a separability in pricing and purchasing decisions across chains, which ensures that a pure-strategy equilibrium exists within a market.

The model that follows builds on the same basic infrastructure as Section 3, but has a more complex set of equilibria and chains between all locations as a result of the larger geography. We choose particular specifications for consumer demand and matching between buying and selling traders that allow the model to be quantified and to be tractable in the presence of the technical challenges just described. This rules out some of the forces introduced in Section 3 – for instance, with a continuum of small traders, there are no pro-competitive effects of entry per se – but the overall conclusions remain the same.

4.1 Environment

We still model the distribution of a single good, which is produced in the "origin", o, and is available there in unlimited supply at a fixed 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 $i \in \{1, \ldots, J\}$. Each location has a fixed measure of consumers, Θ_i, with fixed expenditure on the good, denoted Y_i^c in the aggregate. The good is homogeneous, and routing does not affect the characteristics of the good.

Although the good must at some point be purchased at the origin location, it can also be "re-sold" at other locations. As in the framework from Section 3, it can be traded between any pair of locations i and j (including the origin) with payment of both fixed (F_{ij} and variable (τ_{ij}) trade costs (to move the good from location j to location i), which may but need not conform to triangle inequalities $F_{ij} \leq F_{ik} + F_{kj}$ and $\tau_{ij} \leq \tau_{ik}\tau_{kj}$.

In each location, the good may be sold by a measure of traders, Ω_i, the size of which is endogenously determined in equilibrium. In contrast to Section 3, this now implies that traders necessarily are small relative to the market. Individual traders are indexed by ω in their role as sellers and ψ in their role as buyers, and buy and sell the good without transforming it. Traders can only sell in a single location, but may source from any location where the good is sold if they pay the relevant trade costs to bring the good to their home location. In order to enter to sell the good in market i , a potential trader must pay a fixed cost of entry. After paying this cost the trader will draw characteristics described below.

The optimization decisions of traders are the focus of the model. Sequentially, traders first decide whether or not to enter the market, then choose which market j to source the good from. Next, the trader will choose which individual seller (also a trader) in location j to purchase from. Then, the trader will choose what price(s) to sell the good at in the home market.

4.2 Demand

As in Section 3, the trader faces two types of demand: first, demand from local final consumers; and second, demand from downstream traders who will re-sell the good. In this section, we begin by describing both types of demand at a single location for a single chain, and abstract from the fact that traders will serve multiple chains. We introduce chain-specific notation in the following section.

4.2.1 Consumer demand

Consumers receive utility from consuming the good, and have fixed per capita expenditure on it, $y = \frac{Y^C}{\Theta}$. We assume that if consumers buy the good, they must buy it from traders serving their location.[14] Each consumer makes a discrete choice about which trader in the measure Ω to purchase the good from. The utility associated with purchasing the good from trader ω is

$$u = \ln y - \ln p(\omega) + \mu\varepsilon(\omega)$$

where $p(\omega)$ is the price charged to final consumers by trader ω, $\varepsilon(\omega)$ is a random variable representing the consumer's idiosyncratic taste for purchasing from trader ω, and $\mu > 0$ is a parameter defining the strength of consumer preferences across sellers. We assume that ε is an independently, identically distributed standard Gumbel, so that the probability of choosing each seller is a multinomial logit. This implies that the aggregate local consumer demand faced by each trader is $x(\omega) = \left(\frac{p(\omega)^{-\frac{1}{\mu}-1}}{\int_\Omega p(\omega')^{-\frac{1}{\mu}} d\omega'}\right) Y^c$. This is equivalent to the demand generated by a CES utility function (as noted in Anderson, De Palma and Thisse (1987)) with elasticity of substitution $\sigma_c = \frac{1}{\mu} + 1$.

4.2.2 Intermediary demand

Demand from downstream traders is somewhat more complicated. We consider a trader at an arbitrary step on a chain buying from a trader upstream in that chain. We assume the trader has a payout from buying from selling trader ω given by

$$v(\omega) = \exp\left(\mu^t \zeta(\omega)\right) \pi(\omega)$$

where $\pi(\omega)$ is the variable profits for the buying from matching with seller ω, $\zeta(\omega)$ is independently and identically distributed following a Gumbel distribution with location parameter $-\frac{1}{\mu^t} \ln\left(\Gamma\left(1-\mu^t\right)\right)$ (where Γ is the gamma function) and scale parameter 1, and $0 < \mu^t < 1$ is a parameter defining the strength of preferences across different potential sellers ω.[15]

We show in the Appendix that when demand by final consumers has a CES form (perhaps arising from a discrete choice microfoundation, as in this paper), then all traders in every chain face CES demand. However, the elasticity of demand faced by sellers is higher as the seller is further up the chain. In particular, if a given set of traders face elasticity of demand σ, then their suppliers will face elasticity of demand $\tilde{\sigma} = \sigma\left(1 + \frac{1}{\mu}\right) - \frac{1}{\mu^t}$. The intuition behind this result is that the elasticity of demand faced by a given trader can be decomposed into two parts: a change in the quantity demanded by a matching downstream trader and a change in the matching rate. The change in quantity demanded is just the elasticity of demand faced by downstream traders, so that the match

[14] We make this assumption for simplicity, but note that this is equivalent to assuming that parameters take values such that it will never be cost-effective for a single consumer to pay the fixed costs of sourcing. Alternatively, we could allow consumers to source directly if they prefer that to purchasing from an available local trader; this complicates the set of potential equilibria but does not otherwise change our conclusions.

[15] The location parameter is used to normalize the expected payouts for traders buying from other intermediaries but has no impact on any of the other results.

elasticity leads the elasticity faced by a given trader to be strictly greater than the downstream elasticity. Consequently, a trader will face aggregate demand $x(\omega) = ap(\omega)^{-\tilde{\sigma}}$ where a combines the price index and expenditure, which are constant from the perspective of a small trader.

4.3 Chain-level pricing and profits

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 by 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 $z(1)$ is always the origin, $z(N_z)$ is the location where the good is consumed, and $z(n)$ is the location of the n^{th} transaction in the chain. Traders will end up participating in all chains in which goods move from their chosen source market j to their home market i, and so it will be useful to define this set as $\mathbf{Z}_{ij} = \{z \mid \exists\, n \text{ s.t. } z(n) = i,\, z(n-1) = j\}$.

