Import Sourcing of Chinese Cities: Order versus Randomness*

Keith Head† Ran Jing‡ John Ries§

First version: May 1, 2010
This version: May 12, 2014

Abstract

We utilize detailed import information to assess micro predictions for the import patterns of Chinese cities. Leading trade models yield a hierarchy of suppliers ranked by quality-adjusted costs. When goods transit through a provincial hub, these models predict that every city in a province that imports a good will purchase from the source country offering the lowest-cost variety available at the hub. Our data shows that hierarchy compliance of this type occurs just 66% of the time. A random sourcing model inspired by the balls-and-bins framework fits the data better but over-predicts compliance. We calibrate a modified version of the model in which buyer-seller cost shocks orient some firms towards specific source countries to match observed hierarchy compliance. Our results imply that idiosyncratic buyer-seller shocks play a more important role in explaining variation in costs perceived by importers than variation in the characteristics of supplier countries.

---

*This research has been supported by the Social Sciences and Humanities Research Council of Canada. We thank 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, and Princeton University for helpful suggestions.

†Sauder School of Business, University of British Columbia and CEPR, keith.head@sauder.ubc.ca

‡School of International Trade and Economics, University of International Business and Economics, ran.jing@uibe.edu.cn

§Corresponding author, Sauder School of Business, University of British Columbia, john.ries@sauder.ubc.ca
1 Introduction

Trade economists have made important progress in producing parsimonious, quantitative models that explain the pattern of international trade. Starting from diverse underpinnings—Ricardian productivity differences, national product differentiation, and monopolistic competition with heterogeneous firms—the models have converged on a common form for the equation governing bilateral trade. Because this equation closely resembles the well-established gravity equation, the models are widely viewed as empirically successful. As demonstrated by Arkolakis et al. (2012), these models also comply with the same simple formula for the gains from trade. With greater access to more disaggregated trade data, empirical work can now push these models beyond their aggregate features to inquire whether they can also successfully predict various extensive margins. In this paper, we investigate the accuracy with which standard models predict who buys what from whom by testing predictions about the trade partners of Chinese cities.

We identify conditions under which standard trade models predict hierarchical relationships concerning export destinations and sources. Hierarchy entails that exporters who succeed in selling in a tough (small or distant) market should also be observed to sell in an easy (large or proximate) market. Also, destinations that purchase from a less popular (high cost) exporter should also buy from a more popular (lower cost) exporter. Consistent with observed import patterns, we model firms in cities importing through a provincial transport hub. An implication of hierarchy is that all cities import from the source country offering the lowest-cost variety available at the hub. We test this prediction using highly disaggregated Chinese product-level data and variation in sourcing patterns across cities. Our data indicate widespread non-compliance with this theoretical prediction: Chinese cities only source narrowly defined goods from the top foreign supplier in the province two-thirds of the time.

The failure of the hierarchy prediction points toward the importance of introducing idiosyncratic factors. We develop an alternative model in which cities share common valuations of the relative attractiveness of goods supplied by different countries but individual shipments are influenced by random shocks. Our model can be viewed as a particular economic implementation of the “balls-and-bins” approach of Armenter and Koren (2014). Balls are shipments that randomly fall into bins of different sizes depending on the source country. Along the lines of analysis in Eaton et al. (2012), we establish that the probability a ball falls into a bin can be expressed as a multinomial logit function. We derive our specification from a discrete choice model of cost minimization. The random model with independent shipments has considerable explanatory power in explaining compliance with the hierarchy but it still predicts significantly higher compliance rates than we observe (71% versus 66%).

1See Eaton and Kortum (2002), Anderson and Wincoop (2003), and Chaney (2008).
We extend the random model to allow shipments to be correlated within firms and oriented towards particular source countries. The mechanism is a cost shock that affects all shipments purchased by an importing firm from a particular source country (buyer-seller shock). We show that the model with source orientation results in lower expected compliance and can be calibrated to exactly match the average compliance rate in the data.

Two earlier papers find that exporters do not comply with the hierarchy predictions of the standard models. Eaton et al. (2011) determine 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 market, it should also be profitable in a more popular market. They develop a model where entry and sales of French exporters are influenced by idiosyncratic shocks to unit costs, fixed entry costs, and demand. Bernard et al. (2011) argue that multi-product firms should always ship their more popular products to the destinations where they sell their less popular products but report that the incidence of this is only 67% for US exporters. This observation partly motivates their use of stochastic firm-product-country taste draws in their model of trade with multi-product firms.

We construct a hierarchy test that utilizes narrow product information (8-digit) and geographic areas (cities) to identify hierarchy violations. Our method rules out possible explanations for non-compliance observed in previous studies. In Eaton et al. (2011), a French firm may not sell to the most popular export market, Belgium, if the particular product it exports is something that Belgium does not import. Alternatively, a firm located in southern France might export to the proximate Spanish market, but not to Belgium. In these circumstances, the failure of the firm to export to Belgium would not be a true hierarchy violation. Similarly, Bernard et al. (2011)’s hypothesis that a firm should sell its most popular product in markets where it sells its less popular product would not obtain in situations where the country in question has no demand for the firm’s popular product. Our analysis considers destinations (Chinese cities) that are observed to import a specific product, and standard models predict its import sources should include the top source of that good in the province. The one-third noncompliance rate that we find suggests a large role for random factors in matching buyers to sellers.

As do Armenter and Koren (2014) and Eaton et al. (2012), we predict extensive margin outcomes as the result of finite draws (balls) from a binomial distribution (the probability of a ball falling into the bin of the top source). Under the assumption of independent shipments, we find the balls-and-bins model predicts fewer zeros (too much compliance) than what is observed in the data. 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 the model predicts 72% zeros, there are 82% zeros in the data. The simple balls-and-bins 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.
also find too few zeros. Their random model predicts that on average each country should export to 77% of its 91 partners, which is considerably higher than the 66% observed in the data. We show that the systematic under-prediction of zeros in these types of models can be eliminated by allowing for non-independence between balls.

In assessing variation in the foreign entry decisions of French exporters, Eaton et al. (2011) measure the relative importance of the variation in the “core” efficiency of firms relative to the variation in the entry cost draw. They consider the cost efficiency of exporters located in a single country (France), and find that 57% of the variation across firms in market entry can be attributed to variation in firm efficiency. We measure the relative importance of the variation in the “core” efficiency of supplying countries relative to the variation of the shock orienting a buyer to a seller. This core efficiency reflects supplier factor costs and their trade costs to reach the Chinese provincial hub. We find that variation in traditional determinants of trading patterns—comparative advantage and transport costs—accounts for only 19% of the variation in delivered costs, with the remaining 81% is attributable to idiosyncratic buyer-seller shocks.

This paper makes several contributions to the literature investigating micro predictions of trade models. We formalize the conditions for hierarchy predictions in the data. Our very disaggregated product and geographic information allows us to test hierarchy non-compliance more accurately than was possible with data used in previous studies. Using Monte Carlo methods, we establish the statistical significance of differences between the observed compliance and the ball-and-bins null. Finally, our calibration of the model incorporating buyer-seller shocks reveals that these factors play a dominant role in explaining variation in delivered costs of goods.

