WTO Working Paper ERSD-2014-13 **Date: 16 September 2014**

World Trade Organization

Economic Research and Statistics Division

The Determinants of Quality Specialization

Jonathan I. Dingel

Columbia University
JOB MARKET PAPER

Manuscript date: 18 January, 2014

The Determinants of Quality Specialization*

Jonathan I. Dingel†
Columbia University
JOB MARKET PAPER

January 18, 2014

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 (*JPE* 2011) formalize the Linder (1961) conjecture 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 (Schott, *QJE* 2004). 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 at least as much of the relationship between income and quality as differences in factor usage.

*I am grateful to Don Davis, Amit Khandelwal, Eric Verhoogen, Jonathan Vogel, and David Weinstein for invaluable discussions, guidance, and support. Thanks to Michael Best, Brianna Cardiff, Juan Carlos Hallak, Jessie Handbury, Corinne Low, Kyle Meng, Joan Monras, Eduardo Morales, David Munroe, Paul Piveteau, Bernard Salanie, Daniel Sturm, Allie Tepper, Sebastien Turban, Reed Walker, and Columbia colloquium participants 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 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. All results have been reviewed to ensure that no confidential information is disclosed. Please visit `http://www.columbia.edu/~jid2106/research.html` for the latest draft.

†`jid2106@columbia.edu`

1 Introduction

The Linder hypothesis is the oldest theory of product quality in international trade. 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, Linder conjectured that local consumers' 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, monopolistic-competition 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 export high-quality products.[1] These two 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 high-income locations import high-quality products if preferences are non-homothetic, as indeed they are.[3] Thus, both theories are consistent with the finding that higher-income countries import more from countries exporting products at higher prices (Hallak, 2006). Combining these export and import patterns, both theories predict that countries with more similar incomes trade more intensely with each other, as found by Hallak (2010) and Bernasconi (2013).[4]

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 requires 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 at least as much as factor

[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]Deaton and Muellbauer (1980, p.144) note that homotheticity "contradicts all known household budget studies, not to mention most of the time-series evidence."

[4]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."

abundance.

To guide my empirical investigation, I introduce a theoretical framework that nests the two mechanisms, each of which has been studied separately. Individuals have non-homothetic preferences over a homogeneous and a differentiated good; higher-income individuals are more likely to consume higher-quality varieties of the differentiated good. This demand assumption makes the model consistent with high-income countries importing 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.[5] 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 home-market-effect and factor-abundance hypotheses.

Guided by the model, my empirical investigation yields two main results. First, differences in plants' inputs, which may be induced by either mechanism, explain only a minority of the observed specialization across cities. 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 local quality specialization.

More specifically, in my empirical work I infer quality specialization from two empirical

[5] 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."

measures commonly used in the literature: unit values and estimated 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; 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 estimated demand shifters, but 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.

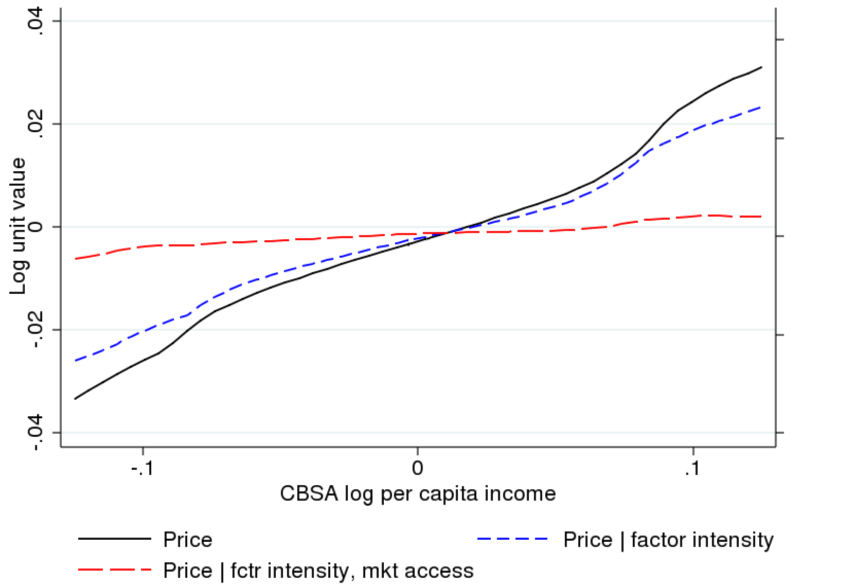
The second empirical finding is that the home-market effect plays a quantitatively significant role in quality specialization, at least as large as differences in factor usage. Using data on cities' incomes and geographic locations, I construct two market-access measures describing the income composition 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 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 than plants' factor inputs. Within-intensity variation in market access explains 54% 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 estimated demand shifters and incomes per capita.[7] I conclude that the home-market effect for quality plays a substantial role in the economic geography of US manufacturing.

Figure 1 summarizes this paper's main results describing within-product price variation.

[6]Using only within-intensity variation is conservative. Unconditionally, variation in market access accounts for 72% of the price-income covariance.

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

Figure 1: Outgoing shipment prices and per capita incomes



NOTES: These are kernel-weighted-local-polynomial versions of columns 4 and 6 of Table 3 and column 6 of Table 4. Log unit value and log per capita income are residuals from regressing each on destination-product-year fixed effects, log shipment mileage, and log origin population. See tables for details.

The black line shows the strong positive relationship between outgoing shipment prices and per capita incomes. This is the common prediction of the two competing hypotheses. The blue line plots this relationship after controlling for price variation attributable to plants' inputs. The modest gap between the black and blue lines illustrates the modest explanatory power of observed factor-usage differences. The red line shows the price-income relationship after also controlling for the income composition of potential customers near the plant. Consistent with a strong home-market effect for quality, this market-access measure explains most of the price variation. The covariance of plants' outgoing shipment prices and cities' per capita incomes reflects factor-intensity differences (27%), within-intensity market-access differences (54%), and residual variation (19%).

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 very different predictions, for instance, for poor countries that have rich neighbors.[8] To the extent that specializing in producing high-quality goods improves a country's growth prospects, the

[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.

strong home-market effect found here suggests an advantage of proximity to high-income countries.[9] And since trade policy can affect 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, 2013) or described the products available to retail consumers without tracking production locations (Handbury and Weinstein, 2011). 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; Diamond, 2012; Handbury, 2012).

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

Linder (1961) started from the proposition that home demand is essential to developing an exportable product. Since higher-income consumers tend to purchase higher-quality products, Linder suggested that demand composition causes higher-income locations to produce higher-quality products. A novel implication was that countries with more similar incomes would trade more intensely with each other. Despite its informal theoretical underpinnings, this trade-flow prediction motivated many empirical investigations (Deardorff, 1984).

Krugman (1980) formalized how economies of scale and trade costs generate a "home-market effect" in which the country with a larger home market for a product is the net exporter of that good. Demand differences determine trade because economies of scale and costly transport mean that a larger home market is a competitive advantage. 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. The Krugman (1980) model featured homothetic preferences and two products produced by different industries,

[9] See Redding (1996), Aghion, Blundell, Griffith, Howitt, and Prantl (2009), 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."

omitting the roles of non-homothetic preferences and product quality emphasized by Linder.

Fajgelbaum, Grossman, and Helpman (2011) recently formalized how income differences can determine the pattern of quality specialization and trade in a general-equilibrium, monopolistic-competition model. They describe a world economy without traditional supply-side determinants of the pattern of trade. The composition of income in a location determines the composition of demand, since higher-income households are more likely to purchase a higher-quality variety. Plants produce higher qualities in higher-income locations because it is more profitable to produce in the larger home market. These mechanics are consonant with the story suggested by Linder (1961).[11] In equilibrium, high-income locations disproportionately produce, export, and import high-quality products.

The canonical factor-abundance theory of comparative advantage yields 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.[13] In these theories, high-income countries both demand higher-quality products and have Ricardian or factor-abundance-driven comparative advantage in producing them, so that “tastes and capabilities are correlated” but not causally linked (Murphy and Shleifer, 1997, p. 6).

Thus, both theories are consistent with the growing body of empirical evidence suggesting that high-income countries export and import high-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 income per capita using data from more than 50 importing countries.[14] Khandelwal (2010) estimates demand shifters using US import

[11]Linder's informal narrative focused on the role of entrepreneurial discovery in bringing products to market. He emphasized the informational costs of distance more than transportation costs and did not explicitly address economics of scale (Linder, 1961, p.89-90).

[12]Markusen (1986, p. 1003) obtains “a Linder-type trading pattern based on a Linder-type demand assumption” with no trade costs and thus no home-market effect. 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]Flam and Helpman (1987) focus on a setting in which the high-wage country, which demands higher qualities, has Ricardian comparative advantage in producing higher-quality varieties. Using aggregate trade flows, Fieler (2011) estimates a two-sector version of the Eaton and Kortum (2002) Ricardian model. She finds that the industry with a greater income elasticity has greater dispersion in idiosyncratic productivities, causing higher-TFP countries to have comparative advantage in these luxuries. Examining variation across 56 broad sectors, Caron, Fally, and Markusen (2012) find a positive correlation between industries' income elasticities of demand and skill intensities.

[14]Torstensson (1996) reports similar results using Swedish imports and more aggregated product categories.

data and finds that they are positively related to exporting countries' GDP per capita and capital abundance. Feenstra and Romalis (2012) and Hallak and Schott (2011), using other methods, also report that higher-income countries export products inferred to be higher quality. High-income countries import narrowly defined products at higher prices (Hallak, 2006), and higher moments of the income and import price distributions are similarly related (Choi, Hummels, and Xiang, 2009).

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.[15] As the model 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) describes exporting-induced quality upgrading by demonstrating that the Mexican peso crisis induced initially more productive plants to become exporters, increase their average wages, and raise their capital-labor ratio. Hallak and Sivadasan (2013) show that, conditional on size, exporting firms in Chile, Colombia, India, and the United States are more capital-intensive and pay higher wages. 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.[16]

As a result, there is no prior empirical evidence distinguishing the home-market effect for quality from factor-abundance-determined quality specialization. There is a large literature on the Krugman (1980) home-market effect, in which a larger home market causes specialization in the industry with greater economies of scale.[17] This empirical work has identified the economies-of-scale home-market effect by using observable sectoral characteristics, such as transport costs and elasticities of substitution (Hanson and Xiang, 2004). But these cross-industry sources of variation are unavailable when considering quality specialization within

[15]In addition to looking at country-level capital abundance, Schott (2004) shows that the unit values 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 any two of 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.

