NBER WORKING PAPER SERIES

THE DETERMINANTS OF QUALITY SPECIALIZATION

Jonathan I. Dingel

Working Paper 22757
http://www.nber.org/papers/w22757

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2016

I am grateful to Donald Davis, Amit Khandelwal, Eric Verhoogen, Jonathan Vogel, and David Weinstein for invaluable discussions, guidance, and support. Thanks to the editor, three anonymous referees, Michael Best, Brianna Cardiff, Arnaud Costinot, Dave Donaldson, Ben Faber, Pablo Fajgelbaum, Gene Grossman, Juan Carlos Hallak, Jessie Handbury, Corinne Low, Kyle Meng, Joan Monras, Eduardo Morales, David Munroe, Paul Piveteau, Bernard Salanie, Daniel Sturm, Allie Tepper, Felix Tintelnot, Sebastien Turban, Reed Walker, Columbia colloquia participants, and numerous seminar audiences for very helpful comments and suggestions. I am grateful to Rob Feenstra and John Romalis for sharing their parameter estimates. Support from the Institute for Humane Studies and the Kathryn and Grant Swick Faculty Research Fund at the University of Chicago Booth School of Business is gratefully acknowledged. This research was conducted while I was a special sworn status researcher of the US Census Bureau. Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the US Census Bureau or the National Bureau of Economic Research. All results have been reviewed to ensure that no confidential information is disclosed.

The Determinants of Quality Specialization
Jonathan I. Dingel
NBER Working Paper No. 22757
October 2016
JEL No. F12,F14,R12

ABSTRACT

A growing literature suggests that high-income countries export high-quality goods. Two hypotheses may explain such specialization, with different implications for welfare, inequality, and trade policy. Fajgelbaum, Grossman, and Helpman (2011) formalize the Linder hypothesis that home demand determines the pattern of specialization and therefore predict that high-income locations export high-quality products. The factor-proportions model also predicts that skill-abundant, high-income locations export skill-intensive, high-quality products. Prior empirical evidence does not separate these explanations. I develop a model that nests both hypotheses and employ microdata on US manufacturing plants' shipments and factor inputs to quantify the two mechanisms' roles in quality specialization across US cities. Home-market demand explains as much of the relationship between income and quality as differences in factor usage.

Jonathan I. Dingel
Booth School of Business
University of Chicago
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
jdingel@chicagobooth.edu

1 Introduction

The Linder hypothesis is the oldest theory of quality specialization in international trade. Staffan Burenstam Linder (1961) posited that profitably exporting a product requires robust demand for that product in the exporter's home market. Since higher-income consumers tend to purchase higher-quality products, he conjectured that local demand causes high-income countries to produce and export high-quality products. This "home-market effect" explanation of quality specialization was recently formalized by Fajgelbaum, Grossman, and Helpman (2011) in a general-equilibrium model. In contrast, the canonical factor-abundance theory of comparative advantage identifies high-income countries' greater supplies of capital and skills as the reason they produce and export high-quality products.[1] These competing theories have distinct implications for welfare, inequality, and trade policy. Empirical work to date has not identified the importance of each mechanism in quality specialization.

The empirical challenge is that the two theories make the same predictions about country-level trade flows. Each predicts that high-income locations export high-quality products, consistent with the finding that higher-income countries export products at higher prices within narrowly defined product categories (Schott, 2004; Hummels and Klenow, 2005).[2] Similarly, each predicts that non-homothetic preferences cause high-income locations to import high-quality products, consistent with the finding that higher-income countries import more from higher-price exporters (Hallak, 2006; Choi, Hummels, and Xiang, 2009). Combining these export and import patterns, each predicts that countries with more similar incomes trade more intensely with each other, as found by Hallak (2010) and Bernasconi (2013).[3]

In this paper, I use theory and data to quantify the roles of the home-market effect and the factor-abundance mechanism in quality specialization across US cities. I develop a model that yields an empirical approach to separate the two mechanisms. It exploits plant-level data on shipments and inputs and location-level data on populations and incomes. I implement the empirical strategy using data on US cities and manufacturing plants and find that the home-market effect influences quality specialization across cities of different income levels as much as factor abundance.

To guide my empirical investigation, I introduce a theoretical framework that nests the

[1] For example, Schott (2004, p. 676) suggests that "high-wage countries use their endowment advantage to add features or quality to their varieties that are not present among the varieties emanating from low-wage countries." Linking quality specialization to relative factor supplies dates to at least Falvey (1981).

[2] Throughout this paper, observed "prices" refer to unit values, which are shipments' value-to-quantity ratios. Like international trade data, the data used in this paper describe transactions' values and quantities.

[3] Hallak (2010, p. 459) notes that "several theories can explain a systematic relationship between per capita income and quality production... The prediction of the Linder hypothesis about the direction of trade can be founded on any of these theories."

two mechanisms, each of which has been studied separately. Individuals have non-homothetic preferences over a homogeneous and a differentiated good; higher-income individuals consume higher-quality varieties of the differentiated good. This demand assumption makes high-income locations import high-quality products and generates the home-market effect when trade is costly. Individuals have heterogeneous skills, and goods can be ranked by their skill intensities. This production assumption allows skill-abundant locations to have a comparative advantage in higher qualities when quality is skill-intensive. The model serves two purposes. First, it confirms that each mechanism alone can generate trade flows consistent with the empirical findings described above. Second, the theory identifies a way to separate the two mechanisms using plant-level data. Factor abundance affects specialization exclusively through plants' factor usage. Conditional on plant-level factor intensity, demand alone determines quality specialization. Thus, plant-level data on shipments and inputs can be combined with data on locations' incomes to identify the home-market effect.

To implement this empirical strategy, I use microdata on US manufacturing plants' shipments and inputs from the Commodity Flow Survey and the Census of Manufactures. These sources provide microdata on plants in many cities with different income levels in a single dataset. In contrast, I am not aware of a source containing plant-level shipment and input data from many countries.[4] I document that US cities exhibit the key patterns found in international data. Both outgoing and incoming shipments exhibit higher prices in higher-income cities, and cities with more similar incomes trade more intensely with each other. I therefore proceed to use these data to distinguish between the two hypotheses by constructing empirical measures of factor inputs and market access.

Guided by the model, my empirical investigation yields two main results. First, observed differences in plants' inputs, which may be induced by either mechanism, explain only a modest share of within-product specialization across cities of different incomes. Most of the variation is within-factor-intensity variation. Second, a market-access measure that describes the income composition of proximate potential customers is strongly related to the pattern of within-intensity specialization. Quantitatively, I find that the home-market effect plays at least as large a role as the factor-abundance mechanism in quality specialization by income.

More specifically, in my empirical work I infer quality specialization from two empirical measures commonly used in the literature: unit values and demand shifters. The first measure is based on the idea that higher-quality products sell at higher prices and has been widely used in the international trade literature (Hummels and Skiba, 2004; Schott, 2004;

[4] My empirical approach thus follows the counsel of Krugman (1991, p.3): "if we want to understand international specialization, a good place to start is with local specialization. The data will be better and pose fewer problems of compatibility, and the underlying economic forces will be less distorted by government policies."

Hallak, 2006; Baldwin and Harrigan, 2011). The second measure follows Sutton (1991, 2012), Berry (1994), Hummels and Klenow (2005), Khandelwal (2010), and others in identifying a product as higher-quality when, conditional on price, it has higher market share. When both measures are available in my data, they yield comparable results.

The first empirical finding is that observed factor-usage differences explain a modest share of within-product specialization. Guided by the model, I construct factor-intensity measures using data on plants' employees, equipment, and wages. Between-intensity variation explains about one quarter of the covariance between locations' per capita incomes and outgoing shipment prices. It explains a larger share of the covariance between incomes and demand shifters, but observed factor-usage differences never explain more than half of the specialization by income per capita in any regression specification. Since the factor-abundance mechanism operates only through between-intensity variation, this finding bounds its explanatory power, at least in terms of observed factor usage.[5]

The second empirical finding is that the home-market effect plays a quantitatively significant role in quality specialization, at least as large as differences in observed factor usage. Using data on cities' incomes and geographic locations, I construct two market-access measures describing the income composition of proximate potential customers. The first omits the residents of the city in which the plant is located, so that it does not reflect any unobserved local supply-side mechanisms. I find that this measure of demand is strongly positively correlated with manufacturing plants' outgoing shipment prices. In fact, this measure explains a larger share of the covariance between income per capita and outgoing shipment prices, 36%, than plant-level factor usage. The second market-access measure follows the model by including residents in the city of production. This demand measure consistently explains a larger share of the observed specialization across cities of different incomes than plants' factor inputs. Within-intensity variation in market access explains 58% of the covariance between product prices and incomes per capita, twice that attributable to factor-usage differences.[6] It explains a similar share, 48%, of the covariance between demand shifters and incomes per capita.[7] Market access is orthogonal to incoming shipment characteristics, so proximity to higher-income consumers is associated with net exporting of higher-quality varieties. I conclude that the home-market effect for quality plays a substantial role in the economic geography of US manufacturing.

[5]Section 5 discusses the particular properties unobserved inputs would need to exhibit in order to account for my findings.

[6]Using only within-intensity variation is conservative. Unconditionally, variation in market access accounts for 77% of the price-income covariance.

[7]Factor-usage differences explain 46% of the covariance between demand shifters and incomes per capita, so there is considerably smaller residual variation in the decomposition of this measure.

These findings are important because the two theories have distinct implications. In predicting the quality of a location's exports, one emphasizes its relative factor supplies while the other stresses its relative proximity to high-income customers. These yield different predictions, for instance, for poor countries that have rich neighbors.[8] To the extent that specializing in producing high-quality goods improves growth prospects, the home-market effect found here suggests an advantage of proximity to high-income countries.[9] And since trade policy governs market access, governments may influence quality specialization.[10]

My empirical strategy of using plant-level data from US cities of different income levels links my results to a number of findings in urban and regional economics. I provide the first characterization of production specialization within product categories across cities. Previous empirical work describing variation in manufacturing across US cities has focused on inter-industry specialization (Henderson, 1991; Holmes and Stevens, 2004; Davis and Dingel, 2014) or described the products available to retail consumers without tracking production locations (Handbury and Weinstein, 2015). The finding that the geography of demand plays a major role in specialization complements a nascent literature describing the consumption benefits of living in cities with high-income populations (Glaeser, Kolko, and Saiz, 2001; Handbury, 2012; Diamond, 2016).

The paper is organized as follows. Section 2 describes the two competing hypotheses. Section 3 introduces a model nesting both and shows how to separate them using plant-level data. Section 4 describes the US microdata and pattern of specialization and exchange. Section 5 reports the empirical results. Section 6 concludes.

2 Background

Burenstam Linder (1961) posited that demand differences can determine production specialization.[11] Krugman (1980) formalized how economies of scale and trade costs can cause a country with a larger home market for a product to be a net exporter of that good. First, economies of scale cause each product to be produced in a single location and sold to many markets. Second, producing in the larger market minimizes transportation costs. Krugman (1980) obtains this result by assuming exogenous differences in countries' demand for

[8]For example, Mexico and Turkey are developing economies that are proximate to high-income customers in the US and EU, respectively. Verhoogen (2008) shows that increased incentive to export caused quality upgrading by Mexican firms.

[9]See Redding (1996) and Lederman and Maloney (2012) on quality and growth.

[10]Helpman and Krugman (1989, p.2): "It is clear that changing one's view of why trade happens, and how international markets work, ought to change one's view of what kind of trade policy is appropriate."

[11]His informal narrative focused on the role of entrepreneurial discovery (p.89-90). He emphasized informational costs of distance more than transport costs and did not explicitly address economics of scale.

different industries' products. Fajgelbaum, Grossman, and Helpman (2011) show how income differences can determine quality specialization within products when preferences are non-homothetic. In their model, the composition of income determines the composition of demand, since higher-income households purchase higher-quality varieties. Plants produce higher qualities in higher-income locations because it is more profitable to produce in the larger home market. In equilibrium, greater demand elicits a more-than-proportionate production response, such that high-income locations are net exporters of high-quality products.

The canonical factor-abundance theory of comparative advantage can yield the same set of predictions when preferences are non-homothetic. An early example is Markusen (1986), in which the income elasticity of demand for capital-intensive manufactures is greater than one, so that high-income, capital-abundant countries specialize in manufactures that are exported to other high-income countries.[12] Many other models make analogous assumptions about the alignment of comparative advantage and relative demand, so that "tastes and capabilities are correlated" but not causally linked (Murphy and Shleifer, 1997, p. 6).[13] In these theories, higher-income countries are net exporters of higher-quality products if the comparative-advantage mechanism exceeds differences in demand.

Thus, both theories are consistent with the growing body of empirical evidence suggesting that higher-income countries export and import higher-quality products. Schott (2004) shows that unit values in product-level US import data are higher for higher-income, more capital- and skill-abundant exporting countries; Hummels and Klenow (2005) find a positive relationship between unit values and exporter incomes using data from 59 importing countries. Khandelwal (2010) estimates demand shifters using US import data and finds that they are positively related to exporting countries' GDP per capita and capital abundance. Feenstra and Romalis (2014) and Hallak and Schott (2011), using other methods, also report that higher-income countries export products inferred to be higher quality.

These common predictions for country-level trade flows motivate this paper's use of plant-level data to separate the two mechanisms. In short, the challenge prior work has faced is that customers and workers are the same people in country-level data.[14] As my model

[12]See also Bergstrand (1990). Strictly speaking, these are general-equilibrium models of intersectoral specialization. Falvey (1981) introduced a partial-equilibrium model of within-industry specialization across qualities by capital intensity consonant with the within-product interpretation of factor-abundance theory suggested by Schott (2004).

[13]Two recent papers study specialization across sectors using models with non-homothetic preferences. Using aggregate trade flows, Fieler (2011) estimates a two-sector version of the Eaton and Kortum (2002) Ricardian model. She infers that the more income-elastic industry has greater dispersion in idiosyncratic productivities, causing higher-TFP countries to have comparative advantage in these luxuries. Examining variation across 56 sectors, Caron, Fally, and Markusen (2014) find a positive correlation between industries' income elasticities of demand and skill intensities. These perfect-competition models do not feature home-market effects.

[14]In addition to looking at country-level capital abundance, Schott (2004) shows that the unit values

demonstrates, assessing the factor-abundance hypothesis requires looking at the factors of production employed by exporting plants. A series of studies using firm-level data have shown that exporters and firms producing higher-quality products use more capital-intensive and skill-intensive production (Verhoogen, 2008; Hallak and Sivadasan, 2013). These firm-level findings are consistent with the factor-abundance explanation of quality specialization. But they do not provide evidence that differences in factor abundance relate to differences in output across locations, since they describe establishments in a single location.

As a result, there is no prior empirical evidence distinguishing the home-market effect for quality from factor-abundance-determined quality specialization. There is an empirical literature on the Helpman and Krugman (1985) home-market effect, in which a larger home market causes specialization in the industry with greater economies of scale (Davis and Weinstein, 1999, 2003; Hanson and Xiang, 2004). This work has relied upon using observable sectoral characteristics, such as transport costs and demand elasticities. Such cross-industry variation is unavailable when considering quality specialization within products. Moreover, since the distributions of income and human capital are closely related, both across countries and cities, it is empirically difficult to distinguish the home-market effect from factor-abundance theories of comparative advantage using aggregate data.

I proceed to introduce a theoretical framework that incorporates both mechanisms and their interaction in equilibrium. This allows me to derive an empirical strategy that relies on observing plants' inputs and outputs.

3 Theory

I introduce a theoretical framework in which both market access and factor abundance may influence the pattern of production and exchange. I use a high-dimensional model with many locations, qualities, and skills.[15] It nests a version of the Fajgelbaum, Grossman, and Helpman (2011, henceforth FGH) model and a factor-abundance model as special cases. Nesting the two mechanisms in one framework allows me to analyze each in isolation and their interaction. For brevity, details and derivations appear in appendix A.

The theory delivers two results key to the empirical investigation. First, it confirms that

of exported products are positively correlated with the capital-labor ratio of the relevant three-digit ISIC industry in the exporting country. However, much of the variation reflects cross-country differences in capital abundance, a fact noted by Dollar, Wolff, and Baumol (1988, p. 33). The mean pairwise correlation between the 28 industries' capital-labor ratios across the 34 countries in the Schott (2003) data is 0.5. Moreover, industry data necessarily aggregate heterogeneous plants and may not represent exporters' factor intensities.

[15] Matching the facts that both outgoing and incoming shipment prices are increasing in average income necessitates a many-location model. Making comparisons across and within qualities of different factor intensities, which is at the heart of my empirical strategy, necessitates many quality levels.

quality specialization is overdetermined. Each mechanism alone can cause higher-income locations to produce, export, and import higher-quality varieties in equilibrium. Second, the theory identifies an important distinction between the two mechanisms. Conditional on plant-level skill intensity, any correlation between local income and plants' output quality is due to the home-market effect. This result is the basis of my empirical approach. The model also informs my construction of empirical measures of the relevant objects.

In the model, there are K locations indexed by k. Location k has a population of size N_k made up of heterogeneous individuals whose skills, indexed by ω, are distributed according to the density $f(\omega, k)$. These skill distributions are exogenous, a standard assumption in trade theory that is innocuous for the purpose of distinguishing the roles of the two mechanisms.[16] I assume that locations can be ranked by their skill abundance in the likelihood-ratio sense. The density $f(\omega, k)$ is strictly log-supermodular, so high-k locations are skill-abundant.[17]

3.1 Preferences

Consumer preferences are non-homothetic, so demand varies with income. As in FGH, individuals consume a homogeneous good and one unit of a differentiated good. Varieties of the latter are indexed by j, with price p_j and quality $q_j \in Q$. For individual h, the utility of consuming z units of the homogeneous good and a unit of variety j is

$$u_{hj} = zq_j + \epsilon_{hj}, \tag{1}$$

with idiosyncratic valuation ϵ_{hj} drawn from a generalized extreme value (GEV) distribution.

A consumer chooses quantity z and variety j to maximize utility. The homogeneous good is the numeraire. A consumer with income y_h therefore chooses j to maximize $(y_h - p_j)q_j + \epsilon_{hj}$, where $z = y_h - p_j$. As FGH show, if ϵ's GEV distribution has dispersion θ_q for q, this yields a nested-logit demand system in which the fraction of consumers with income y buying variety j of quality q can be expressed as $\rho_j(y) = \rho_{j|q} \cdot \rho_q(y)$. $\rho_q(y)$ is the fraction with income y who choose a product of quality q, and $\rho_{j|q} = \exp\left(-p_j q/\theta_q\right) / \sum_{j':q_{j'}=q} \exp\left(-p_{j'} q/\theta_q\right)$ is the fraction buying j among those buying quality q. The latter is income-invariant.

This demand system has two important properties. First, the complementarity between z and q_j in equation (1) makes higher-income consumers more likely to choose higher-quality

[16] Factor mobility would be relevant in considering counterfactuals, since individuals may migrate across cities in response to economic changes. My empirics characterize the equilibrium observed in the data, and there is substantial variation in both skill distributions and income levels across US cities.

[17] I make extensive use of log-supermodularity as an analytical tool; see Costinot (2009) for an introduction. In $\mathbb{R}^2$, a function $f(\omega, k)$ is log-supermodular if $\omega > \omega'$, $k > k' \Rightarrow f(\omega, k)f(\omega', k') \geq f(\omega, k')f(\omega', k)$ and strictly log-supermodular when the inequality is strict. Using educational attainment data, Davis and Dingel (2014) provide evidence that US cities' skill distributions are broadly consistent with this assumption.

varieties. Market share $\rho_q(y)$ varies with income according to $\frac{1}{\rho_q(y)}\frac{\partial \rho_q(y)}{\partial y} = q - q_a(y)$, where $q_a(y)$ is the average quality consumed by individuals with income y. Second, if firms ignore their own effect on the price index, the elasticity of demand is $\frac{\partial \ln \rho_j(y)}{\partial \ln p_j} = -p_j q_j/\theta_q$, so producers of quality q charge a constant additive markup of $m_q \equiv \theta_q/q$.[18]

3.2 Production

Production involves employing workers of heterogeneous skills, so relative factor supplies may be a source of comparative advantage. The homogeneous good is competitively produced and freely traded. Differentiated varieties are produced by monopolistically competitive firms.

Production of the homogeneous good exhibits constant returns to scale, so the total cost of producing quantity $x(z,k)$ at unit cost $c(z,k)$ is $x(z,k)c(z,k)$.[19] Skill ω in location k commands wage $w(\omega,k)$. Hiring $\ell(\omega)$ units of skill ω per unit of output, the unit cost is

$$c(z,k) = \min_{\ell(\omega)} \int_{\omega\in\Omega} \ell(\omega)w(\omega,k)d\omega \quad \text{s.t.} \quad \left(\int_{\omega\in\Omega} b(\omega,z)\ell(\omega)^{\frac{\sigma-1}{\sigma}}d\omega\right)^{\frac{\sigma}{\sigma-1}} \geq 1.$$

The technological coefficients $b(\omega,z)$ describe the contribution of each skill type in production and therefore characterize the homogeneous good's skill intensity. The elasticity of substitution across inputs σ is greater than one and finite. Cost minimization yields per-unit input demands $\ell(\omega,z,k) = w(\omega,k)^{-\sigma}b(\omega,z)^{\sigma}$ wherever $x(z,k) > 0$.

Firms may produce a differentiated variety of quality q by paying fixed cost f_q in units of the numeraire. The constant marginal cost of producing quality q in location k is

$$c(q,k) = \min_{\ell(\omega)} \int_{\omega\in\Omega} \ell(\omega)w(\omega,k)d\omega \quad \text{s.t.} \quad \left(\int_{\omega\in\Omega} b(\omega,q)\ell(\omega)^{\frac{\sigma-1}{\sigma}}d\omega\right)^{\frac{\sigma}{\sigma-1}} \geq 1.$$

Thus, unit input demands are $\ell(\omega,q,k) = w(\omega,k)^{-\sigma}b(\omega,q)^{\sigma}c(q,k)^{\sigma}$, with marginal cost

$$c(q,k) = \left(\int_{\omega\in\Omega} b(\omega,q)^{\sigma}w(\omega,k)^{1-\sigma}d\omega\right)^{\frac{1}{1-\sigma}}.$$

A firm producing $x(q,k)$ units of quality q in location k hires $x(q,k)\ell(\omega,q,k)$ of skill ω.

Firms producing a differentiated variety in location k can export one unit to destination k' at marginal cost $c(q,k) + \tau_{qkk'}$, where the trade cost $\tau_{qkk'}$ is incurred in units of the

[18] I use the nested-logit demand system in part because the constant-additive-markup property makes the model analytically tractable. Only the first property, that high-income consumers are more likely to purchase high-quality varieties, is necessary for the home-market effect to influence the pattern of specialization.

[19] I abuse notation using z to index the homogeneous good. In equation (1), z denotes the quantity of this good.

numeraire. Taking competitors' behavior as given, the profit-maximizing prices charged by firm j producing quality q in k are a constant markup over cost, $p_{jk'} = c(q,k) + \tau_{qkk'} + m_q$.

We can now identify the equilibrium sales level in location k' for a variety of quality q produced in k, which I denote $d_{qkk'}$. If k' has $N_{k'}$ consumers with income distribution $g(y,k')$, this is $d_{qkk'} = N_{k'} \int \rho_j(y) g(y,k') dy$. Denote the number of firms producing varieties of quality q in location k by $n_{q,k}$. Plugging in optimal prices, sales $d_{qkk'}$ can be written in terms of (vectors of) the number of firms ($\mathbf{n}$), unit costs ($\mathbf{c}$), and trade costs (τ).

$$d_{qkk'} = N_{k'} \int \rho_j(q)\rho_q(y) g(y,k') dy = \exp(-(c(q,k) + \tau_{qkk'})/m_q) N_{k'} \Gamma_{k'}(q, \mathbf{n}, \mathbf{c}, \tau)$$

The function $\Gamma_{k'}(q, \mathbf{n}, \mathbf{c}, \tau)$ describes the share of demand in location k' for quality q given the equilibrium prices and locations of all producers. A firm's sales of quality q to k' from k depend on this demand share, population $N_{k'}$, marginal cost $c(q,k)$, and trade cost $\tau_{qkk'}$.[20]

3.3 Equilibrium

In equilibrium, labor markets clear and firms earn zero profits. The full-employment condition for each skill ω in each location k is $f(\omega,k) = x(z,k)\ell(\omega,z,k) + \sum_{q\in Q} n_{q,k} x(q,k)\ell(\omega,q,k)$. Plugging in firms' labor demands and defining $n_{z,k} = 1$, we can write this as

$$f(\omega,k) = w(\omega,k)^{-\sigma} \sum_{r \in z \cup Q} n_{r,k} x(r,k) b(\omega,r)^{\sigma} c(r,k)^{\sigma}, \tag{2}$$

where I use r to sum both the homogeneous good and qualities of the differentiated good.