The next section establishes the hierarchy prediction of a standard heterogeneous firm model. In section 3 we define our hierarchy statistic, describe the data, and measure the extent of hierarchy compliance. We present a random sourcing model in Section 4 and demonstrate that it provides a reasonable fit to the data. Section 5 develops a version of the random model that allows for firms’ 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.

## 2 Hierarchy in models with heterogeneous firms

In popular heterogeneous-firm trade models, all 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 \(d\)th most popular destination.\(^2\) Hierarchy also implies that a destination importing from the \((s + 1)\)th most popular exporter also imports from the \(s\)th most popular exporter.

In models with heterogeneous 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.\(^3\) Not all varieties can be profitably sold into all markets.

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. The hub could be a large seaport, a regional airport, or a geographical feature such 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. The assumption that goods flow through the provincial hub matches our data well. Defining the provincial hub at the province-good level as the port that shipments most frequently flow through, 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. Finally, in 59.4% of the cases, the provincial hub accounts for 100% of a province’s imports of a good.

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. Following the notation of Helpman et al. (2008), there are a continuum of firms indexed by \(a\), which denotes the number of bundles used per unit of output by the firm. Efficiency has bounded support with the most efficient firm in country \(s\) having cost \(a^L_s\).\(^5\) We

\(^2\)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. (2012) 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.

\(^5\)Another interpretation of \(a^L_s\) is that there are a finite number of firms as in Eaton et al. (2012) and \(a^L_s\) is the realization of the lowest cost draw.
choose units such that the quality-adjusted cost of each input bundle in source country \( s \) is \( c_s \) and \( \tau_{sd} \) 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 \( C_{sd}(a) = \tau_{sd}c_s a \).

Under a hub and spoke system, the iceberg trade cost factor can be 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 \( d \).\footnote{Suppose a fraction \( \delta_s \) of production “melts” on the way to the hub, and of the remaining goods, a further fraction \( \delta_d \) melts along the spoke. To deliver one unit to the final destination therefore requires production of \( 1/[(1 - \delta_s)(1 - \delta_d)] \) units. Thus \( T_s = 1/(1 - \delta_s) \) and \( t_d = 1/(1 - \delta_d) \).}

Now \( C_{sd}(a) \) can be expressed as the product of the costs of reaching the provincial hub \( (C_s(a) = c_s a T_s) \) and the cost of transporting the good from the hub to city \( d (t_d) \).

The profits of a firm with unit input requirement \( a \) selling to city \( d \) are given by variable profits minus fixed costs, \( F_{sd} \). Variable profits are a function, \( v() \), of delivered unit costs \( (C_s(a)t_d) \) and expenditures on all varieties \( (Y_d) \). Thus, profits net of fixed costs are given by

\[
\Pi_{sd}(a) = v(C_s(a)t_d,Y_d) - F_{sd},
\]

where the partial derivative of the first argument of \( v() \) is negative and that of the second argument is positive.

Hierarchy predictions require that we solve for a threshold cost level \( C^*_d \) such that \( \Pi_{sd}(a) < 0 \) for all \( a \) such that \( C_s(a) > C^*_d \). This condition implies that a firm that is good enough to sell profitably in city \( d \) with \( C^*_d \) will also be good enough to profitably sell to every other city, \( d' \), with \( C^*_{d'} > C^*_d \). To ensure that \( C^*_d \) is indeed only a function of \( d \)-specific terms we need a separable form for fixed costs, \( F_{sd} \).\footnote{Without separability, the threshold cost for entering a market could depend on \( s \) characteristics, leading to a breakdown of hierarchical ordering. For example, if a source country \( s \) had an advantage over its rivals in serving market \( d \) but that advantage did not apply to the other markets, then it could export to a market that appeared to be hard but fail to export to easier markets.}

We therefore follow \cite{Arkolakis2010} 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 \( \alpha \). We assume that the same home factor prices, \( c_s \), and unit factor requirements, \( a_s \), that govern production costs also apply to fixed costs. The destination-level factor costs are denoted \( w_d \). To avoid excess notation, we also assume that the cost of supplying factor services from home country \( s \) remotely in \( d \) is governed by the same trade costs, \( \tau_{sd} \), that apply to shipments of goods. Taking these assumptions together we obtain

\[
F_{sd} = f_d(C_s(a)t_d)^\alpha w_d^{1-\alpha}.
\]

To obtain a closed form for \( C^*_d \) we we follow \cite{Helpman2008} in specifying the variable profit function \( v() \) as \( \lambda[C_s(a)t_d]^{1-\epsilon} Y_d P_d^{\epsilon} \), where \( \epsilon \) is the elasticity of
substitution, $P_d$ is the price index, and $\lambda \equiv \epsilon^{-\epsilon}(\epsilon - 1)^{\epsilon-1}$. Substituting the variable and fixed costs formulas into equation (1) we obtain

$$
\Pi_{sd}(a) = \lambda[C_s(a) t_d]^{1-\epsilon} Y_d P_d^{\epsilon-1} - f_d(C_s(a) t_d)^{\alpha} w_d^{1-\alpha}.
$$

Setting equation (3) equal to zero and solving for costs determines the critical level of delivered unit costs, $C_d^*$, where profits of serving a particular city $d$ equal zero:

$$
C_d^* = \frac{1}{t_d} \left[ \frac{\lambda Y_d P_d^{\epsilon-1}}{f_d w_d^{1-\alpha}} \right] ^{\frac{1}{\alpha+\epsilon-1}}.
$$

The critical cost level for exporting to city $d$ 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. 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 that is efficient enough to export to a tough destination, will export to all easier destinations. We can empirically infer the most efficient supplier by counting the number of markets to which it exports.

Lacking data on the individual firms who export to city $d$ we focus on which source countries supply which destination cities. 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 delivered unit costs to the provincial hub of the most profitable (lowest cost) firm in each $s$ as $C_s^L = c_s T_s a_s^L$. Therefore, source $s$ sells to city $d$ if $C_s^L \leq C_d^*$.

Figure 1 depicts hierarchy. First consider the left panel. The vertical axis shows profits and the horizontal axis shows delivered unit costs $C_s$. The figure displays the profit schedules 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 source countries 1, 2, 3 and 4. The figure shows that the largest destination imports from all four source countries because $C_s^*(20) > C_s^L$ for $s=1,2,3,4$. Smaller markets import from fewer sources.