[16]These data describe plants in many cities within a single country, but the authors did not exploit cross-city variation.

[17]On identifying the Krugman (1980) home-market effect, see Davis and Weinstein (1999), Davis and Weinstein (2003), and Hanson and Xiang (2004). On the consequences of market access for development, see Redding and Venables (2004a) and World Bank (2009).

products. Moreover, since the composition of income and the composition of human capital are closely related, both across countries and cities, it is empirically difficult to distinguish the home-market effect for quality from factor-abundance theories of comparative advantage using aggregate data.

I proceed to introduce a theoretical framework that incorporates both of these 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 describing an economy in which both the home-market effect and relative factor abundance may influence the pattern of production and exchange. I use a high-dimensional framework with many locations, qualities, and skills.[18] It nests a version of the Fajgelbaum, Grossman, and Helpman (2011) home-market-effect model and a traditional factor-abundance model as special cases. Nesting the two mechanisms within a single framework allows me to analyze each in isolation and their interaction.

The theory delivers two results that are key to the empirical investigation. First, it confirms that quality specialization is overdetermined. Each mechanism alone is sufficient to cause high-income locations to disproportionately produce, export, and import high-quality varieties in equilibrium. Second, the theory identifies an important distinction between the two mechanisms. Conditional on plant-level skill intensity, the correlation between local income and plants' output quality is due solely to the home-market effect. This result is the basis of my empirical approach.

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 $f(\omega, k)$. I take skill distributions as exogenously determined. This is a standard assumption in models of international trade and innocuous for the purpose of distinguishing the roles of the factor-abundance and home-market effect mechanisms.[19] I assume that locations can be ranked by their skill abundance in the likelihood-ratio sense. The skill distribution $f(\omega, k)$ is strictly log-supermodular, so high-k locations are skill-abundant.[20]

[18]The problems at hand necessitate such an approach. 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.

[19]Factor mobility would be relevant in considering counterfactuals, since individuals may migrate across cities in response to economic changes.

[20]My theoretical approach makes extensive use of log-supermodularity as an analytical tool. See Costinot (2009) for an introduction to log-supermodularity in the context of trade theory. In $\mathbb{R}^2$, a function $f(\omega, k)$

3.1 Preferences

Consumer preferences are non-homothetic, so the income distribution influences the composition of demand. As in Fajgelbaum, Grossman, and Helpman (2011), individuals consume a differentiated good and a homogeneous good. Varieties of the differentiated good are indexed by j, and J_q denotes the set of varieties with quality q. For individual h, the utility of consuming z units of the homogeneous good and a unit of variety $j \in J_q$ of the differentiated good is

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

where ϵ_{hj} is the individual's idiosyncratic valuation of the variety.[21] 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 and θ_q governs the strength of idiosyncratic differences among varieties with quality q. This specification yields a nested-logit demand system (McFadden, 1978).

An individual chooses variety j and quantity z to maximize utility. The homogeneous good is the numeraire, and variety j is available at price p_j. A consumer with income y_h therefore chooses the variety j that maximizes $(y_h - p_j)q + \epsilon_{hj}$, where $z = y_h - p_j$ is the amount of the homogeneous good purchased after buying a single unit of differentiated variety j. As Fajgelbaum, Grossman, and Helpman (2011) show, the fraction of individuals with income y who demand variety j of quality q is

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

This demand system has two important properties. First, consumers' incomes are systematically related to the quality of the variety they purchase. The market share of variety j of quality q varies with income according to $\frac{1}{\rho_j(y)}\frac{\partial\rho_j(y)}{\partial y} = q - q_a(y)$, where $q_a(y)$ is the average quality consumed by individuals with income y.[22] The fraction of individuals purchasing

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. Davis and Dingel (2013) provide evidence that cities' skill distributions are consistent with the log-supermodularity assumption.

[21] 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.

[22] $q_a(y) \equiv \sum_{q\in Q} q\frac{\left[\sum_{j'\in J_q}\exp((y-p_{j'})q/\theta_q)\right]^{\theta_q}}{\sum_{q'}\left[\sum_{j'\in J_\omega}\exp((y-p_{j'})\omega/\theta_{q'})\right]^{\theta_{q'}}}$.

variety j rises with income if and only if its quality exceeds $q_a(y)$. Second, the elasticity of demand takes a simple form: holding the terms within summations fixed, $\frac{\partial \rho_j(y)}{\partial p_j}\frac{p_j}{\rho_j(y)} = -\frac{q}{\theta_q}p_j$. This property will cause producers of quality q to charge a constant additive markup, $\frac{\theta_q}{q}$.[23]

3.2 Production

Production involves employing workers of heterogeneous skills, so relative factor supplies may be a source of comparative advantage. Both the homogeneous good and the differentiated good are produced using a constant-elasticity-of-substitution technology. The homogeneous good is freely traded, produced by perfectly competitive firms, and used as the numeraire. Varieties of the differentiated good are produced by monopolistically competitive firms.

3.2.1 Producing the homogeneous good

The freely traded homogeneous good is produced using a continuum of labor inputs, with skill ω available in location k at wage $w(\omega, k)$. Production exhibits constant returns to scale, so the total cost of producing quantity $x(z, k)$ at unit cost $c(z, k)$ is $x(z, k)c(z, k)$.[24] The unit cost resulting from hiring $\ell(\omega)$ units of skill ω per unit of output 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 $x(z, k) > 0$.

3.2.2 Producing varieties of the differentiated good

Firms may enter into the differentiated-good sector by choosing a quality level q that is produced by incurring fixed cost f_q, paid in units of the numeraire. The constant marginal cost of producing units of quality q in location k is

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

[23] I use the nested-logit demand system in part because the constant-additive-markup property makes the model analytically tractable. Only the former 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. A broad class of non-homothetic preferences exhibit this property.

[24] In a slight abuse of notation, I use z to index the homogeneous good. Recall that in the utility function z denotes the quantity of this good.

The resulting 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 numeraire. If the income distribution in location k' is $g(y,k')$, then demand for variety j of quality q by consumers in location k' is

$$d_{jk'}=N_{k'}\int\rho_j(y)g(y,k')dy=N_{k'}\mathbb{E}_{k'}\left[\frac{\exp(-p_{jk'}q/\theta_q)}{\sum_{j'\in J_q}\exp(-p_{j'k'}q/\theta_q)}\frac{\left[\sum_{j'\in J_q}\exp((y-p_{j'k'})q/\theta_q)\right]^{\theta_q}}{\sum_{q'}\left[\sum_{j'\in J_{q'}}\exp((y-p_{j'})q'/\theta_{q'})\right]^{\theta_{q'}}}\right],$$

where $\mathbb{E}_{k'}$ is the expectations operator with respect to the income distribution in location k'. Taking competitors' behavior as given, the optimal prices charged by firm j producing quality q in location k are given by maximizing profits:

$$\begin{aligned}\max_{\{p_{jk'}\}}\pi_j=&\sum_{k'}d_{jk'}(p_{jk'}-c(q,k)-\tau_{qkk'})-f_q\\ \Rightarrow p_{jk'}=&c(q,k)+\tau_{qkk'}-\frac{d_{jk'}}{\frac{\partial d_{jk'}}{\partial p_{jk'}}}=c(q,k)+\tau_{qkk'}+\frac{\theta_q}{q}\end{aligned}$$

I now define the equilibrium demand level for variety j of quality q in location k'. Denote the number of firms producing varieties of quality q in location k by $n_{q,k}$. Plugging in optimal prices, demand $d_{jk'}$ can be written in terms of (vectors of) the number of firms ($\mathbf{n}$), the unit costs in each location ($\mathbf{c}$), and trade costs (τ).

$$\begin{aligned}d_{jk'}=&N_{k'}\exp(-(c(q,k)+\tau_{qkk'})q/\theta_q)\mathbb{E}_{k'}\left[\frac{\exp(yq)\left[\sum_{k''}n_{q,k''}\exp((-c(q,k'')+\tau_{qk''k'})q/\theta_q)\right]^{\theta_q-1}}{\sum_{q'}\exp(yq')\left[\sum_{k''}n_{q',k''}\exp((-c(q,k'')+\tau_{q'k''k'})q'/\theta_{q'})\right]^{\theta_{q'}}}\right]\\ =&N_{k'}\exp(-(c(q,k)+\tau_{qkk'})q/\theta_q)\Gamma_{k'}(q,\mathbf{n},\mathbf{c},\tau)\end{aligned}$$

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.[25] 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 $\tau_{qkk'}$.

[25] Fajgelbaum, Grossman, and Helpman (2011) introduce a similar demand measure 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-price equalization.

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

$$f(\omega, k) = x(z, k)\ell(\omega, z, k) + \int_{q\in Q}\int_{j\in J_q} x(q, k)\ell(\omega, q, k)djdq.$$

Plugging in firms' labor demands and defining $n_{z,k} = 1$, we can write this as

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

where the variable of integration r includes both the homogeneous good and qualities of the differentiated good. The local income distribution density, which depends on equilibrium wages $w(\omega, k)$, is $g(y, k) = \int_{\omega\in\Omega:w(\omega,k)=y} f(\omega, k)d\omega$.

The free-entry condition says that active firms earn zero profits: $\pi_{q,k} \leq 0\ \forall k$ and $n_{q,k} > 0 \Rightarrow \pi_{q,k} = 0$, where

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

Note that zero-profit condition for the homogeneous, numeraire good is $c(z, k) \geq 1\ \forall k$ and $x(z, k) > 0 \Rightarrow c(z, k) = 1$.

3.4 Equilibrium pattern of specialization and trade

Given the distribution of skills across locations, individuals' preferences, and the production technology, the pattern of production and trade in equilibrium is determined by two forces, trade costs and skill intensities. Trade costs, $\tau_{qkk'}$, are important because they shape the pattern of market access and therefore the home-market effect. Skill intensities, governed by $b(\omega, r)$, are important because the relative abundance of skills varies across locations.