The free-entry condition says that the profits from producing quality q in location k are non-positive everywhere and zero where firms are active: $\pi_{q,k} \leq 0 \; \forall k$ and $n_{q,k} > 0 \Rightarrow \pi_{q,k} = 0$.

$$\begin{aligned}\pi_{q,k} =& \sum_{k'} (p_{qkk'} - c(q,k) - \tau_{qkk'}) d_{qkk'} - f_q = m_q \sum_{k'} d_{qkk'} - f_q \\ =& m_q \exp(-c(q,k)/m_q) \sum_{k'} \exp(-\tau_{qkk'}/m_q) N_{k'} \Gamma_{k'}(q, \mathbf{n}, \mathbf{c}, \tau) - f_q \end{aligned} \tag{3}$$

3.4 Equilibrium pattern of specialization and trade

Given the geographic distribution of skills, individuals' preferences, and the production technology, the equilibrium pattern of production and trade depends on two forces. Trade costs

[20] The model therefore implies a gravity equation for exports of q from k to k', but q is not an observable characteristic. Since gravity equations do not aggregate by summation, the model does not deliver a closed-form gravity equation for total exports of differentiated varieties from k to k'.

shape the pattern of market access and therefore the home-market effect. Skill intensities, governed by $b(\omega, r)$, link output composition to skill supplies through labor-market clearing.

I consider two cases for each force. The two trade-cost matrices are costless trade and trade costs that are small but positive, $\tau_{qkk'} > 0$ for $k \neq k'$ and $\tau_{qkk} = 0$. The two skill-intensity cases are uniform skill intensities, $b(\omega, r) = b_1(\omega) b_2(r)$, and skill intensities that are increasing in quality, $b(\omega, r)$ weakly log-supermodular.[21]

I analyze the four cases in turn. When neither mechanism is active, the pattern of production is indeterminate. Section 3.4.1 characterizes equilibrium when only the factor-abundance mechanism is active, while Section 3.4.2 does likewise for the home-market effect. In each case, high-k locations both export and import high-q varieties. Thus, each mechanism alone could account for previously documented empirical patterns. Section 3.4.3 describes equilibrium when both mechanisms are active and shows how to identify the home-market effect after conditioning on plants' skill intensities.

To facilitate the analysis, define a skill-intensity index $i(r)$ such that $i(r) = i(r') \iff b(\omega, r) \propto b(\omega, r')$ and $i(r) > i(r') \Rightarrow r > r'$. This index groups together products so that producers in the higher-i group employ relatively more skilled labor. Let $i(r)$ equal the lowest r in this set of products, so that $b(\omega, i)$ is strictly log-supermodular by definition.

3.4.1 Skill-intensive quality and costless trade

First, consider the case when trade is costless and $b(\omega, r)$ is weakly log-supermodular. Costless trade ($\tau_{qkk'} = 0\ \forall q\ \forall k\ \forall k'$) means that $\pi_{q,k}$ in the zero-profit condition (3) depends on k only through the $c\,(q, k)$ term. In the absence of trade costs, variation in demand across destinations k' is orthogonal to the location of production. Producing quality q is most profitable wherever its unit cost $c(q, k)$ is lowest.

Skill abundance governs the pattern of production through labor-market clearing. Equation (2) and the strict log-supermodularity of $f(\omega, k)$ imply, for $k > k'$ and $\omega > \omega'$,

$$\frac{w(\omega, k)^{-\sigma}}{w(\omega', k)^{-\sigma}} \mathbb{E}_{\omega',k} \left(\frac{b(\omega, i)^\sigma}{b(\omega', i)^\sigma} \right) > \frac{w(\omega, k')^{-\sigma}}{w(\omega', k')^{-\sigma}} \mathbb{E}_{\omega',k'} \left(\frac{b(\omega, i)^\sigma}{b(\omega', i)^\sigma} \right), \tag{4}$$

where $\mathbb{E}_{\omega',k}\left[\alpha(i)\right]$ is an output-share-weighted average of $\alpha(i)$ for production in k, with output shares weighted by use of skill ω'.[22] Since $b(\omega, i)$ is strictly log-supermodular, $\frac{b(\omega,i)^\sigma}{b(\omega',i)^\sigma}$ is strictly increasing in i, and $\mathbb{E}_{\omega',k} \left(\frac{b(\omega,i)^\sigma}{b(\omega',i)^\sigma} \right)$ is a measure of the average skill intensity of output in k.

[21] When $b(\omega, q)$ is log-supermodular, quality is skill-intensive. By allowing z to take any value, I make no assumption on the skill intensity of the homogeneous good, but I assume that there is a value z making $b(\omega, r)$ a log-supermodular function.

[22] For expositional convenience, I assume $f(\omega, k) > 0\ \forall \omega \in \Omega\ \forall k$, so that $k > k, \omega > \omega' \Rightarrow \frac{f(\omega,k)}{f(\omega',k)} > \frac{f(\omega,k')}{f(\omega',k')}$.

This inequality means that more skill-abundant (higher-k) locations produce more skill-intensive (higher-i) products. By skill abundance, products made in k are more skill-intensive ($\mathbb{E}_{\omega',k}\left(\frac{b(\omega,i)^{\sigma}}{b(\omega',i)^{\sigma}}\right)$ is greater) and/or skilled labor in k is relatively cheaper ($\frac{w(\omega,k)^{-\sigma}}{w(\omega',k)^{-\sigma}}$ is greater). The latter implies skill-intensive products' unit costs are relatively lower in k, and thus k's output must be more skill-intensive in equilibrium. When $b(\omega, q)$ is log-supermodular, higher-quality varieties are more skill-intensive, so we interpret inequality (4) as saying that k absorbs its greater supply of higher skills by producing higher-quality varieties.[23]

This result describes the factor-abundance mechanism for quality specialization. Note that specialization across qualities of the same skill intensity is indeterminate in this case, because inequality (4) depends on $i(q)$, not q. Skill-abundant locations produce higher-quality varieties only because such products are more skill-intensive.[24]

What about the equilibrium pattern of demand? Since trade is costless, varieties' prices and income-specific market shares do not vary across locations. Denoting the equilibrium variety counts and factor prices by the vectors $\bar{\mathbf{n}}$ and $\bar{\mathbf{c}}$, sales volumes $\Gamma_k(q, \bar{\mathbf{n}}, \bar{\mathbf{c}}, \mathbf{0})$ vary with location k solely due to differences in the composition of income. Demand for higher-quality varieties is relatively greater in higher-income locations.

Result. When trade is costless, there is no home-market effect. When quality is skill-intensive and skill-abundant locations are higher-income locations, higher-income locations both produce more of and have greater demand for higher-quality varieties in equilibrium.

3.4.2 Uniform skill intensities and costly trade

When skill intensities are uniform, unit costs are multiplicatively separable in (r, k) and can be written as $c(r,k) = b_2(r)^{\frac{\sigma}{1-\sigma}} c(k)$. The labor-market clearing condition (2) becomes

$$f(\omega,k) = b_1(\omega)^{\sigma} w(\omega,k)^{-\sigma} c(k) \sum_{r \in z \cup Q} n_{r,k} x(r,k) b_2(r)^{\frac{\sigma}{1-\sigma}}.$$

Since nothing inside the sum depends on skill ω, the factor-abundance mechanism imposes no restrictions on the equilibrium composition of local production $n_{r,k}x(r,k)$. Any observed re-

[23] This interpretation neglects the skill intensity of the homogeneous good. If the homogeneous good is more skill-intensive, skill-abundant locations may produce more of the homogeneous good rather than higher-quality varieties. When factor intensities vary both across and within goods, the factor-abundance mechanism may operate along both margins. Empirically, Schott (2004) documents that there is little correlation between countries' factor supplies and across-good specialization. Assuming that the homogeneous good is the least skill-intensive product is sufficient to guarantee that high-k locations specialize in high-q varieties.

[24] This result has been derived without any reference to the demand system beyond the fact that costless trade makes consumers' locations irrelevant to the optimal production location. Thus, the empirical investigation of whether the factor-abundance mechanism alone can explain the pattern of specialization does not depend upon the functional form of the preferences in equation (1).

lationship between $f(\omega, k)$ and the pattern of specialization results from the demand channel and reflects the connection between $g(y, k)$ and $f(\omega, k)$.

To characterize how specialization is determined by demand, I follow the approach taken by FGH to determining the equilibrium pattern of production when trade costs are small and locations specialize.[25] With uniform skill intensities, the zero-profit condition is

$$\pi_{q,k} = m_q \exp(-b_2(q)^{\frac{\sigma}{1-\sigma}} c(k)/m_q) \sum_{k'} \exp(-\tau_{qkk'}/m_q) N_{k'} \Gamma_{k'}(q, \mathbf{n}, \mathbf{c}, \tau) - f_q \leq 0.$$

Through this condition, demand governs the location of production in equilibrium. Consider two cases, depending on whether wages vary across locations.

When factor prices equalize, $c(k) = 1 \ \forall k$ and $\pi_{q,k}$ varies only with demand. If trade costs are uniform ($\tau_{qkk'} = \tau_q \ \forall k' \neq k$), then profits vary only with home demand, $\pi_{q,k} > \pi_{q,k'} \iff N_k \Gamma_k(q, \mathbf{n}, \mathbf{c}, \tau) > N_{k'} \Gamma_{k'}(q, \mathbf{n}, \mathbf{c}, \tau)$. Provided that wages are increasing in skill, high-k locations are high-income because they are skill-abundant, and they have greater demand for high-q varieties. This makes producing high-q varieties more profitable in high-k locations. When population sizes are equal, Proposition 6 of FGH describes the resulting equilibrium: if location k produces quality q and location $k' < k$ produces quality q', then $q' < q$. Similarly, since higher-k locations have greater demand for higher-q varieties, their imports are higher-quality (see Proposition 7 of FGH). In the case of $\sigma = \infty$ and $N_k = 1 \ \forall k$, the model under consideration reduces to that in section VII of FGH.

When factor prices do not equalize, the location with the lowest $c(k)$ is the most attractive cost-wise for all producers. Firms locate in higher-cost locations only if these locations have greater demand for their output so that they save on transport costs. In other words, when trade costs are uniform, if $n_{q,k} > 0$ and $c(k) > c(k')$, it must be that $N_k \Gamma_k(q, \mathbf{n}, \mathbf{c}, \tau) > N_{k'} \Gamma_{k'}(q, \mathbf{n}, \mathbf{c}, \tau)$. When population sizes are equal and trade costs are sufficiently low, this difference in demand is due solely to the income composition of the two locations. Thus, higher-income locations specialize in producing higher-quality varieties and export them.

Result. Suppose that population sizes are equal, skill intensities are uniform, and trade costs are uniform and small. If $k > k'$, $n_{q,k} > 0$, and $n_{q',k'} > 0$, then $q > q'$ and $n_{q,k'} = n_{q',k} = 0$. Higher-income locations are net exporters of higher-quality varieties because demand for such qualities is greater in such locations. If consumers in k and k' import varieties of qualities q and q' with $q > q'$, then k imports relatively more of quality q.

Thus, the home-market effect yields equilibrium trade patterns that match the empirical evidence summarized in section 2. Since we obtained the same result in the previous section via the factor-abundance mechanism, quality specialization is overdetermined.

[25] Appendix section A.4.2 discusses the case when trade costs are large and production is diversified.

Result. Higher-income locations both exporting and importing higher-quality varieties is consistent with the factor-abundance mechanism or the home-market effect operating alone.

3.4.3 Skill-intensive quality and costly trade

Now suppose both mechanisms are active. When quality is skill-intensive and trade is costly, the labor-market clearing condition (2) and the zero-profit condition (3) jointly govern the pattern of quality specialization. The critical result that underlies my empirical investigation is that demand alone determines specialization across varieties of the same skill intensity.

First, consider the labor-market-clearing inequality, which is governed by the factor-abundance mechanism. As shown previously, inequality (4) implies that output of higher-i varieties is relatively greater in higher-k locations. Thus, skill intensities govern the broad pattern of production.

Second, consider the zero-profit condition, which depends on potential customers' incomes through demand levels. To summarize demand, define a market-access term

$$M_{q,k}(\tau) \equiv \sum_{k'} \exp(-\tau_{qkk'}/m_q) N_{k'} \Gamma_{k'}(q, \bar{\mathbf{n}}, \bar{\mathbf{c}}, \mathbf{0}),$$

where the costless-trade-equilibrium demand levels $\Gamma_k(q, \bar{\mathbf{n}}, \bar{\mathbf{c}}, \mathbf{0})$ were found in section 3.4.1. When trade costs are small, the profits from producing a variety of quality q in location k are approximately

$$\pi_{q,k} \approx \; m_q \exp(-c(q,k)/m_q) M_{q,k}(\tau) - f_q. \tag{5}$$

With small trade costs, profits are not sensitive to the locational decisions of other firms.[26] This means that all varieties of a given quality are produced in a single location, and we can identify production locations using the profits expression.

Within skill intensities, demand determines where varieties are produced. When two qualities have the same skill intensity, $i(q) = i(q')$, the location that minimizes the cost of producing a variety of quality q also minimizes the cost of a variety of quality q'. Thus, if varieties of the same skill intensity are produced in different locations, this must be due to differences in market access, $M_{q,k}(\tau)$. In particular, if $c(q,k) \neq c(q,k')$, then firms produce in the higher-cost location because its market-access advantage outweighs its cost disadvantage.

Proposition 1 (Within-intensity market access)**.** *When trade costs are small, if* $n_{q,k} > 0$*,* $n_{q',k'} > 0$*, and* $i(q) = i(q')$*, then* $M_{q,k} \geq M_{q,k'}$ *or* $M_{q',k} \leq M_{q',k'}$*.*

Proposition 1, proved in appendix A, establishes that market access alone governs specialization within qualities of the same skill intensity. An important component of $M_{q,k}(\tau)$ is

[26] With large trade costs, $\Gamma_{k'}(q, \mathbf{n}, \mathbf{c}, \tau)$ and thus profits depend on competition through firms' locations $\mathbf{n}$.

demand in the location of production, $N_k\Gamma_k(q, \bar{\mathbf{n}}, \bar{\mathbf{c}}, \mathbf{0})$.[27] This is the home-market effect explanation for why high-income locations specialize in high-quality products.

This yields an empirical strategy for distinguishing the two mechanisms. Both the factor-abundance mechanism and the home-market effect cause high-k locations to specialize in high-q varieties, so variation across skill intensities is overdetermined. Variation within skill intensities is driven by market access alone. We can therefore identify a lower bound on the home-market effect by examining the pattern of specialization conditional on skill intensities.[28] My empirical strategy is to relate the pattern of specialization across locations to variation in market access after controlling for plants' factor usage.[29]

3.5 Taking the theory to plant-level data

The theory describes relationships between product quality (q), location (k), skill intensity (i), and market access ($M_{q,k}(\tau)$). These objects can be inferred from observables using the model and some auxiliary assumptions. The following results are derived in appendix A.

I infer product quality from shipments' prices. Assume that $b(\omega, q)$ is strictly decreasing in q, so that higher qualities have higher costs. If $c(q, k)$ increases in quality faster than m_q declines in quality, the price of a variety $p_{jk'} = c(q, k) + \tau_{qkk'} + m_q$ is informative about its quality. I validate this approach in appendix E.3 by calculating demand shifters to infer product quality. Prices and shifters are strongly positively correlated in my data.

I infer locations' rankings from their per capita incomes, denoted $\bar{y}_k$. Under the assumption that $g(y, k)$ is log-supermodular, average income is a sufficient statistic for k.

I infer skill intensities from the composition and wages of plants' workers. The composition measure assumes that non-production workers are more skilled than production workers. The wage measure assumes that wages are increasing in skill. When factor prices equalize, ranking plants by their share of non-production workers or their average wage is equivalent to ranking them by their factor intensities. When labor is cheaper where it is abundant, plants of all intensities use more skilled workers in skill-abundant locations, so I include the measure and its interaction with $\ln \bar{y}_k$ to control for skill intensity.

My empirical counterpart to the model's market-access term $M_{q,k}(\tau)$ is the average of

[27] When trade costs are uniform, as in FGH, differences in $M_{q,k}(\tau)$ are due solely to differences in demand in the location of production, $M_{q,k}(\tau) > M_{q,k'}(\tau) \iff N_k\Gamma_k(q, \bar{\mathbf{n}}, \bar{\mathbf{c}}, \mathbf{0}) > N_{k'}\Gamma_{k'}(q, \bar{\mathbf{n}}, \bar{\mathbf{c}}, \mathbf{0})$.

[28] The strategy of using variation in demand within a set of goods of the same factor intensity is similar to the approach used by Davis and Weinstein (2003) to integrate factor-abundance and home-market-effect models. We differ when we go to the data. Whereas Davis and Weinstein (2003) assume that factor intensities are fixed within 3-digit ISIC industries, I use plant-level information to infer factor intensities.

[29] Proposition 1 implies cross-sectional comparisons of market access in levels. It does not, for example, imply that I could exploit changes in market access over time, since $n_{q,k}$ is not a continuous function of $M_{q,k}$. If $\pi_{q,k} > \pi_{q,k'}$ both periods, q is not produced in k' both periods, regardless of the change in $\pi_{q,k} - \pi_{q,k'}$.

potential customers' per capita incomes, weighted by population size and distance from the location of production. In the model, per capita income is a sufficient statistic for relative demand for qualities when trade costs are low.[30] Weighting these incomes by population size and distance reflects the fact that it is more profitable to produce in locations that are more proximate to a larger number of consumers due to distance-related trade costs.

I construct two such market-access measures. Denote log income per capita in destination city d in year t by $\ln \bar{y}_{dt}$, population size by N_{dt}, and the mileage distance from origin o by $miles_{od}$. The first measure describes the composition of potential customers not residing in the origin location $M^1_{ot} = \sum_{d \neq o} \frac{N_{dt} miles_{od}^{-\eta}}{\sum_{d' \neq o} N_{d't} miles_{od'}^{-\eta}} \ln \bar{y}_{dt}$. The second market-access measure includes all potential customers, consistent with the model, $M^2_{ot} = \sum_{d} \frac{N_{dt} miles_{od}^{-\eta}}{\sum_{d'} N_{d't} miles_{od'}^{-\eta}} \ln \bar{y}_{dt}$.[31] In constructing each measure, I use a distance elasticity of unity, $\eta = 1$. This is based on the empirical relationship between transaction volumes and distance, which I estimate using a gravity model in appendix C. These gravity estimates for trade between metropolitan areas are consistent with the findings of the vast gravity literature on trade between nations.

I now turn to the data to characterize the empirical relationships linking product qualities, skill intensities, and market access following the model's guidance.

4 Data and empirical setting

This section introduces the empirical setting in which I conduct my investigation. First, I describe the data that I use to characterize the pattern of specialization and exchange between US cities. Additional details are in appendix B. Second, I document that both outgoing shipments and incoming shipments within fine product categories exhibit higher prices in higher-income cities. Thus, this empirical setting is suitable for testing theories of quality specialization.

4.1 Data

I combine microdata on US manufacturing plants' production and shipments with data describing cities and sectors. The two confidential microdata sources used are the 1997,

[30] This exploits spatial variation in income levels to capture demand $\Gamma_{k'}(q, \bar{\mathbf{n}}, \bar{\mathbf{c}}, \mathbf{0})$. With high trade costs, this may be a poor approximation due to $\Gamma_{k'}(q, \mathbf{n}, \mathbf{c}, \tau)$ also depending on firms' equilibrium locations. Trade costs between US cities are lower than those between countries due to the common currency, language, and policy environment. Glaeser and Kohlhase (2004) estimate that incurred transport costs are less than 4% of shipment value for most manufactures. My empirical results suggest that approximation (5) is not underpowered in practice.

[31] M^2_{ot} can be written as a weighted average of M^1_{ot} and $\ln \bar{y}_{ot}$. A city's own income per capita has greater weight in this average when it is more populous than and more distant from other metropolitan areas.

2002, and 2007 Commodity Flow Survey (CFS) and Census of Manufactures (CMF).The CFS describes commodity shipments by a sample of business establishments in terms of their value, weight, destination ZIP code, and transportation mode.[32] Products are described using the Standard Classification of Transport Goods, a distinct scheme that at its highest level of detail defines 512 5-digit product categories. Each quarter of the survey year, plants report a randomly selected sample of 20-40 of their shipments in one week.The CMF describes a plant's location, industry, employees, payroll, material inputs, and revenues. It covers the universe of manufacturing plants, which are classified into 473 6-digit NAICS manufacturing industries.[33]

In most of the analysis, I define a product as the pairing of a 5-digit SCTG commodity code and a 6-digit NAICS industry code. This results in more narrowly defined products when the NAICS industry scheme is more detailed than the SCTG commodity scheme. For example, footwear (SCTG 30400) produced by an establishment in "men's footwear (except athletic) manufacturing" (NAICS 316213) is distinct from footwear produced by an establishment in "women's footwear (except athletic) manufacturing" (NAICS 316214). There are more than 5,000 commodity-industry-year triplets in my estimation sample.[34]

The empirical analysis describes core-based statistical areas (CBSAs), which are 366 metropolitan and 576 micropolitan statistical areas defined by the Office of Management and Budget. I refer to these geographic units as cities. Appendix B describes how data using other geographies were assigned to CBSAs.

I calculate cities' per capita incomes using data on CBSAs' total populations and personal incomes from the Bureau of Economic Analysis's regional economic profiles for 1997, 2002, and 2007. In my baseline specification, I exclude the employees and income of all establishments in the same 6-digit NAICS industry as the shipping plant when calculating the population and per capita income of its CBSA.[35] Since most manufacturing sectors' workforces and payrolls are small relative to the total populations and incomes of the cities in which they are located, the results obtained without making this adjustment to the per capita income and population measures are very similar.

[32]The Census Bureau defines an establishment as "a single physical location where business transactions take place or services are performed."The CFS covers manufacturing, mining, wholesale, and select retail and services establishments. I analyze shipments by manufacturing establishments, which I call "plants."

[33]Information on small establishments is estimated from administrative records, not reported by the establishment. I exclude these administrative records and imputed observations. See appendix B for details.

[34]My results are robust to defining products using only the SCTG commodity codes.

[35]There is therefore variation across industries within a CBSA in the regressors I call "log origin CBSA population" and "log origin CBSA per capita income."

4.2 Pattern of specialization and trade

This section describes variation in manufacturing shipment prices across US cities.[36] The patterns mirror those found in international trade data. First, outgoing shipments exhibit higher prices in higher-income cities. This pattern is consistent with quality specialization in which higher-income cities produce higher-price, higher-quality varieties.[37] Second, incoming shipments exhibit higher prices in higher-income cities. This pattern is consistent with non-homothetic preferences in which higher-income consumers demand higher-price, higher-quality varieties.

One concern with inferring qualities from prices is that products may be horizontally differentiated, as in the model. With horizontal differentiation, two varieties of the same quality can sell at different prices in the same destination, with the high-price variety simply obtaining a smaller market share (Khandelwal, 2010). This raises the concern that high-income locations' specialization in high-price products may only reflect higher costs. However, this objection is unlikely to be problematic for the empirical investigation here.

