Journal of International Economics 105 (2017) 119–129

Contents lists available at ScienceDirect

Journal of International Economics

journal homepage: www.elsevier.com/locate/jie

Import sourcing of Chinese cities: Order versus randomness☆

Keith Head[a,c], Ran Jing[b], John Ries[a,*]

[a]Sauder School of Business, University of British Columbia, Canada
[b]University of International Business and Economics, China
[c]CEPR, United Kingdom

ARTICLE INFO

Article history:
Received 12 May 2014
Received in revised form 13 June 2016
Accepted 2 January 2017
Available online 7 January 2017

JEL classification:
F12
F14

Keywords:
Hierarchy
Trade
China
Granularity
Extensive margin

ABSTRACT

Capitalizing on the geographic detail of Chinese customs data, we show that buyer heterogeneity plays a major role in import sourcing. Hierarchy compliance, a core prediction of supply-focused models, is tested by measuring the frequency with which cities import a narrowly defined good from the country observed to be the preferred source in the province. Hierarchy violation is widespread: 92% of province goods have at least one non-compliant city. We show that introducing granular importers into a standard heterogeneous firm model leads to a prediction of 73% compliance, close to the observed average of 66%. Extending the model to allow buyers from a city to share an orientation towards specific source countries, we calibrate a heterogeneity parameter to match the average observed compliance rate. The results imply that the supply side explains on average 44% of the variance in city-level sourcing probabilities, leaving the majority of variation due to heterogeneity in buyers across cities.

1. Introduction

The availability of disaggregated customs data has expanded the set of phenomena to be understood and incorporated into trade models. Following the empirical insight of Bernard and Jensen (1999) that good firms become exporters, models based on Melitz (2003) have emphasized heterogeneity across exporters, combined with fixed costs, as the drivers of trade outcomes. A supplier hierarchy emerges where better firms export to more markets and to more difficult markets. Bilateral zeros occur when no firm from a given origin is good enough to export to a particular market. With a few exceptions, the literature has neglected the role of heterogeneity across individual importers. Standard models typically assume a representative consumer wishing to buy all the varieties available in her market. Establishing the importance of demand-side heterogeneity requires a methodology and data capable of neutralizing the plausible forms of supply-side variation. The purpose of this paper is to show that a model of heterogeneous, granular consumers can explain key features of micro import patterns.

We model consumers who purchase their preferred variety from heterogenous firms based on relative prices, quality, and idiosyncratic taste shocks. Consistent with the pattern observed in Chinese data, goods reach destination cities by passing through a provincial transport hub. Multiplicative hub-and-spoke transport costs imply that all cities face the same relative prices and have a common ranking of products in terms of quality-adjusted prices.

Our model delivers hierarchical sourcing with a continuum of consumers but admits hierarchy violations in the case of granular consumers. When there is a continuum of tastes across consumers, all importing cities buy from the source country offering the variety with the lowest quality-adjusted price. This is because the continuum assumption leads the aggregate market of heterogeneous consumers to behave like a representative consumer with a love of variety. Thus, the Helpman et al. (2008) model is a special case of our

☆ This research has been supported by the Social Sciences and Humanities Research Council of Canada (410-2010-0922), the National Natural Sciences Foundation of China (71673045) and the Fundamental Research Funds for Central Universities at UIBE (15YQ02). We thank the two referees, Jacques Thisse, Robert Feenstra, seminar participants at the University of Oxford, UC Davis, the North American Meetings of the Regional Science Association International, University of Oregon, UC Santa Cruz, University of British Columbia, Princeton University, and Yale University for their helpful suggestions.

* Corresponding author.
E-mail addresses: keith.head@sauder.ubc.ca (K. Head), ran.jing@uibe.edu.cn (R. Jing), john.ries@sauder.ubc.ca (J. Ries).

http://dx.doi.org/10.1016/j.jinteco.2017.01.003
0022-1996/

model with a continuum of consumers.[1] With love of variety in the aggregate, larger cities import additional, less preferred varieties, as well the best one. The reason is that with larger demand, more firms from more countries will be able to cover fixed market entry costs. The hierarchy prediction we assess is that *whenever there is a non-empty set of sources selected at the city level, it should include the best source*.

The replacement of the continuum with granular consumers leads naturally to hierarchy violations. While supply-side factors influence the likelihood consumers purchase from different sources, idiosyncratic taste shocks may result in no consumers in a city selecting the source that the aggregated provincial outcomes reveal to host the best supplier. Non-compliance with hierarchy, therefore, derives from variation in the granular realizations of taste draws.

Recent work has considered two aspects of demand heterogeneity that are related but quite distinct from the focus of this paper. First, papers such as Halpern et al. (2015) and Antràs et al. (2014) model importing firms as heterogeneous in their productivity. They assume that these firms pay extra fixed costs to add import varieties to the purchase set. This leads to sourcing sets that are increasing in the productivity of the importer. A second type of heterogeneity is considered by Di Comite et al. (2014). In their model, countries can have "taste mismatches" that lead to zero trade flows. The Di Comite et al. (2014) model has representative consumers in each country so it is not designed to capture granular heterogeneity across consumers *within* nations.

We build on a small literature using granularity to explain the incidence of zeros in trade. Armenter and Koren (2014) predict the frequency of zeros in the country-industry trade matrix of the United States and the number of countries to whom US firms export. Eaton et al. (2013) investigate the number of bilateral trade zeros that would occur in a model with finite number of independent firms. Neither of these papers sets out to explain hierarchy violation. Armenter and Koren (2014) do not provide an economic model suitable for assessing hierarchy. Eaton et al. (2013) acknowledge that their model with granular exporters still yields hierarchy, stating in their conclusion, "By stripping out additional sources of heterogeneity, firms from the same source will enter markets according to a strict hierarchy (i.e. a firm will always sell in an easy-to-enter market if it sells in a more difficult market.)" Combining consumer granularity and buyer–seller shocks, our model produces both zeros and hierarchy non-compliance.

Both Armenter and Koren (2014) and Eaton et al. (2013) utilize a simple probability formula for the likelihood of a zero trade flow under independence. Namely, if there are n individual shipments or firms and each has a probability π of selling in a market, then the probability of no sales is $(1-\pi)^n$. Armenter and Koren (2014) use shipment shares to measure these probabilities whereas Eaton et al. (2013) construct them by estimating parameters of a gravity model. Our focus is whether cities import from the top source. If we knew that probability to be π_1, then the probability of compliance by a city receiving n_d shipments is $1-(1-\pi_1)^{n_d}$. The expected value of hierarchy compliance across cities is the average of these probabilities for cities that import the good. Applying this formula and using the realized shipment share of the top source as π_1, we find expected hierarchy compliance to average 72%, remarkably close to the observed compliance rate of 66%. Unfortunately, with a small number of shipments, the realized share of the top source is a biased estimate of π_1. This bias motivates use of a method to calculate expected compliance that does not require an estimate of π_1.

Our method determines the expected number of cities that would be represented in the *realized* shipments of a source country. For each province–good, we draw (without replacement) the observed number of shipments of the realized top source from an urn consisting of all city shipments. Repeating via Monte Carlo simulation and averaging the results yield expected hierarchy compliance of 73%.

The finding that an expectation relying on the assumption of independence predicts excessive compliance—and therefore too few zeros—also appears in the prior work on granularity. Armenter and Koren (2014) find that their random model under-predicts the frequency of zeros in the country-industry trade matrix of the United States. Whereas their model predicts 72% zeros, there are 82% zeros in the data. Their model also under-predicts the fraction of exporters selling to just one country (45% versus 64%). Examining the full bilateral trade matrix of 92 countries, Eaton et al. (2013) also find too few zeros. Their random model predicts that on average each country should export to 77% of its 91 partners, which is higher than the 65% observed in their data.

The under-prediction of zeros in granular models can be eliminated by allowing for non-independence. First, we extend the random model by incorporating a source-specific demand shock that affects all shipments purchased by an importing city. Then we modify the draws-from-urn method to let the probabilities that source s draws a shipment of city d vary according to the data generating process implied by the model. This permits calibration of the parameter governing the variation in the city–source demand shock to exactly match the average compliance rate in the data. This exercise reveals that city–source orientation plays a major role: variation in traditional supply-side determinants of trading patterns accounts for 44% on average of sourcing probabilities of cities, leaving the majority of variation due to heterogeneity in buyers across cities.

Prior work has identified departures from hierarchy structures in exporting and importing. Eaton et al. (2011) find that only 52% of French exporters sell to the most popular export market (Belgium), a violation of the proposition that if a product is profitable in one foreign market, it should also be profitable in easier foreign markets. Antràs et al. (2014) report that 41% of US importers do not import from Canada, the source most commonly used by American importing firms. Looking within exporters, Bernard et al. (2011) find that firms fail to export their most popular product to markets where they export less popular products 33% of the time. Partly to respond to such violations, these papers introduce some form of buyer–seller shocks. While we also conclude that buyer–seller shocks are important, we believe that it is valuable to measure the pervasiveness of hierarchy violations within a data structure that can rule out conventional supply-side explanations. In particular, the absence of a hub-and-spoke transport structure could explain why a firm in southern France might export to Spain, but not Belgium, or why an importer in Texas might choose Mexican suppliers over Canadians. Our disaggregated geographic data, where the hub-and-spoke assumption fits well, combined with conditioning on positive imports of the product, permit us to isolate the role of heterogeneity in the preferred sources of buyers.