I consider two cases of trade costs and two cases of skill intensities. The two potential trade cost matrices are costless trade, in which $\tau_{qkk'} = 0\ \forall q\ \forall k\ \forall k'$, 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 skill intensities that are uniform across products, $b(\omega, r) = b_1(\omega)b_2(r)$, and skill intensities that are increasing in quality, $b(\omega, r)$ weakly log-supermodular.[26]

The resulting four classes of equilibria are summarized in Figure 2. I proceed to analyze each in turn. Section 3.4.1 describes equilibrium when neither mechanism is active. The resulting pattern of production is indeterminate. Section 3.4.2 characterizes equilibrium when only the factor-abundance mechanism is active, while Section 3.4.3 does likewise for the home-market effect. In both cases, high-k locations produce high-q varieties. Together, these two sections demonstrate that each mechanism alone is sufficient to cause high-income locations to produce, export, and import high-quality products. Thus, the existing empirical evidence documenting such patterns does not distinguish between the two mechanisms. Section 3.4.4 describes equilibrium when both mechanisms are active and shows how to identify the home-market effect for quality after conditioning on plants' skill intensities.

Figure 2: Equilibrium pattern of production

	Uniform skill intensities	Quality is skill-intensive
No trade costs	Indeterminate pattern of production	Factor-abundance specialization
Positive trade costs	Home-market-effect specialization	Factor-abundance mechanism + home-market effect

In the following analysis, it is useful to be able to refer to a product's skill intensity. To that end, define a skill-intensity index $i(r)$ with the properties that $i(r) = i(r') \iff b(\omega, r) = h(r, r')b(\omega, r')\ \forall \omega$ for some function $h(r, r')$ and $i(r) > i(r') \Rightarrow r > r'$.[27] 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. In essence, the intensity index $i(r)$ groups together products so that all products in the higher-i group use relatively more skilled labor for any wage schedule.

3.4.1 Uniform skill intensities and costless trade

When trade is costless, the zero-profit condition (3) reduces to

[26] 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.

[27] Therefore $i(r) = i(r') \Rightarrow c(r, k) = h(r, r')^{\frac{\sigma}{1-\sigma}} c(r', k)$ and $i(r) = i(r') \Rightarrow \ell(r, \omega, k) = h(r, r')^{\frac{\sigma}{1-\sigma}} \ell(r', \omega, k)$.

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

In the absence of trade costs, the structure of demand across destinations k' is orthogonal to the location of production. The profits from producing quality q, $\pi_{q,k}$, are highest wherever the unit cost $c(q,k)$ is lowest.

When trade is costless and skill intensities are uniform, unit costs are equal across locations, $c(r,k) = c(r)\ \forall k$.[28] Thus, production of any variety is equally profitable across all locations in equilibrium. To reiterate, the labor-market clearing condition (2) becomes

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

Since nothing inside the integral depends on ω, the composition of local production $n_{r,k}x(r,k)$ is independent of local factor abundance $f(\omega,k)$ in equilibrium.

Result. When trade is costless and skill intensities are uniform, the pattern of production is indeterminate.

3.4.2 Skill-intensive quality and costless trade

Now consider the case when trade is costless and $b(\omega,q)$ is (weakly) log-supermodular. As before, costless trade makes the structure of demand across destinations orthogonal to a location's profitability. The most-profitable location is where production costs are lowest.

When $b(\omega,q)$ is log-supermodular, skill abundance imposes structure on the pattern of production through the labor-market clearing condition. In particular, equation (2) and the strict log-supermodularity of $f(\omega,k)$ imply, for $k > k'$ and $\omega > \omega'$,

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

where $\phi(r,\omega',k) \equiv \frac{n_{r,k}x(r,k)b(\omega',r)^{\sigma}c(r,k)^{\sigma}}{\int_{r\in z\cup Q} n_{r,k}x(r,k)b(\omega',r)^{\sigma}c(r,k)^{\sigma}dr}$ is a density.[29] $\phi(r,\omega',k)$ 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 the density $\phi_i(\omega',k) \equiv \int_{r:i=i(r)} \phi(r,\omega',k)dr$, which describes the output share of products with skill intensity i in location k, and the expectation

[28] Note that when $b(\omega,r) = b_1(\omega)b_2(r)$ unit costs are $c(r,k) = b_2(r)^{\frac{\sigma}{1-\sigma}} \left(\int_{\omega\in\Omega} b_1(\omega)^{\sigma} w(\omega,k)^{1-\sigma} d\omega\right)^{\frac{1}{1-\sigma}}$. If $c(r,k) < c(r,k')$, then $c(r',k) < c(r',k')\ \forall r' \in z\cup Q$. Wages will be bid down until factors are employed.

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

operator $\mathbb{E}_{\omega',k}$ with respect to this density, which is $\mathbb{E}_{\omega',k}[\alpha(i)] \equiv \int_i \alpha(i)\phi_i(\omega',k)di$. 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

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

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.

This labor-market clearing condition implies that more skill-abundant (higher-k) locations produce more skill-intensive (higher-i) products.[30] Since k is skill-abundant relative to k', 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) or skilled labor in k is relatively cheaper ($\frac{w(\omega,k)^{-\sigma}}{w(\omega',k)^{-\sigma}}$ is greater). However, if skilled labor is relatively cheaper in k, then skill-intensive products' unit costs are relatively lower in k, and thus products made in k must be more skill-intensive in equilibrium. Higher-quality varieties are more skill-intensive when $b(\omega,q)$ is (weakly) log-supermodular, so we interpret inequality (4) as saying that k absorbs its greater relative supply of higher skills by producing higher-quality varieties.[31] This result describes the factor-abundance mechanism for quality specialization.

It is important to note that specialization across qualities of the same skill intensity is indeterminate in this case. Inequality (4) depends only on i and imposes no restrictions on $\phi(r,\omega,k)$ conditional on $\phi_i(\omega,k)$. Skill-abundant locations produce higher-quality varieties only because such products are more skill-intensive.[32]

What about the equilibrium pattern of demand? Since trade is costless, varieties' prices do not vary across locations, and therefore the fraction of consumers of a given income level purchasing a variety, $\rho_j(y)$ does not vary across locations. Denoting the equilibrium varieties and factor prices by the vectors $\bar{\mathbf{n}}$ and $\bar{\mathbf{c}}$, demand levels $\Gamma_k(q,\bar{\mathbf{n}},\bar{\mathbf{c}},\mathbf{0})$ 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.

[30] See appendix section A for a more formal derivation of this paragraph's explanation.

[31] This interpretation neglects the skill intensity of the homogeneous good. If the homogeneous good is more skill-intensive, some skill-abundant locations may produce more skill-intensive output by producing 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.

[32] 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).

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.

The demand levels $\Gamma_k(q, \bar{\mathbf{n}}, \bar{\mathbf{c}}, \mathbf{0})$ will also be important when we consider the interaction of the skill-abundance mechanism and the home-market effect. When trade is costly, we will consider the limiting equilibrium that approaches the skill-intensive-quality, costless-trade equilibrium. But first I describe the case in which the skill-abundance mechanism is absent.

3.4.3 Uniform skill intensities and costly trade

When skill intensities are uniform, unit costs are multiplicatively separable in (r, k) and can be written as $c(r, k) = b_2(r)^{\frac{\sigma}{1-\sigma}} c(k)$.[33] The labor-market clearing condition (2) becomes

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

Since nothing inside the integral 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 relationship 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 Fajgelbaum, Grossman, and Helpman (2011) to determining the pattern of specialization when trade costs are small. With uniform skill intensities, the zero-profit condition is

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

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 ($\tau_{qkk'} = \tau_q \ \forall k' \neq k$), profits vary 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 $N_k \Gamma_k(q, \mathbf{n}, \mathbf{c}, \mathbf{0})$.[34] 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), as shown in Lemma 1 in appendix section A. High-k locations are high-income locations because they are skill-abundant, and this causes high-k locations to have relatively greater demand for high-q

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

[34] This equilibrium is distinct from that of the previous subsection.

varieties. This makes producing high-q varieties more profitable in high-k locations. When population sizes are equal and locations specialize, Proposition 6 of Fajgelbaum, Grossman, and Helpman (2011) describes the resulting pattern: if location k produces quality q and location $k' < k$ produces quality q', then $q' < q$. Similarly, since higher-k locations have higher relative demand for higher-q varieties, their imports are higher-quality (see Proposition 7 of Fajgelbaum, Grossman, and Helpman 2011). In the case of $\sigma = \infty$ and $N_k = 1 \; \forall k$, the model under consideration reduces exactly to the model described in section VII of Fajgelbaum, Grossman, and Helpman (2011).

When factor prices do not equalize, the location with the lowest $c(k)$ is the most attractive cost-wise for all producers. Producers are willing to locate in higher-cost locations to the extent that 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 $N_k \Gamma_k(q, \mathbf{n}, \mathbf{c}, \tau) > N_{k'} \Gamma_{k'}(q, \mathbf{n}, \mathbf{c}, \tau)$. Qualities are produced where they are in greater demand. 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. As a result, higher-income locations specialize in higher-quality varieties.

Result. When population sizes are equal, skill intensities are uniform, and trade costs are uniform and small, higher-income locations produce, export, and import higher-quality varieties because demand for such qualities is greater in such locations.

Thus, the home-market effect yields equilibrium patterns of production and trade 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. Each mechanism alone is sufficient to generate the observed patterns.

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

3.4.4 Skill-intensive quality and costly trade

This section analyzes what happens when 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 is that demand alone determines specialization across varieties of the same skill intensity. This result underlies the empirical investigation.