Unit values are likely to be informative about product quality in this context for four reasons. First, investigations of international trade data distinguishing between raw unit values and quality-adjusted prices have shown unit values to be a meaningful, though imperfect, proxy for quality (Khandelwal, 2010; Feenstra and Romalis, 2014). I obtain similar results in section E.3, where I find that demand shifters are positively correlated with unit values. Moreover, these demand shifters exhibit patterns of specialization and factor usage consistent with those found for unit values. Second, my empirical setting allows me to check whether differences in prices across locations only reflect higher costs. Using plant-level data on wages and workers, I can test whether plants shipping from high-income locations charge higher prices only because they have higher labor costs. They don't. Third, consistent with the international evidence presented by Hallak (2006), I find a positive relationship between shipment prices and destinations' per capita income, suggesting that higher-price products are those preferred by higher-income consumers.[38] Fourth, in barcode-level retail data, Hottman, Redding, and Weinstein (2016) and Faber and Fally (2016) find that higher-price products have higher sales or appeal to consumers.

The first feature of the US data matching international findings is that shipments origi-

[36] Recall that all observed "prices" are in fact unit values, the ratio of a shipment's value to its weight in pounds. See data appendix B for details of the sample selection and variable construction.

[37] I follow Schott (2004), who characterized specialization across products using quantities and specialization within products using average prices. The latter is necessary because quantities are reported by product, so by definition we do not observe the quantities of different qualities within narrowly defined products.

[38] A potential concern is that higher-income consumers pay higher prices for identical products because higher-income consumers are less responsive to price changes (Simonovska, 2015). This would be a concern if the observed price variation were primarily within-plant. Table 2 below shows that this is not the case.

nating from higher-income cities exhibit higher prices. To characterize how shipment prices vary with origin characteristics, I estimate linear regressions describing a shipment s of product k by plant j from origin city o to destination city d by transport mode m in year t of the form

$$\ln p_{skjodmt} = \beta \ln \bar{y}_{okt} + \alpha_1 \ln N_{okt} + \alpha_2 \ln miles_{skjodmt} + \gamma_{mt} + \gamma_{kdt} + \epsilon_{skjodt}, \quad (6)$$

where $p_{skjodmt}$ is the shipment's unit value, $\bar{y}_{okt}$ and N_{okt} are per capita income and total population in the origin CBSA excluding the industry of the shipping plant, $miles_{skjodmt}$ is the ZIP-to-ZIP mode-specific mileage distance of the shipment, γ_{mt} are mode-year fixed effects, and γ_{kdt} are product-destination-year fixed effects. Including both per capita income and total population allows me to distinguish between income composition and scale, since these are positively correlated across cities. The mileage and mode covariates allow prices to vary with transport costs.[39] The product-destination-year fixed effects mean that I am comparing prices of the same product shipped to the same metropolitan area in the same year, such as shipments of beer by breweries to Chicago in 1997. This is akin to the comparison of US import prices with product-year fixed effects in Schott (2004).[40]

Table 1 characterizes how variation in outgoing shipments' unit values relates to origin characteristics. The first column reports a large, positive origin-income elasticity of shipment prices of 44%. Higher-income cities specialize in the production of higher-price varieties of products, and this pattern is highly statistically significant. Conditional on the level of per capita income, there is no economically meaningful correlation between origin population size and outgoing shipments' prices.

I proceed to interact the regressors with two measures of the scope for product differentiation. The Sutton (1998) measure, industrial R&D and advertising intensity, proxies the scope for quality differentiation by the cost shares of differentiation-related activities. The Khandelwal (2010) measure infers the scope for quality differentiation from the range of estimated demand shifters in US imports. The second and third columns of Table 1 show that the positive relationship between origin income per capita and outgoing shipment prices is stronger in products with greater scope for quality differentiation, as classified by both measures. These patterns are consistent with higher-income cities specializing in higher-quality products. In products with greater scope for quality differentiation, income differences correspond to greater differences in output prices.

[39] In the model, the constant-markup assumption makes plants' free-on-board prices invariant to transport costs. Including these covariates relaxes that assumption. The main results of this paper are robust to omitting the population and mileage covariates and mode fixed effects.

[40] Note that my unit of observation is a shipment. The paper's main results are robust to aggregating to the origin-destination-industry-product-year level.

Table 1: Outgoing shipment prices

Dep var: Log unit value, $\ln p_{skjodmt}$	(1)	(2)	(3)
Origin CBSA log per capita income, $\ln \bar{y}_{okt}$	0.440**	0.411**	0.430**
	(0.0359)	(0.0491)	(0.0496)
Origin CBSA log population, $\ln N_{okt}$	-0.00764	0.000175	-0.000889
	(0.00421)	(0.00540)	(0.00547)
Log mileage, $\ln miles_{skjodmt}$	0.0404**	0.0538**	0.0528**
	(0.00280)	(0.00357)	(0.00365)
Per capita income (log) × *differentiation*		0.115*	0.191**
		(0.0522)	(0.0733)
Population (log) × *differentiation*		-0.000146	-0.0205*
		(0.00533)	(0.00804)
Mileage (log) × *differentiation*		-0.00858*	-0.00870
		(0.00376)	(0.00539)
Differentiation measure		Sutton	Khandelwal
Within R^2	0.080	0.094	0.094
Number estab-year (rounded)	35,000	20,000	20,000
Number ind-prod-year (rounded)	5250	3000	3000
Observations (rounded)	1,800,000	900,000	900,000

Notes: Manufacturing establishments in the CFS and CMF. All shipments are to a domestic destination CBSA distinct from the origin CBSA. All regressions include SCTG5 × NAICS6 × destination × year fixed effects and mode × year fixed effects. Columns 2 and 3 are estimated on a sample of observations for which industry R&D and advertising intensity, Khandelwal (2010) ladder length, and Rauch (1999) differentiation measures are available. Standard errors, clustered by origin CBSA × year, in parentheses. ** and * denote statistical significance at 1% and 5%, respectively.

The second feature of the US data matching international findings is that shipments destined for higher-income cities exhibit higher prices. To characterize how shipment prices vary with destination characteristics, I estimate linear regressions of the form

$$\ln p_{skjodmt} = \beta \ln \bar{y}_{dt} + \alpha_1 \ln N_{dt} + \alpha_2 \ln miles_{skjodmt} + \gamma_{kt} + \gamma_{mt} + \theta_{ot} + \theta_{kjt} + \epsilon_{skjodt},$$

where $p_{skjodmt}$ is the shipment's unit value, $miles_{skjodmt}$ is the ZIP-to-ZIP mileage distance of the shipment, and $\bar{y}_{dt}$ and N_{dt} are per capita income and total population in the destination CBSA. γ_{kt} and γ_{mt} are product-year and mode-year fixed effects that are included in all specifications. The θ fixed effects, which are mutually exclusive and omitted from some specifications, are origin-year and product-plant-year fixed effects.

Table 2: Incoming shipment prices

Dep var: Log unit value, $\ln p_{skjodmt}$	(1)	(2)	(3)
Destination CBSA log per capita income, $\ln \bar{y}_{dt}$	0.247**	0.159**	0.0509**
	(0.0213)	(0.0168)	(0.00586)
Destination CBSA log population, $\ln N_{dt}$	-0.00448*	-0.00315	0.000142
	(0.00225)	(0.00172)	(0.000701)
Log mileage, $\ln miles_{skjodmt}$	0.0449**	0.0467**	0.0148**
	(0.00308)	(0.00189)	(0.000875)
Commodity × Industry × Year FE	Yes	Yes	
Origin CBSA × Year FE		Yes	
Establishment × Commodity × Year FE			Yes
Within R^2	0.097	0.145	0.029
Number estab-year (rounded)		35,000	
Number ind-prod-year (rounded)		5250	
Observations (rounded)		1,800,000	

Notes: Manufacturing establishments in the CFS and CMF. All shipments are to a domestic destination CBSA distinct from the origin CBSA. All regressions include mode × year fixed effects. Standard errors, clustered by destination CBSA × year, in parentheses. ** and * denote statistical significance at 1% and 5%, respectively.

Table 2 reports regressions characterizing how variation in shipment unit values within products relates to destination characteristics. The first column shows that the per-capita-income elasticity of incoming shipment prices is 25%. Higher-income cities import higher-price varieties, which suggests that preferences are non-homothetic. This pattern is attributable to city income composition, not city size per se, as the coefficient on log population reveals. The distance elasticity of incoming shipment prices is about 4%; longer shipments exhibit higher prices.[41]

The second and third columns show that the large majority of the correlation between income per capita and incoming shipment prices is attributable to cities of different income levels purchasing goods from different cities and plants. The second column introduces fixed effects for cities of origin, θ_{ot}. The destination per capita income elasticity falls by about 10 percentage points, indicating that about 40% of this variation is attributable to

[41] There are at least three possible explanations for the positive coefficient on shipment distance. First, distance-related costs may be included in the reported shipment values. Second, the composition of plants shipping to a destination may vary with distance. Third, plants may charge higher mark-ups when serving more distant destinations. The third column of Table 2 suggests that this last channel could explain at most one-third of such variation, since the within-establishment mileage elasticity is 1.5%.

the composition of cities trading with each other.[42] The coefficients on the other regressors are similar to those in the first column. The third column introduces fixed effects for each plant-product, θ_{kjt}. The within-plant destination-income elasticity of shipment prices is considerably lower, 5.1%. Selling the same product at a higher price therefore accounts for at most one-fifth of price variation across destinations of different income levels. This decomposition suggests that changes in markups are not responsible for the majority of the observed correlation between shipment prices and destination incomes.

These findings demonstrate that the composition of cities' manufactures demand is strongly linked to their income levels. This is consistent with numerous previous empirical studies of both households and countries. Such non-homothetic preferences are necessary for the "home-market effect for quality" hypothesis.[43]

Together, Tables 1 and 2 demonstrate patterns of specialization that are strongly linked to cities' income levels. Within narrowly defined product categories, higher-income locations both export and import higher-price products than lower-income locations. In addition, Appendix C shows that cities with more similar incomes trade more intensely with each other. These findings mirror those found in international trade data and could be generated by the factor-abundance mechanism or the home-market effect. I now use data on plants' factor inputs to empirically distinguish between these potential explanations.

5 Empirical results

This section reports two main bodies of empirical evidence. First, observed factor-usage differences explain about one quarter of the relationship between cities' incomes and the prices of outgoing shipments. This bounds the explanatory power of the factor-abundance mechanism in terms of observed factor inputs. Second, the market-access measures describing the income composition of proximate potential customers are strongly linked to outgoing shipment prices. The estimated home-market effect explains close to half of the observed price-income relationship.

My regression results can be seen as decomposing the covariance between plant j's outgoing shipment price and origin o's per capita income. The regressions in Table 1 demonstrate a positive relationship between prices and incomes but omit both measures of factors employed

[42] Appendix C shows that cities with more similar incomes trade more intensely with each other.

[43] Non-homothetic preferences alone are not sufficient to produce the home-market effect, as discussed in section 2 and shown in section 3.4.1. The home-market effect for quality stems from non-homothetic preferences, economies of scale, and trade costs.

and market access. This "short"specification in equation (6) can be restated as

$$\ln p_{skjodmt} = \beta^S \ln \bar{y}_{okt} + \alpha^S \cdot X_{skjodmt} + \epsilon^S_{skjodmt},$$

where the superscript S denotes the "short" regression and the vector $X_{skjodmt}$ contains origin population, shipment mileage, and destination-product-year fixed effects. The prior empirical literature has estimated $\hat{\beta}^S > 0$, and this result has been interpreted as attributable to differences in factor supplies or demand conditions. In the factor-supplies account, the factor inputs employed in production are omitted variables that explain output prices and are correlated with origin income per capita. In the market-access account, origin income has a causal effect on output prices through its effect on the composition of demand.

My empirical strategy is to introduce observable measures of both factor inputs and market access, described in section 3.5, in order to identify their contribution to the positive coefficient on origin income per capita. The factor-input measures address their omission from the short regression. The market-access measures employ spatial variation in income levels to capture spatial variation in demand, exploiting neighboring cities' contributions to $M_{q,k}(\tau)$.[44] Consider a "long" regression that incorporates factor-employment measures F_{jt} and market-access measure M_{ot},

$$\ln p_{skjodmt} = \beta \ln \bar{y}_{okt} + \alpha \cdot X_{skjodmt} + \delta \cdot F_{jt} + \lambda M_{ot} + \epsilon_{skjodmt}. \tag{7}$$

If this long regression is the correct specification, then the estimated $\hat{\beta}^S > 0$ captures how outgoing shipment prices covary with income per capita both directly (β) and indirectly through factors of production and market access. This can be seen using the omitted variables bias formula:

$$\begin{bmatrix} \beta^S \\ \alpha^S \end{bmatrix} = \begin{bmatrix} \beta \\ \alpha \end{bmatrix} + \mathbb{E}\left[\begin{bmatrix} \ln \bar{y} \\ X \end{bmatrix} \begin{bmatrix} \ln \bar{y} & X \end{bmatrix}\right]^{-1} \mathbb{E}\left[\begin{bmatrix} \ln \bar{y} \\ X \end{bmatrix} \begin{bmatrix} F & M \end{bmatrix}\right] \begin{bmatrix} \delta \\ \lambda \end{bmatrix}$$

$\beta^S > \beta$ to the extent that (1) skill-intensive products have higher prices ($\delta > 0$) and skill-intensive products are produced in higher-income locations ($cov(\ln \bar{y}, F|X) > 0$) and (2) higher-price products are produced where proximate potential customers' per capita incomes are higher ($\lambda > 0$) and higher-income locations are more proximate to higher-income potential customers ($cov(\ln \bar{y}, M|X) > 0$). Each of these channels is a potential explanation because the economic mechanism ($\delta > 0$ or $\lambda > 0$) is plausible, as shown by

[44] While Fajgelbaum, Grossman, and Helpman (2011) derive their theoretical results from a location's own contribution to its market access, I exploit geographic variation in nearby cities' income levels to address concerns about unobservables in the city of production relating to its income level.

the model, and higher-income US metropolitan areas are both populated by more skilled manufacturing workers and more proximate to high-income consumers.[45]

In the model, these mechanisms are the only forces for specialization across income levels, so β should be zero. In practice, prices and and incomes may be correlated after accounting for factor inputs and market access due to additional omitted regressors or causal effects of local income. The former might include unobserved differences in production conditions, like entrepreneurial zeal or exogenous technological advantages, that could plausibly both explain output prices and correlate with local income. The latter would be mechanisms through which higher incomes cause locals to supply higher-quality varieties. Measurement error in the observables F_{jt} and M_{ot} may also leave residual covariance captured by β.

I decompose β^S by first introducing factor inputs F and then introducing market access M. Introducing measures of plants' employment of factors of production decomposes the covariance between outgoing shipment prices and origin per capita income into between-intensity variation and within-intensity variation.[46] While the factor-abundance mechanism operates exclusively through the former, the home-market effect may manifest in both, since higher-income locations have greater demand for higher-quality varieties, regardless of qualities' skill intensities. Section 5.1 shows that the measured across-skill-intensities component is modest, constituting 27% of the covariance between outgoing shipment prices and income per capita.

Introducing a measure of market access yields the long regression specified in equation (7). Following Proposition 1, the coefficient λ describes how outgoing shipment prices covary with potential customers' income levels, conditional on differences in factors employed in production. The coefficient β captures the residual covariance between prices and incomes attributable to neither differences in skill intensity nor the market-access measure. In section 5.2, I find that within-skill-intensity variation in market access accounts for about 58% of the

[45]In terms of skills, higher-income cities' manufacturing plants exhibit higher non-production worker shares and their manufacturing employees exhibit more years of schooling and higher wages. In terms of market access, cities' income levels are spatially correlated, such that regressing M^1_{ot} on log per capita income yields a positive relationship with an R^2 of about 20%. Note that these patterns only imply $\beta < \beta^S$ if the associated mechanism also accounts for price variation conditional on per capita income, that is, if $\delta > 0$ or $\lambda > 0$.

[46]Using the law of total covariance and omitting notation indicating that moments are conditional on X, the covariance between plant j's outgoing shipment price and origin o's per capita income can be written as

$$cov(\ln p_j, \ln \bar{y}_o) = \underbrace{cov\left[\mathbb{E}(\ln p_j|F_j), \mathbb{E}(\ln \bar{y}_o|F_j)\right]}_{\text{between-intensity variation}} + \underbrace{\mathbb{E}\left[cov(\ln p_j, \ln \bar{y}_o|F_j)\right]}_{\text{within-intensity variation}}$$

My regressions are linear projections that approximate these conditional moments.

observed price-income relationship.[47] This makes $\hat{\beta}$ small relative to $\hat{\beta}^S$ and, in a number of specifications, statistically indistinguishable from zero.

After estimating this decomposition, I provide additional empirical evidence to support these results. Section 5.3 summarizes two pieces of further evidence, which are described at length in appendix E, that support an economically large role for home-market demand in determining the pattern of quality specialization. First, the second moment of the local household income distribution is linked to outgoing shipment prices. Second, demand shifters exhibit the same patterns as outgoing shipment prices. Section 5.4 reports a series of robustness checks that yield results consistent with the claim that market access explains as much of the covariance of shipment prices and income levels as observed factor usage.

5.1 The factor-abundance hypothesis

This section identifies the share of within-product specialization attributable to differences in observable plant-level factor usage. The canonical factor-abundance theory posits that differences in locations' outputs are explained by differences in the factors employed by their producers. Within groups of products of the same factor intensity, the location of production is indeterminate. That is, under the null hypothesis that differences in factor supplies are the only source of comparative advantage, there should be no correlation between locational characteristics and plants' outputs after controlling for plant-level factor usage. In fact, there is a very strong relationship between income per capita and outgoing shipments prices after controlling for factor inputs. Observed factor usage explains only 27% of the observed covariance between cities' per capita incomes and outgoing shipment prices.

To characterize how shipment prices vary with origin characteristics, I estimate linear regressions describing a shipment s of product k by plant j from origin city o to destination city d by transport mode m in year t of the form

$$\begin{aligned}\ln p_{skjodmt} = {} & \beta \ln \bar{y}_{okt} + \alpha \cdot X_{skjodmt} + \delta_1 \ln share_{Njt} + \delta_2 \ln \frac{K_{jt}}{L_{jt}} + \delta_3 \ln \bar{w}_{jt} \\ & + \delta_4 \ln share_{Njt} \ln \bar{y}_{ot} + \delta_5 \ln \frac{K_{jt}}{L_{jt}} \ln \bar{y}_{ot} + \delta_6 \ln \bar{w}_{jt} \ln \bar{y}_{ot} + \epsilon_{skjodt} \qquad (8)\end{aligned}$$

[47]Following footnote 46, we can write the within-intensity variation as $\mathbb{E}\left[cov(\ln p_j, \ln \bar{y}_o | F_j)\right] =$

$$\underbrace{\mathbb{E}_F\left[cov_M(\mathbb{E}(\ln p_{jo}|F_j, M_o), \mathbb{E}(\ln \bar{y}_o|F_j, M_o))\right]}_{\text{within-skill-intensity market access}} + \underbrace{\mathbb{E}_F\left[\mathbb{E}_M(cov(\ln p_j, \ln \bar{y}_o|F_j, M_o))\right]}_{\text{residual covariance}}$$

Using regressions to approximate these conditional moments, the market-access component accounts for about 58% of the overall $cov(\ln p_j, \ln \bar{y}_o)$.

where $share_{Njt}$ is the ratio of the plant's non-production workers to total employees, $\frac{K_{jt}}{L_{jt}}$ is gross fixed assets per worker, and $\bar{w}_{jt}$ is average pay per employee.[48] The interactions of plant-level factor-usage measures with origin income per capita address the case in which factor prices do not equalize, as described in section 3.5.[49] In theory, either $\ln share_{Njt}$ and its interaction with $\ln \bar{y}_{ot}$ or $\ln \bar{w}_{jt}$ and its interaction with $\ln \bar{y}_{ot}$ would be sufficient to characterize plants' skill intensities. In practice, I include both and the capital measure to maximize the potential explanatory power of observed factor-usage differences.

Table 3 characterizes how variation in shipment unit values relates to origin characteristics and plant-level observables. The first column relates outgoing shipment unit values to origin characteristics controlling for destination fixed effects, as in Table 1. The next two columns incorporate the plant-level measures of factor usage and their interactions with income per capita. The second column introduces quantity measures of capital intensity ($\frac{K_{jt}}{L_{jt}}$) and labor usage, the non-production employment share ($share_{Njt}$). The third column adds the average wage measure and therefore corresponds to the regression specified in equation (8).

These measures of factor usage are informative predictors of a plant's shipment prices, but they explain only a modest share of the observed origin-income elasticity of outgoing shipment prices. Consistent with the premise that higher-price, higher-quality varieties are more skill-intensive, the coefficients on log non-production worker share and log pay per worker are positive and economically large. The negative coefficient on log assets per worker is inconsistent with a model in which higher-price, higher-quality varieties are more capital-intensive.[50] The observed variation in factor usage helps explain some of the total variation in outgoing shipment prices, but only a small share of the income-linked variation. While introducing the quantity measures in the second column increases the R^2, it only reduces the origin-income elasticity from 44% to 40%. Incorporating the wage measure in the third column reduces this elasticity to 35%. Thus, the observed factor-usage differences can explain about one-fifth of the origin-income elasticity of shipment prices in this specification.

The fourth through sixth columns of Table 3 incorporate the control variables in more flexible functional forms. The mileage, non-production worker share, assets per worker, and pay per worker covariates now enter as cubic polynomials that vary by 3-digit NAICS

[48]The theoretical model emphasized differences in the composition of skill across locations. I also include gross fixed assets per worker as a measure of capital intensity, since this variable has been emphasized in prior empirical work both across countries (Schott, 2004) and across plants (Verhoogen, 2008). Since I cannot construct capital stocks using the perpetual-inventory method with quinquennial data, I use the book value of assets as my measure of plant capital.

[49]Bernard, Redding, and Schott (2013) infer that relative factor prices do not equalize within the US when considering two factors, production and non-production workers.

[50]Using very aggregate data, Torstensson (1996) obtains a negative partial correlation between prices and capital per worker when distinguishing between human and physical capital.

Table 3: Shipment prices and factor usage

Dep var: Log unit value, $\ln p_{skjodmt}$	(1)	(2)	(3)	(4)	(5)	(6)
Origin CBSA log per capita income	0.440**	0.401**	0.353**	0.413**	0.362**	0.311**
	(0.0359)	(0.0350)	(0.0345)	(0.0345)	(0.0334)	(0.0330)
Origin CBSA log population	-0.00764	-0.0111**	-0.0136**	-0.00652	-0.0104**	-0.0133**
	(0.00421)	(0.00408)	(0.00398)	(0.00393)	(0.00379)	(0.00365)
Log mileage (ZIP-ZIP-mode-specific)	0.0404**	0.0418**	0.0413**	$\surd$	$\surd$	$\surd$
	(0.00280)	(0.00275)	(0.00269)			
Non-production worker share (log)		0.135**	0.113**		$\surd$	$\surd$
		(0.00692)	(0.00695)			
Assets per worker (log)		-0.0369**	-0.0540**		$\surd$	$\surd$
		(0.00391)	(0.00404)			
Pay per worker (log)			0.217**			$\surd$
			(0.0180)			
Non-production worker share		0.0571	0.00606		-0.000645	-0.0387
× income per capita		(0.0315)	(0.0327)		(0.0298)	(0.0307)
Assets per worker		-0.0320*	-0.0619**		-0.0106	-0.0253
× income per capita		(0.0155)	(0.0161)		(0.0165)	(0.0168)
Pay per worker			0.336**			0.198**
× income per capita			(0.0686)			(0.0699)
Within R^2	0.080	0.088	0.091	0.084	0.100	0.106
Number estab-year (rounded)			35,000			
Number ind-prod-year (rounded)			5250			
Observations (rounded)			1,800,000			

Notes: Manufacturing establishments in the CFS and CMF. All shipments are to a domestic destination CBSA distinct from the origin CBSA. All regressions include SCTG5 × NAICS6 × destination × year fixed effects and mode × year fixed effects. The fourth through sixth columns include 3-digit-NAICS-specific cubic polynomials in log mileage (4,5,6), log non-production worker share (5,6), log assets per worker (5,6), and log pay per worker (6). Standard errors, clustered by origin CBSA × year, in parentheses. ** and * denote statistical significance at 1% and 5%, respectively.

industry. Since there are 21 3-digit industries, this introduces 63 regressors for each control variable, yielding a total of 252 regressors.[51] I refrain from reporting the coefficients on these controls and indicate their inclusion by $\surd$ in the relevant rows of tables.

