

Within-Industry Specialization and Global Market Power[†]

By AHMAD LASHKARIPOUR*

Export price levels exhibit tremendous cross-national and spatial variation, even within narrowly defined industries. Standard theories attribute this variation to within-industry quality specialization. This paper argues that a significant portion of the export price variation is driven by rich and remote economies specializing in high-market power segments of industries. I also argue that this particular pattern of specialization (i) accounts for 30 percent of the overall gains from trade, and (ii) explains more than 37 percent of the observed cross-national income inequality. (JEL D43, F12, F13, F14, L15, L22, O19)

For some time now, there has been a consensus among trade economists that international specialization occurs predominantly within rather than across industries (Schott 2004). Accordingly, there has been a surge in new theories of within-industry specialization, many of which are motivated by two observations. First, there is tremendous variation in export price levels even within narrowly defined industries. Second, these variations are systematic:¹

- (i) Export price levels increase with the exporter's *income per capita*; and
- (ii) Export price levels increase with *geo-distance* to global markets.

The dominant thesis in explaining the above facts is that high-income and geographically distant economies specialize in high-quality varieties within industries. Quality specialization, in these instances, is fueled by either factor-driven comparative cost advantage (Schott 2004), geography-driven comparative cost advantage (Hummels and Skiba 2004), or the home-market effect (Fajgelbaum, Grossman, and Helpman 2011).² The quality specialization thesis has proven to be quite influential, motivating a vibrant strand of macroeconomic growth theories (Grossman and Helpman 1991), as well as enhancing our understanding of

* Assistant Professor of Economics, Indiana University, 100 S. Woodlawn Ave., Bloomington, IN 47405 (email: alashkar@indiana.edu). John Asker was coeditor for this article. I thank seminar participants at various venues for providing helpful comments and suggestions that improved this paper. Special thanks to Angela G. Campbell, Farid Farrokhi, Cecilia Fieler, James Tybout, and the four anonymous referees. Finally, I owe a debt of gratitude to my former advisors Jonathan Eaton and Stephen Yeaple for their invaluable encouragement and support.

[†]Go to <https://doi.org/10.1257/mic.20170336> to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

¹See Schott (2004), Hummels and Skiba (2004), and Hummels and Klenow (2005).

²See also, among others, Sutton (2007); Hallak (2006); Choi, Hummels, and Xiang (2009); Khandelwal (2010); Baldwin and Harrigan (2011); Crozet, Head, and Mayer (2012); Johnson (2012); Lugovskyy and Skiba (2015); Feenstra and Romalis (2014); Sutton and Treffer (2016); and Dingel (2016), who emphasize the quality specialization channel.

trade-induced inequality (see Verhoogen 2008 and Fajgelbaum, Grossman, and Helpman 2011).

This paper, however, argues that this dominant thesis goes too far in attributing facts (i) and (ii) to exclusively quality specialization. As a matter of arithmetic, price depends on (i) the marginal cost that varies with output quality, and (ii) the markup that varies with the degree of market power. An accumulating body of evidence indicate that the markup component is nontrivial.³ Even more revealing, I present new evidence that aggregate measures of export markup covary with income per capita and geo-distance in a fashion that exactly mirror the variation in export price levels.

Motivated by such evidence, I propose an alternative view of within-industry specialization that emphasizes the market power channel. That is, industries are comprised of multiple segments with differential degrees of market power.⁴ Firms in each country sort into high- and low-market power segments of industries, taking into account aggregate cost measures such as the wage rate or distance to global markets. Collectively, these firm-level decisions lead to international specialization across low-market power and high-market power segments of industries. This pattern of specialization, in turn, provides an alternative explanation for facts (i) and (ii). More importantly, it delivers a set of macro-level implications that are starkly different from the quality specialization thesis.

I establish my theory using a general framework that nests an extensive class of trade models featuring firm heterogeneity, quality-sorting, and variable markups, as a special case. Extending the result in Arkolakis et al. (2019), I show that in this general framework, variable markups alone cannot explain the cross-national heterogeneity in export price levels to a *given* market. However, under a set of testable assumptions, such heterogeneity can arise from within industry specialization across low- and high-markup segments of industries.