The next section establishes the hierarchy prediction in a model with heterogeneous firms and consumers. In Section 3 we define our hierarchy statistic, describe the data, and measure the extent of hierarchy compliance. We present the model with granular consumers in Section 4 and demonstrate that it closely fits the data. Section 5 develops a version of the random model that allows for a cities' shipments to be oriented towards specific source countries. We calibrate the model to match the average amount of compliance observed in the data and assess the importance of buyer–seller idiosyncratic factors. The final section summarizes the results and discusses their implications.

[1] Anderson et al. (1992) fully explore the equivalence between heterogeneous consumers and the representative consumer models.

2. Modelling hierarchy

In standard heterogeneous-firm trade models, exporters generally do not sell to all destinations. An ordering (hierarchy) of suppliers and destinations typically exists in terms of the number of trading partners. Models exhibiting hierarchy predict that an exporter selling to the $(d+1)$th most popular destination also sells to the dth most popular destination.[2] Hierarchy also implies that a destination importing from the $(s+1)$th most popular exporter also imports from the sth most popular exporter.

In models with heterogenous firms linked to varieties, CES preferences and love of variety, varieties of goods can be ordered according to a single firm-specific variable. That variable could be productivity or quality. It can also be a composite of several underlying factors. Here we model the firm variable as quality-adjusted delivered unit costs to a specific market. Another ingredient of hierarchy models is that not all varieties are sold in every market. Varieties are not sold if they are priced above the level that chokes off all demand or if the destination market is sufficiently small and/or distant such that firms cannot cover the fixed costs of exporting to that destination.[3]

We formalize the conditions under which hierarchy obtains by considering a hub-and-spoke transportation system where imports first flow to a hub in a destination province and then travel to individual cities via the spokes. This assumption matches our data well. Defining the provincial hub at the province–good level as the port though which shipments most frequently transit, we calculate that 87.0% of shipments flow through the hub. The provincial hub is the only entry point for 78.8% of city–good combinations.[4] For the remaining 21.2% of cities that import a good through multiple ports or a single port other than the hub, 77.4% of their imports enter through the provincial hub. Even when goods destined for different cities enter through distinct ports, they may have all transited through a natural geographic feature such as the mouth of the Yangtze River. The key assumption is that no source country has a "short cut" it can take to reach the final city destination.

We generate hierarchy in a heterogenous-firm model where consumers purchasing their preferred varieties give rise to CES preferences. We identify the assumptions that lead to hierarchical sourcing patterns. Our analysis focuses on a particular good sold to cities in a specific province. Thus, all variables in the ensuing analysis are province–good specific but, for convenience, we omit notation indicating the good and province.

2.1. Consumers

Consumer utility depends on the amount consumed of a differentiated good (q_j) and a numeraire good (q_0). Firms offer unique varieties and are identified by the variety j they offer. Consumers are located in cities denoted d (destination) and firms are located in a country denoted s (source). To avoid excessive subscripting, we use the most disaggregated subscript applicable and suppress the city d and source country s that are relevant for each i and j. The marginal utility associated with variety j depends on a common term, β_{sd}, capturing the general preference consumers in city d have for all varieties from source s, and an idiosyncratic term, ϵ_{ij}, capturing individual tastes for varieties. Following Anderson et al. (1992, p. 86), the conditional direct utility function is Cobb–Douglas:

$$\tilde{U}_{ij} = (q_{ij}\beta_{sd}\epsilon_{ij})^{\alpha} q_{i0}^{1-\alpha}.$$

Consumer i maximizes utility subject to income y_i and prices p_{dj}, generating (log) conditional indirect utility:

$$\tilde{V}_{ij} = \alpha \ln y_i - \alpha \ln p_{dj} + \alpha \ln \beta_{sd} + \alpha \ln \epsilon_{ij} \tag{1}$$

where p_{dj} is the price of firm j's variety in city d. Consumer i selects the single variety of the differentiated good that leads to the highest $\tilde{V}_{ij}$. The reason consumers in the same market d make different choices is because ϵ_{ij} is a taste draw that causes differences in the indirect utility attached to each variety. It would be isomorphic to allow the variation to enter as an ij price shock, which could arise from the (unobserved) history of prior transactions between the buyer and seller. The key point is that buyers' indirect utilities vary across suppliers.

Dividing by α and assuming ϵ_{ij} is distributed Fréchet with shape parameter θ, the probability consumer i chooses variety j is

$$\pi_{ij} \equiv \mathbb{P}\left[\tilde{V}_{ij} > \tilde{V}_{ih}\ \forall\ h \neq j\right] = \frac{(p_{dj}/\beta_{sd})^{-\theta}}{\sum_{h\in\mathcal{J}_d}(p_{dh}/\beta_{sd})^{-\theta}} \tag{2}$$

where $\mathcal{J}_d$ is the set of firms who offer their products to consumers in city d.

With Cobb–Douglas utility, consumers spend a constant fraction (α) of their income on the preferred variety of the differentiated good. Thus, expected demand for variety j over all consumers in destination d is

$$\mathbb{E}[Q_{dj}] = \alpha Y_d \pi_{ij}/p_{dj} \tag{3}$$

where $Y_d \equiv \sum_i y_i$ is the sum of the incomes of consumers in city d.

When there is a finite number of consumers, the realized volume of sales, Q_{dj}, will vary according to the draws received by the granular consumers in market d. With a continuum of consumers, aggregate demand becomes certain. The key outcome of the Anderson et al. (1992) modelling assumptions is that the continuum of heterogeneous consumers delivers the same demand function as a representative consumer with CES preferences. One merely replaces θ with $\sigma-1$, the elasticity of substitution between varieties for the representative consumer. The novelty here is to consider a granular consumer version of the model.

2.2. Producers

As in Eaton et al. (2013), there are a finite number of firms who receive a unit input requirement draw a_j reflecting the number of bundles used per unit of output by the firm. The realization of the lowest draw of a_j in each country s is denoted a_s^L. The cost of each input bundle in source country s is c_s and τ_{sd} represents an iceberg form transport cost from source s to destination city d. Exporters from s therefore have delivered unit costs to city d given by $\tau_{sd}c_s a_j$.

Firms learn their type prior to the entry and pricing decisions, but the consumers' ϵ_{ij} draws are realized afterwards. The expected profits of firm j selling to city d are given by variable profits minus fixed costs, F_{sd}:

$$\mathbb{E}[\Pi_{dj}] = (p_{ij} - \tau_{sd}c_s a_j)\mathbb{E}[Q_{dj}] - F_{sd}.$$

When firms are small enough to ignore the effect of their price on the price index, maximization of expected profits leads to the price

[2] This definition corresponds to that in Eaton et al. (2011), p.1457.

[3] Helpman et al. (2008) assume a continuum of firms and an upper support for the productivity draw to generate zero trade flows between some countries. Eaton et al. (2013) consider an integer number of firms. Melitz and Ottaviano (2008) do not assume fixed costs but their linear demand model yields hierarchy because marginal costs of some varieties exceed the "choke" price where demand is zero.

[4] These cities account for 47.2% of shipments, indicating that they tend to have fewer shipments than cities that buy goods through multiple hubs.

charged by firm j to consumers in d being a constant markup over marginal costs:

$$p_{dj} = \tau_{sd} c_s a_j[(\theta+1)/\theta]. \tag{4}$$

Substituting the probability Eq. (2) into the demand Eq. (3) and then using the optimal price from Eq. (4), expected profits become

$$\mathbb{E}[\Pi_{dj}] = \lambda[\tau_{sd} c_s a_j/\beta_{sd}]^{-\theta} Y_d P_d^{\theta} - F_{sd}, \tag{5}$$

where $P_d = \left[\sum_{j\in\mathcal{J}_d}(p_{dj}/\beta_{sd})^{-\theta}\right]^{-1/\theta}$ and $\lambda \equiv a\theta^{\theta}(\theta+1)^{-(\theta+1)}$. Firms enter city d, offering their variety at price p_{dj}, if and only if $\mathbb{E}[\Pi_{dj}] \geq 0$.

2.2.1. Separability and hierarchy

Hierarchy predicts that all cities that purchase an imported good will buy from the source country offering the variety possessing the lowest quality-adjusted cost. Also, if a city imports from a firm in a source with higher quality adjusted costs, it would also import from the source with the lowest quality-adjusted costs. As currently specified, the presence of terms β_{sd} and τ_{sd} results in cities having different views of the best variety, implying that hierarchy need not obtain. Moreover, differing values of F_{sd} can cause arbitrary variation in whether sales are profitable between pairs of sources and destinations. For hierarchy, the profit function must not contain *sd* terms. Under assumptions that separate the *sd* terms into *s* and *d* terms, we can solve for the critical level of quality-adjusted delivered costs that generates profitable sales to city d as a function of d characteristics only. The assumptions needed to achieve this separability are listed below.