The results obtained using these more flexible functional forms are similar to those in the first three columns of Table 3. The origin-income elasticity of 41% is reduced to 36% by the introduction of the quantity controls and further to 31% by the full battery of plant-level factor-usage measures. Thus, differences in plants' observed factor usage explain about one-quarter of the correlation between cities' incomes per capita and outgoing shipment prices. This suggests that the factor-abundance hypothesis has meaningful but modest explanatory power for the pattern of within-product specialization across US cities.

[51] Using a 3-digit-NAICS-specific translog approximation with the input measures and a 3-digit-NAICS-specific quadratic in log mileage yields very similar results.

These factor-input covariates may be imperfect measures of plant-level factor usage, leaving residual variation in shipment prices that is correlated with cities' incomes. While I cannot directly rule out measurement error, these plant-level input measures are far more precise than the country-level covariates typically used to evaluate factor-abundance theories of comparative advantage, which aggregate over plants, cities, and industries. Moreover, the observed factor inputs are informative about outgoing shipment prices. There is a considerable increase in the R^2 between the fourth and sixth columns in Table 3. However, these factor inputs do not vary across cities in a way that accounts for the link between city-level income and shipment prices.

Could city income per capita be informative about plant-level factor usage conditional on the plant-level covariates? Plants with observationally equivalent workforces in terms of non-production-to-production-worker ratios may exhibit unobserved differences in worker quality. In particular, prior research has documented weak but systematic sorting of workers across cities on unobservable characteristics correlated with higher wages (Davis and Dingel, 2012; De la Roca and Puga, 2013). However, these differences between workers should appear in the plant-level wage measures included in the third and sixth columns of Table 3. The posited unobserved differences in input factor quality would therefore have to be characteristics of workers that raise output quality, are not priced into their wages, and are systematically correlated with city-level incomes, which seems an unlikely explanation for the findings.

These results are robust to introducing further information on the skills employed in these plants. I construct city-industry-level measures of employees' schooling from public-use microdata from the Census of Population and American Community Survey. These measures are available for a subset of the observations in the main estimation sample. The results are reported in Appendix Table D.1. The partial-correlation origin-income elasticity of 30.5% is quite similar to the 31% obtained in Table 3.

Another potential concern is aggregation bias. Though my data describe hundreds of manufacturing product categories, these are less detailed than the most disaggregated product categories in international trade data. I address this concern using data from the Census of Manufactures product trailer, which describes comparable number of product categories and reports quantities for a subset of them. Appendix Table D.2 describes establishments' average unit values from Census of Manufactures data on products for which quantities are reported and reports results that are consistent with those reported in Table 3.[52] Though the origin-income elasticity is lower than that found in the CFS data, observed plant-level factor usage explains only a small fraction of the total variation.

[52]These plant-level average unit values necessarily include shipments destined for the origin CBSA.

This section has shown that a modest share of the observed within-product variation in outgoing shipment prices across cities of different income levels is attributable to observable differences in plants' factor usage. Under the null hypothesis that differences in factor abundance alone explain within-product specialization, the partial correlation between origin income per capita and outgoing shipment prices conditional on plant-level factor usage would be zero. In the presence of a rich set of plant-level controls, the estimated coefficient $\hat{\beta}_1$ in column six of Table 3 is 31%, roughly three quarters of its value in the absence of plant-level controls. If we were to attribute the full decrease in the value of the coefficient on $\ln \bar{y}_{ot}$ to the factor-abundance mechanism, it would explain about one quarter of the observed variation.[53]

5.2 The market-access hypothesis

This section identifies the share of the covariance between incomes and prices not explained by factor-usage differences that is attributable to home-market demand. I find that cities with greater market access to higher-income households produce higher-price manufactures. This within-intensity market-access variation explains more of the covariance between incomes and prices than differences in plants' factor inputs.

The "home-market" effect in fact depends on the composition of demand in all locations potentially served from a location of production, as described in the model by market access $M_{q,k}(\tau)$. A city that is more proximate to another city with many high-income residents has higher relative demand for higher-quality manufactures, ceteris paribus.[54] Section 3.5 described two market-access measures. The first, $M^1_{ot} = \sum_{d \neq o} \frac{N_{dt} miles_{od}^{-\eta}}{\sum_{d' \neq o} N_{d't} miles_{od'}^{-\eta}} \ln \bar{y}_{dt}$, omits potential customers residing in the location of production. The identifying assumption when using this measure is that variation across locations in neighboring cities' incomes per capita, after conditioning on plants' inputs and income per capita in the city of production, is related to plants' outputs only through variation in the composition of demand. The second market-access measure, $M^2_{ot} = \sum_{d} \frac{N_{dt} miles_{od}^{-\eta}}{\sum_{d'} N_{d't} miles_{od'}^{-\eta}} \ln \bar{y}_{dt}$, includes all potential customers, consistent with the model. The accompanying identifying assumption is that, after conditioning on plants' inputs, variation across locations in potential consumers' incomes, including res-

[53] To the degree that differences in skill intensities are causally induced by differences in demand, this overstates the explanatory power of the factor-abundance hypothesis.

[54] Fajgelbaum, Grossman, and Helpman (2011) assume the cost of exporting to another location is the same across all locations. Thus, in their model the home-market effect depends only on the difference in income composition between the location of production and the rest of the world. When trade costs are not uniform, the home-market effect depends on a production location's access to every other market, as noted by Burenstam Linder (1961, p.87) and Behrens, Lamorgese, Ottaviano, and Tabuchi (2009). Measuring market access has received considerable attention in empirical assessments of the new economic geography (Redding and Venables, 2004). See Lugovskyy and Skiba (2015) for a discussion of market access in the context of quality specialization.

idents in the city of production, is related to plants' outputs only through variation in the composition of demand.[55]

Table 4 demonstrates that market access plays a significant role in explaining the origin-income elasticity of shipment prices. To facilitate comparisons, the first column is identical to the sixth column of Table 3. The second column introduces the first market-access measure, and its coefficient is positive and highly significant. Its inclusion reduces the origin-income elasticity from 31% to 18%. Recall that flexibly controlling for the non-production worker share, assets per worker, and average pay per worker initially reduced the elasticity from 41% to 31%. Thus, the income composition of proximate potential customers other than those in the city of production explains more of the covariance of income per capita and outgoing shipment prices. In locations with better access to high-income customers, plants produce higher-price products. This evidence suggests that the geography of demand influences the pattern of within-product specialization.[56]

The third column uses the second market-access measure, which includes the income of residents in the city of production in the weighted average. This reduces the origin-income elasticity to 9%, a reduction of 22 percentage points compared to the first column. In this specification, market access explains about half of the observed relationship between income per capita and outgoing shipment prices, which is substantially more than that attributable to differences in plants' factor usage.

These results can be succinctly summarized as a decomposition of the covariance between incomes and prices.[57] After controlling for population size and shipment mileage, differences in observed factor usage are responsible for 27% of the covariance between outgoing shipment prices and origin income per capita. Conditional on factor usage, the first market-access measure, which omits residents in the city of production, accounts for 36% of the total covariance, leaving 37% as residual variation. The second market-access measure, which follows the model by including residents in the city of production, accounts for 58% of the total covariance, leaving 15% as residual variation.

The fourth column of Table 4 reports a regression that incorporates the first market-access measure while omitting factor inputs.[58] Because market access is positively correlated with these omitted regressors, the positive coefficient on M^1_{ot} is about 15% larger than in the second column.[59] While within-intensity variation in market access accounts for 36%

[55]This identifying assumption would be violated by unobserved quality-improving inputs or technologies that were correlated with city-level income per capita conditional on my plant-level measures of inputs.

[56]This result reflects income composition, not total income. I have confirmed that the inverse-distance-weighted sum of total incomes, a "market potential" measure, does not explain variation in shipment prices.

[57]Footnotes 46 and 47 report the decomposition that I approximate using linear regressions.

[58]Table D.3 in the online appendix reports further specifications omitting factor inputs.

[59]Similarly, adding classical measurement error to the factor-input covariates raises the estimated market-

Table 4: Shipment prices and market access

Dep var: Log unit value, $\ln p_{skjodmt}$	(1)	(2)	(3)	(4)	(5)
Origin CBSA log per capita income	0.311**	0.179**	0.0929*	0.258**	
	(0.0330)	(0.0357)	(0.0407)	(0.0369)	
Origin CBSA log population	-0.0133**	-0.00298	-0.00847*	0.00510	
	(0.00365)	(0.00376)	(0.00358)	(0.00405)	
Log mileage (ZIP-ZIP-mode-specific)	✓	✓	✓	✓	0.0465**
					(0.00187)
Non-production worker share (log)	✓	✓	✓		
Assets per worker (log)	✓	✓	✓		
Pay per worker (log)	✓	✓	✓		
Market access (excludes origin) M^1_{ot}		1.103**		1.253**	
		(0.112)		(0.115)	
Market access M^2_{ot}			1.015**		
			(0.110)		
Destination CBSA log per capita income					0.165**
					(0.0171)
Destination CBSA log population					-0.00360*
					(0.00173)
Destination market access M^1_{dt}					-0.0535
					(0.0451)
SCTG5 × NAICS6 × Destination × Year FE	Yes	Yes	Yes	Yes	
SCTG5 × NAICS6 × Year FE					Yes
Origin × Year FE					Yes
Within R^2	0.106	0.108	0.108	0.086	0.145
Number estab-year (rounded)			35,000		
Number ind-prod-year (rounded)			5250		
Observations (rounded)			1,800,000		

Notes: Manufacturing establishments in the CFS and CMF. All shipments are to a domestic destination CBSA distinct from the origin CBSA. All regressions include mode × year fixed effects. Standard errors are clustered by origin CBSA × year in columns 1 through 4 and by destination CBSA × year in column 5. Unreported controls in columns 1 through 3 are the interactions of log origin income per capita with the three input variables and 3-digit-NAICS-specific cubic polynomials in log mileage, log non-production worker share, log assets per worker, and log pay per worker. Unreported controls in column 4 are 3-digit-NAICS-specific cubic polynomials in log mileage. Clustered standard errors in parentheses. ** and * denote statistical significance at 1% and 5%, respectively.

of the total covariance, unconditional variation in market access can explain about 47% of the price-income covariance.[60] This result might reflect spurious correlation or an economic relationship. If the correlation between market access and factor input usage is spurious, then the regression in column four is misspecified and overstates the role of market access in explaining the price-income covariance. If differences in market access cause firms to produce varieties of different factor intensities, then market access explains more of the price-income covariance than suggested by the conservative specification in column two that exploits only within-intensity variation in market access.

Market access does not predict incoming shipments' prices. This is shown by the fifth column of Table 4, which introduces the destination city's income level and market access as regressors and uses product-year and origin-year fixed effects rather than destination-product-year fixed effects. Thus, it is the second column of Table 2 augmented by destination market access. While higher-income destinations purchase higher-price incoming shipments, destinations with higher-income neighbors do not exhibit higher incoming prices. The coefficient on M^1_{dt} is a precisely estimated zero.[61] Since better market access predicts higher-price outgoing shipments and unchanged incoming shipments, better market access is associated with net exporting of higher-price varieties.

Appendix section E.1 shows that the patterns found in domestic shipments are also found in export shipments destined for foreign markets. The origin-income elasticity of export prices is 42%. After controlling for plants' factor inputs, this elasticity is 30%. After controlling for both factor inputs and market access, this elasticity becomes negative and statistically indistinguishable from zero.

This section has established the role of market acess in explaining the pattern of outgoing shipment prices. The income composition of proximate potential customers is strongly associated with outgoing shipment prices. Consistent with the model, plants located near higher-income potential customers sell products at higher average prices. The income composition of potential customers other than those in the location of production is quantitatively more important for explaining the origin-income elasticity of outgoing shipment prices than observed plant-level factor usage. When including individuals residing in the city of production, the income composition of potential customers explains about half of the observed origin-income elasticity of shipment prices. Incoming shipment prices are invariant to desti-

access coefficient.

[60]The respective numbers when using M^2_{ot} are 58% and 77%.

[61]This result also demonstrates the point made in footnote 45 that introducing M_{ot} to equation (7) only implies $\beta < \beta^S$ if $\lambda > 0$. The correlation between M^1_{dt} and $\ln \bar{y}_{dt}$ is identical to that between M^1_{ot} and $\ln \bar{y}_{ot}$, yet the coefficient on $\ln \bar{y}_{dt}$ in the fifth column of Table 4 is statistically indistinguishable from its value in the second column of Table 2.

nation market access. This is consistent with a model in which market access plays a large role in quality specialization and makes high-income locations net exporters of high-quality products.

5.3 Further evidence

This section summarizes two further pieces of evidence supporting the inference that home-market demand plays a large role in quality specialization. First, the second moment of the local household income distribution is linked to outgoing shipment prices. Second, I calculate demand shifters and find that they exhibit the same patterns as outgoing shipment prices. These findings are both described in more detail in Appendix E.

Appendix section E.2 uses another moment of the income distribution to identify the role of demand in quality specialization. Conditional on average income, cities with higher dispersion in household income have higher incoming shipment prices. This suggests that second moment of the income distribution is informative about the composition of demand. I then show that cities with greater income dispersion have higher outgoing shipment prices, and this is not due to greater dispersion in the wages or skills of workers employed at the plants shipping these products. This is consistent with the home-market effect under the Fajgelbaum, Grossman, and Helpman (2011) demand system in an equilibrium in which most individuals purchase low-quality varieties.

Appendix section E.3 characterizes the pattern of quality specialization using demand shifters instead of outgoing shipments' unit values as the dependent variable. Due to data constraints, I am only able calculate demand shifters for shipments in 2007. In the absence of exogenous price variation to identify the demand system, I use price-elasticity estimates from Feenstra and Romalis (2014) to calculate demand shifters. The empirical results are consistent with the unit-value findings for the influence of market access, though factor usage exhibits greater explanatory power. The origin-income elasticity of the plant-product demand shifter is 41%. This covariance between income per capita and demand shifter decomposes into factor-intensity differences (46%), within-intensity market-access differences (48%), and residual variation (7%). The greater explanatory power of plants' factor inputs primarily reflects less residual variation, not a dramatically weakened role for the income composition of proximate potential customers. Home-market demand plays a substantial role in quality specialization, as large as that explained by the factor-abundance mechanism.

Table 5: Shipment prices and industrial rents

Dep var: Log unit value, $\ln p_{skjodmt}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Origin CBSA log per capita income	0.446**	0.332**	0.231**	0.322**	0.273**	0.140	-0.290*
	(0.0790)	(0.0760)	(0.0813)	(0.0690)	(0.0731)	(0.0736)	(0.117)
Origin CBSA log population	0.0143	0.00555	-0.00529	-0.00241	-0.0121	0.000298	-0.0257*
	(0.0133)	(0.0129)	(0.0133)	(0.0123)	(0.0139)	(0.0141)	(0.0130)
Log mileage (ZIP-ZIP-mode-specific)	0.0413**	0.0375**	0.0382**	$\surd$	$\surd$	$\surd$	$\surd$
	(0.00637)	(0.00615)	(0.00608)				
Non-production worker share (log)		0.131**	0.133**	$\surd$	$\surd$	$\surd$	$\surd$
		(0.0175)	(0.0175)				
Assets per worker (log)		-0.0545**	-0.0522**	$\surd$	$\surd$	$\surd$	$\surd$
		(0.00848)	(0.00841)				
Pay per worker (log)		0.223**	0.222**	$\surd$	$\surd$	$\surd$	$\surd$
		(0.0413)	(0.0420)				
Asking rent (USD per sqft)			0.0457**		$\surd$	$\surd$	$\surd$
			(0.0119)				
Market access (excludes origin) M^1_{ot}						1.095**	
						(0.219)	
Market access M^2_{ot}							1.625**
							(0.309)
Within R^2	0.065	0.075	0.076	0.100	0.105	0.105	0.106
Number estab-year (rounded)				10,000			
Number ind-prod-year (rounded)				4000			
Observations (rounded)				500,000			

Notes: Manufacturing establishments in the CFS and CMF. All shipments are to a domestic destination CBSA distinct from the origin CBSA. All regressions include SCTG5 $\times$ NAICS6 $\times$ destination $\times$ year fixed effects and mode $\times$ year fixed effects. The fourth through seventh columns include 3-digit-NAICS-specific cubic polynomials in log mileage (4-7), log non-production worker share (4-7), log assets per worker (4-7), log pay per worker (4-7), and industrial asking rent per square foot (5-7). Unreported controls in columns 2-7 are the interactions of log origin income per capita with the three input variables. Standard errors, clustered by origin CBSA $\times$ year, in parentheses. ** and * denote statistical significance at 1% and 5%, respectively.

5.4 Robustness checks

This section reports robustness checks motivated by potentially empirically relevant mechanisms that were omitted from the theoretical model. These include land prices, intermediate inputs, and multi-plant firms. While the measured contributions of factor usage and market access are not completely invariant to addressing these issues, all the robustness checks are consistent with the claim that market access explains at least as much of the covariance of shipment prices and income levels as observed factor usage.

The model omits land, which is an input whose price variation across metropolitan areas is correlated with income levels. While land has not been posited as relevant for quality specialization in the prior literature, shipment prices might covary with land prices if higher-quality varieties are less land-intensive or if land costs are passed through. To address this omission, I use an industrial rent measure from Reis, a commercial real estate information company, as an additional regressor. The measure, asking rent per square foot for industrial properties, is available for only 44 metropolitan areas in the relevant years. Table 5 reports the factor-usage and market-access regressions while controlling for this variation in local industrial rents. Columns 1, 2, and 4 demonstrate that the factor-usage results for this subsample of observations are comparable to those obtained in Tables 3. Columns 3 and 5 introduce the industrial rent covariate, which reduces the income elasticity by about five percentage points in the flexible-control specification. Plants in locations with higher land prices have higher outgoing shipment prices, and this variation in rents is correlated with local income levels such that this explains some of the covariance between shipment prices and incomes. After controlling for industrial rents, market access accounts for most of the remaining covariance between prices and incomes, as the income elasticity becomes statistically indistinguishable from zero. Thus, the previous results are robust to controlling for land prices.

In section 3, all goods are final goods sold to consumers. This absence of intermediate inputs raises two concerns for the empirical evidence presented thus far. First, outgoing shipments in the data are made up of both final and intermediate goods. Second, input market access is an omitted variable that may be correlated with output market access. Each of these concerns warrant additional examination.

The formal theory is narrowly written in terms of quality-differentiated final goods, as in Fajgelbaum, Grossman, and Helpman (2011), while the evidence presented thus far includes all manufacturing industries. Market access is potentially important for a broad class of goods. Burenstam Linder (1961) posited that "there must be a home market for an export good, whether it is a consumer good or a capital good" and that "there is a strong relationship between the level of per capita income, on the one hand, and the types of consumer goods and also capital goods demanded, on the other hand."[62] And Kugler and Verhoogen (2012) suggest a complementarity between input quality and output quality that would generate differences in demand for intermediate inputs. Nonetheless, to narrow the inquiry in line with the final-goods-only model, Table 6 restricts the estimation sample to manufacturing industries selling more than 50% of their output value to final consumers.

[62]The relevance of the home market follows from economies of scale and trade costs. The potential role for per capita income in the composition of capital demand is less familiar. Burenstam Linder (1961, p. 96): "The relative amount of capital also determines the qualitative composition of the demand for new capital goods. A capital-abundant country, i.e., a country which, with some likelihood, finds itself on a high level of per capita income, demands more sophisticated capital equipment than a capital-scarce country. Although there is no direct causal relationship, we might thus expect that the differences in the level of per capita incomes would tell us at least something about what differences there will be in the structure of demand for capital goods."

Table 6: Shipment prices for final consumer goods

Dep var: Log unit value, $\ln p_{skjodmt}$	(1)	(2)	(3)	(4)	(5)
Origin CBSA log per capita income	0.332**	0.246**	0.192**	0.119	0.0102
	(0.0658)	(0.0634)	(0.0656)	(0.0741)	(0.0813)
Origin CBSA log population	0.0103	0.00391	0.00706	0.0129	0.00990
	(0.00765)	(0.00715)	(0.00708)	(0.00770)	(0.00714)
Log mileage (ZIP-ZIP-mode-specific)	0.0260**	0.0258**	$\checkmark$	$\checkmark$	$\checkmark$
	(0.00596)	(0.00584)			
Non-production worker share (log)		0.0686**	$\checkmark$	$\checkmark$	$\checkmark$
		(0.0142)			
Assets per worker (log)		-0.00334	$\checkmark$	$\checkmark$	$\checkmark$
		(0.00766)			
Pay per worker (log)		0.238**	$\checkmark$	$\checkmark$	$\checkmark$
		(0.0351)			
Non-production worker share × income per capita		-0.0841	-0.151	-0.154	-0.154
		(0.0724)	(0.0787)	(0.0786)	(0.0792)
Assets per worker × income per capita		-0.0447	-0.0320	-0.0331	-0.0282
		(0.0275)	(0.0310)	(0.0307)	(0.0305)
Pay per worker × income per capita		0.325	0.0175	0.0143	-0.0184
		(0.166)	(0.151)	(0.150)	(0.149)
Market access (excludes origin) M^1_{ot}				0.622*	
				(0.272)	
Market access M^2_{ot}					0.796**
					(0.219)
Within R^2	0.066	0.077	0.107	0.107	0.108
Number estab-year (rounded)			5000		
Number ind-prod-year (rounded)			750		
Observations (rounded)			400,000		

Notes: Manufacturing establishments in the CFS and CMF. All shipments are to a domestic destination CBSA distinct from the origin CBSA. The estimation sample is restricted to industries for which more than 50% of value produced is sold to final consumers. All regressions include SCTG5 × NAICS6 × destination × year fixed effects and mode × year fixed effects. The third through fifth columns include 3-digit-NAICS-specific cubic polynomials in log mileage, log non-production worker share, log assets per worker, and log pay per worker. Standard errors, clustered by origin CBSA × year, in parentheses. ** and * denote statistical significance at 1% and 5%, respectively.

Table 6 shows that the contribution of market access to the covariance of per capita income and outgoing shipment prices is relatively larger for final consumer goods compared to all manufactures. The income elasticity of 33% reported in the first column falls to 19% in the third column after flexibly controlling for observed plant-level factor usage. Introducing M_{ot}^1, the market-access measure that omits potential consumers in the location of production, reduces the income elasticity to 12% in the fourth column and this point estimate is statistically indistinguishable from zero. Using the M_{ot}^2 measure reduces the point estimate to virtually zero, so that factor usage and market access jointly explain almost all of the observed relationship between cities' income levels and outgoing shipment prices in final goods. Market access accounts for more than half of this covariance.

Table 7 shows that the contribution of market access to the price-income covariance is considerably smaller for shipments of intermediate inputs. For industries that sell less than 10% of their output to final consumers, factor inputs and within-intensity variation in market access each explain about one-third of the price-income relationship. Unlike final consumer goods, intermediate inputs exhibit considerable residual covariation. These findings are consistent with the idea that household income levels better predict the composition of demand for consumer goods than intermediate inputs.

Another concern related to intermediate inputs is that the regressions in Table 4 do not control for input market access. Plants in high-income cities may indirectly employ capital or skill via the factor content of intermediate inputs, which would not be captured by the plant-level factor-usage measures used above. In appendix section E.4, I use plant-level wages and input-output tables to construct measures of upstream human capital to address this concern. The findings of Table are 4 are little changed by controlling for input market access in this way.[63]

The model has single-product, single-plant firms, while in reality most manufacturing output is produced by multi-product firms operating multiple plants. A potential concern is that the prior regression results might reflect shipments by large, multi-plant firms whose decisions are poorly described by the model.[64] Appendix section E.5 addresses this concern by restricting the estimation sample to non-large plants and single-plant firms. The results are very similar to those in Tables 3 and 4. Introducing plant size as an additional regressor also yields very similar results.