Selling traders can set chain-specific prices. Trader ω, who sells at step n on chain z, will choose a price $p_{z,n}(\omega)$ that applies to the good when it is sold to a buyer at step $n+1$ on chain z. Likewise, buyers may choose different sellers for different chains they serve.[16][17][18]

Optimal trader decisions are a function of the elasticities that they face. We describe these elasticities using the recursive formula of section 4.2.2, and show how these formulas can be put in the chain notation of this section. The final step on an arbitrary chain z is N_z, which is always sale to a final consumer in location $i = z(N_z)$. In this case, demand is as shown in section 4.2.1, and so $\sigma_{z,N_z} = \frac{\mu_{z(N_z)}+1}{\mu_{z(N_z)}}$. In section 4.2.2, we showed that in upstream cases (i.e. for $n < N_z$), $\sigma_{z,n}$ can be defined recursively as $\sigma_{z,n} = \sigma_{z,n+1} + \frac{\sigma_{z,n+1}-1}{\mu^t_{z(n)}}$ so long as the objective being maximized by buyer's has a CES form. Consequently, elasticities are defined recursively along all chains. These formulas are an extension of CES demands to a setting with intermediation.

We have already showed that a trader at any step n on chain z faces CES demands, which can be put into chain-specific notation as $x_{z,n}(\omega) = a_{z,n}p_{z,n}(\omega)^{-\sigma_{z,n}}$. Profit maximization implies the usual

[16] These assumptions ensure that there is a pricing equilibrium in pure strategies. If traders could not perfectly price discriminate, then they would face elasticities of demand which based on the composition of their buyers. In particular, as a seller raised prices, this would cause a disproportionate loss of sales to more elastic buyers and drive down the perceived elasticity of demand. Similarly, buyers must be free to choose a different seller for each chain they serve, as if not, their choice of seller would change the composition of their buyers, hence the elasticity perceived by that buyer, and cause the seller they choose to have elasticity which varies in the price set in exactly the same way. Adopting these assumptions yields the standard markup rules and well-behaved demands.

[17] The parameters governing the dispersion of buyer-seller match values, μ and μ^t can be the same or different, and can vary freely across locations (and μ^t can vary across chains within a location). However, it is important that an individual buyer's draws of $\zeta(\omega)$ or $\varepsilon(\omega)$ are i.i.d across seller-chain pairs, not just sellers.

[18] We do not expect that this microfoundation is an literal description of trader behavior; it seems particularly unlikely that traders actually have a different match value with sellers for every chain they serve and therefore buy the same good from many different sellers within a single sourcing location. Instead, these assumptions lead to chain-level decisions that add up to a reasonable approximation to the aggregate sourcing and selling choices of each trader, while maintaining the separability across chains needed to ensure that a pure strategy pricing equlibrium exists. A trader who serves more demand along chains that are relatively elastic downstream will pay lower weighted average unit costs conditional on choosing a sourcing location, and will charge lower weighted average markups.

CES markups, and the price charged at every step on every chain will be $p_{z,n}(\omega) = \left(\frac{\sigma_{z,n}}{\sigma_{z,n}-1}\right) c_{z,n}(\omega)$. The trader's costs are denoted $c_{z,n}(\omega)$, and are determined by the upstream steps on the chain, and the markups and trade costs at those steps. This implies the usual CES variable profits, which in chain notation are $\pi_{z,n}(\omega) = \frac{1}{\sigma_{z,n}} a_{z,n} \left(\left(\frac{\sigma_{z,n}}{\sigma_{z,n}-1}\right) c_{z,n}(\omega)\right)^{1-\sigma_{z,n}}$. Finally, if a trader at step n on the chain faces a measure $\Omega_{z,n-1}$ of upstream traders, then we show in the Appendix that the expected payout to that trader (which incorporates idiosyncratic match value to the upstream trader in addition to variable profits) is $\Omega_{z,n-1}^{\mu_{z,n-1}^{t}} \pi_{z,n}$.

4.4 Sourcing and entry

4.4.1 Sourcing decision

When sourcing the good, traders face a menu of possible locations characterized by the fixed and variable trade costs between that location and their home location, the equilibrium prices for each chain, and the number of traders in each potential source location serving each chain. A trader will serve all chains that have a link from the chosen source to the trader's location. Each trader will choose a single source location to maximize total profits, given by $\mathit{\Pi}_{ij} = \Pi_{ij} - \lambda_i \xi(j)$ where $\Pi_{ij} = \mathbb{E}(\pi_{ij}) - F_{ij}$, and $\mathbb{E}(\pi_{ij}) = \sum_{z,n \in \mathbf{Z}_{ij}} \mathbb{E}(\pi_{z,n})$, which is the sum of expected profits over all chains that have a link going from j to i. The term $\xi(j)$ captures idiosyncratic components of an individual trader's cost of (or taste for) purchasing from location j (or exiting), and λ_i is a local parameter defining the importance of this cost shock in sourcing decisions. The shock is distributed to ensure some traders choose to source from every possible location in equilibrium.

If ξ is an independently, identically distributed standard Gumbel, this again conforms to a standard logit discrete choice framework, and implies that the share of traders in home market i who optimally choose to buy from source market j will be $\beta_{ij} = \frac{e^{\frac{\Pi_{ij}}{\lambda_i}}}{1+\sum_{k \in J} e^{\frac{\Pi_{ik}}{\lambda_i}}}$. Traders can also choose to exit if that is the option that yields the highest total profits, and will do so with probability $\beta_{ie} = \frac{1}{1+\sum_{k \in J} e^{\frac{\Pi_{ik}}{\lambda_i}}}$. Note that in addition to the fixed and idiosyncratic components of the cost of sourcing from each location, traders' choices will also be influenced by the expected variable profits to be earned from sourcing in that location, which depend on variable trade costs, equilibrium prices, the measure of selling traders, and the downstream elasticities (and therefore markups) the trader will face if they serve chains $z \in \mathbf{Z}_{ij}$.

4.4.2 Entry

Every location has an unlimited supply of potential traders who will enter so long as there are positive expected profits net of entry costs. Entry costs in location i are denoted f_i. Traders are ex-ante homogeneous, but after paying an entry cost will draw an idiosyncratic component of the fixed cost they face for sourcing from every potential location. Entry is therefore pinned down by the zero profit condition $\mathbb{E}(\mathit{\Pi}_i) = f_i$. Before observing the draws on on fixed costs, a potential trader's expected profits are $\mathbb{E}\left(\Pi_i\right) = \lambda_i \ln\left(1 + \sum_j e^{\frac{\Pi_{ij}}{\lambda_i}}\right)$.