The hierarchy prediction is more easily visualized in the right panel of Figure 1. If a destination imports from the $(s+1)$th most popular source, it also sources from the $s$th most popular source. Also, if a source finds it profitable to sell to $(d+1)$th the most popular destination in terms of the number of sources that sell there, it also sells to the $d$th most popular destination. All cities import from the low-cost source.
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. The theory laid out in section 2 predicts that this statistic should equal one. The key conditions required to generate hierarchy are that the profit function is decreasing in costs, $C_s(a)$, the hub and spoke nature of trade costs, and the separable form for fixed costs. 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 on national product differentiation, and no fixed costs will also lead to this statistic equal 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. Customs declaration forms ask importers to report the “destination within borders.” The official website for the national exam for customs brokers defines this item as the "known place within China."

---

8The 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.
for consumption, usage, or the final destination of the trip. It need not be the port of entry, which is listed separately.

Table 1: Top Goods, 2006

<table>
<thead>
<tr>
<th>CN8</th>
<th>$bil</th>
<th>#Src</th>
<th>SNA</th>
<th>Rauch</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>27090000</td>
<td>66.4</td>
<td>46</td>
<td>Int</td>
<td>Org</td>
<td>Petroleum oils (crude)</td>
</tr>
<tr>
<td>85422119</td>
<td>39.5</td>
<td>64</td>
<td>Int</td>
<td>Dif</td>
<td>Mon. integ. circuits, digital, $&lt; 0.18 \mu m$</td>
</tr>
<tr>
<td>90138000</td>
<td>25.8</td>
<td>47</td>
<td>Cap</td>
<td>Dif</td>
<td>Liquid crystal display panels</td>
</tr>
<tr>
<td>85422900</td>
<td>15.7</td>
<td>82</td>
<td>Int</td>
<td>Dif</td>
<td>Monolithic integrated circuits, not digital</td>
</tr>
<tr>
<td>85422129</td>
<td>12.2</td>
<td>65</td>
<td>Int</td>
<td>Dif</td>
<td>Mon. int. circ., dig., $0.18 &lt; \text{wid.} \leq 0.35 \mu m$</td>
</tr>
<tr>
<td>26011150</td>
<td>11.8</td>
<td>27</td>
<td>Int</td>
<td>Org</td>
<td>Iron ores and concentrates, non-agglomerated</td>
</tr>
<tr>
<td>85422199</td>
<td>10.5</td>
<td>69</td>
<td>Int</td>
<td>Dif</td>
<td>Mon. integ. circuits, dig., $&gt; 0.35 \mu m$</td>
</tr>
<tr>
<td>27101922</td>
<td>9.0</td>
<td>27</td>
<td>Int</td>
<td>Dif</td>
<td>Fuel oils number 5–7</td>
</tr>
<tr>
<td>12010091</td>
<td>7.5</td>
<td>8</td>
<td>Int</td>
<td>Org</td>
<td>Soya beans, whether or not broken</td>
</tr>
<tr>
<td>84733090</td>
<td>7.1</td>
<td>71</td>
<td>Int</td>
<td>Dif</td>
<td>Computer parts and accessories</td>
</tr>
<tr>
<td>85426000</td>
<td>6.9</td>
<td>59</td>
<td>Int</td>
<td>Dif</td>
<td>Hybrid integrated circuits</td>
</tr>
<tr>
<td>85299020</td>
<td>6.4</td>
<td>42</td>
<td>Int</td>
<td>Dif</td>
<td>Hand-held wireless telephone parts</td>
</tr>
<tr>
<td>85422121</td>
<td>6.3</td>
<td>28</td>
<td>Int</td>
<td>Dif</td>
<td>Mon. int. circ., dig., $0.18 &lt; \text{wid.} \leq 0.35$, orig. film</td>
</tr>
<tr>
<td>29173610</td>
<td>6.1</td>
<td>18</td>
<td>Int</td>
<td>Dif</td>
<td>Terephthalic acid and its salts</td>
</tr>
<tr>
<td>88024010</td>
<td>6.1</td>
<td>4</td>
<td>Cap</td>
<td>Dif</td>
<td>Aircraft between 15 and 45 tons</td>
</tr>
<tr>
<td>84717010</td>
<td>6.1</td>
<td>46</td>
<td>Cap</td>
<td>Dif</td>
<td>Computer hard drives</td>
</tr>
<tr>
<td>26030000</td>
<td>5.9</td>
<td>35</td>
<td>Int</td>
<td>Ref</td>
<td>Copper ores and concentrates</td>
</tr>
<tr>
<td>84798990</td>
<td>5.7</td>
<td>52</td>
<td>Cap</td>
<td>Dif</td>
<td>Machines and mechanical appliances N.E.S.</td>
</tr>
<tr>
<td>74031100</td>
<td>4.9</td>
<td>34</td>
<td>Int</td>
<td>Org</td>
<td>Cathodes of unwrought copper</td>
</tr>
<tr>
<td>52010000</td>
<td>4.8</td>
<td>61</td>
<td>Int</td>
<td>Org</td>
<td>Cotton, not carded/combed</td>
</tr>
</tbody>
</table>

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 (intermediate, capital, or consumption) and the Rauch (1999) classification of differentiated, reference price, or organized exchange products. 8-digit classifications are quite 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 aircrafts between 15 and 45 tons which were sourced from 8 and 4 countries, respectively.

Table 1 indicates that the majority of Chinese imports are intermediate goods. Indeed, Table 2 reveals that the share of intermediates in Chinese imports in 2006

9Details on how we classify CN8 goods into the Rauch categories are available from the authors.
was about 75%. The last column of the table is compiled from the Chinese Customs data used in this study. For comparison, we also show the figures using the United Nations’ Comtrade data base. We observe that information from the two sources closely correspond. The first column shows the breakdown of world exports. Intermediates account for 56% of world exports. Relatively little of Chinese imports are consumption goods—3% compared to 17% for the world. Capital goods account for about 19% of China’s imports and 16% of world imports.

Table 2: Shares of imports by good type (in %)

<table>
<thead>
<tr>
<th>Type</th>
<th>Comtrade World</th>
<th>Customs China</th>
<th>China</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumption</td>
<td>17.0</td>
<td>3.3</td>
<td>3.2</td>
</tr>
<tr>
<td>Intermediate</td>
<td>55.7</td>
<td>74.1</td>
<td>75.6</td>
</tr>
<tr>
<td>Capital</td>
<td>15.5</td>
<td>19.1</td>
<td>18.1</td>
</tr>
<tr>
<td>Unclassified</td>
<td>11.8</td>
<td>3.6</td>
<td>3.2</td>
</tr>
</tbody>
</table>

In our analysis of the sourcing decisions of Chinese cities, we exclude imports into bonded warehouses. 6.1% of 2006 imports are entrepot and not destined for the Chinese market. Another 4.1% go to other types of bonded warehouses and may not be consumed in the city where the warehouse is located. 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.

### 3.2 Hierarchy compliance

We calculate the hierarchy statistic as the share of cities that import from the top provincial source of the good. Since we do not have information on which country is the source of the lowest-cost supplier, we must infer it from the data. Table 3 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.