The intuition behind my theory is the following. Goods manufactured with high-wage labor or transported over long distance exhibit a cost disadvantage. To circumvent such a disadvantage, high-wage and remote economies specialize in industry segments where *aggregate* demand is relatively less sensitive to cost and relatively more sensitive to output quality. I label such segments of an industry as “*quality-intensive*.” From the perspective of a standard trade model, “*quality-intensive*” segments are those that are subject to a lower trade elasticity.

My testable claim is that quality-intensive segments are also those where *individual* firms enjoy more market power. Hence, international specialization across low- and high-quality intensive segments can be interpreted as specialization across low- and high-markup segments. Under this interpretation, international specialization creates a positive link between an economy’s wage rate or distance to global markets and its average export markup levels. This positive link, in turn, provides an alternative markup-based explanation for facts (i) and (ii) presented earlier.

³For instance, according to Berry, Levinsohn, and Pakes (1995), more than 30 percent of the price variation in the auto industry is due to markup differences.

⁴The view that industries consist of segments with differential degrees of market power has deep roots in the literature. See, for example, the case of specialty versus popular-market beer in Hausman, Leonard, and Zona (1994) or luxury versus economy cars in Berry, Levinsohn, and Pakes (1995) and Goldberg (1995).

Employing rich micro-level trade data, I present support for the main assumption underlying my theory. That is, I estimate that the degree of firm-level market power is significantly greater in high quality-intensive segments of industries. Furthermore, I find that the composition of country-level exports aligns with the prediction that high-wage and distant economies export relatively more in high-market power segments of each industry.

Beyond this evidence, the present theory fits well with the observation that markup levels have diverged between rich and poor economies in the past few decades (De Loecker and Eeckhout 2018; Díez, Leigh, and Tambunlertchai 2018). Relatedly, it can explain why local prices and markups have risen in Europe (De Loecker, Fuss, and Van Biesebroeck 2014) but have fallen in China following trade liberalization (Brandt et al. 2017). From the standpoint of standard variable markup theories, the asymmetric effects of import competition on high- versus low-wage economies are difficult to explain. The present theory, however, predicts that import competition will induce European firms to sort into high-market power segments of industries, and Chinese firms to sort into the low-market power segments. These asymmetric effects will raise the average market power and markup of European firms while reducing them for firms in China.

At the *macro level*, my view of within-industry specialization produces distinct implications relative to the quality specialization thesis. The reason being that specialization, here, occurs across industry segments that differ in their (i) degree of market power, (ii) degree of returns to scale, and (iii) trade elasticity. Considering that the latter two characteristics regulate aggregate productivity and the gains from trade, within-industry specialization along these margins can profoundly influence macroeconomic outcomes.

To quantify macro-level implications, I estimate my model using industry-level trade and production data from the World Input-Output Database, covering 35 industries and 32 major economies. Using the estimated model, I find that the gains from trade are (on average) 20 percent in terms of real GDP. Of the total gains, 6 percent is driven by within-industry specialization across low- and high-market power segments, while the remaining 14 percent is driven by traditional forces in the Arkolakis, Costinot, and Rodríguez-Clare (2012) framework.

There is, however, tremendous cross-national heterogeneity in the gains from within-industry specialization. Advanced or remote economies like Australia, France, and Germany gain significantly more as they specialize in high-market power segments where the degree of scale economies and returns to specialization are higher. In comparison, the pure gains from specialization are much smaller (and even negative) for developing economies like Mexico and Indonesia, as they specialize in low-returns-to-scale segments where returns to specialization are relatively low.

At a broader level, the present model sheds new light on cross-country income differences that are believed to be puzzlingly large and have attracted considerable attention in the literature.⁵ From the perspective of the present theory, high-TFP

⁵See, for example, Klenow and Rodríguez-Clare (1997); Prescott (1998); Hall and Jones (1999); Parente and Prescott (1999); Acemoglu, Johnson, and Robinson (2000); Parente and Prescott (2000); Caselli (2005); Jones (2011).

economies tend to specialize in high-market power segments where returns to scale and specialization are higher. This pattern of specialization, in turn, multiplies the fundamental TFP differences between rich and poor economies. Based on my estimated model, within-industry specialization across *high- and low-returns-to-scale* segments can explain up to 37.3 percent of the cross-national real income inequality.