1. β_{sd}: Perceived quality for goods in s does not depend on d and the term becomes β_s.
2. τ_{sd}: We employ a hub-and-spoke transportation system where the iceberg trade cost factor is expressed as $\tau_{sd} = T_s t_d$, where T_s reflects the costs of the good travelling from s to the hub and t_d the costs of going from the hub to d.
3. F_{sd}: We follow Arkolakis (2010) in assuming that fixed marketing entry costs are Cobb–Douglas in home and foreign inputs. In particular, we assume that f_d input bundles are required as fixed costs to support positive levels of exporting. Each fixed cost bundle combines inputs from the home country and inputs from the destination market according to a Cobb–Douglas form with share parameter δ. The same home factor prices, c_s, and unit factor requirements, a_j, that govern production costs also apply to fixed costs. Moreover, the cost of supplying factor services from home country s remotely in d is governed by the same trade costs, $\tau_{sd} = T_s\tau_d$, that apply to shipments of goods. Taking these assumptions together we obtain multiplicatively separable fixed market entry costs:

$$F_{sd} = f_d(T_s\tau_d c_s a_j)^{\delta} w_d^{1-\delta}, \tag{6}$$

where w_d denotes destination-level factor costs.

Substituting the expressions for β_{sd}, τ_{sd}, and F_{sd} into Eq. (5) yields

$$\mathbb{E}[\Pi_{dj}] = \lambda[T_s\tau_d c_s a_j/\beta_s]^{-\theta} Y_d P_d^{\theta} - f_d(T_s\tau_d c_s a_j)^{\delta} w_d^{1-\delta}. \tag{7}$$

Setting profits equal to zero and solving for the quality-adjusted costs of delivering the good to the hub yield the critical level, C_d^*:

$$C_d^* = \frac{1}{t_d}\left[\frac{\lambda Y_d P_d^{\theta}}{f_d w_d^{1-\delta}}\right]^{\frac{1}{\delta+\theta}}. \tag{8}$$

Firms will offer their variety in city d if their quality-adjusted costs delivered to the hub ($C_j \equiv a_j c_s T_s/\beta_s$) are below this critical value, C_d^*. This critical cost depends only on d-specific attributes. In particular, C_d^* is increasing in demand Y_d and the price index P_d but decreasing in local wages and the transport costs from the provincial hub. The set of firms selling to d, denoted $\mathcal{J}_d$ is defined as the set for which $C_j < C_d^*$.

Based on these characteristics we can order destinations within a province from easiest (highest C_d^*) to toughest (lowest C_d^*). The basic idea of hierarchical sourcing is that a supplier who is efficient enough to export to a tough destination will also export to all easier destinations. This idea is predicated on the absence of randomness on the part of consumers. Otherwise a firm could expect to have profitable sales but, because of unfavorable consumer ϵ_{ij} draws it might not *realize* any sales. In order to equate realized market shares with expected shares, we need to follow the literature that, by using CES, makes an implicit assumption of a continuum of consumers. Later we relax the continuum assumption and show how granular consumers can account for hierarchy violations even in a model that preserves the three separability assumptions.

To determine whether source s sells to a city, it is sufficient to focus on whether it is profitable for the lowest cost firm in s to sell there. For each s, we denote the quality-adjusted delivered costs to the provincial hub of the most profitable firm in each s as $C_s^L = c_s T_s a_s^L/\beta_s$. Therefore, source s sells to city d if $C_s^L \leq C_d^*$.

Fig. 1 depicts hierarchy in two ways. The vertical axis of the left panel shows profits and the horizontal axis shows delivered unit costs C_s. Profit schedules are displayed for three cities located in the province with different levels of demand: 5, 10, or 20. The intersection of each profit schedule and the horizontal zero line identifies the critical level of costs that generate zero profits to that particular destination, C_d^*. The figure also identifies the lowest cost firm, C_s^L, in each source country. The figure shows that the largest destination imports from all four source countries because $C^*(20) > C_s^L$ for $s = 1,2,3,4$. Smaller markets import from fewer sources. The hierarchy prediction is more easily visualized in the right panel of Fig. 1. Bigger cities buy from more sources, but all cities buy from source 1.

3. A hierarchy statistic

We develop a hierarchy statistic to measure the extent that import patterns comply with the hierarchical sourcing prediction of the model. It is calculated as the share of cities that import from the top source of the good in the province. Under the conditions laid out in the previous section, this statistic should equal one. A hierarchy statistic of one is also implied by Ricardian comparative advantage where buyers purchase the lowest cost product. Under the hub-and-spoke assumption all cities should buy from the country whose product arrives at the provincial hub at the lowest cost. Likewise, models featuring love of variety, the Armington assumption of national product differentiation, and no fixed costs will also result in this statistic equalling one as cities would buy *all* varieties.

3.1. Data

We examine the predictions of the models using data on import transactions collected by the Chinese Customs Office for 2006. On a monthly basis, we observe each firm's imports by detailed product classification (CN8 level), origin country, port of entry, and destination city in China.[5] Customs declaration forms ask importers to report the "destination within borders". The official website for the

[5] The harmonized system establishes harmonized classifications out to six digits. Thus, the first six digits in the CN8 correspond to the harmonized system. The last two digits are China-specific classifications.

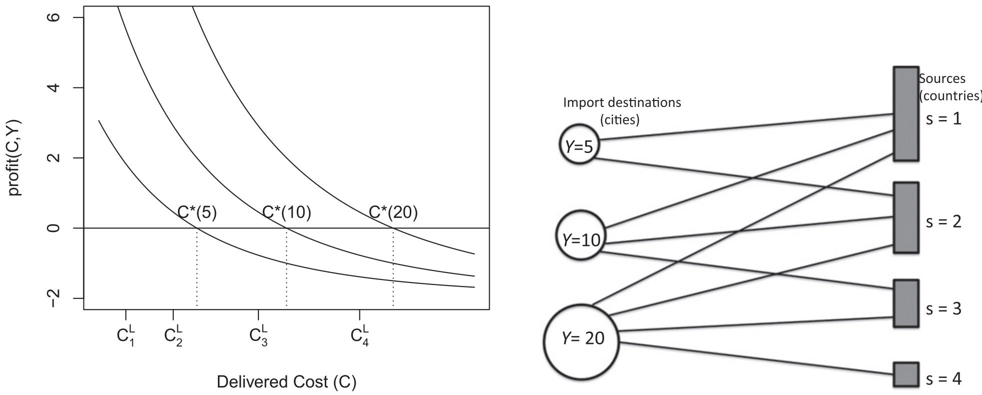

Fig. 1. Hierarchical sourcing.

national exam for customs brokers defines this item as the *known place within China for consumption, usage, or the final destination of the trip*. It need not be the port of entry, which is listed separately.

Table 1 lists information on China's top 20 imported goods according to value. We show the 2006 import value, the number of source countries (#Src), and the detailed product description. We also provide two product classifications: The system of national accounts (SNA) classification of intermediate, capital, and consumption goods and the Rauch (1999) classification of differentiated, reference price, or organized exchange products. 8-digit classifications can be very detailed; the table shows five separate CN8 categories for integrated circuits. The largest imported product is petroleum, and China sourced it from 46 different countries. Most goods were sourced from a large number of countries. Exceptions are soy beans and aircraft between 15 and 45 tons which were sourced from 8 and 4 countries, respectively.

Our primary unit of analysis will be imports of individual cities for specific goods. We have data for 521 cities and 7077 products. The total number of city–product combinations with positive imports is 334,955 and the number of province–good combinations is 82,817. Since our analysis focuses on goods imported by cities from foreign sources for local consumption, we exclude re-imports where the source country is listed as China and imports into bonded warehouses.[6]

Table 1
Top goods, 2006.