First, consider the labor-market-clearing inequality, which is governed by the factor-abundance mechanism. Following section 3.4.2, the inequality is

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

As shown previously, this requires that $\phi_i(\omega',k)$ place more weight on higher-i varieties in higher-k locations. 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

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

where the costless-trade-equilibrium demand levels $\Gamma_k(q,\bar{\mathbf{n}},\bar{\mathbf{c}},\mathbf{0})$ were found in section 3.4.2. When trade costs are small, the profits from producing a variety of quality q in location k are approximately

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

When trade costs are small, profits are not sensitive to the locational decisions of other firms. 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', $c(q,k) < c(q,k') \iff c(q',k) < c(q',k')$. Thus, if varieties of the same skill intensity are produced in different locations, these differences 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* $M_{q',k} \leq M_{q',k'}$*.*

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 $n_{q',k'} > 0$. Hence $M_{q,k} \geq M_{q,k'}$ or $M_{q',k} \leq M_{q',k'}$. □

Proposition 1 establishes that market access alone governs specialization within qualities of the same skill intensity. An important component of $M_{q,k}(\tau)$ is demand in the location of production, $N_k\Gamma_k(q, \bar{\mathbf{n}}, \bar{\mathbf{c}}, \mathbf{0})$.[35] This is the home-market effect explanation for why high-income locations specialize in high-quality products.

Thus, we have an empirical strategy for distinguishing the two mechanisms. When quality is skill-intensive and trade is costly, both the factor-abundance mechanism and the home-market effect cause high-k locations to specialize in producing high-q varieties. Variation across skill intensities is overdetermined with respect to the two mechanisms. 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.[36] My empirical strategy is to relate the pattern of specialization across locations to variation in market access after controlling for plants' factor usage.

3.5 Taking the theory to plant-level data

The predictions above describe relationships between product quality (q), location (k), skill intensity (i), and market access ($M_{q,k}(\tau)$). These objects can be inferred from observables in the data using the model and some auxiliary assumptions. The results summarized here are derived in appendix section A. Denote the plant index j, so that, for example, plant j's skill intensity is $i(j)$.

I infer product quality from shipments' prices. Assume that the schedule of technological coefficients $b(\omega, q)$ is strictly decreasing in q, so that higher-quality varieties are more costly to produce. Since free-on-board prices are $p(q, k) = c(q, k) + \frac{\theta_q}{q}$, if $c(q, k)$ increases in quality faster than θ_q/q declines in quality, the price of a variety is informative about its quality. I validate this approach in appendix section E.1 by using estimated 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. Denote the share of non-production workers employed in a plant with skill intensity i in location k by $share_N(i, k)$. When factor prices equalize, the share of non-production workers

[35] In a many-country world, the "home-market effect" involves an appropriately defined market area, not merely the "home country," as noted at least since Linder (1961, p. 87). When trade costs are uniform, as in Fajgelbaum, Grossman, and Helpman (2011), differences in $M_{q,k}(\tau)$ are due solely to differences in demand in the location of production, $M_{q,k}(\tau) > M_{q,k'}(\tau) \iff N_k\Gamma_k(q, \bar{\mathbf{n}}, \bar{\mathbf{c}}, \mathbf{0}) > N_{k'}\Gamma_{k'}(q, \bar{\mathbf{n}}, \bar{\mathbf{c}}, \mathbf{0})$.

[36] 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.

reveals a plant's skill intensity, $share_N(i,k) > share_N(i',k') \iff i > i'$, so $share_N(j)$ is a sufficient establishment-level control for skill intensity. When labor is cheaper where it is abundant, plants of all intensities use more skilled workers in skill-abundant locations, and $share_N(i,k)$ is increasing in k. I therefore also use $share_N(j) \times \ln \bar{y}_k$ to control for skill intensity.

The wage measures assume that wages are increasing in skill.[37] If wages are increasing in skill, we can infer the skill intensity i of a producer in location k from its average wage, $\bar{w}(i,k)$, average non-production wage, $\bar{w}_N(i,k)$, or average production wage, $\bar{w}_P(i,k)$. These average wages are all increasing in i. When factor prices equalize, ranking plants by their average wages is equivalent to ranking them by their factor intensities, $\bar{w}_j > \bar{w}_{j'} \iff i(j) > i(j')$. When labor is cheaper where it is abundant, $\bar{w}(i,k)$ is increasing in i, increasing in k, and log-supermodular. I therefore use $\ln \bar{w}_j$ and $\ln \bar{w}_j \times \ln \bar{y}_k$ as controls for skill intensity.

To assess the role of the geography of demand in specialization across US cities, I construct empirical counterparts to the model's market-access term $M_{q,k}(\tau)$ in profit expression (5). My market-access measures are weighted averages of potential customers' per capita incomes, in which the weights reflect potential customers' population size and distance from the location of production.[38] Describing the composition of demand using per capita incomes exploits the fact that this is a sufficient statistic for relative demand for qualities under the model's assumptions. 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 between origin o and destination d by $miles_{od}$. The first measure describes the composition of potential customers not residing in the location of production, so it cannot be contaminated by any supply-side mechanism linked to per capita income in the production location. This measure for a plant producing in origin city o is $M^1_{ot} = \sum_{d \neq o} \frac{N_{dt} miles_{od}^{-\eta}}{\sum_{d' \neq o} N_{d't} miles_{od'}^{-\eta}} \ln \bar{y}_{dt}$. I use it to qualitatively establish the relationship between market access and specialization. The second market-access measure includes all potential customers, consistent with the theoretical model of the home-market effect. It is $M^2_{ot} = \sum_{d} \frac{N_{dt} miles_{od}^{-\eta}}{\sum_{d'} N_{d't} miles_{od'}^{-\eta}} \ln \bar{y}_{dt}$. I use this measure to quantify the role of market access in the pattern of within-product specialization. In constructing each measure, I use $\eta = 1$, consistent with the international trade literature and the values estimated in appendix section C, and the values of N_{dt}, $miles_{od}$, and $\bar{y}_{dt}$ in the data.

[37] 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.

[38] This market-access measure is the "home-market effect for quality" analogue to the measure constructed by Redding and Venables (2004b) to test the traditional Krugman (1980) home-market effect.

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 section 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 the characteristics of cities and sectors.

4.1.1 Manufacturing microdata

The two plant-level data sources used in this study are the Commodity Flow Survey and the Census of Manufactures. These sources are components of the quinquennial Economic Census; I use confidential microdata from the 1997, 2002, and 2007 editions.[39]

The Commodity Flow Survey (CFS) describes commodity shipments by business establishments in terms of their value, weight, destination ZIP code, transportation mode, and other characteristics.[40] Products are described using the Standard Classification of Transport Goods, a distinct scheme that at its highest level of detail, five digits, defines 512 product categories. Each quarter of the survey year, plants report a randomly selected sample of 20-40 of their shipments in one week.

The Census of Manufactures (CMF) describes a plant's location, industry, inputs and revenues. This census covers the universe of manufacturing plants, which are classified

[39]The Commodity Flow Survey began in 1993. Atalay, Hortaçsu, and Syverson (2013) use 1993 and 1997 CFS microdata to study vertical integration, Hillberry and Hummels (2003, 2008) use 1997 CFS microdata to study how geographic frictions affect trade volumes, and Holmes and Stevens (2010, 2012) use 1997 CFS microdata to study plant size, geography, and trade. The Census of Manufactures has been used in numerous studies.

[40]The US 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. This paper only analyzes shipments by manufacturing establishments, which I refer to as "plants."

into 473 6-digit NAICS manufacturing industries.[41] The CMF describes establishments' employment of production and non-production workers, production worker hours, production and non-production wages and salaries, book values of equipment and structures, and cost of materials.

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 in cases in which the NAICS industry scheme is more detailed than the SCTG commodity scheme. For example, one product is "footwear" (SCTG 30400) produced by an establishment in "men's footwear (except athletic) manufacturing" (NAICS 316213), which is distinct from "footwear" produced by an establishment in "women's footwear (except athletic) manufacturing" (NAICS 316214). My results are robust to ignoring the NAICS information and using only the SCTG commodity codes to define products.

4.1.2 Geographic data

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.[42] I refer to these geographic units as cities. Appendix section 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.[43] 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 measure are very similar.[44]

[41]Information on small establishments is estimated from administrative records rather than reported by the establishment. I exclude these administrative records and imputed observations. See the data appendix B for details.

[42]More than 93% of the US population lived within a CBSA in 2007.

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

[44]I also obtain similar results when excluding only a plant's own employees and payroll from the per-capita-income calculation.

4.2 Pattern of specialization and trade

This section describes variation in manufacturing shipment prices across US cities.[45] 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. 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. 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 three 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, 2012). I obtain similar results in section E.1, where I find that estimated demand shifters are positively correlated with unit values. Moreover, these estimated 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.[46]

The first feature of the US data matching international findings is that outgoing shipments' prices are systematically higher when originating from higher-income cities. 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

[45] 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.

[46] 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, 2013). This would be a concern if the observed price variation were primarily within-plant. Table 2 below shows that this is not the case.

transport mode m in year t of the form

$$\ln p_{skjodmt} = \beta_1 \ln \bar{y}_{ot} + \beta_2 \ln N_{ot} + \alpha_0 \ln miles_{skjodmt} + \gamma_{mt} + \gamma_{kdt} + \epsilon_{skjodt},$$

where $p_{skjodmt}$ is the shipment's unit value, $miles_{skjodt}$ is the ZIP-to-ZIP mode-specific mileage distance of the shipment, $\bar{y}_{ot}$ and N_{ot} are per capita income and total population in the origin CBSA, γ_{mt} are mode-year fixed effects, and γ_{kdt} are product-destination-year fixed effects.

Table 1: Outgoing shipment prices

Dep var: Log unit value, $\ln p_{skjodmt}$	(1)	(2)	(3)
Log origin CBSA income per capita, $\ln \bar{y}_{ot}$	0.458**	0.448**	0.486**
	(0.0427)	(0.0609)	(0.0607)
Log origin CBSA population, $\ln N_{ot}$	-0.0100*	-0.00519	-0.00737
	(0.00479)	(0.00659)	(0.00657)
Log mileage, $\ln miles_{skjodt}$	0.0400**	0.0521**	0.0496**
	(0.00316)	(0.00406)	(0.00418)
Log orig inc × *differentiation*		0.158**	0.216*
		(0.0555)	(0.0892)
Log orig pop × *differentiation*		-0.00770	-0.0223*
		(0.00575)	(0.0104)
Log mileage × *differentiation*		-0.0109*	-0.0144*
		(0.00446)	(0.00696)
Differentiation measure		Sutton	Khandelwal
R-squared	0.878	0.879	0.879
Obs (rounded)	1,400,000	600,000	600,000
Estab-year (rounded)	30,000	15,000	15,000
Ind-prod-year (rounded)	2,000	1,000	1,000

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 to a domestic destination CBSA distinct from the origin CBSA. All regressions include mode × year fixed effects and SCTG5×NAICS6×destination CBSA×year fixed effects.