In sum, robustness checks addressing land prices, intermediate inputs, and multi-plant

[63] Addressing the intermediate-inputs concern by restricting the estimation sample to establishments with high value-added shares also yields similar results.

[64] This potential concern could not simply be transfer pricing of intermediate inputs, since Atalay, Hortaçsu, and Syverson (2014) show that a small fraction of shipments by vertically integrated establishments are to downstream units in the same firm.

Table 7: Shipment prices for intermediate inputs

Dep var: Log unit value, $\ln p_{skjodmt}$	(1)	(2)	(3)	(4)	(5)
Origin CBSA log per capita income	0.443**	0.364**	0.323**	0.169**	0.120
	(0.0559)	(0.0543)	(0.0538)	(0.0574)	(0.0663)
Origin CBSA log population	-0.00940	-0.0158**	-0.0158**	-0.00384	-0.0109
	(0.00612)	(0.00582)	(0.00576)	(0.00586)	(0.00573)
Log mileage (ZIP-ZIP-mode-specific)	0.0376**	0.0370**	$\checkmark$	$\checkmark$	$\checkmark$
	(0.00395)	(0.00383)			
Non-production worker share (log)		0.163**	$\checkmark$	$\checkmark$	$\checkmark$
		(0.0113)			
Assets per worker (log)		-0.0784**	$\checkmark$	$\checkmark$	$\checkmark$
		(0.00659)			
Pay per worker (log)		0.145**	$\checkmark$	$\checkmark$	$\checkmark$
		(0.0263)			
Non-production worker share $\times$ income per capita		0.0288	0.0245	0.0221	0.0199
		(0.0432)	(0.0431)	(0.0432)	(0.0432)
Assets per worker $\times$ income per capita		-0.0756**	-0.0257	-0.0206	-0.0174
		(0.0266)	(0.0264)	(0.0263)	(0.0268)
Pay per worker $\times$ income per capita		0.597**	0.422**	0.407**	0.391**
		(0.102)	(0.104)	(0.103)	(0.104)
Market access (excludes origin) M^1_{ot}				1.208**	
				(0.182)	
Market access M^2_{ot}					0.918**
					(0.184)
Within R^2	0.082	0.096	0.115	0.116	0.116
Number estab-year (rounded)			15,000		
Number ind-prod-year (rounded)			2250		
Observations (rounded)			800,000		

Notes: Manufacturing establishments in the CFS and CMF. All shipments are to a domestic destination CBSA distinct from the origin CBSA. The estimation sample is restricted to industries for which less than 10% of value produced is sold to final consumers. All regressions include SCTG5 $\times$ NAICS6 $\times$ destination $\times$ year fixed effects and mode $\times$ year fixed effects. The third through fifth columns include 3-digit-NAICS-specific cubic polynomials in log mileage, log non-production worker share, log assets per worker, and log pay per worker. Standard errors, clustered by origin CBSA $\times$ year, in parentheses. ** and * denote statistical significance at 1% and 5%, respectively.

firms yield results consistent with my main empirical finding that both factor abundance and market access shape the pattern of quality specialization, with the composition of demand playing a large role. Attributing these results entirely to factor-usage differences rather than spatial variation in demand composition would involve assuming two conditions. First, my plant-level microdata on factor inputs, which are relied upon in the literature to estimate plant-level productivity and other important measures, would have to omit significant factor usage. Second, these unobserved quality-improving inputs would have to be strongly correlated with surrounding cities' income levels, conditional on the income level in the city of production. While my research design cannot disprove a hypothesis positing such spatially correlated unobserved factor inputs, and my decomposition results should therefore be interpreted with appropriate caution, such an argument would be quite far from the evidence typically marshalled to support the factor-abundance hypothesis.

6 Conclusions

Two prominent theories predict that high-income locations specialize in producing and exporting high-quality products. The Linder hypothesis, formalized by Fajgelbaum, Grossman, and Helpman (2011), emphasizes the role of high-income customers' demand for high-quality products. The canonical factor-proportions theory focuses on the abundant supply of capital and skills in high-income locations. Prior empirical evidence does not separate the contributions of these mechanisms because each implies the same predictions about country-level trade flows.

In this paper, I combine microdata on manufacturing plants' shipments and inputs with data on locations' populations and incomes to quantify each mechanism's role in quality specialization across US cities. I develop a model that nests both mechanisms to guide my empirical investigation. The theory's basic insight is that the factor-abundance mechanism operates exclusively through plants' input usage. Conditional on plant-level factor intensity, demand determines quality specialization. I implement my empirical strategy using US microdata because the Commodity Flow Survey and Census of Manufactures describe plants located in many cities of varying income levels. In doing so, I document that US cities exhibit the same patterns found in international trade data that have been interpreted as evidence of quality specialization by income. My empirical investigation finds that home-market demand explains as much of the specialization across US cities of different income levels as observed differences in plants' factor inputs. Cities with better access to higher-income customers are net exporters of higher-price varieties.

This finding is significant because the two mechanisms have distinct implications for welfare, inequality, and trade policy. The large share of quality specialization attributable to market access suggests that a location's capacity to profitably produce high-quality products depends significantly on the income composition of neighboring locations. As a result, geography influences specialization in part because economic developments in neighboring locations may shift local demand for quality. To the degree that demand shapes entry and product availability, individuals may gain by living in locations where other residents' incomes are similar to theirs. Finally, since market access is affected by trade policy, governments may have scope to influence quality specialization.

References

Atalay, E., A. Hortaçsu, and C. Syverson (2014): "Vertical Integration and Input Flows," *American Economic Review*, 104(4), 1120–48.

Baldwin, R., and J. Harrigan (2011): "Zeros, Quality, and Space: Trade Theory and Trade Evidence," *American Economic Journal: Microeconomics*, 3(2), 60–88.

Behrens, K., A. R. Lamorgese, G. I. Ottaviano, and T. Tabuchi (2009): "Beyond the home market effect: Market size and specialization in a multi-country world," *Journal of International Economics*, 79(2), 259–265.

Bergstrand, J. H. (1990): "The Heckscher-Ohlin-Samuelson Model, the Linder Hypothesis and the Determinants of Bilateral Intra-industry Trade," *Economic Journal*, 100(403), 1216–29.

Bernard, A. B., S. J. Redding, and P. K. Schott (2013): "Testing for Factor Price Equality with Unobserved Differences in Factor Quality or Productivity," *American Economic Journal: Microeconomics*, 5(2), 135–63.

Bernasconi, C. (2013): "Similarity of income distributions and the extensive and intensive margin of bilateral trade flows," Economics Working Papers 115, University of Zurich.

Berry, S. T. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 25(2), 242–262.

Burenstam Linder, S. (1961): *An Essay on Trade and Transformation*. John Wiley & Sons, Ltd.

Caron, J., T. Fally, and J. R. Markusen (2014): "International Trade Puzzles: A Solution Linking Production and Preferences," *The Quarterly Journal of Economics*, 129(3), 1501–1552.

Choi, Y. C., D. Hummels, and C. Xiang (2009): "Explaining import quality: The role of the income distribution," *Journal of International Economics*, 78(2), 293–303.

Costinot, A. (2009): "An Elementary Theory of Comparative Advantage," *Econometrica*, 77(4), 1165–1192.

Davis, D. R., and J. I. Dingel (2012): "A Spatial Knowledge Economy," NBER Working Paper 18188.

——— (2014): "The Comparative Advantage of Cities," NBER Working Paper 20602.

Davis, D. R., and D. E. Weinstein (1999): "Economic geography and regional production structure: An empirical investigation," *European Economic Review*, 43(2), 379–407.

——— (2003): "Market access, economic geography and comparative advantage: an empirical test," *Journal of International Economics*, 59(1), 1–23.

De la Roca, J., and D. Puga (2013): "Learning By Working In Big Cities," .

Diamond, R. (2016): "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000," *American Economic Review*, 106(3), 479–524.

Dollar, D., E. N. Wolff, and W. J. Baumol (1988): "The Factor-Price Equalization Model and Industry Labor Productivity: An Empirical Test across Countries," in *Empirical Methods for International Trade*, ed. by R. C. Feenstra, pp. 23–48. MIT Press.

Eaton, J., and S. Kortum (2002): "Technology, Geography, and Trade," *Econometrica*, 70(5), 1741–1779.

Faber, B., and T. Fally (2016): "Firm Heterogeneity in Consumption Baskets: Evidence from Home and Store Scanner Data," mimeo.

Fajgelbaum, P., G. M. Grossman, and E. Helpman (2011): "Income Distribution, Product Quality, and International Trade," *Journal of Political Economy*, 119(4), 721 – 765.

Falvey, R. E. (1981): "Commercial policy and intra-industry trade," *Journal of International Economics*, 11(4), 495–511.

Feenstra, R. C., and J. Romalis (2014): "International Prices and Endogenous Quality," *The Quarterly Journal of Economics*, 129(2), 477–527.

Fieler, A. C. (2011): "Nonhomotheticity and Bilateral Trade: Evidence and a Quantitative Explanation," *Econometrica*, 79(4), 1069–1101.

Glaeser, E. L., and J. E. Kohlhase (2004): "Cities, regions and the decline of transport costs," *Papers in Regional Science*, 83(1), 197–228.

Glaeser, E. L., J. Kolko, and A. Saiz (2001): "Consumer city," *Journal of Economic Geography*, 1(1), 27–50.

Hallak, J. C. (2006): "Product quality and the direction of trade," *Journal of International Economics*, 68(1), 238–265.

——— (2010): "A Product-Quality View of the Linder Hypothesis," *The Review of Economics and Statistics*, 92(3), 453–466.

Hallak, J. C., and P. K. Schott (2011): "Estimating Cross-Country Differences in Product Quality," *The Quarterly Journal of Economics*, 126(1), 417–474.

Hallak, J. C., and J. Sivadasan (2013): "Product and process productivity: Implications for quality choice and conditional exporter premia," *Journal of International Economics*, 91(1), 53 – 67.

Handbury, J. (2012): "Are Poor Cities Cheap for Everyone? Non-Homotheticity and the Cost of Living Across U.S. Cities," mimeo.

Handbury, J., and D. E. Weinstein (2015): "Goods Prices and Availability in Cities," *The Review of Economic Studies*, 82(1), 258–296.

Hanson, G. H., and C. Xiang (2004): "The Home-Market Effect and Bilateral Trade Patterns," *American Economic Review*, 94(4), 1108–1129.

Helpman, E., and P. R. Krugman (1985): *Market Structure and Foreign Trade*. MIT Press.

——— (1989): *Trade Policy and Market Structure*. MIT Press.

Henderson, J. V. (1991): *Urban Development: Theory, Fact, and Illusion*. Oxford University Press.

Holmes, T. J., and J. J. Stevens (2004): "Spatial distribution of economic activities in North America," in *Handbook of Regional and Urban Economics*, ed. by J. V. Henderson, and J. F. Thisse, vol. 4, chap. 63, pp. 2797–2843. Elsevier.

Hottman, C. J., S. J. Redding, and D. E. Weinstein (2016): "Quantifying the Sources of Firm Heterogeneity," *The Quarterly Journal of Economics*, 131(3), 1291–1364.

Hummels, D., and P. J. Klenow (2005): "The Variety and Quality of a Nation's Exports," *American Economic Review*, 95(3), 704–723.

Hummels, D., and A. Skiba (2004): "Shipping the Good Apples Out? An Empirical Confirmation of the Alchian-Allen Conjecture," *Journal of Political Economy*, 112(6), 1384–1402.

Khandelwal, A. (2010): "The Long and Short (of) Quality Ladders," *Review of Economic Studies*, 77, 1450–1476.

Krugman, P. (1980): "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review*, 70(5), 950–59.

——— (1991): *Geography and trade*. MIT Press.

Kugler, M., and E. Verhoogen (2012): "Prices, Plant Size, and Product Quality," *Review of Economic Studies*, 79(1), 307–339.

Lederman, D., and W. F. Maloney (eds.) (2012): *Does What You Export Matter? In Search of Empirical Guidance for Industrial Policies*. World Bank.

Lugovskyy, V., and A. Skiba (2015): "How geography affects quality," *Journal of Development Economics*, 115, 156 – 180.

Markusen, J. R. (1986): "Explaining the Volume of Trade: An Eclectic Approach," *American Economic Review*, 76(5), 1002–11.

Murphy, K. M., and A. Shleifer (1997): "Quality and trade," *Journal of Development Economics*, 53(1), 1–15.

Rauch, J. E. (1999): "Networks versus markets in international trade," *Journal of International Economics*, 48(1), 7–35.

Redding, S. (1996): "The Low-Skill, Low-Quality Trap: Strategic Complementarities between Human Capital and R&D," *Economic Journal*, 106(435), 458–70.

Redding, S., and A. J. Venables (2004): "Economic geography and international inequality," *Journal of International Economics*, 62(1), 53–82.

Schott, P. K. (2003): "One Size Fits All? Heckscher-Ohlin Specialization in Global Production," *American Economic Review*, 93(3), 686–708.

——— (2004): "Across-product Versus Within-product Specialization in International Trade," *The Quarterly Journal of Economics*, 119(2), 646–677.

Simonovska, I. (2015): "Income Differences and Prices of Tradables: Insights from an Online Retailer," *The Review of Economic Studies*, 82(4), 1612–1656.

Sutton, J. (1991): *Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration*. MIT Press.

——— (1998): *Technology and Market Structure: Theory and History*. MIT Press.

——— (2012): *Competing in Capabilities: The Globalization Process*. Oxford University Press.

Torstensson, J. (1996): "Can factor proportions explain vertical intra-industry trade?," *Applied Economics Letters*, 3(5), 307–309.

Verhoogen, E. A. (2008): "Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector," *The Quarterly Journal of Economics*, 123(2), 489–530.

The Determinants of Quality Specialization – Online Appendix

Jonathan I. Dingel, September 2016

A Theory appendix

This appendix contains details and derivations of the theoretical results summarized in section 3. The order parallels that of the main text.

A.1 Preferences

In equation (1), ϵ_{hj} is the individual's idiosyncratic valuation of the variety. Varieties are thus both vertically and horizontally differentiated (Beath and Katsoulacos, 1991, p.4-6). Conditional on ϵ_h, all consumers prefer higher-q varieties. If all varieties were the same price, consumers would not be unanimous in their ranking of them due to ϵ_h, so these products are horizontally differentiated.

An individual's vector of idiosyncratic valuations, ϵ_h, is drawn from the generalized extreme value distribution, $G_\epsilon(\epsilon) = \exp\left[-\sum_{q\in Q}(\sum_{j\in J_q}\exp(-\epsilon_j/\theta_q))^{\theta_q}\right]$, where Q denotes the set of qualities, $J_q \equiv \{j : q_j = q\}$ is the set of varieties of quality q, and θ_q governs the strength of idiosyncratic differences among varieties in J_q. As shown in Fajgelbaum, Grossman, and Helpman (2011), this specification yields a nested-logit demand system (McFadden, 1978).

$$\rho_j(y) = \rho_{j|q}\cdot\rho_q(y) = \frac{\exp(-p_j/m_q)}{\sum_{j'\in J_q}\exp(-p_{j'}/m_q)}\cdot\frac{\left[\sum_{j'\in J_q}\exp((y-p_{j'})/m_q)\right]^{\theta_q}}{\sum_{q'}\left[\sum_{j'\in J_{q'}}\exp((y-p_{j'})/m_{q'})\right]^{\theta_{q'}}}$$

A.2 Production

Both the homogeneous good and the differentiated good are produced using a constant-elasticity-of-substitution technology, yielding the unit-cost functions given in the main text. A firm producing a variety j of the differentiated good in location k solves the following profit maximization problem:

$$\max_{\{p_{jk'}\}}\pi_j = \sum_{k'} d_{jk'}(p_{jk'} - c_{qk} - c_{qk}\tau_{qkk'}) - f_q$$

with j's sales in k' given by $d_{jk'} = N_{k'}\int\rho_j(y)g(y,k')dy$. Because $\frac{\partial\ln d_{jk'}}{\partial\ln p_{jk'}} = -p_{jk'}/m_q$, the solution to this maximization problem is $p_{jk'} = c(q,k)+\tau_{qkk'}+m_q$. Plugging these solutions into the expression for $\rho_j(y)$ yields the sales for each firm to each destination as a function of trade costs and all other firms' costs.

$$
\begin{aligned}
d_{jk'} =& N_{k'} \int \frac{\exp(-p_j/m_q)}{\sum_{j' \in J_q} \exp(-p_{j'}/m_q)} \cdot \frac{\left[\sum_{j' \in J_q} \exp((y - p_{j'})/m_q)\right]^{\theta_q}}{\sum_{q'} \left[\sum_{j' \in J_{q'}} \exp((y - p_{j'})/m_{q'})\right]^{\theta_{q'}}} g(y, k') dy \\
=& \exp(-(c(q,k) + \tau_{qkk'})/m_q) N_{k'} \Gamma_{k'}(q, \mathbf{n}, \mathbf{c}, \tau) \\
\Gamma_{k'}(q, \mathbf{n}, \mathbf{c}, \tau) \equiv& \int \left[\frac{\exp(yq) \left[\sum_{k''} n_{q,k''} \exp(-(c(q,k'') + \tau_{qk''k'})/m_q)\right]^{\theta_q - 1}}{\sum_{q'} \exp(yq') \left[\sum_{k''} n_{q',k''} \exp(-(c(q,k'') + \tau_{q'k''k'})/m_{q'})\right]^{\theta_{q'}}} \right] g(y, k') dy
\end{aligned}
$$

The function $\Gamma_{k'}(q, \mathbf{n}, \mathbf{c}, \tau)$ describes the share of demand in location k' for quality q given the equilibrium prices and locations of all producers. Fajgelbaum, Grossman, and Helpman (2011) introduce a similar function in their equation (22). Their expression subsumes τ by adjusting $\mathbf{n}$ for trade costs to measure "effective varieties" and does not depend on $\mathbf{c}$ because they assume factor prices equalize.

A.3 Equilibrium

In equilibrium, labor markets clear and firms earn zero profits. These conditions are given in the main text by equations (2) and (3). The zero-profit condition for the homogeneous, numeraire good is $c(z,k) \geq 1 \; \forall k$ and $x(z,k) > 0 \Rightarrow c(z,k) = 1$. The equilibrium local income distribution density, which depends on equilibrium wages $w(\omega, k)$, is

$$g(y, k) = \int_{\omega \in \Omega : w(\omega, k) = y} f(\omega, k) d\omega.$$

A.4 Equilibrium pattern of specialization and trade

The skill-intensity index $i(r)$ is defined such that $i(r) = i(r') \iff b(\omega, r) = h(r, r') b(\omega, r') \; \forall \omega$ for some function $h(r, r')$ and $i(r) > i(r') \Rightarrow r > r'$. It is convenient to choose the labels $i(r)$ such that $i(r)$ is the identity of the lowest r in the set of products with this skill intensity. This allows us to write $b(\omega, r) = h(r, i(r)) b(\omega, i(r))$. It also makes $b(\omega, i)$ strictly log-supermodular by definition, whether $b(\omega, r)$ is multiplicatively separable or weakly log-supermodular.

A.4.1 Skill-intensive quality and costless trade

Trade is costless and $b(\omega, r)$ is weakly log-supermodular. The absence of trade costs ($\tau_{qkk'} = 0 \; \forall q \; \forall k \; \forall k'$) means that the zero-profit condition (3) reduces to

$$\pi_{q,k} = m_q \exp(-c(q,k)/m_q) \sum_{k'} N_{k'} \Gamma_{k'}(q, \mathbf{n}, \mathbf{c}, \mathbf{0}) - f_q \leq 0.$$

This makes demand conditions within the summation orthogonal to the production location.

Skill abundance imposes structure on the pattern of production through the labor-market clearing condition. In particular, equation (2), the strict log-supermodularity of $f(\omega, k)$, and $f(\omega, k) > 0 \; \forall \omega \in \Omega \; \forall k$ together imply, for $k > k'$ and $\omega > \omega'$,

$$\frac{\sum_{r \in z \cup Q} n_{r,k} x(r,k) w(\omega,k)^{-\sigma} b(\omega,r)^{\sigma} c(r,k)^{\sigma}}{\sum_{r \in z \cup Q} n_{r,k} x(r,k) w(\omega',k)^{-\sigma} b(\omega',r)^{\sigma} c(r,k)^{\sigma}} > \frac{\sum_{r \in z \cup Q} n_{r,k'} x(r,k') w(\omega,k')^{-\sigma} b(\omega,r)^{\sigma} c(r,k')^{\sigma}}{\sum_{r \in z \cup Q} n_{r,k'} x(r,k') w(\omega',k')^{-\sigma} b(\omega',r)^{\sigma} c(r,k')^{\sigma}}.$$

This inequality can be rewritten as

$$\frac{w(\omega,k)^{-\sigma}}{w(\omega',k)^{-\sigma}} \sum_{r \in z \cup Q} \frac{b(\omega,r)^{\sigma}}{b(\omega',r)^{\sigma}} \phi(r,\omega',k) > \frac{w(\omega,k')^{-\sigma}}{w(\omega',k')^{-\sigma}} \sum_{r \in z \cup Q} \frac{b(\omega,r)^{\sigma}}{b(\omega',r)^{\sigma}} \phi(r,\omega',k'),$$

where $\phi(r,\omega',k) \equiv \frac{n_{r,k} x(r,k) b(\omega',r)^{\sigma} c(r,k)^{\sigma}}{\sum_{r \in z \cup Q} n_{r,k} x(r,k) b(\omega',r)^{\sigma} c(r,k)^{\sigma} dr}$ is akin to a probability mass function. $\phi(r,\omega',k)$ describes the output share of quality (product) r in location k when shares are weighted by their production costs and use of skill ω'. Similarly, define $\phi_i(\omega',k) \equiv \sum_{r:i=i(r)} \phi(r,\omega',k)$, which describes the output share of products with skill intensity i in location k, and the related expectation operator $\mathbb{E}_{\omega',k}[\alpha(i)] \equiv \sum_i \alpha(i) \phi_i(\omega',k)$. Using the fact that $\frac{b(\omega,r)}{b(\omega',r)}$ is the same for all r with $i(r) = i$, the inequality can then be written as expression (4) in the main text:

$$\frac{w(\omega,k)^{-\sigma}}{w(\omega',k)^{-\sigma}} \mathbb{E}_{\omega',k} \left(\frac{b(\omega,i)^{\sigma}}{b(\omega',i)^{\sigma}} \right) > \frac{w(\omega,k')^{-\sigma}}{w(\omega',k')^{-\sigma}} \mathbb{E}_{\omega',k'} \left(\frac{b(\omega,i)^{\sigma}}{b(\omega',i)^{\sigma}} \right).$$

Inequality (4) says that skill-abundant locations specialize in skill-intensive products. To arrive at the result, we'll first consider the case of factor-price equalization and then describe the more general case.

When wages are equal across locations, $w(\omega,k) = w(\omega) \; \forall k$, this inequality simplifies to

$$\mathbb{E}_{\omega',k} \left(\frac{b(\omega,i)^{\sigma}}{b(\omega',i)^{\sigma}} \right) > \mathbb{E}_{\omega',k'} \left(\frac{b(\omega,i)^{\sigma}}{b(\omega',i)^{\sigma}} \right).$$

In this case, $\phi(r,\omega',k) = \frac{n_{r,k} x(r,k) b(\omega',r)^{\sigma} c(r)^{\sigma}}{\sum_{r \in z \cup Q} n_{r,k} x(r,k) b(\omega',r)^{\sigma} c(r)^{\sigma} dr} = \frac{n_{r,k} x(r,k) \ell(\omega',r)}{\sum_{r \in z \cup Q} n_{r,k} x(r,k) \ell(\omega',r) dr}$ describes ω'-

use-weighted shares of total output of r in location k. The inequality says that the average skill intensity of output is higher in k than k'. This requires that $\phi_i(\omega', k)$ put more weight on some higher values of i than $\phi_i(\omega', k')$ does so that the average skill intensity of output produced in k is higher than in k'.[65] This is the factor-abundance mechanism for quality specialization.