---

10 On the other hand, the Chinese customs data shows that 31% of Chinese exports are consumption goods.

11 We also exclude observations corresponding to re-imports where the source country was listed as China.
Table 3: Top Sources

<table>
<thead>
<tr>
<th>Province</th>
<th>$mn</th>
<th>#(cn8)</th>
<th>#(city)</th>
<th>Top Source</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guangdong</td>
<td>176.1</td>
<td>6184</td>
<td>24</td>
<td>Hong Kong</td>
<td>25.6</td>
</tr>
<tr>
<td>Jiangsu</td>
<td>115.8</td>
<td>5532</td>
<td>27</td>
<td>Japan</td>
<td>40.3</td>
</tr>
<tr>
<td>Shanghai</td>
<td>73.4</td>
<td>6136</td>
<td>22</td>
<td>Japan</td>
<td>41.9</td>
</tr>
<tr>
<td>Shandong</td>
<td>45.2</td>
<td>5109</td>
<td>30</td>
<td>Korea</td>
<td>44.8</td>
</tr>
<tr>
<td>Zhejiang</td>
<td>45.1</td>
<td>5007</td>
<td>24</td>
<td>Japan</td>
<td>32.1</td>
</tr>
<tr>
<td>Beijing</td>
<td>41.1</td>
<td>5584</td>
<td>19</td>
<td>USA</td>
<td>22.7</td>
</tr>
<tr>
<td>Tianjin</td>
<td>26.5</td>
<td>4812</td>
<td>19</td>
<td>Japan</td>
<td>31.7</td>
</tr>
<tr>
<td>Liaoning</td>
<td>21.8</td>
<td>4833</td>
<td>21</td>
<td>Japan</td>
<td>39.3</td>
</tr>
<tr>
<td>Fujian</td>
<td>18.9</td>
<td>4657</td>
<td>12</td>
<td>Taiwan</td>
<td>45.9</td>
</tr>
<tr>
<td>Hebei</td>
<td>8.2</td>
<td>3113</td>
<td>12</td>
<td>Japan</td>
<td>24.2</td>
</tr>
<tr>
<td>Heilongjiang</td>
<td>7.1</td>
<td>2132</td>
<td>21</td>
<td>USA</td>
<td>19.6</td>
</tr>
<tr>
<td>Hubei</td>
<td>6.1</td>
<td>2655</td>
<td>18</td>
<td>Japan</td>
<td>21.4</td>
</tr>
<tr>
<td>Jilin</td>
<td>5.5</td>
<td>2495</td>
<td>17</td>
<td>Korea</td>
<td>26.1</td>
</tr>
<tr>
<td>Anhui</td>
<td>5.4</td>
<td>2317</td>
<td>18</td>
<td>Japan</td>
<td>24.1</td>
</tr>
<tr>
<td>Sichuan</td>
<td>4.8</td>
<td>2592</td>
<td>24</td>
<td>USA</td>
<td>25.3</td>
</tr>
<tr>
<td>Henan</td>
<td>3.7</td>
<td>1887</td>
<td>23</td>
<td>Japan</td>
<td>22.4</td>
</tr>
<tr>
<td>Guangxi</td>
<td>3.7</td>
<td>1648</td>
<td>15</td>
<td>Taiwan</td>
<td>16.6</td>
</tr>
<tr>
<td>Inner Mongolia</td>
<td>3.4</td>
<td>1118</td>
<td>15</td>
<td>USA</td>
<td>22.4</td>
</tr>
<tr>
<td>Yunnan</td>
<td>3.2</td>
<td>1493</td>
<td>21</td>
<td>USA</td>
<td>20.0</td>
</tr>
<tr>
<td>Jiangxi</td>
<td>3.2</td>
<td>1722</td>
<td>13</td>
<td>Japan</td>
<td>18.1</td>
</tr>
<tr>
<td>Xinjiang</td>
<td>3.2</td>
<td>1171</td>
<td>16</td>
<td>USA</td>
<td>30.8</td>
</tr>
<tr>
<td>Shanxi</td>
<td>2.9</td>
<td>1352</td>
<td>12</td>
<td>Germany</td>
<td>20.9</td>
</tr>
<tr>
<td>Hunan</td>
<td>2.8</td>
<td>1787</td>
<td>20</td>
<td>Japan</td>
<td>24.5</td>
</tr>
<tr>
<td>Gansu</td>
<td>2.7</td>
<td>663</td>
<td>13</td>
<td>Germany</td>
<td>25.8</td>
</tr>
<tr>
<td>Shaanxi</td>
<td>2.5</td>
<td>1958</td>
<td>11</td>
<td>USA</td>
<td>23.7</td>
</tr>
<tr>
<td>Hainan</td>
<td>2.2</td>
<td>1315</td>
<td>3</td>
<td>USA</td>
<td>16.6</td>
</tr>
<tr>
<td>Chongqing</td>
<td>2.2</td>
<td>1806</td>
<td>27</td>
<td>Japan</td>
<td>31.6</td>
</tr>
<tr>
<td>Guizhou</td>
<td>0.9</td>
<td>754</td>
<td>10</td>
<td>Japan</td>
<td>21.8</td>
</tr>
<tr>
<td>Ningxia</td>
<td>0.5</td>
<td>451</td>
<td>4</td>
<td>Germany</td>
<td>28.8</td>
</tr>
<tr>
<td>Qinghai</td>
<td>0.4</td>
<td>346</td>
<td>5</td>
<td>Germany</td>
<td>22.5</td>
</tr>
<tr>
<td>Tibet</td>
<td>0</td>
<td>188</td>
<td>5</td>
<td>Nepal</td>
<td>50.0</td>
</tr>
</tbody>
</table>
We define 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, the method may incorrectly identify large countries as lowest cost. To see this, consider equation (6) in Helpman et al. (2008),

\[ M_{sd} = C_{sd}^{1-\epsilon} \bar{P}_d^{1-\epsilon} Y_d N_s V_{sd}, \]

where \( V_{sd} \) is a factor based on the distribution of productivities and \( N_s \) is the number of firms in \( s \). A high volume of imports, \( M_{sd} \), could result from low costs, \( C_{sd} \), peculiarities in the distribution of productivities, \( V_{sd} \), or a large number of varieties in the exporting country, \( N_s \). In a heterogeneous firm, hierarchy model, all destinations will purchase from the source with the lowest cost firm. This will be the source with the lowest \( C^L \), which may not be the source with the greatest volume of exports. 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.\(^{12}\)

Column (5) in Table 3 identifies the country that is most frequently the top source across all the goods imported in the province. 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 \( y_{d} \) as a binary variable equal to 1 if \( d \) imports the good from the top source and zero otherwise. Let \( K \) be number of cities in a province that import the good. We can express the hierarchy statistic, \( h_1 \), a province-good specific measure of the share of all importing cities that source from country 1:

\[ h_1 = \frac{\sum_d y_{d}}{K}. \]

In order to have a sufficient number of cities to reliably identify the top source, we impose the restriction that for each good, there must be at least four cities that

\(^{12}\)This method is consistent with that used by Eaton et al. (2011) and Bernard et al. (2011) in their hierarchy rankings. Like Bernard et al. (2011), we break ties based on highest import value.
import the good in the province. This procedure reduces the number of goods from 7077 to 5239 and the number of province-good combinations to 29,459. This sample accounts for 82.5% of Chinese imports.