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 46%. Higher-income cities specialize in the production of higher-price varieties of products, and this pattern of specialization is quite strong. 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, industries' R&D and advertising intensities, infers the scope for quality differentiation from 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 local quality specialization in which higher-income cities specialize in higher-quality products. In products in which there is greater scope for quality differentiation, income differences generate greater differences in the composition of output.

The second feature of the US data matching international findings is that incoming shipments' prices are systematically higher when destined for higher-income cities. To characterize how shipment prices vary with destination 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

$$\ln p_{skjodmt} = \alpha_1 \ln \bar{y}_{dt} + \alpha_2 \ln N_{dt} + \alpha_0 \ln miles_{skjodt} + \gamma_{kt} + \gamma_{mt} + \theta_{ot} + \theta_{kjt} + \epsilon_{skjodt},$$

where $p_{skjodmt}$ is the shipment's unit value, $miles_{skjodt}$ 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 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 26%. 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 contain higher-price varieties.[47]

[47] There are at least four possible explanations for the positive correlation between shipment distances and free-on-board prices. First, shipping costs may shift relative demand toward higher-price varieties through the Alchian-Allen effect. Second, non-shipping costs of distance may be included in the fob price. Third, survey respondents may fail to exclude shipping costs from their reported prices. Fourth, plants may charge higher mark-ups when serving more distant/remote locations. The third column of Table 2 suggests that this last hypothesis could explain at most one-third of such variation, since the within-establishment mileage

Table 2: Incoming shipment prices

Dep var: Log unit value, $\ln p_{skjodmt}$	(1)	(2)	(3)
Log destination CBSA income per capita, $\ln \bar{y}_{dt}$	0.256**	0.165**	0.0481**
	(0.0247)	(0.0191)	(0.00715)
Log destination CBSA population, $\ln N_{dt}$	-0.00450	-0.00336	0.00123
	(0.00266)	(0.00201)	(0.000831)
Log mileage, $\ln miles_{skjodt}$	0.0437**	0.0466**	0.0141**
	(0.00344)	(0.00221)	(0.00103)
R-squared	0.818	0.830	0.916
SCTG5 × NAICS6 × Year FE	Yes	Yes	
Origin CBSA × Year FE		Yes	
Establishment × SCTG5 × Year FE			Yes
Obs (rounded)		1,400,000	
Estab-year (rounded)		30,000	
Ind-prod-year (rounded)		2,000	

Standard errors, clustered by destination 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 mode × year fixed effects.

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 the composition of cities trading with each other.[48] 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, 4.8%. 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. Similarly, the small within-plant-product distance elasticity is evidence against the mark-up explanation for the positive correlation between bilateral distances and free-on-board prices.

These findings demonstrate that the composition of cities' manufactures demand is

elasticity is 1.4%.

[48]Appendix section C shows that cities with more similar income levels trade more intensely with each other.

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.[49]

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 bodies of empirical evidence. First, observed factor-usage differences explain only 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. Second, the market-access measures describing the income composition of proximate potential customers are strongly linked to outgoing shipment prices. A conservative estimate of the home-market effect explains more than half of the price-income relationship.

My regression results can be understood as decomposing the covariance between plant j's outgoing shipment price and origin o's per capita income, $cov(\ln p_{jo}, \ln \bar{y}_o)$. Omitting notation indicating that all these moments are conditional on shipment mileage, origin population size, and destination-product-year fixed effects, we can decompose this covariance into two terms using the law of total covariance:

$$cov(\ln p_{jo}, \ln \bar{y}_o) = \underbrace{cov\left[\mathbb{E}(\ln p_{jo}|i(j)), \mathbb{E}(\ln \bar{y}_o|i(j))\right]}_{\text{across-intensity variation}} + \underbrace{\mathbb{E}_i\left[cov(\ln p_{jo}, \ln \bar{y}_o|i(j))\right]}_{\text{within-intensity variation}}$$

Higher-income origins have higher outgoing shipment prices to the extent that (1) skill-intensive products have higher prices and skill-intensive products are produced in higher-income locations and (2) products of the same skill intensity have higher prices in higher-income locations. The factor-abundance mechanism operates exclusively through the first component, variation across skill intensities. The home-market effect may appear in both components, since higher-income locations have greater demand for higher-quality varieties,

[49]Non-homothetic preferences alone are not sufficient to produce the home-market effect, as discussed in section 2 and shown in section 3.4.2. The home-market effect for quality stems from non-homothetic preferences, economies of scale, and trade costs.

regardless of qualities' skill intensities.

Section 5.1 shows that the across-skill-intensities component is small, constituting 27% of the total variation. To introduce a market-access measure M_o, we can further decompose the within-skill-intensity variation into two terms, yielding

$$\begin{aligned} cov(\ln p_{jo}, \ln \bar{y}_o) = & \underbrace{cov\left[\mathbb{E}(\ln p_{jo}|i(j)), \mathbb{E}(\ln \bar{y}_o|i(j))\right]}_{\text{factor usage} = 27\%} \\ & + \underbrace{\mathbb{E}_i\left[cov(\mathbb{E}(\ln p_{jo}|i(j), M_o), \mathbb{E}(\ln \bar{y}_o|i(j), M_o))\right]}_{\text{within-intensity market access} = 54\%} + \underbrace{\mathbb{E}_i\left[\mathbb{E}(cov(\ln p_{jo}, \ln \bar{y}_o|i(j), M_o))\right]}_{\text{residual covariance} = 19\%}. \end{aligned}$$

The first new term is the share of within-skill-intensity variation attributable to higher-price products being produced where proximate potential customers' per capita incomes are higher and higher-income locations being proximate to higher-income potential customers. The second new term is the share of the covariance explained by neither differences in skill intensity nor the market-access measure. Note that this decomposition is conservative with respect to the home-market effect because it restricts attention to within-skill-intensity variation. In section 5.2, I find that within-intensity variation in market access accounts for 54% of the observed price-income relationship, while the residual constitutes 19%. Hence, home-market demand explains much of the within-product price-income covariance, and its influence is at least twice as large as observed factor-usage differences.

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

$$\begin{aligned}\ln p_{skjodmt} =& \beta_1 \ln \bar{y}_{ot} + \beta_2 \ln N_{ot} + \alpha_0 \ln miles_{skjodt} + \gamma_{kt} + \gamma_{mt} + \gamma_{kdt} \\ &+ \alpha_1 \ln share_{Njt} + \alpha_2 \ln \frac{K_{jt}}{L_{jt}} + \delta_1 \ln share_{Njt} \ln \bar{y}_{ot} + \delta_2 \ln \frac{K_{jt}}{L_{jt}} \ln \bar{y}_{ot} \\ &+ \alpha_3 \ln \bar{w}_{jt} + \alpha_4 \ln \bar{w}_{Njt} + \alpha_5 \ln \bar{w}_{Pjt} \\ &+ \delta_3 \ln \bar{w}_{jt} \ln \bar{y}_{ot} + \delta_4 \ln \bar{w}_{Njt} \ln \bar{y}_{ot} + \delta_5 \ln \bar{w}_{Pjt} \ln \bar{y}_{ot} + \epsilon_{skjodt}\end{aligned} \tag{6}$$

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, $\bar{w}_{jt}$ is average pay per employee, $\bar{w}_{Njt}$ is average pay per non-production worker, and $\bar{w}_{Pjt}$ is average pay per production worker.[50] 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.[51] 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 a battery of plant-level factor-usage controls 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 three wage measures and therefore corresponds to the regression specified in equation (6).

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, log pay per worker, and log pay per non-production 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.[52] The positive coefficient on the interaction

[50] 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.

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

[52] 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: Outgoing shipment prices w/ plant-level factor usage

Dep var: Log unit value, $\ln p_{skjodmt}$	(1)	(2)	(3)	(4)	(5)	(6)
Log origin CBSA income per capita	0.458**	0.421**	0.370**	0.430**	0.381**	0.323**
	(0.0427)	(0.0415)	(0.0397)	(0.0411)	(0.0396)	(0.0381)
Log origin CBSA population	-0.0100*	-0.0139**	-0.0172**	-0.00881	-0.0132**	-0.0161**
	(0.00479)	(0.00464)	(0.00448)	(0.00451)	(0.00437)	(0.00414)
Log mileage	0.0400**	0.0416**	0.0410**			
	(0.00316)	(0.00310)	(0.00302)			
Log non-production worker share		0.139**	0.128**			
		(0.00797)	(0.00937)			
Log assets per worker		-0.0385**	-0.0552**			
		(0.00446)	(0.00470)			
Log non-production worker share		0.0672	0.0398		0.00833	-0.00793
× log per capita income		(0.0355)	(0.0397)		(0.0341)	(0.0390)
Log assets per worker		-0.0194	-0.0579**		0.00127	-0.0246
× log per capita income		(0.0180)	(0.0197)		(0.0195)	(0.0212)
Log pay per worker			0.202**			
			(0.0420)			
Log pay per production worker			-0.0258			
			(0.0330)			
Log pay per non-production worker			0.0489**			
			(0.0146)			
Log pay per worker			0.0106			-0.0887
× log per capita income			(0.120)			(0.122)
Log pay per production worker			0.351**			0.332**
× log per capita income			(0.0930)			(0.0877)
Log pay per non-production worker			0.0915			0.103
× log per capita income			(0.0495)			(0.0555)
R-squared	0.878	0.879	0.879	0.878	0.881	0.882
Obs (rounded)				1,400,000		
Estab-year (rounded)				30,000		
Ind-prod-year (rounded)				2,000		
Note				miles flex	ctrl qnt flex	ctrl all flex

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 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), log pay per worker (6), log pay per production worker (6), and log pay per non-production worker (6).

of log pay per production worker and log origin per capita income is consistent with the model-predicted behavior when factor prices are not fully equalized. However, the observed variation in factor usage explains only a small share of the cross-city variation in outgoing shipment prices. Introducing the quantity measures in the second column reduces the origin-income elasticity from 46% to 42%. Incorporating the wage measures in the third column reduces this elasticity to 37%. Thus, these observed factor-usage differences can explain about one-fifth of the origin-income elasticity of shipment prices.