Aside for the vast literature on quality specialization, this paper is closely related to a growing literature on variable markups and pricing-to-market in international settings. Generally speaking, this literature is either focused on (i) the procompetitive effects of trade (Melitz and Ottaviano 2008; Spearot 2013; Holmes, Hsu, and Lee 2014; de Blas and Russ 2015; Edmond, Midrigan, and Xu 2015; De Loecker et al. 2016), or (ii) export price discrimination *across* markets (e.g., Hummels and Lugovskyy (2009), Alessandria and Kaboski 2011, Simonovska 2015, Bertoletti and Etro 2017). In comparison, this paper concerns (i) the heterogeneity in export price levels *within* markets as well as (ii) the asymmetric effects of trade liberalization on markups in low- versus high-income economies.

My theory also shares common elements with Fieler (2011).⁶ In both models, rich economies specialize in low trade elasticity product categories. The present theory, however, exhibits two key differences. First, it identifies a new source of heterogeneity in the trade elasticities, whereby differences in the segment-level trade elasticities are driven by differences in the degree of quality-intensity. Second, my theory establishes a systematic link between patterns of specialization, degree of global market power, and export price levels, which are absent in Fieler (2011). That being the case, my model delivers strictly different macro-level predictions relative to the perfectly competitive model in Fieler (2011).

Finally, the present paper is related to a growing literature concerning aggregation bias in the estimated gains from trade (e.g., Costinot and Rodríguez-Clare 2014, Levchenko and Zhang 2014, Ossa 2015, Brooks and Pujolas 2016). My paper contributes to this literature by identifying a systematic source of aggregation bias that is triggered by specialization across low- and high-returns-to-scale segments of industries.

I. Suggestive Evidence

As noted in the introduction, many theories of within-industry specialization are motivated by the fact that export price levels increase systematically with the exporting economy's (i) income per capita and (ii) geo-distance to global markets. Moreover, the thesis underlying these theories is that these two facts are driven by quality specialization within industries—see Schott (2004) and Hummels and Skiba (2004), among others.

In this section, I present suggestive evidence that facts (i) and (ii) can be partially attributed to cross-national differences in export markup levels. That is, they arise due to high-income and remote exporters either charging higher markups for the same good or specializing in the production of high-markup varieties within industries.

⁶Relatedly, see Kraay and Ventura (2007); Coibion, Einav, and Hallak (2007); McCalman and Spearot (2013); and Holmes and Stevens (2014) for other papers that emphasize specialization across more- and less-differentiated goods.

As a useful benchmark, I first present the evidence corresponding to facts (i) and (ii). I then present additional evidence that highlights the importance of the previously overlooked markup channel. My analysis here employs two different datasets. To analyze export price levels, I use the the COMTRADE-BACI database. To study aggregate markup levels, I use the the CEPII TRADEPROD database (De Sousa, Mayer, and Zignago 2012). To make the two datasets compatible, I restrict my analysis to 149 economies and 27 three-digit ISIC industries that are represented in both datasets. Even after making the two datasets compatible, the COMTRADE-BACI database remains considerably more granular, as it reports bilateral trade statistics at the 6-digit Harmonized System (HS) level of aggregation. By comparison, the TRADEPROD database only reports production statistics at the 3-digit ISIC level of aggregation—a more comprehensive description of both datasets is provided in Appendix A.⁷

The effect of income per capita and geographical remoteness on export price levels can be estimated using the following equation:

$$(1) \quad \ln \bar{p}_{jkt} = \beta_1 \ln \overline{Dist}_j + \beta_2 \ln GDPcap_{jt} + \beta_3 \ln GDP_{jt} + \delta_{kt} + \varepsilon_{jkt},$$

where \bar{p}_{jkt} denotes the (quantity-weighted) unit price of country j 's exports in industry k in year t . On the right-hand side, \overline{Dist}_j is country j 's population-weighted distance to global markets; $GDPcap_{jt}$ and GDP_{jt} control for country j 's GDP per capita and market size in year t ; and, finally, δ_{kt} controls for industry-year fixed effects.