4.5 Within and cross-location equilibrium

An equilibrium is characterized by five conditions following from the previous sections:

1. Sellers set chain prices to maximize profits
2. Buyers choose a seller within a source market to maximize profits for each chain they serve
3. Traders choose which location to source from to maximize expected profits
4. Zero expected profits from entry
5. Chain-level market clearing – expenditures at one step add up to revenues at the previous one.

The full equilibrium of the model has both within- and across-location components. Conditions 1 and 2 define the equilibrium within a market. Conditional on the measure of traders in every location, the sourcing decision of traders in those locations, and the expenditures flowing along each chain, there is a unique equilibrium within each location. Conditions 3, 4, and 5 determine the cross-location equilibrium. Because there are agglomeration economies in intermediation, there may be multiple equilibria that feature different combinations of flows across locations.

In order to understand the impact of intermediation chains on consumers, we need to consider the characteristics of chains that end in a given location in equilibrium. The (infinite) set of chains that end in each location is fixed, and we have assumed that parameters are such that there are positive flows on each in equilibrium. The "characteristics" are therefore more properly the weighted average characteristics of all the chains terminating in a location, where the weights are the revenue in the final location along each.

The price of the good at the end of any chain can be decomposed into cost and markup components: $p_z\left(N_z\right) = p_o T\left(z\right) M\left(z\right)$, where we define $T\left(z\right) = \prod_{n=2}^{N_z} \tau_{z(n)z(n-1)}$ and $M\left(z\right) = \prod_{n=2}^{N_z} m_{z(n)}$. Note that these prices are exogenous – they depend only on fundamentals, and do not vary across equilibria. The price index and the weighted average characteristics of chains, however, depend on the share of consumer expenditure that flows along each chain, and therefore on endogenous entry and sourcing decisions in each destination.

We define $\mathbf{Z}_{ij} = \{z | z\left(N_z\right) = i,\ z\left(N_z - 1\right) = j\}$ as the set of chains terminating in location i whose previous step is in j, and Ω_{ij} as the measure of traders from i sourcing from j. Therefore, $\alpha_{ij} = \Omega_{ij}\left(\frac{P_{ij}}{P_i}\right)^{1-\sigma_i}$ is the share of consumer expenditure in i that goes to chains whose previous step is in location j. We can then define weighted average total variable trade costs incurred along chains serving final consumers in location i as:

$$T_i = \sum_j \alpha_{ij} T_{ij}$$

where $T_{ij} = \sum_{z \in \mathbf{Z}_{ij}} \left(\frac{p_z(N_z)}{P_{ij}} \right)^{1-\sigma_i} T(z)$. Similarly, the weighted average total markups incurred along chains serving final consumers in location i will be:

$$M_i = \sum_j \alpha_{ij} M_{ij}$$

where $M_{ij} = \sum_{z \in \mathbf{Z}_{ij}} \left(\frac{p_z(N_z)}{P_{ij}} \right)^{1-\sigma_i} M(z)$. Finally, the weighted average length of chains serving final consumers in location i will be:

$$L_i = \sum_j \alpha_{ij} L_{ij}$$

where $L_{ij} = \sum_{z \in \mathbf{Z}_{ij}} \left(\frac{p_z(N_z)}{P_{ij}} \right)^{1-\sigma_i} N_z$.

For each of these definitions, the T_{ij},M_{ij}, and L_{ij} components are entirely exogenous. However, the expenditure weights are endogenous, depending both on the shares of traders in i choosing to source from each j, and on the consumer price index in i, which itself also depends on the sourcing shares of traders. Traders' sourcing decisions depend on not only the prices available in each location, but also the fixed and variable trade costs and the measure of traders active in each j.

4.6 Consumer welfare

Expected utility for a consumer in location i is

$$\mathbb{E}[u] = \left(\frac{1}{\sigma_i - 1} \right) \left(\frac{y_i}{P_i} \right)^{\sigma_i - 1}$$

where $P_i \equiv \left(\int_{\omega \in \Omega_i} p(\omega)^{1-\sigma_i} d\omega \right)^{\frac{1}{1-\sigma_i}}$ is the CES price index over final consumer prices on chains terminating in location i.

Because aggregate consumer welfare is a monotonic transformation of $\frac{Y_i^c}{P_i}$ as in CES, we will focus on the determination of the price index. As noted in the previous section, the price charged by traders along any particular chain is fixed. However, the measure of traders serving a given chain is endogenous, and consumers have ideal variety preferences and thus benefit from entry. Consequently, the consumer price index is

$$P_i^{1-\sigma_i} = \sum_j \Omega_{ij} P_{ij}^{1-\sigma_i}$$

where

$$P_{ij}^{1-\sigma_i} \equiv \sum_{\{z | z(N_z) = i,\ z(N_z - 1) = j\}} p_z(N_z)^{1-\sigma_i}$$

P_{ij} is effectively the price index for chains ending in i coming from location j. (Importantly, this is not the same as the consumer price index for consumers in j.)

All else equal, entry lowers the price index, as it must increase Ω_{ij} for at least one source j. Given the total measure of traders in a location, the price index will be lower when sourcing is weighted toward locations with lower P_{ij}. Holding entry fixed, consumers are better off when a larger fraction of the measure of traders sources from lower cost locations. All else equal, a location is lower cost when variable trade costs are lower and traders are more inclined to source in low cost places when the fixed costs of sourcing there are low and the measure of traders in that location is large.

4.7 Discussion

The model in the previous section imbeds endogenous intermediation chains into a model that otherwise shares many features of a standard trade model with monopolistically competitive sellers. Although the ways in which our model departs from a standard framework are conceptually simple – essentially, allowing for re-sale – they yield several new forces driving patterns in trade.

First, because there may be multiple, independent intermediaries between a producer and consumer, each of whom operates in an imperfectly competitive market, there is potential for double marginalization along the distribution chain. On chains that are indirect – i.e. that do not go directly from the origin to final consumers – markups will compound at each transaction point, so that downstream sellers are marking up over upstream markups. Suppose the intermediaries along any given chain were able to reach an agreement to coordinate. The price of purchase at the origin would be the same as in the equilibrium laid out in Section 4, as would the total trade costs incurred along the comparable chain. However, there would only be a markup at the final step of the chain, in sales to consumers. Holding entry and the price along all other chains fixed, the quantity of the good purchased by consumers from that particular chain would be higher, and the profits earned by sellers would be higher than the sum of profits earned along the entire chain in the baseline equilibrium. We rule out this kind of coordination – or integration to serve consumers in multiple locations – by assumption in the model. This is a reasonable approximation in the context of Nigerian consumer goods trade, where it is empirically true that nearly all traders serve only a single location. However, relaxing this assumption is likely to be an important avenue for future work, to understand the implications of distribution chains in contexts where integration is more prevalent.