![Histogram of hierarchy statistic](image)

Figure 2: Distribution of hierarchy statistics ($h_1$) for 29,459 province-goods

Figure 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 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. The results are also inconsistent with a Ricardian model where all cities purchase exclusively from the low-cost supplier. In the next section, we evaluate our observed hierarchy statistics relative to a model incorporating an element of randomness in the sourcing decision.

---

13 The number of cities in a province ranges from 4 to 28 with a mean and median of 10.
4 Random sourcing model

A natural way to model non-compliance with hierarchy is to introduce a random element into the discrete choice of which source to use for a particular shipment. A model with randomness does not have to be truly stochastic so long as it contains an idiosyncratic term in the buyer’s objective function. The most straightforward way to model this is to employ results from logit random utility models analyzed by [Anderson et al.](1992).

Individual shipments of goods can be thought of as the smallest potentially independent units of trade. We imagine a procurement agent in each city $d$ who selects the lowest cost source country for each shipment $j$. The city-specific number of shipments is denoted $n_{d}$. The delivered cost perceived by the agent for shipment $j$ in city $d$ is the product of a common term and an idiosyncratic term. To keep the model as simple as possible, we assume that within each 8-digit product category, every country $s$ competitively supplies a homogeneous good with a production cost of $c_{s}$. The common delivered cost also includes a transport cost, $\tau_{sd}$.

The key assumption of the random sourcing model is that individual shipments have idiosyncratic costs, $\varepsilon_{js}$, associated with each source. We can think of the $\varepsilon_{js}$ as shipment-specific transaction costs. The delivered cost inclusive of the random transaction cost is given by

$$
C_{sd}^{j} = c_{s}T_{s}t_{d}\varepsilon_{js}.
$$

In the benchmark formulation of random sourcing, the $\varepsilon_{js}$ are independent draws from a Weibull distribution with shape parameter $\theta$. The probability, $\pi$, that country $s$ is viewed as a lower cost source than any alternative $s'$ for any shipment $j$ in city $d$ is given by

$$
\pi_{sd} = \mathbb{P}[C_{sd}^{j} < C_{s'd}^{j} \forall s' \neq s] = \left(\frac{c_{s}T_{s}}{\sum_{h}(c_{h}T_{h})}\right)^{-\theta}.
$$

This expression lacks $d$-specific terms because the supplier who has the lowest cost at the provincial hub maintains its advantage when the cost of transporting from the hub to the spoke city, $t_{d}$, is incorporated. As a result $\pi_{sd} = \pi_{s}$ for all $d$. That is, no matter which city in a province an order emanates from, it has the same probability of being filled by a supplier from source country $s$. The likelihood that at least one of $n_{d}$ shipments to city $d$ will be supplied from source $s$ is therefore

$$
1 - (1 - \pi_{s})^{n_{d}}.
$$

The random sourcing model presented here gives a simple microeconomic foundation for the balls-and-bins model that [Armenter and Koren](2014) use to explain the incidence of zeros in United States product-country and firm-country trade flows. Our application measures the likelihood that a city will import a good

\[14\] Eaton et al. (2012) employ a more complex model but it also generates a multinomial logit expression for the probability that a firm sells to a destination country. They use their model to estimate the probability that countries transact with each other.
from the top provincial source. In our analogy, cities randomly assign (throw) shipments (balls) to source countries (bins).

For each province-good combination, we denote the probability of a shipment coming from the top source as \( \pi_1 \). The top source is the country from which the largest number of cities import (recall the discussion of Table 3 above). We calibrate \( \pi_1 \) for each province-good as \( x_1 \), defined as source 1’s share of total shipments of that good at the province level.\(^{15}\) The expected \( h_1 \) for a province with \( K \) cities is

\[
\mathbb{E}[h_1] = \frac{\sum_d 1 - (1 - x_1)^{n_d}}{K}. \tag{8}
\]

This expected value is increasing in both \( x_1 \) and \( n_d \). In order to calculate \( x_1 \) and \( n_d \), we need to measure shipments. 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 will be 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. Excluding export trade, our 2006 data contain 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 $3,221, about twice the $1,800 value in the US data.\(^{16}\)

Given this definition of shipments, we calculate the probability of choosing source 1 under randomness by summing over all cities in a province:

\[
x_1 = \frac{\sum_d n_{1d}}{\sum_s \sum_d n_{sd}} = \frac{n_1}{\sum_s n_s}.
\]

We plug \( x_1 \) and \( n_d \) (measured as city shipments) into equation (8) to calculate \( \mathbb{E}h_1 \) defined at the province-good level. This allows us to calculate \( \mathbb{E}h_1 \) and compare it to the actual sourcing behavior. We can compare \( \mathbb{E}h_1 \) to the hierarchy statistic, \( h_1 \) generated earlier.

We show average \( h_1 \) and \( \mathbb{E}h_1 \) across the 29,459 province-good combinations in Table 4. We consider 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. The table reveals that compliance ranges from 0.62 to 0.73, being highest for organized exchange goods and lowest for consumption goods. Hierarchy violations are common for all types of goods.

\(^{15}\)Armenter and Koren (2014) also use observed shares in place of the true probabilities.

\(^{16}\)We thank Miklos Koren for providing us the US shipment size data.
Table 4: Hierarchy statistics, their expected values and confidence intervals

<table>
<thead>
<tr>
<th>Type of good</th>
<th>#obs(\dagger)</th>
<th>(h_1)</th>
<th>(Eh_1)</th>
<th>(x_1)</th>
<th>#shm</th>
<th>&lt;(Q_{.05})</th>
<th>&gt;(Q_{.95})</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>29,459</td>
<td>0.66</td>
<td>0.71</td>
<td>0.39</td>
<td>7.57</td>
<td>0.14</td>
<td>0.02</td>
</tr>
<tr>
<td>Intermediate</td>
<td>18,995</td>
<td>0.67</td>
<td>0.74</td>
<td>0.40</td>
<td>8.96</td>
<td>0.17</td>
<td>0.01</td>
</tr>
<tr>
<td>Capital</td>
<td>6,084</td>
<td>0.63</td>
<td>0.65</td>
<td>0.36</td>
<td>4.86</td>
<td>0.07</td>
<td>0.02</td>
</tr>
<tr>
<td>Consumption</td>
<td>4,203</td>
<td>0.62</td>
<td>0.66</td>
<td>0.38</td>
<td>5.30</td>
<td>0.10</td>
<td>0.02</td>
</tr>
<tr>
<td>Differentiated</td>
<td>24,700</td>
<td>0.66</td>
<td>0.71</td>
<td>0.39</td>
<td>7.46</td>
<td>0.13</td>
<td>0.02</td>
</tr>
<tr>
<td>Reference</td>
<td>3,509</td>
<td>0.66</td>
<td>0.75</td>
<td>0.40</td>
<td>8.26</td>
<td>0.19</td>
<td>0.01</td>
</tr>
<tr>
<td>Organized</td>
<td>505</td>
<td>0.73</td>
<td>0.79</td>
<td>0.47</td>
<td>7.52</td>
<td>0.13</td>
<td>0.01</td>
</tr>
</tbody>
</table>

\(h_1\), \(Eh_1\), and \(x_1\) portray averages across province-goods. #shm is the average across province-goods of the city-level median number of shipments. \(Q_{.05}/Q_{.95}\) are the 5th/95th percentiles of the \(Eh_1\) distribution. Counts do not add up to the total because there are 29 and 134 cn8 categories missing SNA and Rauch classifications.