Since the BACI data reports trade statistics at HS6 level of aggregation, I conduct the above estimation independently with both HS6-year and ISIC-year fixed effects. Estimation results are reported in Table 1, reestablishing the widely held belief that export price levels increase systematically with the exporting economies GDP per capita and remoteness.

Noting that export price levels, $P = m\mathcal{C}(\varphi)$, are composed of a markup component, m , and a marginal cost component, $\mathcal{C}(\varphi)$, there are two ways to interpret the cross-national export price variation. The standard “quality specialization” thesis attributes the variation to cross-national quality differences. That is, the patterns documented in Table 1 are attributed to high-income and remote economies specializing in high- φ (high-quality) varieties within industries.⁸

Below, I present evidence in favor of the alternative view that the observed export price variation is driven partly by markup differences. To present my claim, I first calculate aggregate markup levels by closely following the methodology in De Loecker and Eeckhout (2018). That is, I assume a conventional production function for country j in industry k in year t . Namely, $Q_{jkt} = T_{jkt} V_{jkt}^{\alpha_{kt}} K_{jkt}^{\beta_{kt}}$, where Q_{jkt} denotes the units of output produced, V_{jkt} denotes the variable input (e.g., labor, intermediate inputs, etc.), K_{jkt} denotes capital, and T_{jkt} denotes Total Factor Productivity (TFP). As in De Loecker and Eeckhout (2018), this production

⁷The TRADEPROD database has a more comprehensive longitudinal coverage. It spans 26 years from 1980 to 2006. The COMTRADE-BACI database, by comparison, covers the years 2003 to 2015.

⁸To be more specific, existing theories attribute the effect of GDP per capita to *factor-driven* or *demand-driven* quality specialization, and attribute the effect of remoteness to the “*Washington Apples*” effect or *quality-sorting*.

TABLE 1—CROSS-NATIONAL VARIATION IN EXPORT PRICE AND MARKUP LEVELS

	Dependent (log)		
	Unit price		Markup
Remoteness (log)	0.145 (0.002)	0.242 (0.029)	0.290 (0.016)
GDP per capita (log)	0.141 (0.002)	0.162 (0.014)	0.026 (0.003)
GDP (log)	−0.058 (0.001)	−0.062 (0.008)	0.001 (0.003)
Observations	4,345,390		50,851
Fixed effects	HS6 × year	ISIC × year	ISIC × year
R ²	0.64	0.37	0.53

Notes: The estimating equations are (1) and (2). In the markup regression, each observation is weighted by the number of represented firms, which is proportional to the number of workers based on the forthcoming model. Robust standard errors are reported in parentheses.

structure implicitly assumes that countries vary in their TFP and the chosen inputs, but not in the technology Q .

The Lagrangian of the cost minimization problem is thus given by $\mathcal{P}_{jkt}^V V_{jkt} + \mathcal{P}_{jkt}^K K_{jkt} - \nu_{jkt} (T_{jkt} V_{jkt}^{\alpha_{kt}} K_{jkt}^{\beta_{kt}} - \bar{Q}_{jkt})$, where \mathcal{P}_{jkt}^V and \mathcal{P}_{jkt}^K denote input prices, ν_{jkt} is the Lagrangian multiplier, and \bar{Q}_{jkt} is a constant. The first-order condition of this problem with respect to V_{jkt} implies that $\nu_{jkt} = \mathcal{P}_{jkt}^V V_{jkt} / \alpha_{kt} Q_{jkt}$. Given that the shadow price of the Lagrangian is, by construction, equal to the marginal cost, the markup $m_{ji,k} = \mathcal{P}_{jkt} / \nu_{jkt}$ can, therefore, be expressed as

$$m_{jkt} = \alpha_{kt} \frac{\mathcal{P}_{jkt} Q_{jkt}}{\mathcal{P}_{jkt}^V V_{jkt}}.$$