Second, the elasticity of demand faced by sellers at each step of a chain is decreasing as the chain goes downstream. As shown in Section 4.2.2, demand from downstream intermediaries reflects both the elasticity of their own sales to buyers at the next step of the chain, and any factors that make sellers imperfectly substitutable from their perspective. This implies that the total elasticity at any given step is always greater than the elasticity at the step below it on the chain. Although we allow the dispersion of buyer-seller match factors to vary freely across locations in model, to simplify the expression of the following intuition suppose that $\mu_i^t = \mu_j^t \; \forall i,j$. Then defining $m = \frac{\mu^t+1}{\mu^t}$, and plugging this into the recursive definition of the elasticity at each step from Section 4.3, we see that $\sigma_{z,n} = m^n \left(\sigma_{z,N_z} - 1\right) + 1$ for any step n on chain z. Rearranging, $\sigma_{z,n} - \sigma_{z,n-1} = \left(m^n - m^{n-1}\right)\left(\sigma_{z,N_z} - 1\right) > 0$ since by definition $m > 1$ and $\sigma_{z,N_z} > 1$. Therefore, $\sigma_{z,n}$ is increasing in n, and the elasticity is decreasing.

Third, there is an entrepôt effect associated with locations that serve a large amount of downstream demand. One mechanism is that increased demand in a location (either from local consumers or downstream traders) induces entry, which increases the expected value of the best seller match for buyers who source from that market. This brings more demand to the location, which induces more entry, and so on. A second mechanism, which is in operation if the increase in demand also increases the average elasticity, is that sellers will shift their sourcing toward relatively low variable cost source locations, which also makes the location more appealing to downstream traders.

5 Quantification for Nigeria

We now return to the Nigerian context, and apply our many-location model to quantify the implications of long distribution chains. We will calibrate the model to broadly capture the empirical features of the distribution of Chinese-made apparel to locations across Nigeria. Because we lack data on domestic Nigerian trade outside of Lagos, our goals are more modest than full estimation. They are, first, to show that the model is able to capture nuanced empirical patterns in the part of the trade network that we do observe, at reasonably realistic parameter values; and second, to provide a sense of the potential magnitude of implications for consumer welfare and trade cost measurement.

5.1 Baseline calibration

In our baseline calibration, we will consider the distribution of Chinese-made apparel to consumers in Dubai, Lagos, and seven additional cities within Nigeria. The eight Nigerian cities are state capitals chosen to illustrate a variety of possible domestic sourcing patterns related to their location and size. Lagos is the largest by an order of magnitude, and is the main port of entry for goods in Nigeria. The other seven include two cities in the southwest near Lagos (one large, Ibadan, and one smaller, Abeokuta), and five cities in the north (Kano, Kaduna, Katsina, Sokoto, and Bauchi) of varying size and relative proximity to Lagos. We treat "apparel" as a single good, and take each location (including the origin, "China") to be a single market.[19]

To simulate the model, we need values for three groups of exogenous parameters: those governing local consumer demand in each downstream location, trade costs between locations, and parameters related to sellers' businesses. Some of these we take from the Lagos Trader Survey (LTS) data, some from other standard data sources, and some we calibrate to roughly capture the patterns of outcomes we observe in our LTS data. Table 5 shows the definition and source for each, and example values for a subset of locations. Parameters for the full set of locations are shown in the Appendix.

Consumer demand in each location is determined by the total value of consumer expenditure on the good, Y_i^C, and the dispersion of consumer preferences across sellers, μ^r. The former is

[19] Clearly, not all sellers or buyers are located at the same place in a large city like Lagos. There is no conceptual limitation to breaking these "locations" up into multiple smaller markets, but adding more locations creates both additional computational burden and data requirements (e.g. some measure of trade costs across sub-markets within Lagos). The level of aggregation chosen is not innocuous, however, since it has implications for the ability of buyers arriving at a location to benefit from variety.

calculated based on the overall fraction of Nigerian GDP spent by consumers on apparel in 2010 (from the World Bank Global Consumption Database), state-level GDP (from the Nigerian Bureau of Statistics), and the fraction of each state's population that lives in the capital city. The dispersion of preferences, which determines the retail elasticity of demand, is set to match the retail markups observed in the LTS data for apparel.

Trade costs consist of two matrices describing the fixed (F_{ij}) and variable (τ_{ij}) costs of trading goods between each pair of locations, and a parameter governing the dispersion of the idiosyncratic component of the fixed cost individual traders face to buy from each source location, λ^f. The fixed costs of sourcing in China and Dubai from Lagos are drawn from average travel costs reported by traders sourcing from those locations in the LTS data. Ideally, the fixed costs between locations within Nigeria would be calibrated to match sourcing shares across these locations. Since we do not observe this in our current data, they are set very approximately to generate reasonable sourcing patterns. The variable trade costs of bringing goods from China and Dubai to Lagos are also drawn from reported costs in the LTS data, and the cost between Lagos and Ibadan is taken from an additional survey of used clothing traders in those locations. Variable costs in the rest of Nigeria are set by scaling the Lagos-Ibadan cost by relative travel times between each pair of cities.[20] The parameter λ^f governing trader-source specific variation in fixed costs is calibrated approximately to match sourcing shares from Lagos. Further details on the sources and assumptions behind trade costs are provided in the Appendix.

Finally, we need three more parameters: the fixed cost of entry in each location, f_i, and the dispersion of traders' preferences across upstream wholesalers, μ^t,and the price of the good in the origin, p_o. The first does not have a good empirical analog. However, since this parameter plays a major role in governing entry, we can compare our outcomes to estimates of the number of wholesale and retail traders from the Small and Medium Enterprises Development Agency of Nigeria (SMEDAN), and test sensitivity. The μ^t parameter would ideally be calibrated to match reported wholesale markups in Lagos in the LTS data. However, we will show that since this parameter also drives the strength of agglomeration economies in intermediation, we are not able to perfectly match both markups and sourcing shares from Lagos, and so we adjust it to fit the latter rather than the former. Finally, we set the price of the good in the origin to the median unit cost reported for apparel in China in the LTS data.