\(\dagger\): An observation is a province-good combination.

Column (3) of Table 4 shows that \(Eh_1\) generated by the random sourcing model varies across good types. It is relatively high when there is a dominant supplier in the province (large bin as measured by \(x_1\)) or there are many shipments (balls) in cities. The table show average \(x_1\) and the median number of city shipments. Organized exchange goods have high \(x_1\) and this leads to high values of \(Eh_1\). Comparing actual \(h_1\) to the expectation, we observe that the goods expected to have highest compliance (intermediates, organized) indeed have the highest actual compliance. In every category, compliance with the hierarchy is less than what is expected in the random sourcing model. Thus while the random model of independent shipments provides a reasonable prediction of actual compliance, it appears to be biased upwards.

The last two columns provide information on the statistical significance of the deviations of \(h_1\) from \(Eh_1\). Columns (7) and (8) show the share of \(h_1\) observations that fall below the 5th percentile and above the 95th percentile of the theoretical distribution of \(Eh_1\). To compute the values for these percentiles of \(Eh_1\), we conduct the following Monte Carlo experiment. For each city, we draw from the binomial distribution using \(x_1\) and \(n_d\) as arguments and identify cities where at least one ball lands in source 1. We code “complying” cities as 1 and others as 0. Taking the average across cities for each province-good generates realized \(h_1\). Averaging across province-goods gives us a single measure for that round of the experiment. We repeat this process 1000 times to produce a distribution of \(Eh_1\), and we can recover its value at the 5% and 95% percentile.

Columns (7) and (8) show the share of actual \(h_1\)'s that fall below and above these levels. If the random sourcing null is correct, we would expect the share in
Table 5: Hierarchy statistics and their expected values, Robustness

<table>
<thead>
<tr>
<th>Sample constraint</th>
<th>#obs†</th>
<th>$h_1$</th>
<th>$Eh_1$</th>
<th>$x_1$</th>
<th>#shm</th>
<th>$&lt; Q_{0.05}$</th>
<th>$&gt; Q_{0.95}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥ 3 cities</td>
<td>37,957</td>
<td>0.66</td>
<td>0.71</td>
<td>0.41</td>
<td>6.78</td>
<td>0.12</td>
<td>0.01</td>
</tr>
<tr>
<td>≥ 5 cities</td>
<td>23,730</td>
<td>0.66</td>
<td>0.72</td>
<td>0.37</td>
<td>8.28</td>
<td>0.16</td>
<td>0.02</td>
</tr>
<tr>
<td>≥ 4 cities &amp; hub only</td>
<td>23,085</td>
<td>0.65</td>
<td>0.71</td>
<td>0.40</td>
<td>6.80</td>
<td>0.15</td>
<td>0.01</td>
</tr>
<tr>
<td>≥ 4 cities &amp; no frequency ties</td>
<td>21,931</td>
<td>0.70</td>
<td>0.73</td>
<td>0.41</td>
<td>7.71</td>
<td>0.11</td>
<td>0.02</td>
</tr>
</tbody>
</table>

$h_1$, $Eh_1$, and $x_1$ portray averages across province-goods. #shm is the average across province-goods of the city-level median number of shipments. $Q_{0.05}$/$Q_{0.95}$ are the 5th/95th percentiles of the $Eh_1$ distribution.

†: An observation is a province-good combination.

These tails to be about 5%. They are not: For the full sample, 14% are below the 5th percentile value and only 2% are above the 95th percentile value. These results indicate that $h_1$ is significantly lower than the prediction of the random sourcing model. This is the case for the full sample as well as all the subsamples shown in Table 4.

To investigate robustness, Table 5 reports average $h_1$ and $Eh_1$ 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 $Eh_1$ do not change very much across these samples. Average compliance is around two-thirds and always significantly less than $Eh_1$.

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. This situation is demonstrated in Figure 3 where goods from two countries flow through the provincial hub and are stored in warehouses before distribution. In this illustration, city $C$ is buying from the top source, source 1, but it is not observed because it obtains source 1’s good via city $A$. The logic of the depicted distribution strategy is that it is more economical to use a central warehouse and distribute to local cities than open up separate warehouses.

From a theoretical perspective, the situation shown in the figure seems unlikely. Low volume goods are the ones where central warehouses are likely to
be preferred to more numerous city-specific warehouses. Since source 2 sells less attractive (and lower volume) goods than source 1, it should be source 2 goods that are shipped via a central warehouse. If source 2 mimicked source 1’s distribution, then we would not observe non-compliance because city C would not appear to be importing from anyone. If they switched distribution strategies, we would find compliance because city C would be importing from the top source.

Figure 3 suggests that false non-compliance may occur when cities are close together and it is economical to serve them with a single warehouse. As a robustness check, we calculate $h_1$ and $E h_1$ for the municipalities Beijing, Chongqing, Shanghai, and Tianjin. In our data, they are provinces containing (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 6 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
Table 6: Robustness to central warehouses and trade intermediaries

<table>
<thead>
<tr>
<th>Sample:</th>
<th>#obs†</th>
<th>$h_1$</th>
<th>$Eh_1$</th>
<th>$x_1$</th>
<th>#shm</th>
<th>$&lt; Q_{0.05}$</th>
<th>$&gt; Q_{0.95}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of province</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Municipality</td>
<td>10,117</td>
<td>0.63</td>
<td>0.70</td>
<td>0.38</td>
<td>6.63</td>
<td>0.15</td>
<td>0.01</td>
</tr>
<tr>
<td>Non-municipality</td>
<td>19,342</td>
<td>0.67</td>
<td>0.72</td>
<td>0.39</td>
<td>8.06</td>
<td>0.14</td>
<td>0.02</td>
</tr>
<tr>
<td>Share of imports handled by intermediaries</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$&gt; 20%$</td>
<td>14,835</td>
<td>0.66</td>
<td>0.73</td>
<td>0.39</td>
<td>8.68</td>
<td>0.16</td>
<td>0.02</td>
</tr>
<tr>
<td>$\leq 20%$</td>
<td>14,624</td>
<td>0.65</td>
<td>0.70</td>
<td>0.38</td>
<td>6.45</td>
<td>0.12</td>
<td>0.02</td>
</tr>
</tbody>
</table>

$h_1$, $Eh_1$, and $x_1$ portray averages across province-goods. #shm is the average across province-goods of the city-level median number of shipments. $Q_{0.05}$/$Q_{0.95}$ are the 5th/95th percentiles of the $Eh_1$ distribution.