In the absence of factor-price equalization, a slightly longer chain of reasoning delivers the same conclusion. Suppose that inequality (4) were true but $\phi_i(\omega', k)$ did not place greater weight on higher values of i in higher-k locations (i.e. $\phi_i(\omega', k) = \phi_i(\omega', k')$). If so, $w(\omega, k)^{-\sigma}$ must be strictly log-supermodular to satisfy this inequality. If $w(\omega, k)^{-\sigma}$ is log-supermodular, then $c(r, k)^{-\sigma}$ is log-supermodular, since $c(r, k)^{1-\sigma} = \int_{\omega\in\Omega} b(\omega, r)^{\sigma} w(\omega, k)^{1-\sigma} d\omega$ and $b(\omega, r)$ is log-supermodular (Lehmann, 1955). In the absence of trade costs, varieties of skill intensity i are only produced in location k when $c(i, k) = \min_{k'} c(i, k')$. Thus, log-supermodularity of $c(i, k)^{-\sigma}$ and the zero-profit condition imply that $\phi_i(\omega', k)$ is log-supermodular in (i, k). As a result, $\mathbb{E}_{\omega',k}\left(\frac{b(\omega,i)^{\sigma}}{b(\omega',i)^{\sigma}}\right) > \mathbb{E}_{\omega',k'}\left(\frac{b(\omega,i)^{\sigma}}{b(\omega',i)^{\sigma}}\right)$. Therefore skill-abundant locations specialize in skill-intensive products.

A.4.2 Uniform skill intensities and costly trade

When skill intensities are uniform, $b(\omega, r) = b_1(\omega)b_2(r)$, unit costs are multiplicatively separable in (r, k) and can be written as

$$c(r, k) = b_2(r)^{\frac{\sigma}{1-\sigma}} \left(\int_{\omega\in\Omega} b_1(\omega)^{\sigma} w(\omega, k)^{1-\sigma} d\omega\right)^{\frac{1}{1-\sigma}} \equiv b_2(r)^{\frac{\sigma}{1-\sigma}} c(k)$$

The main text considers the case of trade costs that are small and uniform, $\tau_{qkk'} = \tau_q \ \forall k' \neq k$ and $\tau_{qkk} = 0$. When factor prices equalizes, profits vary only with home demand, $\pi_{q,k} > \pi_{q,k''} \iff N_k\Gamma_k(q, \mathbf{n}, \mathbf{c}, \tau) > N_{k''}\Gamma_{k''}(q, \mathbf{n}, \mathbf{c}, \tau)$. When trade costs are sufficiently low, demands approach their costless-trade equilibrium levels $\Gamma_k(q, \bar{\mathbf{n}}, \bar{\mathbf{c}}, \mathbf{0})$. Provided that wages are increasing in skill, demand share $\Gamma_k(q, \mathbf{n}, \mathbf{c}, \mathbf{0})$ is strictly log-supermodular in (q, k).

Lemma 1. *When factor prices equalize and wages are increasing in skill, $\Gamma_k(q, \mathbf{n}, \mathbf{c}, \mathbf{0})$ is strictly log-supermodular in (q, k).*

Proof. The proof of this lemma is quite similar to the proof of Lemma 1 in Fajgelbaum,

[65] Note that the inequality of expectations holds for an arbitrary $\omega > \omega'$. Therefore, $\{\frac{b(\omega,i)^{\sigma}}{b(\omega',i)^{\sigma}}\}_{\omega,\omega'\in\Omega,\omega>\omega'}$ is a class of strictly increasing functions. If this class were all increasing functions, we would conclude that $\phi_i(\omega', k)$ stochastically dominates $\phi_i(\omega', k')$, since $\mathbb{E}_{\tilde{x}}(u(x)) \geq \mathbb{E}_{\tilde{x}'}(u(x)) \ \forall u'(x) > 0 \iff F_{\tilde{x}}(x) \leq F_{\tilde{x}'}(x) \ \forall x$, where F is the cumulative distribution function.

Grossman, and Helpman (2011). The demand level is

$$\Gamma_k(q, \mathbf{n}, \mathbf{c}, \mathbf{0}) = \int \underbrace{\frac{\exp(yq)\left[\sum_{k''} n_{q,k''} \exp(-c(q,k'')q/\theta_q)\right]^{\theta_q - 1}}{\sum_{q'} \exp(yq')\left[\sum_{k''} n_{q',k''} \exp(-c(q,k'')q'/\theta_{q'})\right]^{\theta_{q'}}}}_{\equiv \Psi(y,q,\mathbf{n},\mathbf{c},\mathbf{0})} g(y,k)dy$$

$\exp(yq)$ is strictly log-supermodular (SLSM), so $\Psi(y, q, \mathbf{n}, \mathbf{c}, \mathbf{0})$ is SLSM in (y, q). Since $w(\omega)$ is increasing and $f(\omega, k)$ is SLSM, $g(y, k)$ is SLSM. Since $\Psi(y, q, \mathbf{n}, \mathbf{c}, \mathbf{0})$ is SLSM in (y, q) and $g(y, k)$ is SLSM in (y, k), $\Gamma_k(q, \mathbf{n}, \mathbf{c}, \mathbf{0})$ is strictly log-supermodular in (q, k) (Lehmann, 1955). □

This property yields the result reported in the main text that high-k locations are high-income because they are skill-abundant, and this causes them to have greater demand for high-q varieties.

Consider another case, in which trade costs are large and uniform. When trade costs are sufficiently high, production is diversified, with every quality segment produced in every location. With an additional assumption, that θ_q is increasing in q, Fajgelbaum, Grossman, and Helpman (2011) characterize the pattern of trade that arises when factor prices equalize. It can be shown that all firms producing a quality q make the same local sales and export sales. Therefore, k is a net exporter of q if it has more firms producing q than the world average, $n_{q,k} > \frac{1}{K}\sum_{k'} n_{q,k'}$. Lemma 2 and Proposition 5 in FGH show that, in this diversified equilibrium, higher-income locations consume relative more of higher-quality varieties and therefore host relatively more firms producing higher-quality varieties, making them net exporters.

A.4.3 Skill-intensive quality and costly trade

Proof of Proposition 1:

Proof. Suppose not. That is, suppose $M_{q,k} < M_{q,k'}$ and $M_{q',k} > M_{q',k'}$. If $c(q, k) \geq c(q, k')$, then by approximation (5) $\pi_{q,k'} > \pi_{q,k}$, which contradicts $n_{q,k} > 0$ by the free-entry condition. Similarly, if $c(q, k) \leq c(q, k')$, then by $\pi_{q',k} > \pi_{q',k'}$, which contradicts $n_{q',k'} > 0$. □

A.5 Taking the theory to plant-level data

Denote the plant index j, so that, for example, plant j's skill intensity is $i(j)$.

If $b(\omega, q)$ is strictly decreasing in q, higher-quality varieties are more costly to produce,

$c(q,k) > c(q',k) \iff q > q'$.

$$\frac{\partial c(q,k)}{\partial q} = \frac{-\sigma}{\sigma-1} c(q,k)^{\sigma} \int_{\omega\in\Omega} b(\omega,q)^{\sigma-1} w(\omega,k)^{1-\sigma} \frac{\partial b(\omega,q)}{\partial q} d\omega$$
$$\frac{\partial b(\omega,q)}{\partial q} < 0 \ \forall\omega \Rightarrow \frac{\partial c(q,k)}{\partial q} > 0$$

If $g(y,k)$ is log-supermodular, average income is a sufficient statistic for k. $k > k'$ if and only if $g(y,k)$ likelihood-ratio dominates $g(y,k')$, so $\mathbb{E}_k(y) > \mathbb{E}_{k'}(y) \iff k > k'$.

The composition measure assumes that non-production workers are more skilled than production workers. Denote the fraction of workers of skill ω labeled as non-production by $l(N,\omega)$ and the fraction labeled production as $l(P,\omega) = 1 - l(N,\omega)$. Denote the share of non-production workers employed in a plant with skill intensity i in location k by $share_N(i,k) \equiv \frac{\int_\omega \ell(\omega,i(q),k) l(N,\omega) d\omega}{\int_\omega \ell(\omega,i(q),k) d\omega}$. If $l(N,\omega)$ is strictly increasing in ω, then $share_N(i,k)$ is strictly increasing in i.[66] Inside the factor-price equalization (FPE) set, $\ell(\omega,q,k) = \ell(\omega,q) \ \forall k$ and therefore $share_N(i,k) = share_N(i) \ \forall k$. Outside the FPE set, if $w(\omega,k)^{-\sigma}$ is log-supermodular, $share_N(i,k)$ is strictly increasing in k.[67] I therefore use $share_N(j) \times \ln \bar{y}_k$ as an additional control for skill intensity.

The wage measure assumes that wages are increasing in skill, ω.[68] If wages are increasing in skill, we can infer the skill intensity of a plant's variety from its average wage. The average wage at a plant producing quality q with skill intensity $i(q)$ in location k is

$$\bar{w}(i,k) = \bar{w}(q,k) = \int_\omega \frac{w(\omega,k)\ell(\omega,i(q),k)}{\int_{\omega'} \ell(\omega',i(q),k) d\omega'} d\omega = \int_\omega w(\omega,k) \varphi(\omega,i,k) d\omega,$$

where $\varphi(\omega,i(q),k) \equiv \frac{\ell(\omega,i(q),k)}{\int_{\omega'} \ell(\omega',i(q),k) d\omega'}$ is a density that is strictly log-supermodular in (ω,i), which means that $\varphi(\omega,i,k)$ likelihood-ratio dominates $\varphi(\omega,i',k)$ if and only if $i > i'$.[69] The average wage $\bar{w}(i,k)$ is therefore strictly increasing in skill intensity i. We can similarly

[66] This assumption is analogous to Property (28) in Costinot and Vogel (2010), which connects observable and unobservable skills. If $l(N,\omega)$ is strictly increasing in ω, then choosing the labeling scheme of worker type $t = N$ or $t = P$ with $N > P$ makes $l(t,\omega)$ a strictly log-supermodular function. Since $\ell(\omega,i,k)$ is strictly log-supermodular in (ω,i) and strict log-supermodularity is preserved by integration, the integral $\int_\omega \ell(\omega,i,k) l(t,\omega) d\omega$ is strictly log-supermodular in (t,i). As a result, the ratio $\int_\omega \ell(\omega,i,k) l(N,\omega) d\omega / \int_\omega \ell(\omega,i,k) l(P,\omega) d\omega$ is strictly increasing in i. Therefore $share_N(i,k)$ is strictly increasing in i.

[67] This follows from $\ell(\omega,i,k)$ log-supermodular in (ω,k) and $l(t,\omega)$ log-supermodular in (t,ω).

[68] A sufficient condition is $\frac{\partial \ln w(\omega,k)}{\partial \omega} = \frac{\partial \ln b(\omega,q)}{\partial \omega} - \frac{1}{\sigma} \frac{\partial \ln f(\omega,k)}{\partial \omega} > 0 \ \forall q \forall k$. Informally, more skilled individuals have greater absolute advantage than local abundance.

[69] $\bar{w}(i,k)$ and $\varphi(\omega,i,k)$ can be written in terms of the skill intensity $i(q)$ because $i(q) = i(q') \Rightarrow \frac{\ell(\omega,q,k)}{\int_{\omega'} \ell(\omega',q,k) d\omega'} = \frac{\ell(\omega,q',k)}{\int_{\omega'} \ell(\omega',q',k) d\omega'}$.

define the average wage of production workers by

$$\bar{w}_P(i,k) = \int \frac{w(\omega,k)\ell(\omega,q,k)l(P,\omega)}{\int \ell(\omega',q,k)l(P,\omega')d\omega'}d\omega = \int w(\omega,k)\varphi_P(\omega,i,k)d\omega,$$

and an analogous average wage $\bar{w}_N(i,k)$ for non-production workers with density $\varphi_N(\omega,i,k)$. $\varphi_P(\omega,i,k)$ and $\varphi_N(\omega,i,k)$ are strictly log-supermodular in (ω,i), so these average wages are strictly increasing in skill intensity i.

Inside the FPE set, wages equalize across locations, $\bar{w}(q,k) = \bar{w}(q)\ \forall k$, and ranking plants by their averages wages is equivalent to ranking them by their factor intensities, $\bar{w}_j > \bar{w}_{j'} \iff i(j) > i(j')$. $\bar{w}_P(q,k)$ and $\bar{w}_N(q,k)$ also have these properties. This motivates using these average wages as establishment-level controls for skill intensity. Outside the FPE set, if $w(\omega,k)^{1-\sigma}$ is log-supermodular, $w(\omega,k)\varphi(\omega,i,k)$ is strictly log-supermodular in (ω,i) and in (ω,k). As a result, $\bar{w}(i,k)$ is increasing in i, increasing in k, and log-supermodular. I therefore use $\bar{w}_j \times \ln \bar{y}_k$ as an additional establishment-level control for skill intensity.

The approximation $M_{q,k}(\tau)$ and its empirical counterparts describing the composition of demand using per capita incomes exploit the fact that this is a sufficient statistic for relative demand for qualities under the model's assumptions. Lemma 1 in Fajgelbaum, Grossman, and Helpman (2011) shows that, when $g(y,k)$ is log-supermodular and trade costs are small, relative demand for higher-q varieties is greater in the higher-k location. That is, when $g(y,k)$ is log-supermodular, $\Gamma_{k'}(q,\bar{\mathbf{n}},\bar{\mathbf{c}},\mathbf{0})$ is log-supermodular in (q,k). As mentioned previously, when $g(y,k)$ is log-supermodular, income per capita is a sufficient statistic for k.

If trade costs were large, expression (5) may not be a good approximation of profits. The approximation characterizes market access $M_{q,k}(\tau)$ using the composition of demand $\Gamma_k(q,\bar{\mathbf{n}},\bar{\mathbf{c}},\mathbf{0})$ when trade is costless. When trade costs are high, $\Gamma_{k'}(q,\mathbf{n},\mathbf{c},\tau)$ also varies across k' due to the equilibrium locations of firms. To the degree that trade costs are large, $\Gamma_{k'}(q,\mathbf{n},\mathbf{c},\tau)$ may not be well predicted by $\Gamma_k(q,\bar{\mathbf{n}},\bar{\mathbf{c}},\mathbf{0})$ and therefore not well predicted by variation in income per capita alone. In my empirical work, I find that my market-access measures explain much of the variation in outgoing shipment prices that is related to origin incomes, so they do not seem to be underpowered in practice.

M^2_{ot} can be written as a weighted average of M^1_{ot} and $\ln \bar{y}_{ot}$. A city's own income per capita has greater weight in this average when it is more populous than and more distant from other metropolitan areas.

$$M^2_{ot} = \frac{1}{\sum_d N_{dt} miles_{od}^{-\eta}} \left(\left[\sum_{d' \neq o} N_{d't} miles_{od'}^{-\eta} \right] M^1_{ot} + \left[N_{ot} miles_{oo}^{-\eta} \right] \ln \bar{y}_{ot} \right)$$

B Data appendix

B.1 Public data

B.1.1 Geography

All the reported results describe core-based statistical areas (CBSAs). ZIP-code tabulation areas (ZCTAs), counties, and public-use microdata areas (PUMAs) were assigned to OMB-defined CBSAs using the MABLE Geocorr2K geographic correspondence engine from the Missouri Census Data Center. I use the CBSA geographies as defined in November 2008. These consist of 366 metropolitan areas and 574 micropolitan statistical areas.

In the gravity regressions and in constructing the market-access measures, I define the mileage distance between two CBSAs as the geodetic distance between their population centers. These population centers are the population-weighted average of the latitude and longitude coordinates of all ZCTAs within the CBSA, using population counts from the 2000 Census. I define the mileage distance from a CBSA to itself as the population-weighted average of the pairwise geodetic distances between all the ZCTAs within the CBSA. For five CBSAs containing only one ZCTA, I generated this mileage distance to self using the predicted values obtained by projecting mileage distance to self for the other 935 CBSAs onto their land areas.

B.1.2 Locational characteristics

Data on CBSAs' aggregate populations and personal incomes come from the BEA's regional economic profiles for 1997, 2002, and 2007, data series CA30.

Data on the distribution of household incomes for the 1997 and 2002 samples were constructed from county-level estimates reported in U.S. Census Bureau, 2000 Census Summary File 3, Series P052. Data on the distribution of household incomes at the CBSA level for the 2007 sample were obtained from U.S. Census Bureau, 2005-2009 American Community Survey 5-Year Estimates, Series B19001.

City-industry college shares were constructed from the 2000 Census and 2005-2009 American Community Surveys microdata made available via IPUMS-USA (Ruggles, Alexander, Genadek, Goeken, Schroeder, and Sobek, 2010). City-industry means and standard deviations of years of schooling and wages from the 2000 Census and 2005-2009 American Community Surveys microdata made available via IPUMS-USA (Ruggles, Alexander, Genadek, Goeken, Schroeder, and Sobek, 2010).

Locations' latitudes and longitudes were compiled from various sources. Latitude and longitude coordinates for US ZIP-code Tabulation Areas (ZCTAs) were obtained from the

2000 Census.[70] The geodetic distances for export shipments in appendix section E.1 were calculated using latitude and longitude coordinates for major Canadian cities, which were constructed by aggregating Canadian dissemination areas' populations and coordinates in the 2006 Census Geographic Attribute File from Statistics Canada, and coordinates of each nation's capital or main city from Mayer and Zignago (2011).

B.1.3 Industrial and product characteristics

The Sutton (1998) R&D and advertising intensity measure of scope for vertical differentiation is provided at the SIC72 level in Federal Trade Commission (1981). These were mapped to 1987 SIC codes using the Bartlesman, Becker, and Gray concordance from Jon Haveman's website and to 1997, 2002, and 2007 (via 2002) NAICS codes using concordances from the US Census. For industries to which multiple SIC72 industries were mapped, I calculated the weighted average of intensities, using 1975 sales as weights.

The Khandelwal (2010) 'ladder' measure of scope for vertical differentiation was mapped from HS10 product codes to 6-digit NAICS codes using the Pierce and Schott (2012) concordance. For industries to which multiple commodities were mapped, I calculated the weighted average of ladder lengths, using the initial period import values reported by Khandelwal as weights.

I use the Bureau of Economic Analysis Benchmark Input-Output Data to distinguish between final consumer goods and intermediate inputs and to construct input-market access measures. I use the 1997 and 2002 Standard Make and Use Tables at the detailed level to compute each industry's fraction of value produced that is sold to final consumers. I also use them to construct input-market-access measures based on industries' expenditure shares for purchased inputs. I calculate share-weighted average wages using the input-output table shares and local wages from the Longitudinal Business Data microdata.

In calculating demand shifters, I mapped the Feenstra and Romalis (2014) estimates of $\hat{\sigma}$ and $\hat{\lambda}$ from SITC revision 2 commodity codes to 5-digit SCTG product codes using a United Nations SITC-HS concordance and a Statistics Canada HS-SCTG concordance. When multiple parameter estimates mapped to a single 5-digit SCTG product code, I used the median values of $\hat{\sigma}$ and $\hat{\lambda}$. The results are robust to using the arithmetic mean.

B.2 Confidential Census data

I use establishment-level microdata from the 1997, 2002, and 2007 editions of the Commodity Flow Survey (CFS) and Census of Manufactures (CMF).

[70]Downloaded from http://www.census.gov/tiger/tms/gazetteer/zcta5.txt in December 2012.

B.2.1 Commodity Flow Survey

I use data describing shipments by manufacturing plants from the Commodity Flow Survey, a component of the quinquennial Economic Census. Each quarter of the survey year, establishments report a randomly selected sample of 20-40 of their shipments from a given week and describe them in terms of commodity content, value, weight, destination, transportation mode, and other characteristics. The approximately 100,000 establishments sampled by the CFS were selected using a stratified sampling design reflecting the Commodity Flow Survey's objectives (Bureau of Transportation Statistics and US Census Bureau, 2010); of the approximately 350,000 manufacturing establishments in the United States, about 10,000 per year appear in my estimation sample.[71]

These data are analogous to firm-level customs data with four important distinctions. First, the data describe shipments at the establishment level rather than at the firm level. Second, the geographic detail of ZIP-to-ZIP shipments is orders of magnitude more precise than the distance measures used to describe international transactions. Each shipment's mileage was estimated by BTS/Census using routing algorithms and an integrated, intermodal transportation network developed for that purpose. Third, establishments report a sample of their shipments in the survey, not a complete record of all transactions. Each quarter of the survey year, establishments report a randomly selected sample of 20-40 of their shipments in one week. The CFS data include statistical weights that can be used to estimate aggregate shipment flows. Fourth, the CFS uses a distinct product classification scheme, the Standard Classification of Transport Goods, that is related to the Harmonized System used in international trade data. At its highest level of detail, five digits, the SCTG defines 512 product categories.[72]

The Commodity Flow Survey microdata include statistical weights so that observations can be summed to obtain estimated totals that are representative. Each shipments' associated "tabulation weight" is the product of seven component weights (Bureau of Transportation Statistics and US Census Bureau, 2010, Appendix C). The products of four of these weights (shipment weight, shipment nonresponse weight, quarter weight, and quarter nonresponse weight) scale up an establishment's shipments to estimate the establishments' total annual shipments. The other three component weights (establishment-level adjustment weight, establishment weight, industry-level adjustment weight) scale up establishments' total shipments to estimate national shipments. The 1997 and 2002 microdata report only the

[71] The 2002 CFS sample is roughly half that of the 1997 and 2007 surveys, sampling about 50,000 establishments in total and a proportionate number that appear in my estimation sample (Bureau of Transportation Statistics and US Census Bureau, 2004).

[72] By comparison, the HS scheme has 97 2-digit and about 1400 4-digit commodity categories.

tabulation weights, while the 2007 microdata report all seven component weights.

The demand estimation performed in section E.3 requires measures of establishments' market shares, which are calculated from estimates of their total sales of that product in a destination market. These shares are estimated using the first four component weights. These establishment-level measures should not be scaled up by the latter three component weights, such as the probability of the establishment being selected into the CFS sample. As a result, it is only possible to calculate the plant-product demand shifters using the 2007 microdata, which include the component weights required to estimate market shares.

The CFS classifies shipments' commodity contents using the Standard Classification of Transported Goods (SCTG), a coding system based on the Harmonized System (HS) classification that was introduced in the 1997 CFS. At its highest level of detail, five digits, the SCTG defines 512 product categories. By comparison, the HS scheme has 97 2-digit and about 1400 4-digit commodity categories. I exclude from my analysis all SCTG product categories whose 5-digit product codes end in 99, since these are catch-all categories such as 24399 "Other articles of rubber."

I calculate shipment unit values by dividing shipment value by shipment weight. All my analyses of these unit values are within-product comparisons or regressions incorporating product fixed effects. These unit values are proxies for producer prices, because they do not include shipping costs or shelving costs that may appear in the retail consumer price.

Each shipment is reported to have been sent by any combination of eight transportation modes. In much of the analysis, including all the shipment price regressions, I restrict attention to unimodal shipments and the five most common multimodal shipments. Unimodal shipments constitute more than 80% of shipments by value and 90% by weight (Bureau of Transportation Statistics and US Census Bureau, 2010).

B.2.2 Census of Manufactures

Each manufacturing plant appearing in the Commodity Flow Survey also appears in the Census of Manufactures, which describes plant-level characteristics such as wage bills, production and non-production employees, and capital stocks. The CMF also identifies the county in which an establishment is located. Very small establishments do not report detailed production data to the CMF. Instead, the Census Bureau uses data from administrative records from other agencies, such as tax records, to obtain information on revenues and employment. It then imputes other variables.