Given the above expression, I can calculate m_{jkt} up to an *industry* × *year*-specific output elasticity, using data on total output, $\mathcal{P}_{jkt} Q_{jkt}$, and variable input cost, $\mathcal{P}_{jkt}^V V_{jkt}$, from the CEPII TRADEPROD database. Doing so, I can run the following regression that is analogous to (1), but includes m as the dependent variable:

$$(2) \quad \ln m_{jkt} = \beta_1 \ln \overline{Dist}_j + \beta_2 \ln GDPcap_{jt} + \beta_3 \ln GDP_{jt} + \delta_{kt} + \varepsilon_{jkt}.$$

As before, δ_{kt} , in the above equation, controls for industry-year fixed effects and absorbs all the variation in the output elasticity, α_{kt} .⁹

The last column in Table 1 reports the estimation results corresponding to the markup regression. Evidently, the output sold by high-income and remote economies exhibits a systematically higher-markup level. Moreover, the elasticity at which markups increase with GDP per capita and remoteness closely resembles the corresponding elasticity for export prices. The higher number of observations in the

⁹It should be reemphasized that the assumption that α_{kt} is *industry* × *year*-specific has precedent in the literature. De Loecker and Eeckhout (2018) and Díez, Leigh, and Tambunlerchai (2018), for instance, adopt a similar assumption when analyzing global market power using firm-level data.

first two columns is simply reflective of the COMTRADE-BACI data reporting export price levels at a much finer level of aggregation.

While revealing, the evidence presented in Table 1 is subject to an important caveat. I use country-level production data to identify the aggregate markup level in each economy. So, if firm-level markup distribution is highly dispersed within each economy, the results presented in Table 1 may be subject to aggregation bias; the extent of which depends on how the firm-level markup dispersion covaries with GDP per capita and remoteness.¹⁰

II. A Model of Within-Industry Specialization

Consider a world economy consisting of $i = 1, \dots, \mathcal{N}$ countries. Country i is populated by L_i individuals, each endowed with one unit of labor, which is the sole factor of production. All individuals are perfectly mobile across the production of different goods but are immobile across countries; and are paid a wage, w_i , in country i . There are \mathcal{K} industries indexed by k , each comprised of multiple segments that differ in characteristics such as degree of market power and quality intensity. Segments are indexed by z , with \mathcal{Z}_k denoting the set of all segments pertaining to industry k —for example, k can be the auto industry, which is composed of luxury, economy, and cross-over vehicle segments.

Consumers and Preferences.—All individuals have similar preferences that are characterized by a general utility function, $U(\mathcal{U}_1, \dots, \mathcal{U}_{\mathcal{K}})$, where \mathcal{U}_k is the subutility corresponding to industry k . The industry-level subutility is a function of the composite consumption bundle, which combines goods from various segments in that industry. In particular,

$$\mathcal{U}_k = \mathcal{U}_k(C_1, \dots, C_{\mathcal{Z}}); \quad 1, \dots, \mathcal{Z} \in \mathcal{Z}_k$$

where the segment-level subutility, C_z , is a function of the quantity $\mathbf{q} \equiv [q_\omega]_{\omega \in \Omega_z}$ and quality $\varphi \equiv [\varphi_\omega]_{\omega \in \Omega_z}$ of firm-level varieties (indexed by ω) available to the consumer in segment z . Contrary to the dominant approach in the literature, I impose only weak isomorphism between quality and quantity, which is outlined by the following assumption.

ASSUMPTION A1: *Quality and quantity are only weakly isomorphic. In particular, C_z is a function of the quantity and quality of the goods in one's consumption basket:*

$$C_z(\varphi^{\rho_z} \mathbf{q}^{1-\rho_z}).$$

Correspondingly, $\rho_z \in [0, 1]$ denotes the relative importance of product quality in segment z , which I hereafter refer to as the degree of quality intensity.¹¹

¹⁰Importantly, running regression (2) using the markups estimated by De Loecker and Eeckhout (2018) with firm-level data for a sample of 40 countries produces qualitatively similar results.

¹¹In the CES case, Assumption A1 can be stated as $C_z(\varphi^{\rho_z} \mathbf{q}^{1-\rho_z}) = \left[\int_{\omega \in \Omega_z} (\varphi_\omega^{\rho_z} q_\omega^{1-\rho_z})^{\eta_z} d\omega \right]^{1/\eta_z(1-\rho_z)}$.