Given these exogenous parameters, we solve the model for the equilibrium number of firms and sourcing shares in each location.[21]In this baseline, we compute chains with up five intermediaries

[20]We assume that the triangle inequality holds with equality for variable costs of importing for cities within Nigeria. This implies, for instance, that the cost to move goods from China to Ibadan is exactly the cost to move them from China to Lagos and then Lagos to Ibadan. 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.

[21]We choose starting values for the number of firms and sourcing shares, and iterate over simulations of the model until we find a fixed point. Because the model features agglomeration economies in intermediation, there can be multiple equilibria. In practice, we find that there is some sensitivity to the starting values for non-Lagos locations in Nigeria, but very little for Dubai or Lagos. This is consistent with the fact that, given the exogenous parameters, Dubai and Lagos traders have more clearly dominant strategies, while smaller downstream locations are relatively indifferent about which domestic hub to use based on trade cost fundamentals.

between the producer and final consumer.[22] The results are shown in Table 6. Lagos is the only modeled city in Nigeria that does substantial importing. It in turn serves as a hub for smaller Nigerian cities in both the north and south. In the south, both Ibadan and Abeokuta source primarily from Lagos. In the north all cities do some purchasing from Lagos, but Kano also serves as a hub. In Bauchi and Kaduna, which are small and comparatively closer to Kano than to Lagos, more traders buy from Kano, although the total value of goods coming from Lagos is still larger.

The average length of chains serving consumers in each destination is increasing in the remoteness of the location from the origin, ranging within Nigeria from about 4.5 in Lagos (implying between two and three intermediaries between the manufacturer and the final consumer) to just over 6 links in Katsina. The average retail price is also increasing. Total trade costs incurred along the route from the origin to the destination are increasing as a percent of the final price, ranging from only 5 percent in Dubai to almost 35 percent in Katsina. Total markups as a percent of the final price are relatively constant across locations, at around one-third. However, this is an artifact of one feature of the data that our baseline calibration clearly does not capture well: keeping sourcing patterns realistic involves setting the wholesale demand elasticity unrealistically low. Lagos traders report in the LTS that wholesale prices are on average around 10 percent lower than retail ones, while our calibration implies that they are almost one-third lower, featuring almost no markup over cost. This lack of realism will have implications for our conclusions about double marginalization and the impact of counterfactuals that induce more direct sourcing, which we discuss below.

The model also yields a number of outcomes of interest that have empirical counterparts in our Lagos data. We compare these in Table 7, and see that the quantification does capture the broad patterns of trade in Lagos. One set of moments – the share of traders sourcing from each location – was informally targeted in choosing the parameters. The model also captures average firm size in Lagos relatively well, although this was not targeted. For the remaining moments, it is useful to compare the data both to the baseline calibration and to the calibration with chain position "truncated" to match the data. In the latter case, we top-code both upstream and downstream links at two, which is the maximum we can observe in the data. As should be expected, the truncated version of the calibration does better at matching the chain position of traders in the data. Finally, we can also using the calibration to quantify the same patterns relating firm outcomes to chain position that we showed using the survey data in Table 4. The model captures the direction and approximate magnitude of all four patterns: traders who are closer to the origin are also further from customers (i.e. upstreamness and downstreamness are negatively related), traders who are closer to the origin are larger, those closer to final consumer are smaller, and markups are increasing in downstreamness.

These comparisons between the data and the quantification of the model are not and should not be exact, however, for several reasons. First, because the LTS survey focused on commercial areas of Lagos, we should expect small retailing to be underrepresented in the data relative to its actual prevalence in the city as a whole. To the extent that the model captures the retail sector

[22] While in theory there are infinite chains serving every location, chains beyond some length will carry infinitesimally small values of the good.

more fully, we should expect the calibration to feature smaller firms and more domestic sourcing relative to importing. On the flip side, in reality many traders in Lagos source from intermediaries rather than manufacturers in China. Since the model assumes that producers sell directly in China, the calibration should overstate the upstreamness of importers. Finally, in reality there are many origin points for multiple goods, and one of the purposes served by hubs like Dubai and Lagos is presumably the ability to do "one-stop shopping". Since the calibration has only one good from one origin point, we should expect it to underestimate the relative appeal of these hubs.[23]

5.2 Counterfactuals

5.2.1 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 – how does the equilibrium with chains of intermediation compare to one in which only "direct" sourcing from the origin location is possible? In our model, there is still one intermediary (a retailer) between a producer and a consumer even under direct sourcing, because consumers are assumed to buy only from traders serving their home location. However, there is a close – although not exact – correspondence between our model with only direct sourcing and a Melitz (2003) type model in which all firms have a shared cost of production.

In practice, we arrive at a direct-sourcing-only equilibrium by setting trade costs between consumer locations and all sources other than the origin to prohibitive levels. 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. Table 8 compares outcomes under direct sourcing to our baseline.

5.3 Measurement of trade costs

In Section 3.3, we noted that assuming direct trade when in reality goods reach consumers via chains of intermediation may lead us to overestimate distance costs of trade. The first reason is simply that when goods travel on indirect routes, they will incur weakly higher trade costs. When estimates of these costs are projected onto proxies like shortest distance or travel time, it will appear that distance-related trade costs are higher in places that tend to be served by longer chains. We illustrate this in Table 9 by comparing weighted average actual trade costs incurred to reach each destination in our baseline calibration to the costs that would be incurred under direct sourcing. The third column shows that the difference between direct and actual variable trade costs is higher in locations that are more remote from the origin and are served by longer chains.

[23] Of course, there are many additional features of the real environment that are also not captured here and should be expected to affect the comparison between data and model outcomes. Some particularly salient ones are the level of aggregation chosen as a single market, the existence of credit constraints that affect a given firm's ability to pay a fixed cost to access a source location, and the choice of product scope carried by a single firm in a multi-product world. We expect capturing these important nuances to be a productive direction for future work.

6 Conclusion

We began by asking what might be lost in the implicit simplifying assumption common to the trade literature that goods go directly from producers to consumers. We have shown that, both in theory and in practice in the context of imported Nigerian consumer goods, this simplication may matter a great deal. First, in a world with fixed costs of trade and imperfect competition, allowing for the existence of chains carries different implications for the welfare of consumers in across locations than a standard model. The option for traders to source via resale markets, rather than the original production location, may increase or decrease consumer welfare, depending on whether the value of increased entry in the consumer market is larger or smaller than the increase in passed through variable costs. Consumers in entrepôt locations, however, unambiguously gain from increased downstream demand flowing through their location.