The product trailer of the CMF describes the products produced at each establishment. For all products, establishments report the total value of their annual output. For a subset of products, establishments report both values and quantities.[73] I calculate unit values by dividing product value shipped by product quantity shipped; these unit values are used in appendix Table D.2.

B.2.3 Longitudinal Business Database

I also use information from the Longitudinal Business Database (LBD), which is a census of US business establishments and firms with paid employees. Microdata from the 1997, 2002, and 2007 editions are a combination of survey and administrative records. I use the LBD for three purposes. First, I use the records in linking establishments across the 2002 CFS and CMF data sets. Second, I use an establishment's first year appearing in the LBD, which is a comprehensive census, to calculate plants' ages. Third, I use the LBD in constructing input-market-access measures.

B.2.4 Combining the CFS and CMF

I matched shipment-level observations in CFS data to establishment-level characteristics in CMF data using unique establishment identifiers called Census File Numbers (1997) and Survey Unit Identifiers (2002, 2007). I also used information in the LBD to address the switch from CFNs to SUIs in 2002.

Each data source contains information on the geographic location of an establishment. The CMF reports the county in which an establishment is located. The CFS reports the ZIP code and state in all survey years and some information about core-based statistical areas in 2002 and 2007.

B.2.5 Sample selection

Though the CFS and CMF data are very rich descriptions of establishments and their shipments, some of the observations exhibit limitations that warrant their exclusion from the estimation sample.

CFS & CMF: I exclude SCTG5-NAICS6 pairs in which fewer than five establishments report shipping a commodity.

CFS: I restrict the sample to shipments whose mode of transportation is unimodal or one of the five most common multimodal combinations. I exclude shipments with unit values more than two standard deviations from the product mean. I exclude destination-product pairs for which only one establishment ships that product to that destination. I

[73] The set of products for which product quantity shipped data are collected has shrunk over time.

exclude shipments that are the unique instance of that commodity being shipped by that establishment.

CMF: I exclude establishments not belonging to a OMB-defined CBSA as identified by their county location in the CMF and the November-2008-vintage CBSA definitions. I exclude establishments whose information in the Census of Manufactures are derived from administrative records rather than directly reported. I exclude establishments whose employment levels or wage bills are imputed in the 2002 and 2007 CMF.[74] I exclude establishments with wages that lie below the 1st percentile or above the 99th percentile of the wage distribution for manufacturing establishments.

C Gravity regressions appendix

This appendix characterizes the pattern of manufactures shipments between US cities using a gravity model of shipment volumes. Gravity regressions relate the volume of trade to the origin's economic size in terms of output produced, the destination's economic size in terms of consumer expenditure, and trade frictions between the origin and destination.[75] Hallak (2010) and Bernasconi (2013) use gravity models of trade flows between countries to assess whether locations with more similar income levels trade more with each other, controlling for origin characteristics, destination characteristics, and bilateral trade frictions. They find that countries with more similar income distributions trade more with each other, as predicted by Burenstam Linder (1961). I find a similar pattern of trade flows between US cities.

The baseline gravity specification is

$$\ln X_{odt} = -\eta \cdot \ln \text{miles}_{od} + \beta \cdot |\ln \bar{y}_{ot} - \ln \bar{y}_{dt}| + \gamma_{ot} + \gamma_{dt} + \epsilon_{odt}$$

where X_{odt} is the volume of manufactures shipments from origin o to destination d in year t, miles_{od} is the distance between the two locations, $|\ln \bar{y}_{ot} - \ln \bar{y}_{dt}|$ is the difference in their log per capita incomes, γ_{ot} and γ_{dt} origin-year and destination-year fixed effects, and ϵ_{odt} is a residual reflecting both random sampling and potential measurement error.[76]

This baseline specification is estimated using observations with strictly positive shipment

[74] The 1997 CMF does not identify variables that have been imputed, so I am unable to exclude such observations.

[75] See Anderson (2011) and Bergstrand and Egger (2011) for surveys of the gravity literature. Importantly, the correct notions of economic size account for "multilateral resistance" terms that depend on the locations' bilateral trade frictions with all trading partners. These are captured by fixed effects in my regressions.

[76] Observations of X_{odt} include sampling error since a representative sample of shipments is used to estimate the total shipment volume. See the data appendix B for details. Noise in the dependent variable will not bias the estimated coefficients of interest.

volumes. In fact, there are many zeros in the trade matrix, and these non-positive shipment volumes reflect economic mechanisms, like trade costs. For example, every city ships a positive amount to itself $X_{oot} > 0 \; \forall o \forall s \forall t$.

I use two approaches to correct for the non-random nature of zeros. First, I implement the Heckman (1979) two-step selection correction. The first-step probit regression, which has the same regressors on the right-hand side, yields an estimated probability of a strictly positive shipment volume for each origin-destination-sector-year. The second-step estimating equation is

$$\ln X_{odt} = -\eta \cdot \ln \text{miles}_{od} + \beta \cdot |\ln \bar{y}_{ot} - \ln \bar{y}_{dt}| + \delta \frac{\phi(\Phi^{-1}(\hat{\rho}_{odt}))}{\hat{\rho}_{odst}} + \gamma_{ot} + \gamma_{dt} + \epsilon_{odt},$$

where $\hat{\rho}_{odt}$ is the predicted probability from the probit regression, ϕ and Φ are the probability and cumulative density functions of the normal distribution, and $\frac{\phi(\Phi^{-1}(\hat{\rho}_{odt}))}{\hat{\rho}_{odt}}$ is the inverse Mills ratio. The second approach I use to address zeros is the Poisson pseudo-maximum likelihood estimator introduced by Silva and Tenreyro (2006) to estimate the gravity equation in levels.

$$\mathbb{E}(X_{odt}) = \text{miles}_{od}^{-\eta} |\ln \bar{y}_{ot} - \ln \bar{y}_{dt}|^{\beta} \exp(\gamma_{ot} + \gamma_{dt})$$

This estimation approach allows me to include observations for which $X_{odt} = 0$.[77]

Table C.1 reports the result of estimating this gravity regression for aggregate manufactures shipment volumes. The first column reports the result of estimating the baseline gravity specification via OLS. The second column uses the squared difference in incomes, $(\ln \bar{y}_{ot} - \ln \bar{y}_{dt})^2$, as a regressor instead of the absolute difference. The third column adds the two-step Heckman (1979) correction for selection into strictly positive trade flows. The fourth column uses the Silva and Tenreyro (2006) Poisson pseudo-maximum likelihood estimator for all observations. This final estimator is my preferred specification.

In all specifications, the difference between two cities' average income levels is negatively correlated with the level of trade between them. The Linder pattern of trade, in which locations with more similar incomes trade more intensely with each other, holds true for manufactures shipments between US cities. The finding that locations disproportionately demand products that are produced in locations of similar income levels suggests two elements that any model explaining within-product specialization must incorporate. First, preferences are non-homothetic, so the composition of demand varies with locations' incomes.

[77] Silva and Tenreyro (2006) emphasize that their estimation procedure addresses a concern that higher moments of $\exp(\epsilon_{odt})$ are correlated with the regressors, which would cause the estimated coefficients in the log-linear specification to be inconsistent.

Table C.1: Shipment volumes (2007)

	(1)	(2)	(3)	(4)
Estimator	OLS	OLS	Heckman	PPML w/ zeros
Dependent variable	$\ln X_{od}$	$\ln X_{od}$	$\ln X_{od}$	X_{od}
Mileage, $\ln \text{miles}_{od}$	-0.916**	-0.917**	-1.773**	-0.952**
	(0.00745)	(0.00715)	(0.0104)	(0.0170)
Income difference, $\lvert \ln \bar{y}_o - \ln \bar{y}_d \rvert$	-1.696**		-1.074**	-0.430**
	(0.0413)		(0.0454)	(0.0840)
Income difference squared, $(\ln \bar{y}_o - \ln \bar{y}_d)^2$		-2.657**		
		(0.0854)		
Inverse Mills ratio, $\frac{\phi(\Phi^{-1}(\hat{\rho}_{od}))}{\hat{\rho}_{od}}$			2.434**	
			(0.0222)	
R^2	0.347	0.347	0.378	
Observations (rounded)	175,000	175,000	175,000	850,000
Origin CBSAs (rounded)			900	
Destination CBSAs (rounded)			950	

Standard errors in parentheses

** p<0.01, * p<0.05

NOTES: Aggregate shipment volume by establishments in the CFS and CMF between distinct CBSAs in 2007. All regressions include origin and destination fixed effects. Standard errors are bootstrapped with 50 repetitions in columns 1-3 and heteroskedastic-robust in columns 4-5.

Second, high-income locations have comparative advantage in producing products that are particularly attractive to high-income consumers. As described in section 2, these patterns are compatible with the factor-abundance mechanism or home-market effect determining the pattern of within-product specialization.

The distance elasticity of trade is estimated to be near negative one, although the two-step Heckman specification implies that shipment volumes are notably more sensitive to the distance between origin and destination. An elasticity close to negative one is consistent with the central tendency of the vast international literature summarized by Disdier and Head (2008) and the elasticity of domestic shipments reported in Table 1 of Hillberry and Hummels (2008) for geographically aggregate shipment volumes. In section 3.5, I use the fact that sales volumes are inversely proportionate to distance to construct market-access measures.

D Tables appendix

Table D.1: Outgoing shipment prices and city-industry schooling measures

Dep var: Log unit value, $\ln p_{skjodmt}$	(1)	(2)	(3)	(4)	(5)	(6)
Origin CBSA log per capita income	0.521**	0.424**	0.402**	0.492**	0.379**	0.305**
	(0.0435)	(0.0425)	(0.0444)	(0.0412)	(0.0400)	(0.0428)
Origin CBSA log population	-0.0180**	-0.0224**	-0.0204**	-0.0165**	-0.0213**	-0.0171**
	(0.00583)	(0.00553)	(0.00586)	(0.00535)	(0.00486)	(0.00548)
Log mileage	0.0389**	0.0397**	0.0396**	√	√	√
	(0.00314)	(0.00302)	(0.00305)			
Non-production worker share (log)		0.115**	0.114**		√	√
		(0.00809)	(0.00815)			
Assets per worker (log)		-0.0496**	-0.0499**		√	√
		(0.00443)	(0.00442)			
Pay per worker (log)		0.225**	0.224**		√	√
		(0.0211)	(0.0210)			
Mean years of schooling (log)			0.172*			√
			(0.0711)			
Within R^2	0.076	0.086	0.086	0.079	0.102	0.104
Number estab-year (rounded)	30,000					
Number ind-prod-year (rounded)	5250					
Observations (rounded)	1,400,000					

Standard errors, clustered by origin CBSA × year, in parentheses

** p<0.01, * p<0.05

NOTES: Manufacturing establishments in the CFS and CMF. All shipments are to a domestic destination CBSA distinct from the origin CBSA. All regressions include SCTG5 × NAICS6 × destination × year fixed effects and mode × year fixed effects. Unreported controls are the interactions of income per capita with log non-production worker share (columns 2,3,5,6), log assets per worker (2,3,5,6), log pay per worker (2,3,5,6), and log mean years of schooling (3,6), and 3-digit-NAICS-specific cubic polynomials in log mileage (4,5,6), log non-production worker share (5,6), log assets per worker (5,6), log pay per worker (5,6), and log mean years of schooling (6).

Table D.2: Establishments' prices and origin characteristics

Dep var: CMF Log unit value	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log origin CBSA income per capita	0.218**	0.204**	0.193**	0.192**	0.155**	0.142**	0.0591
	(0.0384)	(0.0384)	(0.0392)	(0.0409)	(0.0373)	(0.0392)	(0.0459)
Log origin CBSA population	-0.000772	-0.00120	-0.00151	-0.000374	0.00380	0.00345	-0.00274
	(0.00484)	(0.00481)	(0.00473)	(0.00501)	(0.00483)	(0.00492)	(0.00490)
Log non-production worker share		0.0564**	0.0471**	√		√	√
		(0.00539)	(0.00652)				
Log assets per worker		-0.0125**	-0.0159**	√		√	√
		(0.00395)	(0.00404)				
Log non-production share		0.0261	0.0320	0.0147		0.0125	0.00937
× log per capita income		(0.0268)	(0.0343)	(0.0311)		(0.0310)	(0.0310)
Log assets per worker		-0.0229	-0.0183	-0.0229		-0.0226	-0.0181
× log per capita income		(0.0226)	(0.0213)	(0.0198)		(0.0197)	(0.0198)
Log pay per worker			0.109**	√		√	√
			(0.0281)				
Log pay per production worker			-0.0411*	√		√	√
			(0.0187)				
Log pay per non-production worker			-0.0257*	√		√	√
			(0.0124)				
Log pay per worker			-0.0729	0.00495		0.00157	-0.0108
× log per capita income			(0.170)	(0.147)		(0.146)	(0.147)
Log pay per production worker			-0.0408	-0.0751		-0.0752	-0.0922
× log per capita income			(0.120)	(0.101)		(0.100)	(0.101)
Log pay per non-production worker			-0.00298	0.00839		0.00891	0.00244
× log per capita income			(0.0575)	(0.0531)		(0.0527)	(0.0531)
Market access (excl origin) M^1_{ot}					0.564**	0.469**	
					(0.136)	(0.136)	
Log std dev household income							0.0804
							(0.0460)
Market access M^2_{ot}							0.403**
							(0.113)
R-squared	0.914	0.914	0.914	0.917	0.914	0.917	0.917
Observations (rounded)				100000			
Estab-year (rounded)				27500			
Ind-prod-year (rounded)				8000			
Note				ctrl all flex		ctrl all flex	ctrl all flex

Standard errors, clustered by CBSA × year, in parentheses
** p<0.01, * p<0.05

NOTES: Manufacturing establishments in the CMF reporting a product with quantity shipped. All regressions include NAICS6×year and NAICS product code×year fixed effects. The fourth, sixth, and seventh columns include 3-digit-NAICS-specific third-order polynomials in log non-production worker share, log assets per worker, log pay per worker, log pay per producer worker, and log pay per non-production worker.

Table D.3: Shipment prices and market access (no factor input covariates)

Dep var: Log unit value, $\ln p_{skjodmt}$	(1)	(2)	(3)	(4)	(5)	(6)
Origin CBSA log per capita income	0.440**	0.266**	0.149**	0.413**	0.258**	0.151**
	(0.0359)	(0.0378)	(0.0434)	(0.0345)	(0.0369)	(0.0424)
Origin CBSA log population	-0.00764	0.00517	-0.00153	-0.00652	0.00510	-0.00101
	(0.00421)	(0.00432)	(0.00412)	(0.00393)	(0.00405)	(0.00385)
Log mileage (ZIP-ZIP-mode-specific)	0.0404**	0.0476**	0.0454**	$\checkmark$	$\checkmark$	$\checkmark$
	(0.00280)	(0.00274)	(0.00272)			
Market access (excludes origin) M^1_{ot}		1.374**			1.253**	
		(0.117)			(0.115)	
Market access M^2_{ot}			1.311**			1.194**
			(0.118)			(0.116)
Within R^2	0.080	0.083	0.083	0.084	0.086	0.086
Number estab-year (rounded)			35,000			
Number ind-prod-year (rounded)			5250			
Observations (rounded)			1,800,000			

Standard errors, clustered by origin CBSA × year, in parentheses
** p<0.01, * p<0.05

NOTES: Manufacturing establishments in the CFS and CMF. All shipments are to a domestic destination CBSA distinct from the origin CBSA. All regressions include SCTG5 × NAICS6 × destination × year fixed effects and mode × year fixed effects. The third through fifth columns include 3-digit-NAICS-specific cubic polynomials in log mileage.

E Empirical results appendix

This appendix reports further empirical evidence consistent with the results presented in section 5. First, export shipments exhibit patterns consistent with those found in domestic transactions. Second, cities with greater income dispersion have higher outgoing shipment prices, consistent with the model's demand system in an equilibrium in which most individuals purchase low-quality varieties. Third, demand shifters exhibit the same patterns as outgoing shipment prices. Fourth, the reported results are robust to controlling for input market access. Fifth, the reported results hold in samples that exclude large plants or multi-plant firms.

E.1 Export shipments

This section examines export shipments. My empirical investigation was motivated in part by a growing international trade literature on quality specialization. I implemented my empirical strategy using plant-level data from US cities of varying income levels. The analysis above described shipments destined for US cities, which account for the vast majority of US manufactures output, to characterize how shipments' characteristics are related to the characteristics of their production locations. This section shows that the patterns found in domestic shipments are also found in export shipments destined for foreign markets.

Export shipments by US manufacturing plants exhibit price patterns consistent with those observed in domestic shipments. Table E.1 presents results for regressions analogous to those presented in Tables 3 and 4 using shipments sent to foreign destinations. The sample size is considerably smaller, since exports represent less than 8% of shipments by value and 4% by weight (Bureau of Transportation Statistics and US Census Bureau, 2010).[78] I calculate the mileage distance from origin CBSA to foreign destination using latitude and longitude coordinates.[79]

The estimated coefficients are consistent with those reported for shipments to domestic destinations. The origin-income elasticity of export prices is 42%. After controlling for plant-level factor usage, the origin-income elasticity is 30%. Upon introduction of the market-access measures, the origin-income elasticity becomes negative and statistically indistinguishable from zero. The dispersion of household income in the origin CBSA is positively related to export shipment prices, though this relationship is statistically insignificant, presumably due to the small sample size. Thus, the empirical relationships between export shipment prices, factor usage, and the demand measures are in line with those found for shipments to domestic destinations.

[78] This small sample size prevents me from calculating demand shifters using export shipments.

[79] For Canadian destinations, I use the coordinates of major Canadian cities. For other countries, I use the coordinates of the capital or main city. See data appendix B for details.

Table E.1: Export shipments

Dep var: Log unit value, $\ln p_{skjodmt}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Log origin CBSA income per capita	0.434**	0.401**	0.334**	0.419**	0.369**	0.304**	0.305*	0.218	0.312**	-0.0724	-0.215
	(0.117)	(0.115)	(0.114)	(0.117)	(0.107)	(0.105)	(0.151)	(0.135)	(0.106)	(0.114)	(0.138)
Log origin CBSA population	0.00876	0.00335	0.00333	0.00920	-0.000623	-0.00218	0.00104	-0.00998	-0.00337	0.0288*	0.0112
	(0.0129)	(0.0125)	(0.0124)	(0.0127)	(0.0119)	(0.0115)	(0.0134)	(0.0129)	(0.0116)	(0.0118)	(0.0116)
Log mileage	0.0787**	0.0812**	0.0815**	√	√	√	√	√	√	√	√
	(0.0205)	(0.0202)	(0.0198)								
Log non-production worker share		0.169**	0.141**		√	√		√	√	√	√
		(0.0222)	(0.0271)								
Log assets per worker		-0.0451**	-0.0601**		√	√		√	√	√	√
		(0.0128)	(0.0134)								
Log non-production worker share		-0.00972	-0.0296		-0.105	-0.0258					
× log per capita income		(0.0825)	(0.0969)		(0.0839)	(0.0978)					
Log assets per worker		-0.0432	-0.0787		-0.0811	-0.119*					
× log per capita income		(0.0418)	(0.0455)		(0.0503)	(0.0541)					
Log pay per worker			0.283**			√		√	√	√	√
			(0.0961)								
Log pay per production worker			-0.0633			√		√	√	√	√
			(0.0623)								
Log pay per non-production worker			0.00207			√		√	√	√	√
			(0.0395)								
Log pay per worker			-0.0302			-0.485					
× log per capita income			(0.379)			(0.423)					
Log pay per production worker			0.430			0.590*					
× log per capita income			(0.238)			(0.279)					
Log pay per non-production worker			0.153			0.400*					
× log per capita income			(0.145)			(0.159)					
Log std dev orig household income							0.187	0.153			
							(0.146)	(0.137)			
Log market access (excl orig) M^1_{ot}										2.477**	
										(0.327)	
Log market access M^2_{ot}											1.961**
											(0.375)
R-squared	0.821	0.823	0.824	0.823	0.828	0.831	0.823	0.831	0.831	0.832	0.831
Note				miles flex	qnt flex	all flex	miles flex	all flex	all flex	all flex	all flex
Observations (rounded)						64,000					
Estab-year (rounded)						10,000					
Ind-prod-year (rounded)						2,000					

Standard errors, clustered by CBSA × year, in parentheses
** p<0.01, * p<0.05

Notes: Manufacturing establishments in the CFS and CMF. All shipments are exports to a foreign destination. All regressions include SCTG5 × NAICS6 × year fixed effects and destination × year fixed effects. The fourth through eleventh columns include 3-digit-NAICS-specific third-order polynomials in log mileage (4-11), log non-production worker share (5, 6, 8-11), log assets per worker (5, 6, 8-11), log pay per worker (6, 8-11), log pay per production worker (6, 8-11), and log pay per non-production worker (6, 8-11).

E.2 Income dispersion

This section documents the relationship between the second moment of the income distribution and shipment prices. Cities with greater dispersion in household income exhibit higher prices for both incoming and outgoing shipments. The latter is not explained by dispersion in workers' wages or skills. These findings are consistent with a demand-side mechanism linking local income distributions to the pattern of quality specialization.

In the Fajgelbaum, Grossman, and Helpman (2011) model, income inequality is linked to quality specialization because the income distribution determines the composition of local demand for quality. In general, the effect of greater income dispersion on relative demand for quality is ambiguous. The authors' Proposition 2(iii) shows that, when there are two qualities and the majority of individuals at all income levels consume the low-quality variety, a mean-preserving spread of the income distribution raises local relative demand for the high-quality variety in their demand system. Since the converse would hold if a majority of individuals consumed high-quality varieties, there is no general theoretical result for the correlation between income dispersion and relative demand for quality.

A few theories link income dispersion to specialization through supply-side mechanisms that are absent from the model in section 3. In Grossman and Maggi (2000) and Bombardini, Gallipoli, and Pupato (2012), locations with more diverse skill distributions have comparative advantage in sectors in which skills are more substitutable.[80] In Grossman (2004), imperfect labor contracting causes locations with more diverse skill distributions to have comparative advantage in sectors in which the most talented individuals' contributions are more easily identified. Applying these models to the question at hand involves reinterpreting them as theories of intrasectoral specialization. For example, if different skills were less substitutable in the production of higher-quality products, these models would predict that locations with greater skill dispersion would specialize in lower-quality varieties.[81] If they were more substitutable, the reverse prediction would result.

In light of these theoretical ambiguities, I rely on the distinction between income dispersion among local consumers and skill dispersion among local workers to empirically distinguish between the demand-side and supply-side mechanisms. Income dispersion among all potential customers influences the demand channel. In the supply-side theories (appro-

[80] Though both papers describe locations with greater skill dispersion specializing in sectors with greater substitutability of skills, these two papers differ considerably. Grossman and Maggi (2000) compare two countries and two sectors, one in which output is supermodular in the two workers' talents and another in which output is submodular in talents. They assume that talent is perfectly observed. Bombardini, Gallipoli, and Pupato (2012) describe imperfectly observed skills and CES production functions that vary in their elasticities of substitution between skills.

[81] Grossman and Maggi (2000, p.1255,1271) cite quality control as an example of supermodular production in which less dispersion yields comparative advantage.

priately reinterpreted to describe specialization within sectors), only skill dispersion among those working in the industry in question is relevant. Thus, I construct two types of empirical measures: the standard deviation of household income within each city and the standard deviations of years of schooling and weekly wages within each city-industry pair. The former proxies for the demand-side mechanism; the latter for the supply-side. I proceed to include these measures, which are available for most but not all metropolitan areas in the sample, in linear regressions describing shipment prices.

Table E.2 documents how shipment prices are related to income and skill dispersion in the destination and origin cities. The first two columns report the result of adding the standard deviation of household income in the shipment destination to multivariate regressions like those appearing in Table 2. The first column includes observable origin-city characteristics; the second column includes origin-city fixed effects. In each case, the standard deviation of household income is strongly positively related to the price of incoming shipments. This is consistent with an equilibrium in which a more dispersed income distribution has more households in the right tail of the distribution who purchase higher-price, higher-quality varieties.