†: An observation is a province-good combination.

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.

Eaton et al. (2011) also find widespread departures from hierarchy in their study of French exporters. However, they report that there is considerably more compliance in the exporter data than would be predicted by independent choices. Their benchmark for independence is very different from the one used here as they calculate the probabilities of various export-market combinations assuming all exporters have identical probabilities of entering each market. Our method takes into account the fact that some exporters are much bigger than others (and therefore have larger “bins”).

5 Source orientation

The random sourcing model predicts significantly more compliance with hierarchy than is observed. To better match the data, we need to modify the model to lower $Eh_1$. The random model assumes that firm shipments are influenced by common perceptions about the fundamental attractiveness of individual supplying countries, represented by the $c_s T_s$ term in equation (7). There is also an idiosyncratic shipment-source specific term, $\varepsilon_{js}$, that is assumed to be an identical independent draw (IID). Here we show that in a model incorporating a firm-source shock that is common to all shipments by the same firm, expected hierarchy compliance can be reduced to the level observed in the data.

There are a number of reasons why importing firms may be oriented towards specific source countries. The most obvious reason is the tendency of foreign-owned firms to prefer 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. In addition to relationships with foreign firms, importing firms could have developed particular familiarity with the business environment in a specific source country, leading to a bias towards transacting with suppliers from that country.

Source orientation lowers expected compliance due to the non-linearity of $E h_1$. To provide intuition, we first consider a simple probabilistic approach. Suppose that firms have either a high or low probability of choosing source 1, with the unconditional probability equal to $x_1$, the province-level share of source 1. Each firm $f$ has $\pi_{1f} = \pi_1^H > x_1$ with probability $p$ and $\pi_{1f} = \pi_1^L < x_1$ with probability $1 - p$. The unconditional probability of source 1 is

$$p\pi_1^H + (1 - p)\pi_1^L = x_1.$$  

The probability of a type $H$ firm with $n_f$ balls not placing any orders from source 1 is $(1 - \pi_1^H)^{n_f}$ with the analogous expression for type $L$. Hence, the unconditional firm-level probability of no shipments from source 1 is

$$\mathbb{P}(f \text{ does not comply})_{\text{orient}} = p(1 - \pi_1^H)^{n_f} + (1 - p)(1 - \pi_1^L)^{n_f}. \quad (9)$$

When firms are not oriented, the IID case, $\pi_1^H = \pi_1^L = x_1$, and the probability of a type $H$ firm with $n_f$ balls not landing in source 1 is

$$\mathbb{P}(f \text{ does not comply})_{\text{iid}} = (1 - x_1)^{n_f}. \quad (10)$$

Since the second derivative of $(1 - x_1)^{n_f}$ with respect to $x_1$ is positive, the definition of a convex function implies that for any $\pi_1^H > x_1 > \pi_1^L$,

$$\mathbb{P}(f \text{ does not comply})_{\text{orient}} > \mathbb{P}(f \text{ does not comply})_{\text{iid}}.$$

Next, consider the probability that none of the firms in a city import from source 1. If each firm in a city chooses sources independently from every other firm, this probability is simply the product of the probabilities of non-compliance of the individual firms:

$$\mathbb{P}(\text{no firms comply}) = \prod_f \mathbb{P}(f \text{ does not comply}). \quad (10)$$

Since each factor in the multiplication is higher in the orientation case than the IID case, the probability that none of the firms in a city complies is higher in the case of orientation. Hence, the probability of compliance is lower, $\mathbb{E}[h_1]_{\text{iid}} \geq \mathbb{E}[h_1]_{\text{orient}}$.

An extreme case of source orientation is depicted in Figure 4. There are six firms denoted by the dashed circles that are distributed across four cities. Shipments per importing firm range from one to three. Here firms are completely oriented towards different source countries as portrayed by color matches (implying $\pi_1^H = 1$.
and $\pi_f = 0$. In the figure, half the cities comply with hierarchy ($h_1 = 0.5$), less than the IID expectation of $E[h_1]^{iid} = 0.71$.

We can introduce firm orientation into the economic model of random sourcing by modifying the cost function developed in Section 4. The cost to firm $f$ located in city $d$ of importing a shipment from source $s$ becomes $C_{fs} = c_s T_s u_{fs} \varepsilon_s$, where $\varepsilon_s$ remains an IID Weibull shipment-source shock (with shape parameter $\theta$) and $u_{fs}$ is a firm-source shock. The probability that a shipment of firm $f$ is filled by $s$ is

$$\pi_{fs} \equiv P[C_{fs} < C_{fs}', \forall s' \neq s] = \frac{(c_s T_s u_{fs})^{-\theta}}{\sum_h (c_h T_h u_{fh})^{-\theta}}. \quad (11)$$

As was the case in the IID random model, this probability does not depend on the firm’s location within the province (because all goods are assumed to flow through the provincial transport hub).

Following Eaton et al. (2012), we define $\Lambda_s$ as the probability of choosing $s$ evaluated with the firm-level shocks set constant (and hence eliminated).

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

It is not generally the case that $E[\pi_{fs}] = \Lambda_s$. This is because the $u_{fs}$ shocks appear in the denominator as well as the numerator of equation (11). Eaton et al. (2012)
establish that assuming $\nu_{fs} \equiv u_{fs}^{-\theta}$ to be distributed Gamma with shape and scale parameters $\Lambda_s/\eta^2$ and $\eta^2/\Lambda_s$ leads to $E[\pi_s] = \Lambda_s$.\footnote{These parameters imply $E[\nu_{fs}] = 1$ and $\text{var}(\nu_{fs}) = \eta^2/\Lambda_s$.} Using footnote 29 of Eaton et al. (2012) to determine the probability of a zero, the probability that a firm does not source from country 1 is given by

$$P(\text{f does not comply})_{\text{orient}} = \frac{\Gamma(\frac{1}{\eta^2})\Gamma(n_f + \frac{1-\Lambda_1}{\eta^2})}{\Gamma(\frac{1-\Lambda_1}{\eta^2})\Gamma(n_f + \frac{1}{\eta^2})}.$$ \hfill (12)

Raising the number of shipments increases the probability of hierarchy compliance, just as in the IID case. The difference here is that higher $\eta^2$, which raises the variance of the firm-source shock, causes the probability of non-compliance to converge to $1 - \Lambda_1 < 1$.\footnote{The proof relies on the Gamma function’s recurrence property that $\Gamma(z) = \Gamma(z + 1)/z$.}

Since this limiting value is independent of $n_f$, we see that extreme dispersion of the idiosyncratic firm-source shock converts firms into single balls in the balls-in-bins analogy, a case we refer to as “firm balls.” If firms’ draws in a destination city are independent of each other, the probability a destination complies becomes $1 - (1 - \Lambda_1)^{N_d}$ where $N_d$ is the number of firms in city $d$. The case where shipment balls become firm balls due to high $\eta^2$ is illustrated in Figure 4 where all the balls of each firm land in the same source. We can compute the average probability a city does not comply, $\mathbb{E}[h_1]_{\text{orient}}$, based on information in this example. The province level $x_1$ provides the method of moments estimate for $\mathbb{E}[h_1] = \Lambda_s$. Using $x_1 = .4$ and given that half the cities host two firms and half host one firm, $\mathbb{E}[h_1]_{\text{orient}} = 0.5[1 - (1 - 0.4)^2] + 0.5[1 - (1 - 0.4)] = 0.52$.