Note that Assumption A1 imposes weaker conditions on preferences than conventional theories, which typically assume exact isomorphism between quality and quantity (see Melitz and Redding 2014).¹² To elaborate, exact isomorphism corresponds to a special case of Assumption A1, where ρ_z is uniform across all segments. In such a case, there is no real distinction between quality and quantity, and productivity differences across producers can be interpreted as quality differences and *vice versa*.¹³ There is, however, exhaustive evidence from consumer surveys that individuals care relatively more about quality (and relatively less about price) when purchasing certain product categories (see Fornell et al. 1996). By relaxing the uniformity restriction on ρ_z , Assumption A1 accommodates this feature of the data. Accordingly, I will hereafter refer to high- ρ segments as “high quality-intensive” segments of the industry.

According to Assumption A1, consumers maximize their utility along both the quality and quantity dimensions. To streamline the presentation, however, I can convert the consumer problem into a one-dimensional one that is expressed in terms of *effective* (or *quality-adjusted*) quantities and prices. To this end, I use the *tilde* notation to denote effective values, with $\tilde{q} = (\varphi^{1-\rho_z} q^{\rho_z})^{1/\rho_z}$ and $\tilde{p} \equiv (p^{\rho_z}/\varphi^{1-\rho_z})^{1/\rho_z}$, respectively, denoting effective quantity and price. Considering this choice of notation, I assume that $U_i(\cdot)$ and $U_{i,k}(\cdot)$ are additively separable and $C_z(\cdot)$ exhibits a parameterization that delivers the following class of demand functions.

ASSUMPTION A2 (Generalized Gorman-Pollak Demand): *For a consumer with income y , facing a schedule of effective prices $\tilde{\mathbf{p}} \equiv [\tilde{p}_\omega]_{\omega \in \Omega}$, the Marshallian demand for variety $\omega \in \Omega$ assumes the following formulation:*

$$\tilde{q}_\omega = \mathcal{D}_z(F_z(\Lambda_z) \tilde{p}_\omega/y) \mathcal{Q}_z(\Lambda_z)$$

where $\mathcal{D}_z(\cdot)$ is a strictly decreasing function; $F_z(\cdot)$ and $\mathcal{Q}_z(\cdot)$ are differentiable functions; and $\Lambda_z \equiv \Lambda_z(\tilde{\mathbf{p}}, y)$ is a scalar demand shifter that firms take as given and that implicitly solves the budget constraint.¹⁴

The above demand system nests (i) the directly separable preferences in Arkolakis et al. (2019) as a case where $\mathcal{Q}_z(\Lambda_z)$ is constant and $F_z(\Lambda_z) = \Lambda_z$, (ii) the indirectly additive preferences in Bertolotti, Etro, and Simonovska (2018) as a case where $F_z(\Lambda_z)$ is constant and $\mathcal{Q}_z(\Lambda_z) = \Lambda_z$, and (iii) the homothetic single aggregator in Matsuyama and Ushchev (2017) as a case where $F_z(\Lambda_z) = \mathcal{Q}_z(\Lambda_z) = \Lambda_z$. Another well-known special case of Assumption A2 is the CES demand system, which arises when $F_z(\Lambda_z)$ is constant, $\mathcal{Q}_z(\Lambda_z) = \Lambda_z$, and $\mathcal{D}_z(x) = x^{-\epsilon_z}$.¹⁵

¹²Hallak and Sivadasan (2013) and Sutton (2007) have challenged the *exact* quality-quantity isomorphism from a theoretical perspective, while Bils and Klenow (2001); Baldwin and Ito (2008); Fan, Li, and Yeaple (2015); and Flach (2017) have challenged it based on empirical evidence.

¹³As we will see shortly, the multi-industry Melitz-Pareto model in Chaney (2008) corresponds to a special case of the present model where ρ_z is uniform and equal to 1/2 in all industries.