Accounting for chains of intermediation also has implications for the measurement of trade costs and our understanding of price gaps between producers and consumers. These price gaps 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. Decomposing price gaps into trade costs and markups is therefore particularly relevant for understanding which policy levers are likely to 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 markets to increase competition (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).

The intermediaries who feature in the quantitative model we have presented look quite similar to those we actually observe in distribution chains in Nigeria. Different traders source the same or similar goods from different upstream locations, and often from upstream intermediaries rather than producers. They deal in a narrow range of goods. They sell to both downstream intermediaries and to consumers, and they price discriminate across different buyers. They generally operate out of a single shop in a single market, rather than serving multiple locations.

While modeling a single good and restricting intermediaries to serve a single location therefore appears to be a reasonable approximation to the reality of the current setting, we have abstracted from two 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. The second force we have left out is the motive to integrate and serve consumers in multiple locations. Integration would allow traders 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 are more likely to face constraints on the scale they can achieve 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 key to 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. 2014. "Information Frictions in Trade." *Econometrica*, 82(6): 2041–2083.

Anderson, James E., and Eric van Wincoop. 2004. "Trade Costs." *Journal of Economic Literature*, 42(3): 691–751.

Anderson, Simon P., Andre De Palma, and Jacques-Francois Thisse. 1987. "The CES is a discrete choice model?" *Economics Letters*, 24(2): 139–140.

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

AntrÃ s, Pol, and Alonso de Gortari. 2017. "On the Geography of Global Value Chains." National Bureau of Economic Research Working Paper 23456.

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.

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

Bernard, Andrew B., Andreas Moxnes, and Karen Helene Ulltveit-Moe. 2018. "Two-Sided Heterogeneity and Trade." *The Review of Economics and Statistics*, 100(3): 424–439.

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."

Brancaccio, Giulia, Myrto Kalouptsidi, and Theodore Papageorgiou. 2017. "Geography, Search Frictions and Endogenous Trade Costs." National Bureau of Economic Research Working Paper 23581.

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

Chatterjee, Shoumitro. 2018. “Market Power and Spatial Competition in Rural India.”

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

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

Emran, M. Shahe, Dilip Mookherjee, Forhad Shilpi, and M. Helal Uddin. 2017. “Credit Rationing and Pass-Through in Supply Chains: Theory and Evidence from Bangladesh.”

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.

Fally, Thibault, and Russell Hillberry. 2015. “A Coasian Model of International Production Chains.” *National Bureau of Economic Research.*

Ganapati, Sharat. 2018. “The Modern Wholesaler: Global Sourcing, Domestic Distribution, and Scale Economies.”

Johnson, Robert, and Andreas Moxnes. 2019. “GVCs and Trade Elasticities with Multistage Production.” C.E.P.R. Discussion Papers CEPR Discussion Papers 13827.

Lim, Kevin. 2018. “Endogenous production networks and the business cycle.” *Work. Pap.*

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

Startz, Meredith. 2018. “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.

Wong, Woan Foong. 2018. “The Round Trip Effect: Endogenous Transport Costs and International Trade.”

Young, Alwyn. 1999. “Transport, processing and information: value added and the circuitous movement of goods.”

Table 1: Fraction of suppliers who are wholesalers, by region

Africa		**Asia**		**Other**		**Europe**	
Benin	100%	China	62%	Turkey	64%	Germany	41%
Ghana	100%	Hong Kong	88%	UAE	71%	Italy	37%
Togo	89%	India	81%	US	66%	Switzerland	43%
		Japan	13%			UK	74%

Table 2: Chain Length

	Chain length	Steps upstream	Steps downstream
Mean	4.11	1.52	1.58
Apparel	4.07	1.51	1.54
Electronics	4.20	1.67	1.48
Beauty	4.25	1.44	1.78

Table 3: Relationship between indirect sourcing and business size and costs

	(1) Revenue ($US)	(2) Number of workers	(3) Log unit cost ($US)
Fraction of purchases from wholesalers	-72187.58**	-0.47**	0.23
	(31746.07)	(0.20)	(0.17)
Obs	220	403	403
Product FEs	x	x	x

Table 4: Relationship between traders' chain position and outcomes

	(1) % of purchases from wholesaler	(2) Revenue ($US)	(3) Number of workers	(4) Log unit cost ($US)	(5) Log markup	(6) Log markup
% of sales that are	-0.18**	104424.74*	0.53	-0.74**	-0.22**	-0.32***
wholesale	(0.09)	(56127.23)	(0.35)	(0.31)	(0.10)	(0.09)
% of purchases from		-64470.80**	-0.44**	0.19	-0.07	-0.05
wholesaler		(31835.56)	(0.20)	(0.17)	(0.06)	(0.05)
Log unit cost ($US)						-0.13***
						(0.01)
Obs	403	220	403	403	403	403
Mean	1.52	67,687	1.45	5.2	.46	.46
Product FEs	x	x	x	x	x	

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

Table 5: Baseline calibration parameter sources and values

Symbol	Meaning	Source		Value		
				Dubai	Lagos	Kano
Y_i	Consumer spending on apparel ($US mil.)	World Bank Global Consumption Database (2010 data), Nigerian Bureau of Statistics		3,000	1,541.6	48.6
μ	Scale parameter of consumer preference across sellers	Retail markups in LTS		0.5	0.5	0.5
F_{ij}	Fixed trade cost to buy in i and sell in j ($US)	LTS	From j =China	2,000	2,000	2,200
			From j =Lagos	2,000	200	200
τ_{ij}	Variable trade cost to buy in j and sell in i	LTS and Google Maps travel times	From j =China	1.07	1.20	1.58
			From j =Lagos	1.15	1.03	1.32
λ^f	Scale parameter of trader-specific fixed cost draws	Calibrated		150		
f_i	Fixed cost for trader to enter in i ($US)	Calibrated		15000	12000	3000
μ^t	Scale parameter of trader preference across sellers	Calibrated		0.03		
p_o	Price in the origin	LTS		$5		

Table 6: Results from baseline calibration

	Outcome								
		Share of traders sourcing from[24]							
Destination	Number of traders (thousand)	China	Dubai	Lagos	Ibadan	Kano	Average retail price	Average chain length	Export sales, mil (sales to other locations)
Dubai	88.4	62%	38%	0%	0%	0%	8.18	3.45	272
Lagos	61.0	22%	11%	67%	0%	0%	9.50	4.42	82.3
Ibadan	7.44	0%	0%	67%	30%	0%	9.89	5.25	29.1
Abeokuta	2.18	0%	0%	65%	16%	0%	10.0	5.27	13.0
Kano	8.35	1%	2%	45%	4%	40%	12.7	5.74	4.82
Bauchi	1.40	0%	0%	22%	6%	27%	13.1	5.83	4.18
Kaduna	5.76	0%	0%	22%	5%	7%	12.2	5.79	9.85
Katsina	2.07	0%	0%	11%	7%	37%	13.2	6.04	0.629
Sokoto	2.78	0%	0%	119%	11%	6%	12.9	5.86	0.175

Note: All averages reflect chain-level outcomes weighted by expenditure along each chain.