The fourth through sixth columns of Table 3 incorporate the control variables in more flexible functional forms. Each variable with an α coefficient in estimating equation (6) is entered as a cubic polynomial that varies by 3-digit NAICS industry. Since there are 21 3-digit industries, this introduces 63 regressors for each control variable, yielding a total of 378 regressors.[53] I refrain from reporting the coefficients on these controls.

The results obtained using these more flexible functional forms are quite similar to those in the first three columns of Table 3. The origin-income elasticity of 43% is reduced to 38% by the introduction of the quantity controls and further to 32% 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 modest explanatory power for the pattern of within-product specialization across US cities.

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 34% is quite similar to the 32% obtained in Table 3.

A potential alternative interpretation of these results would be that the plant-level variables are imperfect controls for plant-level factor usage and that city-level income per capita is informative about plant-level factor usage. Suppose that plants' factor inputs exhibit unobserved differences in quality that are correlated with city-level average income, conditional on plant-level observables. It is plausible that plants with observationally equivalent workforces in terms of non-production-to-production-worker ratios may differ 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, 2012). However, these differences between workers should appear in

[53] Using a 3-digit-NAICS-specific translog approximation with the five input measures and a 3-digit-NAICS-specific quadratic in log mileage, for a total of 462 regressors, yields very similar results.

the plant-level wage data, and the third and sixth columns of Table 3 includes plant-level wage measures. The posited unobserved differences in input factor quality would therefore have to be characteristics of worker 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.

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.[54] Though the origin-income elasticity is lower than that found in the CFS data, observed plant-level factor usage explains only about 12% of the total variation.

Finally, one may worry that some other dimension of plant heterogeneity has been omitted. Appendix Table D.3 reports results from CFS data while controlling for plant size. This yields a partial-correlation origin-income elasticity of 30%, similar to that in Table 3.

This section has shown that only 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 after controlling for plant-level factor usage should 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 32%, roughly 3/4 of its value in the absence of any 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.[55]

5.2 The home-market effect for quality

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 home-market effect alone explains twice as much of the covariance between incomes and prices as differences in plants' factor inputs.

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

[55] 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.

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.[56] 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 second, $M^2_{ot} = \sum_{d} \frac{N_{dt} miles_{od}^{-\eta}}{\sum_{d'} N_{d't} miles_{od'}^{-\eta}} \ln \bar{y}_{dt}$, includes all potential customers, consistent with the model.

Table 4 demonstrates that market access plays a significant role in explaining the origin-income elasticity of shipment prices. The first column reports a regression that flexibly controls for the quantity measures of factor usage, the non-production worker share and assets per worker, like the fifth column of Table 3. The second column adds the first market-access measure. This reduces the origin-income elasticity from 38% to 24%, which is a large change. Controlling for factor quantities only reduced the elasticity from 43% to 38%. This implies that the income composition of proximate potential customers *other than those in the city of production* is more than twice as quantitatively important as the capital intensity and the non-production employee share in explaining the pattern of 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. The third column uses the second market-access measure, which includes the income of potential customers in the city of production in the weighted average. This reduces the origin-income elasticity to 13.6%, a reduction of nearly 25 percentage points compared to column 1.

Table 4 reports similar results when controlling for factor usage using the wage measures. The fourth column is a regression that flexibly controls for the quantity and wage measures of factor usage, like the sixth column of Table 3. The fifth column introduces the first market-access measure, reducing the origin-income elasticity from 32.3% to 19.5%. The factor-usage measures reduced this elasticity by 10 percentage points, so the income composition of proximate potential customers other than those in the city of production has considerable explanatory power. The sixth column uses the second market-access measure that includes residents in the city of production. This reduces the origin-income elasticity to 12%, a

[56] Fajgelbaum, Grossman, and Helpman (2011) derive their results in a setting in which 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, i.e. the matrix of bilateral trade costs. Measuring multilateral market access has received considerable attention in empirical assessments of the new economic geography, e.g. Redding and Venables (2004b). See Lugovskyy and Skiba (2012) for a discussion of market access in the context of quality specialization.

Table 4: Outgoing shipments and market access

Dep var: Log unit value	(1)	(2)	(3)	(4)	(5)	(6)
Log origin CBSA income per capita	0.381**	0.240**	0.136**	0.323**	0.195**	0.120*
	(0.0396)	(0.0412)	(0.0495)	(0.0381)	(0.0396)	(0.0481)
Log origin CBSA population	-0.0132**	-0.00233	-0.00814	-0.0161**	-0.00591	-0.0118**
	(0.00437)	(0.00435)	(0.00425)	(0.00414)	(0.00412)	(0.00410)
Market access (excl orig) M^1_{ot}		1.144**			1.075**	
		(0.129)			(0.125)	
Market access M^2_{ot}			1.040**			0.899**
			(0.131)			(0.126)
R-squared	0.881	0.881	0.881	0.882	0.882	0.882
Obs (rounded)			1,400,000			
Estab-year (rounded)			30,000			
Ind-prod-year (rounded)			2,000			
Note	ctrl qnt flex	ctrl qnt flex	ctrl qnt flex	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 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 include 3-digit-NAICS-specific cubic polynomials in log mileage (1-6), log non-production worker share (1-6), log assets per worker (1-6), log pay per worker (4-6), log pay per production worker (4-6), and log pay per non-production worker (4-6). Also unreported are the interactions of log origin income per capita with the five input variables.

reduction of more than 20 percentage points compared to column 4. Thus, market access explains substantially more of the observed relationship between income per capita and outgoing shipment prices than differences in plants' factor usage.[57]

These results can be succinctly summarized as a decomposition of the covariance between incomes and prices.[58] 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 34% of the total covariance, leaving 39% as residual variation. The second market-access measure, which follows the model by including residents in the city of production, accounts for 54% of the total covariance, leaving 19% as residual variation.

This section has established the role of home-market demand 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 lo-

[57]If the demand-side mechanism induces differences in the factor intensities used to produce different qualities, then market access *causally* explains an even greater fraction of the observed correlation. I am conservative in controlling for observed factor usage prior to attributing the change in coefficients to the introduction of the market-access measure.

[58]Comparing the estimated elasticities yields shares very similar to this decomposition. However, that back-of-the-envelope calculation does not account for the changing set of regressors. The law of total covariance is an identity, so the numbers reported in this paragraph are an exact decomposition.

cated 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 at least half of the observed origin-income elasticity of shipment prices. This is consistent with a model in which the home-market effect plays a large role in determining the pattern of quality specialization.

5.3 Additional evidence

This section briefly summarizes a series of results in Appendix E that are consistent with the results reported above.

Appendix section E.1 characterizes the pattern of quality specialization using estimated demand shifters instead of outgoing shipments' unit values as the dependent variable. Due to data constraints, these are only available in 2007. 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 estimated demand shifter is 41%, remarkably similar to the 43% origin-income elasticity of outgoing shipment prices. This covariance between income per capita and estimated 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 weakened role for the income composition of proximate potential customers. Home-market demand plays a substantial role in quality specialization, at least as large as that explained by the factor-abundance mechanism.

Appendix section E.2 uses another moment of the income distribution to illustrate the role of demand in quality specialization. Conditional on average income, cities with higher variance in household income have higher incoming shipment prices. 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 shows that the patterns founds 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. Home-market demand explains a greater share of within-product variation in export prices than differences in factor usage.

6 Conclusions

Two prominent theories predict that high-income locations specialize in producing and exporting high-quality products. The Linder (1961) conjecture, 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 makes 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 income 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. My empirical investigation finds that home-market demand explains at least as much of the quality specialization across US cities as differences in plants' factor inputs.

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

Aghion, P., R. Blundell, R. Griffith, P. Howitt, and S. Prantl (2009): "The Effects of Entry on Incumbent Innovation and Productivity," *The Review of Economics and Statistics*, 91(1), 20–32. 5

Anderson, J. E. (2011): "The Gravity Model," *Annual Review of Economics*, 3(1), 133–160. 52

Atalay, E., A. Hortaçsu, and C. Syverson (2013): "Vertical Integration and Input Flows," mimeo. 21

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

Beath, J., and Y. Katsoulacos (1991): *The Economic Theory of Product Differentiation*. Cambridge University Press. 9

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. 6

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. 52

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. 29

Bernasconi, C. (2013): "Similarity of income distributions and the extensive and intensive margin of bilateral trade flows," Economics Working Papers 115, University of Zurich. 1, 52

Berry, S. T. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 25(2), 242–262. 3, 58

Bombardini, M., G. Gallipoli, and G. Pupato (2012): "Skill Dispersion and Trade Flows," *American Economic Review*, 102(5), 2327–48. 61, 62

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. 49

——— (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. 49, 50, 51, 65

Caron, J., T. Fally, and J. R. Markusen (2012): "Skill Premium and Trade Puzzles: a Solution Linking Production and Preferences," NBER Working Paper 18131. 6

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. 7

Costinot, A. (2009): "An Elementary Theory of Comparative Advantage," *Econometrica*, 77(4), 1165–1192. 8

Costinot, A., and J. Vogel (2010): "Matching and Inequality in the World Economy," *Journal of Political Economy*, 118(4), 747–786. 45

Davis, D. R., and J. I. Dingel (2012): "A Spatial Knowledge Economy," NBER Working Papers 18188, National Bureau of Economic Research, Inc. 31

——— (2013): "The Comparative Advantage of Cities," mimeo. 5, 9

Davis, D. R., and D. E. Weinstein (1999): "Economic geography and regional production structure: An empirical investigation," *European Economic Review*, 43(2), 379–407. 7

——— (2003): "Market access, economic geography and comparative advantage: an empirical test," *Journal of International Economics*, 59(1), 1–23. 7, 19

De la Roca, J., and D. Puga (2012): "Learning by working in big cities," mimeo. 31

Deardorff, A. V. (1984): "Testing trade theories and predicting trade flows," in *Handbook of International Economics*, ed. by R. W. Jones, and P. B. Kenen, vol. 1 of *Handbook of International Economics*, chap. 10, pp. 467–517. Elsevier. 5