Now we consider whether this extreme depiction of firm orientation produces the expected amount of compliance observed in the data. We calculate expected compliance for each of the 29,459 province-good combinations using $\Lambda_1 = x_1$ and measuring balls as the number of firms importing the good in each Chinese city. The exercise yields average expected compliance of 0.59, well below average $h_1 = 0.66$, indicating that the assumption that firms always or never purchase shipments from source 1 is too extreme.

Since expected compliance under the extreme assumptions of shipment balls (IID shipment-source shocks) and firm balls (all shipments from same firm going

$$\lim_{\eta^2 \to \infty} P(\text{f does not comply})_{\text{orient}} = \frac{\Gamma(1)\Gamma(1 + n_f)(1 - \Lambda_1)(n_f)}{\Gamma(1)\Gamma(1 + n_f)(n_f)} = 1 - \Lambda_1$$
to the same source) bracket actual compliance, we can use data on $x_1$ and firm shipments ($n_f$) and equations (10) and (12) to search for the value of $\eta^2$ that matches average $h_1 = 0.66$. We find that expected and actual hierarchy compliance are equal out to three decimal points when $\eta^2 = 0.9$.

To assess the economic importance of firm-source shocks, we decompose the variance of the cost of delivering a good to the provincial hub, $C_{fs}^{ij} = C_s u_{fs} \varepsilon^{ij}_s$ (where $C_s = c_s T_s$). Since the individual elements of delivered costs are independent by construction, we can express the variance of the natural logarithm of delivered cost as

$$\text{var}[\ln(C_s u_{fs} \varepsilon^{ij}_s)] = \text{var}(\ln C_s) + \text{var}(\ln u_{fs}) + \text{var}(\ln \varepsilon^{ij}_s)$$

We focus on the relative size of the variation of the first two factors, the country-specific term ($\ln C_s$) and the firm-source idiosyncratic term ($\ln u_{fs}$). We define

$$\rho_s = \frac{\text{var}(\ln u_{fs})}{\text{var}(\ln C_s) + \text{var}(\ln u_{fs})}$$

as the relative share of the firm-source idiosyncratic term. It varies between 0, when only source country-level determinants matter, and 1, when source costs have no systematic component and all that matters is the quality of the match between an importing firm and a source country.

We previously defined $\nu_{fs} = \frac{u_{fs}^{\theta}}{\Lambda_s}$ and therefore $\text{var}(\ln u_{fs}) = \text{var}(\ln \nu_{fs})/\theta^2$. The model implies that the variance of the systematic country component of costs, $\text{var}(\ln C_s)$, is equal to $\text{var}(\ln x_s)/\theta^2$.[19] Substituting, we obtain

$$\rho_s = \frac{\text{var}(\ln \nu_{fs})}{\text{var}(\ln x_s) + \text{var}(\ln \nu_{fs})}.$$ 

Using data on $x_s$ and our calibrated values of $\eta^2$, we are able to evaluate this expression. The variance of the log of the Gamma shock $\nu_{fs}$ with shape scale parameters $\Lambda_s/\eta^2$ and $\eta^2/\Lambda_s$ is given by evaluating the trigamma function at $\Lambda_s/\eta^2$. Using $\eta^2 = 0.9$ and $\Lambda_1 = x_1$, we compute values of $\rho_1$ for each of our 29,459 province-goods. We find an average value of $\rho_1 = 0.81$. This implies that variation in firm-source factors influence sourcing to a much greater extent than variation in supplier-country factors. Specifically, variation in the firm-source component accounts for 81% of the variation in the costs perceived by importers attributable to these two factors. Variation of influences such as comparative advantage and proximity to China therefore explain only 19% of this variation.

The large role of buyer-seller factors in influencing trade patterns is not surprising given the orientation of firms to specific source countries we observe. In our data, among the 51% of the firms with more than one shipment, the mean shipment share from the top source is 83%. Evidence of the importance of firm

[19] The log shipment share of $s$ is given by $\ln x_s = \ln (C_s^{-\theta}/(\sum_h C_h^{-\theta})) = -\theta \ln C_s - \ln(\sum_h C_h^{-\theta})$. Since the last term is a constant, $\text{var}(\ln x_s) = \theta^2 \text{var}(\ln C_s)$. 

22
orientation is also found in Blum et al. (2010) who find that trade intermediaries in Chile obtain the vast majority of their imports from one or two countries.

To summarize, a simple adjustment to the random sourcing model—oriented shipments—can replicate the observed level of hierarchy compliance. Allowing for oriented balls should also allow for a better fit of the random model to the data in Armenter and Koren (2014). They find that the ball and bins model under-predicts the frequency of zeros in the country-industry trade matrix of the United States (72% zeros in the model versus 82% in actual data) and under-predicts the fraction of exporters selling to only one country (45% versus 64%). If the export shipments of US firms displayed idiosyncratic orientation towards specific destination countries, they would find more zeros in the trade matrix and more single-country exporters. The balls and bins model provides a remarkably good first pass to the data but its fit improves when we relax the extreme assumption of IID shipments.

6 Conclusion

Prominent trade models predict that importers will buy from multiple countries and exporters can be ranked according to a hierarchy in which all buyers purchase from the top ranked source of supply. This implies that all Chinese cities import from the top provincial source, a prediction violated about one-third of the time in our data. Of course, stark theoretical predictions rarely hold exactly in the data. More surprisingly, we find that hierarchy is observed significantly less often than predicted by the ball-and-bins model of Armenter and Koren (2014). We introduce firm-level cost shocks into a random sourcing model inducing correlation between firm shipments. The modified model can be calibrated to match the average compliance with hierarchy observed in the data.

Our results reveal that random factors play a major role in determining who buys from whom. Moreover, these factors are not independent and identically distributed. Instead, they exhibit strong correlation within firms. Calibration reveals that 81% of the variation in the costs of exporting to China can be attributed to variation in buyer-supplier idiosyncratic factors and only 19% is due to variation in source-country factors and trade costs. The findings of this paper motivate further research on formation of bilateral connections between buyers and sellers as a key determinant of international trade patterns.
References