CN8	\$bil	#Src	SNA	Rauch	Description
27090000	66.4	46	Int	Org	Petroleum oils (crude)
85422119	39.5	64	Int	Dif	Mon. integ. circuits, digital, ≤0.18 μm
90138030	25.8	47	Cap	Dif	Liquid crystal display panels
85422900	15.7	82	Int	Dif	Monolithic integrated circuits, not digital
85422129	12.2	65	Int	Dif	Mon. int. circ., dig., 0.18 < wid. ≤ 0.35μm
26011120	11.8	27	Int	Org	Iron ores and concentrates, non-agglomerated
85422199	10.5	69	Int	Dif	Mon. integ. circuits, dig., >0.35μm
27101922	9.0	27		Dif	Fuel oils number 5–7
12010091	7.5	8	Int	Org	Soya beans, whether or not broken
84733090	7.1	71	Int	Dif	Computer parts and accessories
85426000	6.9	59	Int	Dif	Hybrid integrated circuits
85299020	6.4	42	Int	Dif	Hand-held wireless telephone parts
85422121	6.3	28	Int	Dif	Mon. int. circ., dig., 0.18 < wid. ≤ 0.35, orig. film
29173610	6.1	18	Int	Dif	Terephthalic acid and its salts
88024010	6.1	4	Cap		Aircraft between 15 and 45 tons
84717010	6.1	46	Cap	Dif	Computer hard drives
26030000	5.9	35	Int	Ref	Copper ores and concentrates
84798990	5.7	52	Cap	Dif	Machines and mechanical appliances N.E.S.
74031100	4.9	34	Int	Org	Cathodes of unwrought copper
52010000	4.8	61	Int	Org	Cotton, not carded/combed

3.2. Hierarchy compliance

In the model, a city complies with hierarchy when it imports from the source offering the lowest quality-adjusted price at the hub. Lacking data on quality, we infer the top supplying country ("source 1") for a good in a province as the source that is most often chosen by the cities in the province. While it may seem more natural to identify the top source as the country with the highest import market share, that method confounds source country size with cost. To see this, multiply firm-level demand and price equations (Eqs. (3) and (4)), and sum across the set of firms who offer their products to consumers in city d (set $\mathcal{J}_{sd}$) to obtain the value of exports to city d by source s:

$$M_{sd} = \sum_{h \in \mathcal{J}_{sd}} p_{hd} Q_{hd} = \gamma Y_d P_d^{\theta} [\tilde{a}_{sd} \tau_{sd} c_s / \beta_{sd}]^{-\theta}$$

where $\tilde{a}_{sd} \equiv \left(\sum_{h \in \mathcal{J}_{sd}} a_h^{-\theta}\right)^{-1/\theta}$ is the inverse of a CES index of the productivity of the firms from country s that have entered city d and $\gamma \equiv \alpha[(1+\theta)/\theta]^{-\theta}$. To simplify, suppose that all N_{sd} firms in set $\mathcal{J}_{sd}$ have the same input requirements, equal to a_s^L. This implies $\tilde{a}_{sd} = a_s^L N_{sd}^{-1/\theta}$. Applying the hierarchy assumptions and substituting into the equation yields $M_{sd} = \gamma Y_d P_d^{\theta} t_d N_{sd} [C_s^L]^{-\theta}$. Since M_{sd} increases proportionately with N_{sd}, it is possible for source country s to have the highest volume of exports to d because it hosts the most firms even though it does not host the lowest cost exporter of the good. By counting the frequency with which cities source positive amounts from each source, we have a popularity rating that orders countries reliably in terms of their least cost suppliers.[7]

Table 2 summarizes information on sources of goods for each province. The first column lists the provinces ordered by total imports in 2006, shown in column (2). Guangdong is the largest importer, importing \$171 billion. Column (3) and column (4) contain the number of goods imported by the province and the number of cities that import goods. We observe that provinces with more cities tend to import more goods with a higher total value.

[6] Of 2006 imports, 6.1% pass though bonded warehouses on their way to other countries. Another 4.1% go to other types of bonded warehouses and may not be consumed in the city where the warehouse is located.

[7] Eaton et al. (2011) and Bernard et al. (2011) also use popularity to determine hierarchy rankings.

Table 2
Top sources.

Province	\$mn	#(cn8)	#(city)	Top source	%
Guangdong	176.1	6184	24	Hong Kong	25.6
Jiangsu	115.8	5532	27	Japan	40.3
Shanghai	73.4	6136	22	Japan	41.9
Shandong	45.2	5109	30	Korea	44.8
Zhejiang	45.1	5007	24	Japan	32.1
Beijing	41.1	5584	19	USA	22.7
Tianjin	26.5	4812	19	Japan	31.7
Liaoning	21.8	4833	21	Japan	39.3
Fujian	18.9	4657	12	Taiwan	45.9
Hebei	8.2	3113	12	Japan	24.2
Heilongjiang	7.1	2132	21	USA	19.6
Hubei	6.1	2655	18	Japan	21.4
Jilin	5.5	2495	17	Korea	26.1
Anhui	5.4	2317	18	Japan	24.1
Sichuan	4.8	2592	24	USA	25.3
Henan	3.7	1887	23	Japan	22.4
Guangxi	3.7	1648	15	Taiwan	16.6
Inner Mongolia	3.4	1118	15	USA	22.4
Yunnan	3.2	1493	21	USA	20.0
Jiangxi	3.2	1722	13	Japan	18.1
Xinjiang	3.2	1171	16	USA	30.8
Shanxi	2.9	1352	12	Germany	20.9
Hunan	2.8	1787	20	Japan	24.5
Gansu	2.7	663	13	Germany	25.8
Shaanxi	2.5	1958	11	USA	23.7
Hainan	2.2	1315	3	USA	16.6
Chongqing	2.2	1806	27	Japan	31.6
Guizhou	0.9	754	10	Japan	21.8
Ningxia	0.5	451	4	Germany	28.8
Qinghai	0.4	346	5	Germany	22.5
Tibet	0	188	5	Nepal	50.0

Column (5) in Table 2 identifies the country that is most frequently the top source across all the goods imported in the province.[8] The last column shows the frequency for which that country is the top source across the goods. Overall, the top sources tend to be the United States, Japan, and Germany. We observe some economic geography influencing the choice of top source as Nepal is the top source for Tibet and Hong Kong is the top source in Guangdong. Guangdong provides a case where measuring top source based on frequency generates different results than a definition based on highest market share. Based on the latter method, Japan is the top source. Arguably, Hong Kong tends to host the lowest cost suppliers but the larger number of Japanese exporting firms results in higher market shares for Japan. Cities in Guangdong with small markets would import from the low-cost Hong Kong suppliers before they would import from more numerous, higher cost Japanese suppliers.

We now calculate the share of cities that import a narrowly defined good from the top source of that good in the province. Define K as the count of complying cities, i.e. those that import the good from the top source at the province level. Let D be the number of cities in a province that import the good from any source. The hierarchy statistic is simply $h_1 = K/D$. In order to have a sufficient number of cities to reliably identify the top source, we restrict the sample for most of our analyses to $D \geq 4$. The number of cities in a province ranges from 4 to 28 with a mean and median of 10. The $D \geq 4$ requirement reduces the number of goods from 7077 to 5239 and the number of province–good combinations to 29,459. Nevertheless, the reduced sample still accounts for 82.5% of Chinese imports.

Fig. 2 presents the histogram of the hierarchy statistic. In models predicting hierarchy, the expected value of hierarchy compliance is one. As depicted in the figure, the incidence of all cities importing from the top source of a good in the province is only 8%. For the other

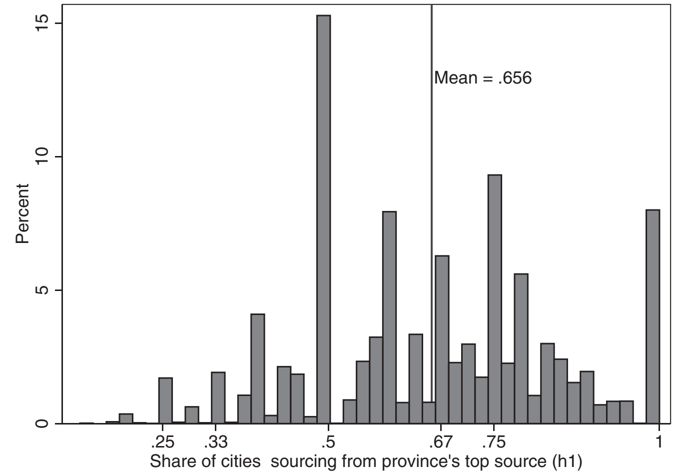

Fig. 2. Distribution of hierarchy statistics (h_1) for 29,459 province-goods.

92% of the province–good observations, one or more cities do not comply. Mean compliance is 0.66. Because there are many cases with relatively few cities importing goods in a province, spikes appear in the distribution at 0.25, 0.33, 0.5, 0.67, and 0.75. The data reveals substantial deviation from the prediction that all firms should import from the top source in the province.

4. Granular consumers

A straightforward way to generate non-compliance with hierarchy is to replace the continuum assumption with granularity in the demand side of the model. With a finite number of consumers, the top source may not realize trade with all cities due to randomness. Accordingly, we now abandon the continuum of consumers that has been implicit in past work using the Dixit–Stiglitz model of demand.

In this version of the model, we assume $\beta_{sd} = \beta_s$ and replace city-level fixed costs with hub-level fixed costs. Firms offer goods in the hub if the sum of the expected value of variable profits exceeds the hub fixed costs. This will generate a set of $\mathcal{J}_s$ firms from each country s that sell to all the cities connected to a given hub. The hub-specific fixed costs can rationalize why not all countries sell in all provinces.