The next three columns of Table E.2 relate outgoing shipment prices to income and skill dispersion in the shipment origin. These regressions all include destination-product-year fixed effects and control variables with industry-specific third-order polynomials, like those regressions appearing in the last column of Table 3. The third column introduces the standard deviation of household income as a regressor alongside origin characteristics, shipment mileage, and the three plant-level factor-usage controls. The standard deviation of household income in the shipment origin is strongly, positively related to the outgoing shipment price. This is consistent with models in which income dispersion generates demand for high-price, high-quality varieties or skill dispersion generates comparative advantage in high-price, high-quality varieties. The fifth and sixth columns introduce city-industry-level measures of skill dispersion are available so that we can contrast income dispersion amongst potential customers with skill dispersion amongst workers employed in production. The key result is that controlling for the logs of the standard deviations of years of schooling and weekly wages at the city-industry level leaves the coefficients on origin characteristics virtually unaltered. Skill dispersion on the supply side appears unrelated to outgoing shipment prices. These findings suggest that a demand-side mechanism links the local income distribution to specialization.

Table E.2: Shipment prices and income dispersion

Dep var: Log unit value	(1)	(2)	(3)	(4)	(5)
Destination CBSA log per capita income	0.144**	0.101**			
	(0.0224)	(0.0195)			
Destination CBSA log population	-0.00662**	-0.00724**			
	(0.00211)	(0.00181)			
Destination CBSA log of std dev household income	0.103**	0.0961**			
	(0.0215)	(0.0186)			
Origin CBSA log per capita income	0.446**		0.287**	0.281**	0.286**
	(0.0160)		(0.0524)	(0.0525)	(0.0524)
Origin CBSA log population	-0.00726**		-0.0274**	-0.0292**	-0.0272**
	(0.00149)		(0.00552)	(0.00575)	(0.00555)
Origin CBSA log of std dev household income			0.147*	0.139*	0.147*
			(0.0595)	(0.0598)	(0.0595)
Log mileage (ZIP-ZIP-mode-specific)	0.0432**	0.0462**	$\surd$	$\surd$	$\surd$
	(0.00258)	(0.00188)			
Log of std dev weekly wage (CBSA-INDNAICS)				0.0160	
				(0.00989)	
Log of std dev years schooling (CBSA-INDNAICS)					-0.000837
					(0.00166)
Within R^2	0.102	0.145	0.102	0.102	0.102
Standard errors clusterd by	cbsa dest × year	cbsa dest × year	cbsa × year	cbsa × year	cbsa × year
Number estab-year (rounded)	35,000	35,000	30,000	30,000	30,000
Number ind-prod-year (rounded)	5250	5250	5250	5250	5250
Observations (rounded)	1.8m	1.8m	1.4m	1.4m	1.4m
Origin CBSA x Year FE		Yes			
SCTG5 x NAICS6 x dest CBSA x Year FE			Yes	Yes	Yes

Standard errrors, clustered as indicated, in parentheses
** p<0.01, * p<0.05

NOTES: Manufacturing establishments in the CFS and CMF. All shipments are to a domestic destination CBSA distinct from the origin CBSA. All regressions include mode × year fixed effects. Unreported controls in columns 3 through 5 include 3-digit-NAICS-specific cubic polynomials in log mileage, log non-production worker share, log assets per worker, and log pay per worker. Also unreported in columns 3 through 5 are the interactions of log origin income per capita with the three input variables.

E.3 Demand shifters

This section characterizes the pattern of within-product specialization across US cities and its determinants using estimated demand shifters. As previously described, consumer love of variety in the presence of horizontal differentiation breaks the price-quality mapping by allowing high-cost varieties to sell alongside low-cost varieties of the same quality. Section 5 addresses this concern by including a variety of plant-level cost measures, which were not available to researchers analyzing aggregate trade flows between countries. This section addresses the concern a second time by estimating demand shifters for each plant-product pair. The empirical results are consistent with the unit-value findings.

The demand-shifter approach assigns higher quality valuations to products that have higher market shares, conditional on price (Berry, 1994; Khandelwal, 2010; Sutton, 2012). As described in the data appendix, it is only possible to calculate plants' market shares in the 2007 edition of the Commodity Flow Survey. This considerably reduces the number of observations compared to the number underlying the previously presented results.

To compute demand shifters in the absence of instruments for producer prices, I use the "non-homothetic CES preferences" of Feenstra and Romalis (2014).[82] In this specification, the sales volume s_{jd} of product j in destination market d in 2007 is described by

$$\ln s_{jd} = (\sigma - 1)(\ln q_j + \lambda \ln \bar{y}_d \ln q_j - \ln p_{jd}) + \gamma_d + \epsilon_{jd}$$

where q_j is the product-quality shifter, p_{jd} is price, $\bar{y}_d$ is per capita income in market d, γ_d captures both aggregate expenditure and the price index in the destination market, and ϵ_{jd} captures both idiosyncratic demand shocks and measurement error. The parameter λ governs how consumer valuation of quality varies with income; the parameter σ is the elasticity of substitution and price elasticity of demand. In the demand system, p_{jd} is the price paid by the consumer, while in my data the observed price $\check{p}_{jd}$ excludes shipping costs.[83] I therefore include the shipment mileage from establishment to destination and the shipment mileage interacted with price, $\ln miles_{jd}$ and $\frac{1}{\check{p}_{jdt}} \ln miles_{jd}$, as additional regressors to control for shipping costs.[84] I use sectoral estimates of $\hat{\lambda}$ and $\hat{\sigma}$ from Feenstra and Romalis (2014) in order to estimate q_j in the linear regression

$$\frac{\ln s_{jd} + (\hat{\sigma} - 1) \ln \check{p}_{jd}}{(1 + \hat{\lambda} \ln \bar{y}_d)(\hat{\sigma} - 1)} = \ln q_j + \eta_1 \ln miles_{jd} + \eta_2 \frac{1}{\check{p}_{jd}} \ln miles_{jd} + \tilde{\gamma}_d + \tilde{\epsilon}_{jd}$$

where $\tilde{\gamma}_d$ and $\tilde{\epsilon}_{jd}$ are rescaled versions of γ_d and ϵ_{jd}. These regressions are estimated product-by-product, for products defined by SCTG5-NAICS6 codes, for the 2007 sample.[85]

Table E.3 describes how these estimated demand shifters relate to the observable characteristics of products, plants, and cities. The first column reports that the estimated demand shifters are strongly positively correlated with plants' prices. This validates the use of prices

[82] One merit of this demand system is its computational simplicity. Because the nested-logit demand system used in the Fajgelbaum, Grossman, and Helpman (2011) model uses quality levels as nests, its estimation would require a computationally intensive iterative approach. Products must be assigned to quality nests in order to estimate the demand system, and product qualities must be inferred by estimating demand.

[83] In international trade parlance, demand depends on the "cost-insurance-and-freight" price while my data reports the "free-on-board" price. If the consumer price reflects both multiplicative and additive trade costs, τ^m_{jd} and τ^a_{jd}, then $p_{jd} = \check{p}_{jd}\tau^m_{jd} + \tau^a_{jd}$ and $\ln p_{jd} \approx \ln \check{p}_{jd} + (\tau^m_{jd} - 1) + \frac{1}{\check{p}_{jd}}\tau^a_{jd}$. Assuming that τ^m_{jd} and τ^a_{jd} are functions of the shipment distance motivates the inclusion of $\ln miles_{jd}$ and $\frac{1}{\check{p}_{jd}} \ln miles_{jd}$as regressors.

[84] Omitting these regressors has very little impact on the estimated demand shifters and the subsequent results relating these shifters to city and plant characteristics.

[85] I obtain very similar results if I define products using only 5-digit SCTG codes. See the appendix section B for details of how I mapped the Feenstra and Romalis (2014) parameter estimates to these product codes.

Table E.3: Estimated demand shifters

Dep var: $\ln \hat{q}_j$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Log unit value	0.746**										
	(0.0102)										
Log CBSA income per capita		0.412**	0.270**	0.157	0.0856	0.265**	0.164	0.237**	0.124	0.0451	0.0401
		(0.0840)	(0.0819)	(0.0831)	(0.105)	(0.0857)	(0.112)	(0.0793)	(0.0814)	(0.101)	(0.114)
Log CBSA population		0.00577	-0.00259	0.00629	0.00197	0.0170	0.0119	-0.000985	0.00793	0.00375	-0.0154
		(0.0120)	(0.0114)	(0.0109)	(0.0113)	(0.0115)	(0.0119)	(0.0109)	(0.0106)	(0.0109)	(0.0107)
Market access (excl origin) M^1_{ot}				0.807**		1.034**			0.822**		
				(0.241)		(0.250)			(0.241)		
Market access M^2_{ot}					0.681*		0.904**			0.715*	
					(0.277)		(0.319)			(0.278)	
Log std dev origin hh income											0.317**
											(0.119)
Log non-production worker share			0.0907**	0.0892**	0.0907**			√	√	√	√
			(0.0135)	(0.0136)	(0.0135)						
Log assets per worker			-0.0496**	-0.0485**	-0.0484**			√	√	√	√
			(0.0109)	(0.0109)	(0.0109)						
Log non-production worker share			0.0106	0.00320	-0.00394			0.0212	0.0116	0.00263	0.0152
× log per capita income			(0.0727)	(0.0735)	(0.0728)			(0.0784)	(0.0793)	(0.0786)	(0.0780)
Log assets per worker			0.0119	0.0136	0.0230			0.0277	0.0300	0.0417	0.0294
× log per capita income			(0.0592)	(0.0594)	(0.0590)			(0.0623)	(0.0624)	(0.0620)	(0.0622)
Log pay per worker			0.409**	0.406**	0.407**			√	√	√	√
			(0.0485)	(0.0486)	(0.0493)						
Log pay per production worker			0.0881**	0.0844*	0.0844*			√	√	√	√
			(0.0330)	(0.0333)	(0.0331)						
Log pay per non-production worker			0.0376	0.0363	0.0358			√	√	√	√
			(0.0232)	(0.0232)	(0.0232)						
Log pay per worker			0.198	0.201	0.170			0.0201	0.0224	-0.00358	0.0309
× log per capita income			(0.224)	(0.224)	(0.231)			(0.220)	(0.221)	(0.226)	(0.220)
Log pay per production worker			0.196	0.180	0.170			0.0874	0.0687	0.0571	0.0819
× log per capita income			(0.148)	(0.149)	(0.148)			(0.166)	(0.167)	(0.167)	(0.166)
Log pay per non-production worker			-0.0965	-0.104	-0.115			-0.127	-0.132	-0.145	-0.145
× log per capita income			(0.120)	(0.119)	(0.121)			(0.113)	(0.112)	(0.114)	(0.110)
R^2 (including fixed effects)	0.857	0.614	0.626	0.627	0.626	0.615	0.615	0.646	0.647	0.647	0.646
Standard errors	cl estab	cl cbsa	cl cbsa	cl cbsa	cl cbsa	cl cbsa	cl cbsa	cl cbsa	cl cbsa	cl cbsa	cl cbsa
Observations (rounded)						13,000					
Establishments (rounded)						10,000					
Ind-prod (rounded)						1,000					

Clustered standard errors in parentheses
** p<0.01, * p<0.05

NOTES: Manufacturing establishments in the CFS and CMF. All regressions include SCTG5 × NAICS6 × year fixed effects. Unreported controls in columns 8-11 are 3-digit-NAICS-specific cubic polynomials in log non-production worker share, log assets per worker, log pay per worker, log pay per production worker, and log pay per non-production worker.

in inferring the pattern of quality specialization earlier in the paper. The second column shows that plants with higher estimated demand shifters are located in cities with higher per capita incomes. The 41% origin-income elasticity of the estimated demand shifter is similar to the 44% origin-income elasticity of outgoing shipment prices. The third and eighth columns demonstrate that this positive relationship persists after controlling for plants' input usage. Qualitatively consistent with the result in section 5.1, observed differences in plant-level factor usage explain less than half of the observed correlation between plants' output characteristics and per capita incomes. The fourth and ninth columns replicate the finding that the income composition of proximate potential customers, excluding those in the city of production, is strongly positively associated with a plant's output profile. The 11-percentage-point decline in the origin-income elasticity caused by introducing the first-market access measure after controlling for factor-usage differences suggests that proximity to these customers explains at least one-quarter of the observed variation. The sixth column demonstrates that introducing the first market-access measure prior to controlling for factor usage would result in a change in the origin-income elasticity of essentially the same magnitude as controlling for factor usage. The fifth, seventh, and tenth columns demonstrate that the second market-access measure, which includes residents in the city of production, has greater explanatory power. The eleventh column replicates the findings of section 5.2 by demonstrating a positive relationship with origin CBSA income dispersion and the market-access measure.

These results can be succinctly summarized as a decomposition of the covariance between incomes and shifters. After controlling for population size, differences in observed factor usage are responsible for 46% of the covariance between per capita incomes and estimated demand shifters. Conditional on factor usage, the conservative market-access measure that omits residents in the city of production accounts for 25% of the total covariance, leaving 30% as residual variation. The model-consistent market-access measure that includes residents in the city of production accounts for 48% of the total covariance, leaving 7% as residual variation.[86] Thus, the observed pattern of specialization and inferences about its determinants obtained using estimated demand shifters are similar to those obtained by examining unit values.

E.4 Input market access

This section demonstrates that the regression results in section 5.2 are robust to controlling for spatial differences in the factor content potentially embedded in intermediate in-

[86]Numbers sum to 101% due to rounding.

Table E.4: Shipment prices and input market access

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Origin CBSA log per capita income	0.353**	0.348**	0.351**	0.206**	0.311**	0.322**	0.305**	0.178**
	(0.0345)	(0.0350)	(0.0346)	(0.0368)	(0.0330)	(0.0340)	(0.0327)	(0.0353)
Origin CBSA log population	-0.0136**	-0.0139**	-0.0135**	-0.00230	-0.0133**	-0.0141**	-0.0132**	-0.00296
	(0.00398)	(0.00398)	(0.00396)	(0.00404)	(0.00365)	(0.00365)	(0.00363)	(0.00372)
Log mileage (ZIP-ZIP-mode-specific)	0.0413**	0.0413**	0.0413**	0.0475**				
	(0.00269)	(0.00268)	(0.00269)	(0.00264)				
Non-production worker share (log)	0.113**	0.113**	0.114**	0.112**				
	(0.00695)	(0.00696)	(0.00696)	(0.00700)				
Assets per worker (log)	-0.0540**	-0.0541**	-0.0541**	-0.0528**				
	(0.00404)	(0.00404)	(0.00404)	(0.00400)				
Pay per worker (log)	0.217**	0.215**	0.216**	0.205**				
	(0.0180)	(0.0182)	(0.0180)	(0.0179)				
Log of input-weighted average wage (CBSA-NAICS6)		0.0140						
		(0.0195)						
Log of input-weighted average wage in CBSAs within 200 miles			0.639*	0.300				
			(0.291)	(0.289)				
Market access (excludes origin) M^1_{ot}				1.187**				1.098**
				(0.114)				(0.114)
Within R^2	0.091	0.091	0.091	0.093	0.106	0.107	0.108	0.109
Number estab-year (rounded)				35000				
Number ind-prod-year (rounded)				5250				
Observations (rounded)				1,800,000				

Standard errors, clustered by origin CBSA × year, in parentheses
** p<0.01, * p<0.05

NOTES: Manufacturing establishments in the CFS and CMF. All shipments are to a domestic destination CBSA distinct from the origin CBSA. All regressions include SCTG5 × NAICS6 × destination × year fixed effects and mode × year fixed effects. Unreported controls in all columns are the interactions of log origin income per capita with the three input variables.

puts purchased from suppliers. To account for "upstream human capital", I construct a city-industry-level measure of the average wage at potential input suppliers, using pay per worker in the Longitudinal Business Database and the input shares reported in the BEA's Benchmark Input-Output data. Since the factor content embedded in inputs may be sourced from other cities, I also calculate such an average wage using all plants in cities within 200 miles to account for input market access.

Table E.4 introduces these input-market-access measures as additional regressors and shows that the prior results are largely unchanged. The first column, included to facilitate comparisons, is identical to the first column of Table 4. The average wage in upstream industries in the city of production, the regressor aimed at controlling for upstream human capital that is introduced in the second column, has little explanatory power. The input-market-access measure, introduced in the third column, is statistically significant as a regressor but does not meaningfully alter the estimated income elasticity of outgoing shipment prices. The fourth column shows that the coefficient on M^1_{ot}, the output-market-access measure, is not sensitive to controlling for this input market access. The fourth through eighth columns demonstrate that these findings hold when flexibly controlling for plant-level factor usage and input market access.

E.5 Non-large plants and single-plant firms

Table E.5: Shipment prices for non-large and single-plant firms

Dep var: Log unit value	(1)	(2)	(3)	(4)	(5)	(6)
Origin CBSA log per capita income	0.372**	0.285**	0.153**	0.356**	0.238**	0.0814
	(0.0553)	(0.0536)	(0.0567)	(0.0726)	(0.0706)	(0.0803)
Origin CBSA log population	-0.00620	-0.0131*	-0.00308	-0.0124	-0.0178*	-0.00548
	(0.00640)	(0.00605)	(0.00616)	(0.00793)	(0.00755)	(0.00832)
Log mileage (ZIP-ZIP-mode-specific)	0.0403**	0.0413**	0.0463**	0.0147**	0.0149**	0.0218**
	(0.00351)	(0.00347)	(0.00337)	(0.00534)	(0.00521)	(0.00525)
Non-production worker share (log)		0.0991**	0.0982**		0.107**	0.106**
		(0.0100)	(0.0101)		(0.0161)	(0.0163)
Assets per worker (log)		-0.0526**	-0.0519**		-0.0488**	-0.0482**
		(0.00576)	(0.00572)		(0.00742)	(0.00730)
Pay per worker (log)		0.260**	0.251**		0.273**	0.263**
		(0.0255)	(0.0256)		(0.0328)	(0.0328)
Non-production worker share × income per capita		0.0267	0.0223		-0.144*	-0.148*
		(0.0400)	(0.0398)		(0.0674)	(0.0673)
Assets per worker × income per capita		-0.0642**	-0.0641**		-0.0927**	-0.0942**
		(0.0202)	(0.0202)		(0.0293)	(0.0288)
Pay per worker × income per capita		0.230*	0.231*		0.473**	0.471**
		(0.0946)	(0.0930)		(0.123)	(0.121)
Market access (excludes origin) M^1_{ot}			0.961**			1.143**
			(0.161)			(0.247)
Within R^2	0.072	0.081	0.083	0.060	0.069	0.071
Number estab-year (rounded)	20,000	20,000	20,000	15,000	15,000	15,000
Number ind-prod-year (rounded)	5000	5000	5000	4500	4500	4500
Observations (rounded)	900,000	900,000	900,000	550,000	550,000	550,000
Sample description	Non-large	Non-large	Non-large	SPF	SPF	SPF

Standard errors, clustered by origin CBSA × year, in parentheses
** p<0.01, * p<0.05

NOTES: Manufacturing establishments in the CFS and CMF. All shipments are to a domestic destination CBSA distinct from the origin CBSA. All regressions include SCTG5 × NAICS6 × destination × year fixed effects and mode × year fixed effects. Columns 1 through 3 restrict the sample to plants below the 75th percentile in plant sales and firms below the 75th percentile in firm sales. Columns 4 through 6 restrict the sample to "single-plant firms" (SPF), which are plants that are the only plant in the Census of Manufactures owned by their parent.

This section attempts to address the concern that the regression results might be driven by shipments from plants whose decisions are poorly described by the model if, for example, large firms impose uniform pricing across plants irrespective of local conditions. Table E.5 restricts the estimation sample to non-large plants and single-plant firms. The first through third columns show that findings for non-large plants owned by non-large firms are similar to those found in the sample as a whole. The fourth through sixth columns show that relative contributions of factor usage and market access are also similar in the subsample of single-plant firms.

References

Anderson, J. E. (2011): "The Gravity Model," *Annual Review of Economics*, 3(1), 133–160.

Beath, J., and Y. Katsoulacos (1991): *The Economic Theory of Product Differentiation*. Cambridge University Press.

Bergstrand, J. H., and P. Egger (2011): "Gravity Equations and Economic Frictions in the World Economy," in *Palgrave Handbook of International Trade*, ed. by D. Bernhofen, R. Falvey, D. Greenaway, and U. Kreickemeier. Palgrave Macmillan.

Bernasconi, C. (2013): "Similarity of income distributions and the extensive and intensive margin of bilateral trade flows," Economics Working Papers 115, University of Zurich.

Berry, S. T. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 25(2), 242–262.

Bombardini, M., G. Gallipoli, and G. Pupato (2012): "Skill Dispersion and Trade Flows," *American Economic Review*, 102(5), 2327–48.

Bureau of Transportation Statistics, and US Census Bureau (2004): "2002 Economic Census: 2002 Commodity Flow Survey," EC02TCF-US, US Department of Transportation and US Department of Commerce, https://1bts.rita.dot.gov/pdc/user/products/src/products.xml?p=1836.

——— (2010): "2007 Economic Census: 2007 Commodity Flow Survey," EC07TCF-US, US Department of Transportation and US Department of Commerce, https://1bts.rita.dot.gov/pdc/user/products/src/products.xml?p=3193.

Burenstam Linder, S. (1961): *An Essay on Trade and Transformation*. John Wiley & Sons, Ltd.

Costinot, A., and J. Vogel (2010): "Matching and Inequality in the World Economy," *Journal of Political Economy*, 118(4), 747–786.

Disdier, A.-C., and K. Head (2008): "The Puzzling Persistence of the Distance Effect on Bilateral Trade," *The Review of Economics and Statistics*, 90(1), 37–48.

Fajgelbaum, P., G. M. Grossman, and E. Helpman (2011): "Income Distribution, Product Quality, and International Trade," *Journal of Political Economy*, 119(4), 721 – 765.

Federal Trade Commission (1981): *Statistical Report: Annual Line of Business Report 1975.*

Feenstra, R. C., and J. Romalis (2014): "International Prices and Endogenous Quality," *The Quarterly Journal of Economics*, 129(2), 477–527.

Grossman, G. M. (2004): "The Distribution of Talent and the Pattern and Consequences of International Trade," *Journal of Political Economy*, 112(1), pp. 209–239.

Grossman, G. M., and G. Maggi (2000): "Diversity and Trade," *American Economic Review*, 90(5), 1255–1275.

Hallak, J. C. (2010): "A Product-Quality View of the Linder Hypothesis," *The Review of Economics and Statistics*, 92(3), 453–466.

Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153–61.

Hillberry, R., and D. Hummels (2008): "Trade responses to geographic frictions: A decomposition using micro-data," *European Economic Review*, 52(3), 527–550.

Khandelwal, A. (2010): "The Long and Short (of) Quality Ladders," *Review of Economic Studies*, 77, 1450–1476.

Lehmann, E. (1955): "Ordered Families of Distributions," *The Annals of Mathematical Statistics*, 26(3), 399–419.

Mayer, T., and S. Zignago (2011): "Notes on CEPII's distances measures: The GeoDist database," Working Papers 2011-25, CEPII.

McFadden, D. (1978): "Modelling the Choice of Residential Location," in *Spatial Interaction Theory and Planning Models*, ed. by A. Karlvist, L. Lundqvist, F. Snickars, and J. Weibull, pp. 75–96. North-Holland.

Pierce, J., and P. Schott (2012): "A Concordance between Ten-Digit U.S. Harmonized System Codes and SIC/NAICS Product Classes and Industries," *Journal of Economic and Social Measurement*, 37(1-2), 61–96.

Ruggles, S., J. T. Alexander, K. Genadek, R. Goeken, M. B. Schroeder, and M. Sobek (2010): "Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]," Minneapolis, MN: Minnesota Population Center.

Silva, J. M. C. S., and S. Tenreyro (2006): "The Log of Gravity," *The Review of Economics and Statistics*, 88(4), 641–658.

Sutton, J. (2012): *Competing in Capabilities: The Globalization Process.* Oxford University Press.