Deaton, A., and J. Muellbauer (1980): *Economics and Consumer Behavior*, no. 9780521296762 in Cambridge Books. Cambridge University Press. 1

Diamond, R. (2012): “The Determinants and Welfare Implications of US Workers’ Diverging Location Choices by Skill: 1980-2000,” mimeo. 5

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. 54

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, chap. 2, pp. 23–48. MIT Press. 7

Eaton, J., and S. Kortum (2002): “Technology, Geography, and Trade,” *Econometrica*, 70(5), 1741–1779. 6

Fajgelbaum, P., G. M. Grossman, and E. Helpman (2011): “Income Distribution, Product Quality, and International Trade,” *Journal of Political Economy*, 119(4), 721 – 765. 1, 6, 8, 9, 11, 16, 17, 19, 33, 35, 36, 45, 46, 58, 61

Falvey, R. E. (1981): “Commercial policy and intra-industry trade,” *Journal of International Economics*, 11(4), 495–511. 1, 6

Federal Trade Commission (1981): *Statistical Report: Annual Line of Business Report 1975*. 48

Feenstra, R. C., and J. Romalis (2012): “International Prices and Endogenous Quality,” NBER Working Papers 18314, National Bureau of Economic Research, Inc. 7, 23, 48, 58, 59

Fieler, A. C. (2011): “Nonhomotheticity and Bilateral Trade: Evidence and a Quantitative Explanation,” *Econometrica*, 79(4), 1069–1101. 6

Flam, H., and E. Helpman (1987): “Vertical Product Differentiation and North-South Trade,” *American Economic Review*, 77(5), 810–22. 6

Glaeser, E. L., J. Kolko, and A. Saiz (2001): “Consumer city,” *Journal of Economic Geography*, 1(1), 27–50. 5

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. 62

Grossman, G. M., and G. Maggi (2000): “Diversity and Trade,” *American Economic Review*, 90(5), 1255–1275. 61, 62

Hallak, J. C. (2006): "Product quality and the direction of trade," *Journal of International Economics*, 68(1), 238–265. 1, 3, 7, 23

——— (2010): "A Product-Quality View of the Linder Hypothesis," *The Review of Economics and Statistics*, 92(3), 453–466. 1, 52

Hallak, J. C., and P. K. Schott (2011): "Estimating Cross-Country Differences in Product Quality," *The Quarterly Journal of Economics*, 126(1), 417–474. 7

Hallak, J. C., and J. Sivadasan (2013): "Product and Process Productivity: Implications for quality choice and conditional exporter premia," . 7

Handbury, J. (2012): "Are Poor Cities Cheap for Everyone? Non-Homotheticity and the Cost of Living Across U.S. Cities," Columbia University working paper. 5

Handbury, J., and D. E. Weinstein (2011): "Is New Economic Geography Right? Evidence from Price Data," NBER Working Papers 17067, National Bureau of Economic Research, Inc. 5

Hanson, G. H., and C. Xiang (2004): "The Home-Market Effect and Bilateral Trade Patterns," *American Economic Review*, 94(4), 1108–1129. 7

Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153–61. 53

Helpman, E., and P. R. Krugman (1989): *Trade Policy and Market Structure*. MIT Press. 5

Henderson, J. V. (1991): *Urban Development: Theory, Fact, and Illusion*. Oxford University Press. 5

Hillberry, R., and D. Hummels (2003): "Intranational Home Bias: Some Explanations," *The Review of Economics and Statistics*, 85(4), 1089–1092. 21

——— (2008): "Trade responses to geographic frictions: A decomposition using microdata," *European Economic Review*, 52(3), 527–550. 21, 54

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. 5

——— (2010): "An Alternative Theory of the Plant Size Distribution with an Application to Trade," NBER Working Papers 15957, National Bureau of Economic Research, Inc. 21

——— (2012): "Exports, borders, distance, and plant size," *Journal of International Economics*, 88(1), 91–103. 21

Hummels, D., and P. J. Klenow (2005): "The Variety and Quality of a Nation's Exports," *American Economic Review*, 95(3), 704–723. 1, 3, 6

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. 3

Khandelwal, A. (2010): "The Long and Short (of) Quality Ladders," *Review of Economic Studies*, 77, 1450–1476. 3, 6, 23, 25, 48, 58

Krugman, P. (1980): "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review*, 70(5), 950–59. 5, 7, 20

——— (1991): *Geography and trade*. MIT Press. 2

Lederman, D., and W. F. Maloney (eds.) (2012): *Does What You Export Matter? In Search of Empirical Guidance for Industrial Policies*. World Bank. 5

Lehmann, E. (1955): "Ordered Families of Distributions," *The Annals of Mathematical Statistics*, 26(3), 399–419. 44, 45

Linder, S. B. (1961): *An Essay on Trade and Transformation*. John Wiley & Sons, Ltd. 1, 5, 6, 19, 36, 52

Lugovskyy, V., and A. Skiba (2012): "How Geography Affects Quality," mimeo. 33

Markusen, J. R. (1986): "Explaining the Volume of Trade: An Eclectic Approach," *American Economic Review*, 76(5), 1002–11. 6

Mayer, T., and S. Zignago (2011): "Notes on CEPII's distances measures: The GeoDist database," Working Papers 2011-25, CEPII. 48

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. 9

Murphy, K. M., and A. Shleifer (1997): "Quality and trade," *Journal of Development Economics*, 53(1), 1–15. 6

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. 48

Redding, S. (1996): "The Low-Skill, Low-Quality Trap: Strategic Complementarities between Human Capital and R&D," *Economic Journal*, 106(435), 458–70. 5

Redding, S., and A. Venables (2004a): "Geography and Export Performance: External Market Access and Internal Supply Capacity," in *Challenges to Globalization: Analyzing the Economics*, ed. by R. E. Baldwin, and L. A. Winters, pp. 95–130. 7

Redding, S., and A. J. Venables (2004b): "Economic geography and international inequality," *Journal of International Economics*, 62(1), 53–82. 20, 33

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. 47, 48

Schott, P. K. (2003): "One Size Fits All? Heckscher-Ohlin Specialization in Global Production," *American Economic Review*, 93(3), 686–708. 7

——— (2004): "Across-product Versus Within-product Specialization in International Trade," *The Quarterly Journal of Economics*, 119(2), 646–677. 1, 3, 6, 7, 15, 29

Silva, J. M. C. S., and S. Tenreyro (2006): "The Log of Gravity," *The Review of Economics and Statistics*, 88(4), 641–658. 53

Simonovska, I. (2013): "Income Differences and Prices of Tradables," UC Davis mimeo. 23

Sutton, J. (1991): *Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration*. MIT Press. 3

——— (1998): *Technology and Market Structure: Theory and History*. MIT Press. 25, 48

——— (2012): *Competing in Capabilities: The Globalization Process*. Oxford University Press. 3, 58

Torstensson, J. (1996): "Can factor proportions explain vertical intra-industry trade?," *Applied Economics Letters*, 3(5), 307–309. 6, 29

Verhoogen, E. A. (2008): "Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector," *The Quarterly Journal of Economics*, 123(2), 489–530. 4, 7, 29

World Bank (2009): *World Development Report 2009: Reshaping Economic Geography.* 7

A Theory appendix

Equilibrium pattern of specialization and trade

Skill-intensive quality and costless trade

Inequality (4) says that skill-abundant locations specialize in skill-intensive products. Recall:

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

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, $\phi(r,\omega',k) = \frac{n_{r,k}x(r,k)b(\omega',r)^{\sigma}c(r)^{\sigma}}{\int_{r\in z\cup Q} n_{r,k}x(r,k)b(\omega',r)^{\sigma}c(r)^{\sigma}dr} = \frac{n_{r,k}x(r,k)\ell(\omega',r)}{\int_{r\in z\cup Q} n_{r,k}x(r,k)\ell(\omega',r)dr}$ is a density that 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'.[59] 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 q 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, high-quality varieties.

[59] 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.

Uniform skill intensities and costly trade

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, 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). □

Taking the theory to plant-level data

If $b(\omega, q)$ is strictly decreasing in q, higher-quality varieties are more costly to produce, $c(q, k) > c(q', k) \iff q > q'$.

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

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.[60] Inside the factor-price equalization (FPE) set, $\ell(\omega, q, k) = \ell(\omega, q) \; \forall k$ and there-

[60]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 in-

fore $share_N(i,k) = share_N(i)\ \forall k$. Outside the FPE set, if $w(\omega,k)^{-\sigma}$ is log-supermodular, $share_N(i,k)$ is strictly increasing in k.[61] I therefore use $share_N(j) \times \ln \bar{y}_k$ as an additional control for skill intensity.

The wage measures assume that wages are increasing in skill, ω.[62] 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 $i > i'$.[63] The average wage $\bar{w}(i,k)$ is therefore strictly increasing in skill intensity i. We can similarly 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.

Describing the composition of demand using per capita incomes exploits 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-

creasing in i.

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

[62] 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.

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

supermodular in (q,k). As mentioned previously, when $g(y,k)$ is log-supermodular, income per capita is a sufficient statistic for k.

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.[64] The geodetic distances for export shipments in appendix section E.3 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.

In estimating demand shifters, I mapped the Feenstra and Romalis (2012) 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).

[64] 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.[65]

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.[66]

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

[65]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).

[66]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.1 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 estimate 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, I restrict attention to unimodal shipments, which account for more than 80% of shipments by value and 90% by weight (Bureau of Transportation Statistics and US Census Bureau, 2010, Table 1b).

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. 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.[67] I calculate unit values by

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

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 two 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.

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 establishments not belonging to a OMB-defined CBSA. I exclude a small number of establishments for which the CFS and CMF do not report the same CBSA. I exclude SCTG5-NAICS6 pairs in which fewer than five establishments report shipping a commodity.

CFS: I restrict the sample to unimodal shipments, which constitute more than 80% of shipments by value and 90% by weight (Bureau of Transportation Statistics and US Census Bureau, 2010). 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 exclude shipments that are the unique instance of that commodity being shipped by that establishment.

CMF: 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.[68] I exclude establishments with wages that lie below the 1st percentile or above the 99th percentile of the wage distribution for manufacturing establishments. I exclude CBSA-NAICS6 pairs in which total employment is reported to be more than 10% of residential population.