With shipments representing consumers, the probability that firm j fills a shipment order i is

$$\pi_{ij} = \frac{(p_{dj}/\beta_s)^{-\theta}}{\sum_{h \in \mathcal{J}_s}(p_{dh}/\beta_h)^{-\theta}}.$$

Substituting $p_{dj} = [\tau_{sd} c_s a_j](\theta + 1)/\theta$ and maintaining hub-and-spoke transportation costs ($\tau_{sd} = T_s t_d$) yield,

$$\pi_{ij} = \frac{(c_s T_s a_j/\beta_s)^{-\theta}}{\sum_{h \in \mathcal{J}_s}(c_h T_h a_h/\beta_h)^{-\theta}}. \tag{9}$$

This expression lacks d-specific terms because hauling costs from the hub to a city do not change the relative prices of products (t_d factors out). As a result $\pi_{ij} = \pi_j$ for all shipments i in city d. That is, no matter which city in a province an order emanates from, it has the same probability of being filled by firm j.

The probability that a shipment order is filled by any of the $\mathcal{J}_s$ firms from source s can be expressed as

$$\pi_s = \sum_{j \in \mathcal{J}_s} \pi_j = \frac{(\tilde{a}_s c_s T_s/\beta_s)^{-\theta}}{\sum_h(\tilde{a}_h c_h T_h/\beta_h)^{-\theta}} \tag{10}$$

[8] We use the value of imports to break ties in cases where sources are equally popular.

where $\tilde{a}_s = \left(\sum_{h\in\mathcal{J}_s} a_h^{-\theta}\right)^{-1/\theta}$ is the inverse of a CES index of the productivity of the firms from country s that sell to the hub. With these foundations for a common π_s, we can now proceed to the calculation of the expected value of hierarchy compliance with independent, identical draws.

4.1. The draws-from-urn method

Under the assumption of independent and identically distributed (IID) shipments, the likelihood that at least one of n_d shipments to city d will be supplied from source s is $1-(1-\pi_s)^{n_d}$. Armenter and Koren (2014) and Eaton et al. (2013) use versions of this formula to predict zeros in trade. Armenter and Koren (2014) liken the process to throwing balls into bins.

For each province–good combination, we denote the probability of a shipment coming from the top source as π_1. The expected h_1 for a province with D cities is

$$\mathbb{E}[h_1] = \frac{\sum_d 1-(1-\pi_1)^{n_d}}{D}. \tag{11}$$

This expectation, which we refer to as the "balls-and-bins" formula, is increasing in both π_1 and n_d.

We wish to test whether the average h_1 observed in the data is consistent with expected compliance predicted by the granular model. The simple formula shown in Eq. (11) appears promising for this purpose—if we had a measure of π_1. Unfortunately, with granular shipments, the fraction of balls received by the observed top source, which we denote as x_1, is a biased estimate of π_1, the true probability of the observed top source. Furthermore, h_1 is a biased estimate of expected compliance with the observed top source. Order statistics bias poses a problem for both measures.[9] The top source is determined by the realizations of the shipments. The source observed to serve the most cities is most likely to be the true top source but there will be cases when another source's luck of the draw yields the highest compliance rate and shipment share. This implies that the fraction of balls received by the observed top source (x_1) will exceed π_1 on average. Using observed x_1 rather than π_1 raises expected compliance generated by Eq. (11). At the same time, the observed compliance, h_1, on average will exceed true expected compliance with the observed top source. Simulations we conducted to investigate the two biases reveal that they do not cancel each other. This compels us to use an alternative method to determine expected compliance with the observed top source.

$\mathbb{E}[h_1]$ can be calculated without knowing π_1 if we condition on the number of realized shipments. Under independence, each shipment (ball) is equally likely to be one of the shipment orders filled by any given source. Knowing the distribution of shipments across the cities in the province, we can use combinatorics to calculate the probability that exactly k cities are represented by at least one shipment among the shipments in the source. To understand the intuition, imagine that the shipments of cities are different coloured balls in an urn and the top source receives n_1 shipments. To calculate expected compliance for a given city, we simply need to compute the likelihood that at least one of the n_1 draws from the urn has the color corresponding to that city.

We express the computation of conditional $\mathbb{E}[h_1]$ as

$$\mathbb{E}[h_1 \mid n_1] = \sum_{k=1}^{D} (k/D)\mathbb{P}(k \mid n_1) \tag{12}$$

where $\mathbb{P}(k \mid n_1)$ depends on the vector of city balls (n_d) for the province–good in question. The term k/D reflects the different levels of compliance and ranges between $1/D$ and 1. $\mathbb{E}[h_1 \mid n_1]$ is given by these compliance possibilities weighted by their probability of occurring (which may be zero is some cases). The probabilities $\mathbb{P}(k \mid n_1)$ are given by the ratio of the number of ways that k cities could buy from source 1 divided by the total number of ways to achieve n_1 shipments to the top source, i.e. $\binom{N}{n_1}$.

It may be helpful to illustrate the determination of the $\mathbb{P}(k \mid n_1)$ with a simple example which nevertheless matches the moments of our overall dataset fairly closely. Suppose that there are three cities with 1 red, 2 green, and 4 blue balls, respectively. Suppose the top source has received 3 balls. For this outcome to be associated with $h_1 = 1/3$, requires that 3 of the 4 blue balls land in source 1. There are four possible ways to obtain this "BBB" configuration out of $\binom{7}{3} = 35$ total ways to have 3 of 7 balls land in source 1. Thus $\mathbb{P}(k = 1 \mid n_1 = 3) = 4/35$. Compliance of all three cities requires an "RGB" configuration in source 1, which can be obtained 8 different ways. There are 23 ways to have 2/3 compliance: "RGG" (1), "RBB" (6), "GBB" (12), "GGB" (4). Thus $\mathbb{E}[h_1 \mid n_1 = 3] = [(1/3)\times 4 + (2/3)\times 23 + (3/3)\times 8]/35 = 0.705$. This is 0.04 higher than the 0.665 which arises from applying Eq. (11) prediction treating the outcome $x_1 = 3/7$ as if it were π_1, the true probability of choosing source 1.

Using combinatorics as in the above example is difficult to code into an algorithm applicable to the myriad set of shipment-city distributions of n_d present in our 29,459 province-goods. Fortunately, it is straightforward to obtain the expected values—and the whole distribution of potential compliance rates—via Monte Carlo simulation using a sampling function. For each province–good, we draw n_1 (the shipment count of the observed top source) balls from a set of all $N = \sum_d n_d$ balls, keeping track of which balls come from which cities. A city complies if at least one of its balls is drawn. The fraction of cities included in the drawn sample gives h_1 for that repetition. We repeat 100,000 times and take the average as $\mathbb{E}[h_1]$.

4.2. Comparing of observed and expected h_1

To implement the method, we need to define a shipment, our concept of the granular consumer. As described in the appendix of Armenter and Koren (2014), a US exporter's shipping declaration defines shipments as "all merchandise sent from one USPPI [firm] to one foreign consignee, to a single foreign country of ultimate destination, on a single carrier, on the same day.". Each shipment has a unique product code. We utilize the available information in the Chinese customs data to define shipments as similarly as possible. Thus, we define a shipment by disaggregating imports by month, country of origin, CN8 good classification, importing firm, route, transport mode, and city-zone. Thus, shipments of a narrowly defined good from source s to city d would be counted separately if they occurred in a different month, were received by a different firm, entered via a different port, were routed through a different country along the way to China, were transported by a different mode (air, sea, ground), or ended up in a different zone in the city (e.g. Shenzhen SEZ vs Shenzhen city). This measure is more aggregated than the individual customs declarations used by Armenter and Koren (2014) since it lumps together all shipments that occurred in the same month. Our 2006 data contains 7.9 million import shipments (as we define them) compared to 21.6 million customs declarations for the US in 2005. The median size of our shipments is \$3221, about twice the \$1800 value in the US data.[10]

[9] We thank an anonymous referee for pointing out the bias associated with using x_1 as an estimate of π_1.

[10] We thank Miklos Koren for providing us the US shipment size data.

At our level of analysis—province-goods—the data exhibits extreme granularity. The draws-from-urn method draws the number of shipments realized by the top source (n_1) from an urn consisting of total shipments (N). In our sample of province-goods with four or more importing cities, the average, median and 25th percentile value of n_1 are 84, 20, and 8. The corresponding statistics for N are 255, 57, and 23. The distribution of shipments across cities (n_d) is also very skewed: On average across province-goods, the within province–good mean and median n_d are 19.6 and 7.6. One quarter of cities in our sample import a single shipment. These statistics underscore the importance of a method that is robust to random variation.

Table 3 provides information about the correspondence between h_1 and $\mathbb{E}h_1$ for the full sample of 29,459 province–good combinations and different subsets of the data based on types of goods. The goods are defined as consumption, intermediate and capital goods according to the SNA as well as differentiated, reference, and organized exchange goods as classified by Rauch (1999). The first column of the table reveals that h_1 is highest for organized exchange goods and lowest for consumption goods. Hierarchy violations are common for all types of goods, with average non-compliance ranging from 27% to 38%.