[24] These do not add up to 100% due to exit, sources not shown, and rounding.

Table 7: Lagos data to baseline calibration comparison

Outcome		LTS data	Calibration	Truncated calibration
Average firm revenues	All traders	\$107,581	\$43,662	n/a
Steps upstream	All traders	1.51	2.40	1.78
Steps downstream	All traders	1.54	1.55	1.38
Sourcing share	From China	27.6%	21.7%	n/a
	From Dubai	9.5%	10.8%	n/a
OLS coefficients	Steps upstream & downstream	-0.18	-0.32	-0.26
	Revenue & steps downstream	\$104,425	\$58,034	\$118,149
	Revenue & steps upstream	-\$64,471	-\$17,524	-\$29,746
	Average markup & steps downstream	-0.32	-0.24	-0.50

Note: Steps up- and downstream and OLS coefficients from LTS data are calculated based on importers only, because we do not have data on suppliers for those who purchase domestically.

Table 8: Comparison between baseline and direct sourcing equilibria

Destination	Baseline Price (\$US)	Direct Sourcing Price (\$US)	Welfare ratio (direct/baseline)
Dubai	8.18	8.03	99.76%
Lagos	9.50	9.00	99.25%
Ibadan	9.89	9.39	90.18%
Abeokuta	10.00	9.48	84.29%
Kano	12.66	11.87	87.24%
Bauchi	13.10	12.01	69.17%
Kaduna	12.17	11.29	81.78%
Katsina	13.20	12.17	71.78%
Sokoto	12.94	11.98	76.53%

Table 9: Comparison between actual and direct variable trade costs from China

Destination	Direct trade cost	Average actual trade cost	Average chain length	Ratio of actual to direct
Dubai	0.35	0.42	3.45	1.20
Lagos	1.00	1.34	4.42	1.34
Ibadan	1.26	1.71	5.25	1.36
Abeokuta	1.32	1.82	5.27	1.38
Kano	2.91	5.37	5.74	1.85
Bauchi	3.01	5.89	5.83	1.96
Kaduna	2.52	4.52	5.79	1.79
Katsina	3.11	6.18	6.04	1.99
Sokoto	2.99	5.93	5.86	1.98

Appendix

Derivations

There are fewer firms under direct sourcing

In Section 3.2 of the text, we claim that there will be exit when traders are forced to shift from an indirect sourcing to a direct sourcing equilibrium. This arises from revealed preference. If there is an indirect sourcing equilibrium, then if a single trader switched to sourcing from the origin, it must obtain weakly negative profits (otherwise the indirect sourcing equilibrium would not exist). However, since goods cost weakly more in the origin, it must be more profitable to be the only trader sourcing from the origin than to source from the origin when the full set of traders (held fixed relative to the indirect sourcing equilibrium) is also sourcing from the origin. Therefore, if the set of traders stayed the same relative to the indirect sourcing equilibrium but all traders sourced from the origin, they must all earn weakly negative profits. Thus there must (weakly) be exit to maintain the zero profit condition when an indirect sourcing equilibrium is forced to switch to a direct sourcing equilibrium.

Demand along chains is CES with increasing elasticity along the chain

In the text, we claim that the when traders have the given payout, wholesalers will face CES demand with increasing elasticity going up the chain.

Consider a given buying trader choosing among sellers $\omega \in \Omega$ which offer prices $p(\omega)$. Suppose that its profit function for buying from seller ω takes the form $\pi(\omega) = ap(\omega)^{1-\sigma}$. This trader will choose a seller to maximize its payout; since natural log is a monotone transformation, it will also act to maximize the natural log of its payout. The natural log of its payout from seller ω is therefore

$$\begin{aligned} \ln[v(\omega)] &= \mu^t \zeta(\omega) + \ln[\pi(\omega)] \\ &= \mu^t \zeta(\omega) + \ln a - (\sigma - 1)\ln[p(\omega)] \\ &= \zeta(\omega) + \frac{\ln a}{\mu^t} - \frac{\sigma - 1}{\mu^t}\ln[p(\omega)] \end{aligned}$$

This is isomorphic to the payout from Anderson, Thisse, and De Palma (1987) which is shown to yield CES demand in expectation for the seller. However, the elasticity faced by the seller is now

$$\tilde{\sigma} = \frac{\sigma - 1}{\mu^t} + 1$$

CES demands along the chain now follows by induction. Consumers preferences leads the final seller on the chain to face CES demands. And as we show above, if the downstream seller faces CES demands, then the upstream seller faces CES demands as well.

Finally, elasticities will increase going up the chain because $\mu^t < 1$ as long as consumer elasticity is at least 2 (i.e. $\mu < 1$). When μ^t is the same at all points along the chain, then it is possible to derive a formula for the elasticity faced by sellers at step n (where step 0 is consumer sales)

$$\sigma_n = (\sigma_c - 1)\left(\frac{1}{\mu^t}\right)^n + 1$$

Expected trader payouts

In the text, we provide the expected payout (incorporating match value in addition to variable profits) for a trader at step n on the chain faces a measure $\Omega_{z,n-1}$ of upstream traders as $\Omega_{z,n-1}^{\mu_{z,n-1}^t}\pi_{z,n}$.