C Gravity 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.[69] 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 Linder (1961). I find a similar pattern of trade flows between US cities.

The baseline gravity specification is

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

where X_{odst} is the volume of shipments in sector s sent from origin o to destination d in year t, miles_{odt} 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_{ost} and γ_{dst} origin-sector-year and destination-sector-year fixed effects, and ϵ_{odst} is a residual reflecting both random sampling and potential measurement error.[70]

This baseline specification is estimated using observations with strictly positive shipment 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

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

[69] 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.

[70] Observations of X_{odst} 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.

positive amount to itself in every sector $X_{oost} > 0 \ \forall o \forall s \forall t$.

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

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

where $\hat{\rho}_{odst}$ is the predicted probability from the probit regression, ϕ and Φ are the probability and cumulative density functions of the normal distribution, and $\frac{\phi(\Phi^{-1}(\hat{\rho}_{odst}))}{\hat{\rho}_{odst}}$ 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.

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

This estimation approach allows me to include observations for which $X_{odst} = 0$.[71]

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 and fifth columns use the Silva and Tenreyro (2006) Poisson pseudo-maximum likelihood estimator for all observations and strictly positive trade flows, respectively.

In all the relevant 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. 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

[71] Silva and Tenreyro (2006) emphasize that their estimation procedure addresses a concern that higher moments of $\exp(\epsilon_{odst})$ 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)	(5)
	OLS	OLS	Heckman	PPML w/ zeros	PPML w/o zeros
Dep var:	$\ln X_{od}$	$\ln X_{od}$	$\ln X_{od}$	X_{od}	X_{od}
$\ln \text{miles}_{od}$	-0.916**	-0.917**	-1.773**	-0.952**	-0.831**
	(0.00745)	(0.00715)	(0.0104)	(0.0170)	(0.0170)
$\lvert \ln \bar{y}_o - \ln \bar{y}_d \rvert$	-1.696**		-1.074**	-0.430**	-1.241**
	(0.0413)		(0.0454)	(0.0840)	(0.0934)
$(\ln \bar{y}_o - \ln \bar{y}_d)^2$		-2.657**			
		(0.0854)			
Inverse Mills $\frac{\phi(\Phi^{-1}(\hat{\rho}_{od}))}{\hat{\rho}_{od}}$			2.434**		
			(0.0222)		
R-squared	0.347	0.347	0.378		
Obs (rounded)	175,000	175,000	175,000	850,000	175,000
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.

are compatible with the factor-abundance mechanism or home-market effect determining the pattern of within-product specialization.

The distance elasticity of trade, η, is estimated to be near one. Most of the estimates are about 0.9, while the two-step Heckman specification implies that shipment volumes are notably more sensitive to the distance between origin and destination. These results are 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.

D Tables appendix

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

Dep var: Log unit value	(1)	(2)	(3)	(4)	(5)
Log origin CBSA income per capita	0.529**	0.472**	0.392**	0.404**	0.340**
	(0.0546)	(0.0529)	(0.0517)	(0.0587)	(0.0570)
Log origin CBSA population	-0.0217**	-0.0241**	-0.0253**	-0.0209**	-0.0232**
	(0.00672)	(0.00635)	(0.00607)	(0.00731)	(0.00688)
Log non-production worker share		-0.0149	-0.0536	-0.0171	-0.0580
× log per capita income		(0.0450)	(0.0527)	(0.0454)	(0.0529)
Log assets per worker		0.00456	-0.0250	0.00485	-0.0251
× log per capita income		(0.0251)	(0.0295)	(0.0245)	(0.0290)
Log pay per worker			0.0680		0.0933
× log per capita income			(0.146)		(0.146)
Log pay per production worker			0.265*		0.249*
× log per capita income			(0.110)		(0.113)
Log pay per non-production worker			0.0207		0.0184
× log per capita income			(0.0661)		(0.0665)
City-industry mean years schooling				0.0191	-0.000797
× log per capita income				(0.0325)	(0.0332)
R-squared	0.883	0.885	0.886	0.885	0.887
Note	flex miles	ctrl qnt flex	ctrl all flex	ctrl qnt flex + school flex	ctrl all flex + school flex
Obs (rounded)			1,000,000		
Estab-year (rounded)			22,500		
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 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 3-digit-NAICS-speific third-order polynomials in log mileage (columns 1-5), log non-production worker share (2-5), log assets per worker (2-5), log pay per worker (3,5), log pay per producer worker (3,5), log pay per non-production worker (3,5), and city-industry means years of schooling (4-5).

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 hh 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
Obs (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.

Table D.3: Outgoing shipment prices w/ plant size controls

Dep var: Log unit value	(1)	(2)	(3)
Log origin CBSA income per capita	0.412**	0.366**	0.298**
	(0.0415)	(0.0402)	(0.0384)
Log origin CBSA population	-0.00823	-0.0112*	-0.0150**
	(0.00454)	(0.00441)	(0.00413)
Log non-production worker share		0.0253	0.0194
× log per capita income		(0.0345)	(0.0398)
Log assets per worker		0.000542	-0.0246
× log per capita income		(0.0191)	(0.0210)
Log pay per worker			-0.0819
× log per capita income			(0.121)
Log pay per production worker			0.363**
× log per capita income			(0.0876)
Log pay per non-production worker			0.133*
× log per capita income			(0.0555)
R-squared	0.880	0.882	0.883
Note	miles flex + size	ctrl qnt flex + size	ctrl all flex + size
Obs (rounded)		1,400,000	
Estab-year (rounded)		30,000	
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 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 3-digit-NAICS-speific third-order polynomials in log mileage (columns 1-3), log establishment size (1-3), log non-production worker share (2-3), log assets per worker (2-3), log pay per worker (3), log pay per producer worker (3), and log pay per non-production worker (3).

E Supplementary appendix

This appendix reports further empirical evidence consistent with the results presented in section 5. First, estimated demand shifters exhibit the same patterns as outgoing shipment prices. 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, export shipments exhibit patterns consistent with those found in domestic transactions.

E.1 Estimated 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 estimate demand shifters, I use the "non-homothetic CES preferences" of Feenstra and Romalis (2012).[72] 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

[72] 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.

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.[73] 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.[74] I use sectoral estimates of $\hat{\lambda}$ and $\hat{\sigma}$ from Feenstra and Romalis (2012) 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 1000 products defined by SCTG5-NAICS6 codes, for the 2007 sample.[75]

Table E.1 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 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 remarkably similar to the 43% 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

[73] 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.

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

[75] 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 (2012) parameter estimates to these product codes.

Table E.1: 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 orig) 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 orig 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-squared	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
Obs (rounded)						13,000					
Estab (rounded)						10,000					
Ind-prod (rounded)						1,000					
Note								ctr all flex	ctr all flex	ctr all flex	ctr all flex

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 mileage, 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.

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.[76] 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.2 Income dispersion

This section documents the relationship between the second moment of the income distribution and shipment prices. Cities with greater variation in household income exhibit higher prices for both incoming and outgoing shipments. The latter is not explained by variation 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

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

advantage in sectors in which skills are more substitutable.[77] 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 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.[78] 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 (appropriately 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 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 omits any 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 five 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 three columns of Tables 3. The third column introduces

[77] 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.

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

the standard deviation of household income as a regressor alongside origin characteristics and shipment mileage. The fourth column adds controls for plant-level factor usage in quantities and wages. 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 repeat the third and fourth columns for the subset of observations for which 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 sixth column introduces an additional control, city-industry mean years of schooling, that is available for these observations. Among this subsample, the coefficients on the log standard deviation of household income are more than double their values for the full sample. The key result, appearing in the seventh column, 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, $\ln p_{skjodmt}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log destination CBSA income per capita	0.140**	0.0988**					
	(0.0267)	(0.0232)					
Log destination CBSA population	-0.00720**	-0.00806**					
	(0.00248)	(0.00205)					
Log origin CBSA income per capita	0.455**		0.363**	0.259**	0.387**	0.194**	0.195**
	(0.0184)		(0.0492)	(0.0464)	(0.0691)	(0.0724)	(0.0725)
Log origin CBSA population	-0.00936**		-0.0137**	-0.0212**	-0.0321**	-0.0332**	-0.0345**
	(0.00169)		(0.00502)	(0.00468)	(0.00744)	(0.00735)	(0.00774)
Log mileage	0.0421**	0.0461**					
	(0.00288)	(0.00219)					
Log std dev dest hh income	0.118**	0.109**					
	(0.0256)	(0.0224)					
Log std dev orig hh income			0.112*	0.110*	0.230**	0.231**	0.219**
			(0.0516)	(0.0478)	(0.0812)	(0.0792)	(0.0818)
Log std dev orig-ind weekly wage							0.00521
							(0.0143)
Std dev orig-ind years schooling							0.00561
							(0.00998)
R-squared	0.819	0.830	0.878	0.882	0.883	0.887	0.887
Standard errors			cl cbsa x year	cl cbsa x year	cl cbsa x year	cl cbsa x year	cl cbsa x year
Origin CBSA x Year FE		Yes					
SCTG5 x NAICS6 x Year FE	Yes	Yes					
SCTG5 x NAICS6 x dest CBSA x Year FE			Yes	Yes	Yes	Yes	Yes
Obs (rounded)	1,400,000	1,400,000	1,400,000	1,400,000	1,000,000	1,000,000	1,000,000
Estab-year (rounded)	30,000	30,000	30,000	30,000	22,500	22,500	22,500
Ind-prod-year (rounded)	2,000	2,000	2,000	2,000	2,000	2,000	2,000
Note			flex miles	flex ctrl	flex miles	flex ctrl + school flex	flex ctrl + school flex

Robust standard errors 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. The third through seventh columns include 3-digit-NAICS-specific third-order polynomials in log mileage (3-7), log non-production worker share (4,6,7), log assets per worker (4,6,7), log pay per worker (4,6,7), log pay per production worker (4,6,7), log pay per non-production worker (4,6,7), and city-industry mean years of schooling (6,7).

E.3 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.3 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).[79] I calculate the mileage distance from origin CBSA to foreign destination using latitude and longitude coordinates.[80]

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.

[79] This small sample size prevents me from estimating demand shifters using export shipments.

[80] 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.3: 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 hh 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
Obs (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).