We compare average h_1 to two measures of $\mathbb{E}h_1$ in Table 3. Column (3) reports the average values generated by the Monte Carlo simulations of the draws-from-urn method whereas the next column uses the formula shown in Eq. (11) assuming $\pi_1 = x_1$ (labelled "b&b").[11] Across all province-goods, the former has an average value of 0.73 and the latter 0.72. The two measures of $\mathbb{E}h_1$ vary similarly across the subsamples with the b&b formula always producing a slightly lower (or equivalent after rounding) $\mathbb{E}h_1$ than what is obtained in the draws-from-urn method. It appears that even in a sample with significant granularity such as ours, one can use the observed shipment share as the true probability of receiving a shipment and obtain a remarkably similar but very slightly downwardly biased estimate of $\mathbb{E}[h_1]$.

Table 3 also shows the average shipment share of the observed top source, x_1 and the average of the median number of city balls, n_d. In the balls-and-bins formula, both these variables increase $\mathbb{E}h_1$. For example, expected compliance is high for organized exchange goods due to a high value of x_1. It is high for intermediates and differentiated goods because the median city has a large number of shipments, making it likely that shipments from many different cities will be represented in the shipments realized by the top source.

The table reveals $\mathbb{E}h_1$ explains variation in h_1: Goods predicted to have the highest compliance (intermediates, organized) have the highest actual compliance. However, in every category, hierarchy compliance is less than the expectation from the random sourcing model. Thus, while the random model with independent shipments provides a reasonable prediction of actual compliance, it appears to be upwardly biased.

The last two columns provide statistical evidence that observed compliance is significantly lower than the expectation of the granular model. For each province–good, there is a minimum and maximum feasible h_1 conditional on the number of shipments filled by source 1. The minimum may be one divided by the number of cities: a situation where all the shipments realized by the top source came from the same city. There will be cases, however, where any single source does not have sufficient shipments to account for the realized total shipments and the minimum h_1 will be higher. The maximum h_1 is often 1 but will be less when there are more cities than top-source shipments (not all cities could contribute a shipment). With 100,000 repetitions, the simulation generates the minimum and maximum feasible values of h_1 and we record the frequency that these values obtain and take their average across our province-goods. We also know the frequency with which the actual data corresponds to the minima and maxima. We define $\mathbb{R}_{\min}$ and $\mathbb{R}_{\max}$ as the ratios of the observed frequency of minima and maxima to their frequencies predicted in the simulation. If the IID model over-predicts compliance, there would be too many actual values corresponding to the minimum and $\mathbb{R}_{\min}$ will be greater than one. Also, $\mathbb{R}_{\max}$ would be less than one, indicating that perfect compliance occurs less often than randomness would predict. Columns (7) and (8) show these ratios and provide clear evidence that the $\mathbb{E}h_1$ from the granular sourcing model with IID shipments over-predicts observed h_1.

Table 3
Hierarchy statistics and their expected values.

Type of good	obs	h_1	$\mathbb{E}h_1$		n_d	x_1	$\mathbb{R}_{\min}$	$\mathbb{R}_{\max}$
			Urn	b&b				
All	29,459	0.66	0.73	0.72	0.40	7.6	2.48	0.57
Intermediate	18,995	0.67	0.76	0.75	0.41	9.0	3.19	0.49
Capital	6084	0.63	0.67	0.66	0.37	4.9	1.45	0.77
Consumption	4203	0.62	0.68	0.67	0.38	5.3	1.81	0.70
Differentiated	24,700	0.66	0.72	0.72	0.40	7.5	2.33	0.59
Reference	3509	0.66	0.77	0.76	0.41	8.3	3.75	0.42
Organized	505	0.73	0.80	0.80	0.47	7.5	2.56	0.64

Observations (obs) do not add up to the total because there are 29 and 134 cn8 categories missing SNA and Rauch classifications. h_1, $\mathbb{E}h_1$, and x_1 average across province-goods. Expectations for urn based on draws-from-urn simulation, "b&b" uses Eq. (11) with $\pi_1 = x_1$. n_d is the average of the median (across cities) number of shipments for the province–good. $\mathbb{R}_{\min}$ and $\mathbb{R}_{\max}$ are the ratios of observed to predicted shares of h_1 that take on the minimum and maximum feasible values.

This over-prediction is a robust feature of the data. Table 4 reports average h_1 and $\mathbb{E}h_1$ for different subsets of cities and methods of identifying the top source. In the first two rows, we consider samples with at least 3 or 5 importing cities for each province–good combination. In the third row, we confine the analysis to the cities that import a particular good exclusively through the provincial hub. In the last row, we only consider province–good combinations for which the frequency method of determining top source does not result in a tie for the top source. The table reveals that average h_1 and $\mathbb{E}h_1$ do not change very much across these samples. Average compliance is around two-thirds and always significantly less than $\mathbb{E}h_1$.

The use of central warehouses located in one city to distribute products to other cities may cause us to incorrectly identify non-compliance. This occurs when a city directly imports a good from a source country other than source 1 and accesses the good from source 1 via a central warehouse located in a different city. Replacing city-specific warehouses with a single centralized warehouse would

Table 4
Robustness to number of cities, hub-use, ties.

Type of good	obs	h_1	$\mathbb{E}h_1$		n_d	x_1	$\mathbb{R}_{\min}$	$\mathbb{R}_{\max}$
			Urn	b&b				
≥3 cities	37,957	0.66	0.73	0.72	0.42	6.8	1.85	0.62
≥5 cities	23,730	0.66	0.73	0.73	0.38	8.3	3.77	0.52
≥4 cities & hub only	23,085	0.65	0.73	0.72	0.41	6.8	2.67	0.56
≥4 cities & no frequency ties	21,931	0.70	0.74	0.73	0.41	7.7	1.75	0.68

Observations (obs) do not add up to the total because there are 29 and 134 cn8 categories missing SNA and Rauch classifications. h_1, $\mathbb{E}h_1$, and x_1 average across province-goods. Expectations for urn based on draws-from-urn simulation, "b&b" uses Eq. (11) with $\pi_1 = x_1$. n_d is the average of the median (across cities) number of shipments for the province–good. $\mathbb{R}_{\min}$ and $\mathbb{R}_{\max}$ are the ratios of observed to predicted shares of h_1 that take on the minimum and maximum feasible values.

[11] Henceforth, we use shipments to break ties when multiple sources sell goods to the same number of cities. This is different than what we used compiling top source information in Table 2 where we broke ties based on value. Since shipments reflect consumers in our model, it makes sense to break ties based on the number of shipments.

be most economical for cities that are geographically clustered. As a robustness check, we calculate h_1 and $\mathbb{E}h_1$ for the municipalities Beijing, Chongqing, Shanghai, and Tianjin. In our data, they are defined as provinces and contain geographically proximate cities. We conjecture that if false non-compliance is an issue, it will be most prevalent in these places. We also classify goods according to whether trade intermediaries account for more than 20% of imports, anticipating that these agents are most likely to use central warehouses.

The results shown in Table 5 do not indicate that warehouse trade is responsible for the low levels of compliance we observe in the full sample. There is less compliance within municipalities, 0.63 as compared to 0.67 for non-municipalities, but half that difference is explained by lower expected compliance. The last two rows show that goods characterized by 20% intermediaries trade or more behave very similarly to the remaining set of goods. Overall, compliance is significantly lower than the random model prediction in all cases shown in the table.

5. Source orientation

The independent shipments granular model predicts significantly more compliance with hierarchy than is observed. In this section we incorporate a city–source shock which allows us to reduce $\mathbb{E}h_1$ to the level observed in the data. The calibrated parameter for this shock implies substantial cross-city variation in buyer preferences.

Importing cities may be oriented towards specific source countries because of correlations in procurement preferences across shipments from the same importing firm and across firms in the same city. Firms appear to prefer particular source countries. Blum et al. (2010) find that trade intermediaries in Chile obtain the vast majority of their imports from just one or two countries. Our data also indicates that firm's shipments tend to go to specific source countries. At the city–good level, among the 51% of the firms with more than one shipment, the mean and median shipment share going to the firm's top source are 60%. This is a significantly higher rate than the average 40% shipment shares for the countries we identify as the top source (see Table 3). One explanation for firms' attraction towards particular source countries is preferences of foreign-owned firms to import from their home country (often from the parent firm). In our data, 39% of importing firms are wholly foreign-owned. A further 17% involve some foreign equity or cooperation. Taken together, these firms account for 56.4% of total Chinese imports. Even domestic, Chinese importers may develop particular familiarity with the business environment in a specific source country, leading to a bias towards transacting with suppliers from that country.

Table 5
Robustness to central warehouses and trade intermediaries.