Consider a given trader which offers payout $P(\omega) = \exp\left(\mu_{z,n-1}^t \zeta(\omega)\right)\pi(\omega)$ where $\zeta(\omega)$ is Gumbel with unit scale parameter and location parameter $\alpha = \frac{1}{\mu_{z,n-1}^t}\ln\left(\frac{1}{\Gamma\left(1-\mu_{z,n-1}^t\right)}\right)$. Then

$$\begin{aligned}
\Pr\left(P(\omega) \le P\right) &= \Pr\left(\zeta(\omega) \le \frac{1}{\mu_{z,n-1}^t}\ln\left(\frac{P}{\pi(\omega)}\right)\right) \\
&= \exp\left(-\exp\left(-\frac{1}{\mu_{z,n-1}^t}\ln\left(\frac{P}{\pi(\omega)}\right) + \frac{1}{\mu_{z,n-1}^t}\ln\left(\frac{1}{\Gamma\left(1-\mu_{z,n-1}^t\right)}\right)\right)\right) \\
&= \exp\left(-\beta\left(\frac{P}{\pi(\omega)}\right)^{-\frac{1}{\mu_{z,n-1}^t}}\right)
\end{aligned}$$

where $\beta \equiv \left(\Gamma\left(1-\mu_{z,n-1}^t\right)\right)^{-\frac{1}{\mu_{z,n-1}^t}}$ for convenience, so that the CDF (and consequently PDF) of the maximum payout across the measure of traders is

$$\begin{aligned}
\Pr\left(\max_\omega P(\omega) \le P\right) &= \exp\left(-\beta\int_\omega \left(\frac{P}{\pi(\omega)}\right)^{-\frac{1}{\mu_{z,n-1}^t}} d\omega\right) \\
&= \exp\left(-\beta\int_\omega \left(\frac{\pi(\omega)}{P}\right)^{\frac{1}{\mu_{z,n-1}^t}} d\omega\right) \\
\Pr\left(\max_\omega P(\omega) = P\right) &= \frac{\partial}{\partial P}\Pr\left(\max_\omega P(\omega) \le P\right) \\
&= \frac{\beta}{\mu_{z,n-1}^t}\cdot\frac{1}{P}\cdot\int_\omega \left(\frac{\pi(\omega)}{P}\right)^{\frac{1}{\mu_{z,n-1}^t}} d\omega \exp\left(-\beta\int_\omega \left(\frac{\pi(\omega)}{P}\right)^{\frac{1}{\mu_{z,n-1}^t}} d\omega\right) \\
\mathbb{E}\left[\max_\omega P(\omega)\right] &= \beta\int_{-\infty}^{\infty} P\cdot\frac{1}{\mu_{z,n-1}^t}\cdot\frac{1}{P}\cdot\left(\int_\omega \left(\frac{\pi(\omega)}{P}\right)^{\frac{1}{\mu_{z,n-1}^t}} d\omega\right)\exp\left(-\beta\int_\omega \left(\frac{\pi(\omega)}{P}\right)^{\frac{1}{\mu_{z,n-1}^t}} d\omega\right) dP \\
&= \beta\int_{-\infty}^{\infty}\left(\int_\omega \left(\frac{\pi(\omega)}{P}\right)^{\frac{1}{\mu_{z,n-1}^t}} d\omega\right)\exp\left(-\beta\int_\omega \left(\frac{\pi(\omega)}{P}\right)^{\frac{1}{\mu_{z,n-1}^t}} d\omega\right) dP \\
&= \left(\beta\int_\omega \pi(\omega)^{\frac{1}{\mu_{z,n-1}^t}} d\omega\right)^{\mu_{z,n-1}^t}\int_0^\infty P z^{1-\mu_{z,n-1}^t}\exp(-z)\, dz \\
&= \left(\int_\omega \pi(\omega)^{\frac{1}{\mu_{z,n-1}^t}} d\omega\right)^{\mu_{z,n-1}^t}
\end{aligned}$$

and when all sellers offer the same profit (because all sellers on a chain are symmetric), then $\pi(\omega) = \pi_{z,n}\ \forall\omega$ so that

$$\mathbb{E}\left[\max_\omega P(\omega)\right] = \Omega_{z,n-1}^{\mu_{z,n-1}^t}\pi_{z,n}$$

Additional details on the simulation

Parameters for the full set of locations

In the text here and in the table below we provide the full set of parameters used in the simulation. The scale parameter for the firm match shock (which is identical across all locations) is 0.03, and the scale parameter for the consumer preference shock across sellers is 0.5 in all locations. The scale parameter for the fixed cost of sourcing shock is 150 for all sources and destinations. We assume 120,000 firms in China. The remaining parameters to summarize are the consumer expenditures, trade costs, fixed costs of sourcing, and fixed cost of entry. These parameters are summarized in Table 10, below.

Table 10: Additional parameters for the baseline simulation

Destination	Consumer Expenditure (mil)	Fixed cost entry	Trade cost from: China	Dubai	Lagos	Ibadan	Abeokuta	Kano	Bauchi	Kaduna	Katsina	Sokoto
Dubai	3,000	15000	1.07	1.03	1.15	1.2	1.21	1.52	1.54	1.44	1.55	1.53
Lagos	1,542	12000	1.2	1.15	1.03	1.04	1.05	1.32	1.33	1.25	1.35	1.33
Ibadan	53	3500	1.25	1.2	1.04	1.02	1.04	1.29	1.33	1.25	1.31	1.3
Abeokuta	13	3000	1.26	1.21	1.05	1.04	1.02	1.3	1.34	1.26	1.33	1.33
Kano	48	3000	1.58	1.52	1.32	1.29	1.3	1.02	1.07	1.07	1.06	1.15
Bauchi	5.4	2000	1.6	1.54	1.33	1.33	1.34	1.07	1.02	1.13	1.14	1.23
Kaduna	23	2000	1.5	1.44	1.25	1.25	1.26	1.07	1.13	1.02	1.13	1.17
Katsina	7.9	2000	1.62	1.55	1.35	1.31	1.33	1.06	1.14	1.13	1.02	1.1
Sokoto	11	2000	1.6	1.53	1.33	1.3	1.33	1.15	1.23	1.17	1.1	1.02
Destination			Fixed cost from: China	Dubai	Lagos	Ibadan	Abeokuta	Kano	Bauchi	Kaduna	Katsina	Sokoto
Dubai			2000	400	2000	2000	2000	2000	2000	2000	2000	2000
Lagos			2000	800	200	250	250	300	350	400	400	400
Ibadan			2200	900	150	100	150	250	300	400	400	400
Abeokuta			2200	900	150	150	100	250	300	400	400	400
Kano			2200	1000	200	250	250	100	150	200	200	200
Bauchi			2200	1000	350	350	350	100	100	200	200	200
Kaduna			2300	1000	400	400	400	100	200	100	250	250
Katsina			2300	1000	400	400	400	100	200	250	100	250
Sokoto			2300	1000	400	400	400	100	250	200	200	100