Type of good	obs	h_1	$\mathbb{E}h_1$ Urn	$\mathbb{E}h_1$ b&b	n_d	x_1	$\mathbb{R}_{min}$	$\mathbb{R}_{max}$
Type of province								
Municipality	10,117	0.63	0.72	0.71	0.39	6.6	2.83	0.54
Non-municipality	19,342	0.67	0.74	0.73	0.41	8.1	2.35	0.58
Share of imports handled by intermediaries								
<20%	14,835	0.66	0.74	0.74	0.40	8.7	2.71	0.54
≥20%	14,624	0.65	0.72	0.71	0.40	6.4	2.28	0.60

Observations (obs) do not add up to the total because there are 29 and 134 cn8 categories missing SNA and Rauch classifications. h_1, $\mathbb{E}h_1$, and x_1 average across province-goods. Expectations for urn based on draws-from-urn simulation, "b&b" uses Eq. (11) with $\pi_1 = x_1$. n_d is the average of the median (across cities) number of shipments for the province–good. $\mathbb{R}_{min}$ and $\mathbb{R}_{max}$ are the ratios of observed to predicted shares of h_1 that take on the minimum and maximum feasible values.

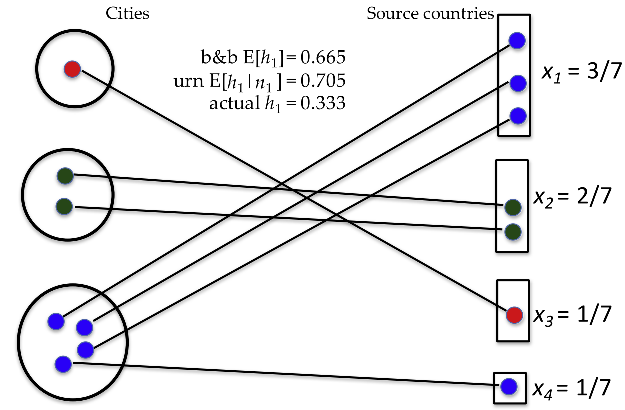

Fig. 3. City–source cost shocks lower $\mathbb{E}[h_1]$. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

Firms of the same nationality have been observed to agglomerate.[12] To investigate the presence agglomeration in our data, we adapt the Ellison and Glaeser (1997) measure of industry geographic concentration to generate (foreign) nationality geographic concentration. For each province, we compute $G_s = \sum_d (z_{sd} - z_d)^2$ where z_{sd} is city d's share of source s's firms and z_d is d's share of *all* foreign firms in the province. The customs data does not list the nationality of foreign firms, and we confine analysis to the 70% of foreign firms for whom we can identify foreign nationality using the 2013 *China Foreign Enterprise Directory*. Ellison and Glaeser (1997) provide a test statistic for whether the concentration is greater than a dartboard null (firms allocated randomly based on city size as measured by the total number of foreign firms). Our measure of concentration is at the province–source level (firms of a particular nationality may or may not exhibit geographic concentration relative to the dartboard null). We find that nationalities are significantly concentrated within provinces in 82% of the 861 province–source cases (5% significance level).[13]

Fig. 3 illustrates the idea of how source orientation lowers expected hierarchy compliance. Each city is associated with a different color ball, reds in the smallest city, greens in the medium city, and blues in the largest. We imagine that blue balls are oriented towards source 1 whereas green balls prefer source 2 and red balls source 3. In an IID model we would expect relatively high compliance given the observation that 3/7 of the shipments selected source 1. The IID formula from Eq. (11) predicts about two thirds compliance (0.665) and combinatorics give a higher amount (0.705). With the high degree of orientation depicted in the figure, balls of particular colors are clustered in different sources, leading to lower compliance than what is expected under independence (one-third in this case).

Our model allows for city orientation to particular source countries through the β_{sd} term. Using Eq. (9) and substituting β_{sd} for β_s,

$$\pi_{sd} = \frac{(\tilde{a}_s c_s T_s/\beta_{sd})^{-\theta}}{\sum_h (\tilde{a}_h c_h T_h/\beta_{hd})^{-\theta}}. \tag{13}$$

As with the ϵ_{ij} term in the individual indirect utility, we model the variation across buyers captured in β_{sd} as a preference term. Just as in that case, it would be isomorphic to replace or combine taste shocks with sd-level cost shocks. In either case, there will be variation in

[12] Head et al. (1995) find evidence at the level of US states for Japan-based manufacturers.

[13] Within province-goods, the data is very sparse and difficult to identify significance. We find that 12.5% of province–good–source combinations exhibit significant concentration.

city-level probabilities of choosing a source—even after controlling for the supply-side term $\tilde{a}_s c_s T_s$.

It is useful to concentrate the source-specific terms and the idiosyncratic terms separately in order to re-express the city–source probability. Following Eaton et al. (2013), define Λ_s as the probability of choosing s evaluated with the city–source shocks held constant (which eliminates them). This returns the probability derived in the sourcing model with independent random shocks:

$$\Lambda_s \equiv \frac{(\tilde{a}_s c_s T_s)^{-\theta}}{\sum_h (\tilde{a}_h c_h T_h)^{-\theta}}.$$

Letting $\nu_{sd} \equiv \beta_{sd}^{\theta}$, π_{sd} can be expressed concisely as

$$\pi_{sd} = \frac{\Lambda_s \nu_{sd}}{\sum_h \Lambda_h \nu_{hd}}. \tag{14}$$

Ideally the unconditional expected probability, $\mathbb{E}[\pi_{sd}]$ should equal the probability in the absence of heterogeneity, i.e. Λ_s. Surprisingly, this reasonable requirement is only met by making particular assumptions on the functional form and parameterization of the ν distributions. Eaton et al. (2013) footnote 21 establishes that assuming ν_{sd} are distributed Gamma with shape and scale parameters Λ_s/η^2 and η^2/Λ_s implies $\boldsymbol{\pi} \sim \mathrm{Dir}(\boldsymbol{\Lambda}/\eta^2)$. "Dir" denotes the Dirichlet distribution, which is the standard formulation employed to generate a vector of probabilities that are themselves random, but have expectations given by the $\boldsymbol{\Lambda}$ vector.[14]

Using footnote 29 of Eaton et al. (2013) to determine the probability of a zero, the probability that a city does not buy from source 1 is given by

$$\mathbb{P}(d \text{ does not comply})_{\text{orient}} = \frac{\Gamma\left(\frac{1}{\eta^2}\right)\Gamma\left(n_d + \frac{1-\Lambda_1}{\eta^2}\right)}{\Gamma\left(\frac{1-\Lambda_1}{\eta^2}\right)\Gamma\left(n_d + \frac{1}{\eta^2}\right)}. \tag{15}$$

Raising the number of shipments increases the probability of hierarchy compliance, just as in the IID case. The difference here is that higher η^2, which raises the variance of the city–source shock, causes the probability of non-compliance to converge to $1-\Lambda_1 < 1$.[15] Since this limiting value is independent of n_d, we see that extreme dispersion of the idiosyncratic city–source shock treats cities as having only a single shipment. In the other extreme, where $\eta^2 = 0$, we obtain the IID probability of non-compliance, $(1-\Lambda_1)^{n_d}$.

If we knew Λ_1, we could use Eq. (15) to calibrate η^2 to match average observed hierarchy across our sample of 29,459 province-goods. We only know *observed* shipment share x_1, a statistic that overstates the true probability for reasons articulated in Section 4.1.

However, we can apply the draws-from-urn approach to calibrate η^2. As with the previous exercise, for each province–good, we draw the number of shipments received by the top source (n_1) from an urn of N total shipments. Unlike before, the probability of drawing a ball from a particular city is not the same across sources. This probability can be obtained as a function of the original city–choice probabilities, π_{sd}, by applying Bayes' theorem:

$$\mathbb{P}(d \mid s) = \frac{\mathbb{P}(s \mid d) \times \mathbb{P}(d)}{\mathbb{P}(s)}.$$

Substituting $\mathbb{P}(s \mid d) = \pi_{sd}$, $\mathbb{P}(d) = n_d/N$ and using the law of total probability to set $\mathbb{P}(s) \equiv \mathbb{P}(s \mid d)\mathbb{P}(d) = \sum_i \pi_{si}(n_i/N)$, we have

$$\mathbb{P}(d \mid s) = \frac{\pi_{sd} n_d}{\sum_i \pi_{si} n_i}. \tag{16}$$

In the absence of idiosyncratic heterogeneity, $\beta_{sd} = \beta_s$ and $\pi_{sd} = \pi_s$. Then, Eq. (16) reduces to $\mathbb{P}(d \mid s) = \mathbb{P}(d) = n_d/N$, which are the same probabilities we used in the draws-from-urn exercise under independence. As η^2 increases, $\mathbb{P}(d \mid s)$ diverges away from the shipment shares and towards a more even $1/D$ share for each city.

In our Monte Carlo simulation, we choose an η^2 and draw from the Dirichlet distribution to generate the matrix of π_{sd} using observed x_s as the maximum likelihood estimates for Λ_s. Plugging in π_{sd} and the number of shipments of each city into Eq. (16) allows us to identify the probability that source 1 draws a ball from city d, $\mathbb{P}(d \mid 1)$. As before, for each province–good we draw the number of shipments realized by source 1 using these probabilities. A city complies if at least one of its balls is drawn and we can calculate h_1. We repeat 1000 times and take the average as $\mathbb{E}[h_1]$. We vary η^2 until the mean h_1 produced by the Monte Carlo experiment matches the observed mean h_1 of 0.66 from the 29,459 province goods with four or more cities. We find that expected and actual hierarchy compliance equal 0.66 when $\eta^2 = 0.70$.

To assess the economic importance of city–source shocks, we calculate the share of variation in π_{sd} that can be accounted for by country fixed effects. Variation in π_{sd} will reflect variation in Λ_s (which captures the whole supply side of the choice) and variation in the source–city preference parameter, β_{sd}, as reflected in η^2. In the above Monte Carlo simulation with $\eta^2 = 0.70$, for each of the 1000 repetitions, we calculate the R^2 for the regression of π_{sd} on source-country fixed effects.[16] We find an average R^2 of 0.44, indicating that majority of variation in city–source probabilities comes from city–source shocks, with s-specific factors (comparative advantage, transport costs) accounting for less than half the variation.

6. Conclusion

Prominent trade models predict that exporters can be ranked according to a hierarchy in which all buyers purchase from the top ranked source of supply. Our model, incorporating heterogeneous buyers and hub-and-spoke transport costs, generates hierarchy compliance with a continuum of consumers and hierarchy violations with granular consumers. The Chinese customs dataset is uniquely suited to investigating hierarchy non-compliance because the prevalent pattern of import flows is through provincial transport hubs into destination cities. Since the hubs neutralize major sources of supply heterogeneity such as comparative advantage and trade costs, differences in choices made by cities can be attributed to demand-side variation.

We find that Chinese cities import from the top provincial source only two-thirds of the time. Of course, stark theoretical predictions rarely hold exactly in the data. More surprisingly, we find that hierarchy is observed significantly less often than by the granular sourcing model where random consumer taste shocks are independent and

[14] The value of these distributional assumptions comes from being able to use data to calibrate π_{sd} without having to estimate θ. Since θ is subsumed in Λ_s and ν_{sd}, we only need data on expected probabilities, which we take to be the province-level data on shipment shares, x_s.

[15] The proof relies on the Gamma function's recursive property that $\Gamma(z) = \Gamma(z+1)/z$.

$$\begin{aligned}
\mathbb{P}(f \text{ does not comply})_{\text{orient}} &= \frac{\Gamma\left(\frac{1}{\eta^2}\right)\Gamma\left(n_d + \frac{1-\Lambda_1}{\eta^2}\right)}{\Gamma\left(\frac{1-\Lambda_1}{\eta^2}\right)\Gamma\left(n_d + \frac{1}{\eta^2}\right)} \\
&= \frac{\Gamma\left(1+\frac{1}{\eta^2}\right)\Gamma\left(1+n_d + \frac{1-\Lambda_1}{\eta^2}\right)\left(\frac{1-\Lambda_1}{\eta^2}\right)\left(n_d+\frac{1}{\eta^2}\right)}{\Gamma\left(1+\frac{1-\Lambda_1}{\eta^2}\right)\Gamma\left(1+n_d + \frac{1}{\eta^2}\right)\left(\frac{1}{\eta^2}\right)\left(n_d+\frac{1-\Lambda_1}{\eta^2}\right)} \\
\lim_{\eta^2 \longrightarrow \infty} \mathbb{P}(f \text{ does not comply})_{\text{orient}} &= \frac{\Gamma(1)\Gamma(1+n_f)(1-\Lambda_1)(n_f)}{\Gamma(1)\Gamma(1+n_f)(n_f)} = 1-\Lambda_1
\end{aligned}$$

[16] Equivalently, we divide the sum of squared errors between sources by the total sum of squares.

identically distributed. This under-prediction of zeros is also evident in previous research investigating random outcomes and granularity such as Armenter and Koren (2014) and Eaton et al. (2013). Introducing a city–source shock into the model generates orientation of cities to particular source countries and reduces expected compliance. We calibrate the parameter governing the dispersion of this shock to match the average hierarchy compliance in the data. We find that city–source orientation plays a major role in influencing the sourcing decisions. The variation in traditional supply-side determinants of trading patterns accounts for less than half the variation in the sourcing probabilities of cities.

The importance of granularity on the buyer side of import transactions has implications for the estimation of trade models, for example to obtain trade cost elasticities. Granular buyers give rise to a new source of zeros. While the upper bound to productivity in Helpman et al. (2008) or the finite distribution of export productivities will result in zero trade flows in the presence of fixed market entry costs, granularity on the buyer side leads to zeros even if exporters could offer their wares in all markets without incurring entry costs. This implies that one should be cautious in using methods that treat zeros as if they were emerging entirely from a fixed cost truncation process.

With granular consumers, zeros arise in the realized trade data even if the expected amount of bilateral trade is positive. This suggests that the correct econometric approach to these zeros is to use a model such as the Multinomial Pseudo-Maximum Likelihood method proposed by Eaton et al. (2013) and shown to have good robustness properties by Head and Mayer (2014). In contrast, one should be wary of methods involving ratios of trade flows such as Head et al. (2010) or Caliendo and Parro (2014). Such models exploit features of CES preferences that are true with a continuum of consumers but will lead to non-random selection via division by zero in settings where granular consumers are important.

Pervasive buyer heterogeneity also raises issues for calculating the consumer gains from trade. Ex ante (before preference draws are realized) the results of Anderson et al. (1992, p. 89) imply that we can still use the standard CES price index to determine changes in expected utility. Ex post a lower import price from some source would benefit buyers inclined towards that source, but be neutral for those whose taste shocks lead them to choose other sources before and after the shock. (Anderson et al., 1992, pp. 97–100) show that in scenarios where some prices rise and others fall, a reduction in the CES price index is not sufficient to guarantee that the winning consumers can compensate the losers. The importance of heterogeneous, granular consumers suggests that quantitative analyses of the consumer gains for trade should be wary of the representative consumer assumption.

References

Anderson, S.P., de Palma, A., Thisse, J.-F., 1992. Discrete Choice Theory of Product Differentiation. MIT Press, Cambridge and London.

Antràs, P., Fort, T.C., Tintelnot, F., 2014. The margins of global sourcing: theory and evidence from U.S. firms. NBER Working Paper No. 20772.

Arkolakis, C., 2010. Market penetration costs and the new consumers margin in international trade. J. Polit. Econ. 118 (6), 1151–1199.

Armenter, R., Koren, M., 2014. A balls-and-bins model of trade. Am. Econ. Rev. 104 (7), 2127–2151.

Bernard, A.B., Jensen, J.B., 1999. Exceptional exporter performance: cause, effect, or both? J. Int. Econ. 47 (1), 1–25.

Bernard, A.B., Redding, S.J., Schott, P.K., 2011. Multiproduct firms and trade liberalization. Q. J. Econ. 126 (3), 1271–1318.

Blum, B.S., Claro, S., Horstmann, I., 2010. Facts and figures on intermediated trade. Am. Econ. Rev. Pap. Proc. 100 (2), 419–423. May.

Caliendo, L., Parro, F., 2014. Estimates of the trade and welfare effects of NAFTA. Rev. Econ. Stud. 82 (1), 1–44.

Di Comite, F., Thisse, J.-F., Vandenbussche, H., 2014. Verti-zontal differentiation in export markets. J. Int. Econ. 93 (1), 50–66.

Eaton, J., Kortum, S., Kramarz, F., 2011. An anatomy of international trade: evidence from French firms. Econometrica 79 (5), 1453–1498.

Eaton, J., Kortum, S.S., Sotelo, S., 2013. International trade: linking micro and macro. Advances in Economics and Econometrics. vol. 1. Cambridge University Press., 329–370.

Ellison, G., Glaeser, E.L., 1997. Geographic concentration in US manufacturing industries: a dartboard approach. J. Polit. Econ. 105 (5), 889–927.

Halpern, L., Koren, M., Szeidl, A., 2015. Imported inputs and productivity. Am. Econ. Rev. 105 (12), 3660–3703,

Head, K., Mayer, T., 2014. Gravity equations: workhorse, toolkit, and cookbook. In: Helpman, E., Gopinath, G., Rogoff, K. (Eds.), Handbook of International Economics. vol. 4. Elsevier., 131–195.

Head, K., Mayer, T., Ries, J., 2010. The erosion of colonial trade linkages after independence. J. Int. Econ. 81 (1), 1–14.

Head, K., Ries, J., Swenson, D., 1995. Agglomeration benefits and location choice: evidence from Japanese manufacturing investments in the United States. J. Int. Econ. 38 (3), 223–247.

Helpman, E., Melitz, M., Rubinstein, Y., 2008. Estimating trade flows: trading partners and trading volumes. Q. J. Econ. 123 (2), 441–487.

Melitz, M., Ottaviano, G., 2008. Market size, trade, and productivity. Rev. Econ. Stud. 75 (1), 295–316.

Melitz, M.J., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. Econometrica 71 (6), 1695–1725.

Rauch, J.E., 1999. Networks versus markets in international trade. J. Int. Econ. 48 (1), 7–35.