¹⁴Specifically, Λ_z is an implicit solution to $\int_{\omega \in \Omega} p_\omega \mathcal{D}_z(F_z(\Lambda_z) \tilde{p}_\omega/y) \mathcal{Q}_z(\Lambda_z) d\omega = e_z \alpha_k y$, where α_k and e_z , respectively, denote the share of expenditure on industry k and segment $z \in \mathcal{Z}_k$.

¹⁵See Fally (2019, 2018) for a more thorough discussion regarding the generalized Gorman-Pollak demand system.

It is important to note that the theoretical propositions that follow apply to several demand systems that are not readily nested by Assumption A2. For instance, the theory presented here applies equally to (iv) the quadratic but nonseparable demand system (Ottaviano, Tabuchi, and Thisse 2002; Melitz and Ottaviano 2008); (v) the QMOR expenditure function, as well as (vi) Kimball preferences (e.g., Kimball 1995; Klenow and Willis 2016).

To simplify the notation, and given that the schedule of prices, $\tilde{\mathbf{p}}$, and income, y , vary across markets, I henceforth use $\Lambda_{i,z} \equiv \Lambda_z(\tilde{\mathbf{p}}_i, y_i)$; $\mathcal{Q}_{i,z} \equiv \mathcal{Q}_z(\Lambda_{i,z})$ and $F_{i,z} \equiv F_z(\Lambda_{i,z})$ to denote demand shifters associated with country i 's market in segments z .

Firms and Production.—Every industry k is populated with $N_{j,k}$ firms from country j , which compete under monopolistic competition. There is free entry, whereby a large pool of ex ante identical firms in country j can pay an entry cost, $w_j f_k^e$, to enter the industry. The free entry assumption introduces scale effects that magnify the extent of international specialization, but is not consequential to the main predictions of the model.

Firms operating in segment z of industry k are heterogeneous in their output quality, φ , which is the realization of a random variable drawn independently across firms from a distribution $G_{j,z}$. As a key assumption, I impose that $G_{j,z}$ is Pareto with the same shape parameter $\gamma_k > 0$ in all countries and in all segments pertaining to industry k .

ASSUMPTION A3: For all φ and all $z \in \mathcal{Z}_k$, $G_{j,z}(\varphi) = 1 - \mathcal{A}_{j,z} \varphi^{-\gamma_k}$, with $\gamma_k > 0$.¹⁶

The Pareto parametrization specified by Assumption A3 is by far the most common distributional assumption in heterogeneous firm models (e.g., Chaney 2008, Melitz and Ottaviano 2008, Simonovska 2015, Arkolakis et al. 2019). As we will see shortly, the main advantage of Assumption A3 is delivering the gravity equation for aggregate trade flows. The more specific assumption that γ_k is uniform within industries is not consequential for the results that follow. Instead, it will allow me to highlight a previously unknown source of variation in the trade elasticities. Moreover, if ρ_z is also uniform across industries (i.e., quality and quantity are exactly isomorphic), the quality heterogeneity specified by Assumption A3 can be trivially reinterpreted as productivity heterogeneity (see Melitz and Redding 2014). In that regard, the present framework also nests an extensive class of models featuring heterogeneous productivity firms.

Considering Assumptions A1–A3, firm ω is uniquely characterized by its output quality, φ . So, to simplify the notation without risk of confusion, I henceforth drop firm indices and use lowercase letters to denote firm-level variables as a function of firm-level quality and aggregate variables.

¹⁶The support of distribution $G_{j,z}(\varphi)$ is $\varphi \geq \mathcal{A}_{j,z}^{1/\gamma_k}$.

Upon entry, a firm can supply a differentiated good in each segment of industry k .¹⁷ The supply of differentiated goods is governed by constant returns to scale technologies, whereby firms from country j face the following linear cost function when serving segment $z \in \mathcal{Z}_k$ in market i :

$$\mathcal{C}(q; \tau_{ji,k}, f_{i,z}, w_j, w_i) = \tau_{ji,k} a_z w_j q + w_i f_{i,z}.$$

In the above expression q is the quantity supplied, with $c_z(\tau_{ji,k}, w_j) \equiv \tau_{ji,k} a_z w_j$ denoting the marginal cost of production and transportation that is paid in terms of the domestic labor wage, w_j . The marginal cost is governed by a segment-specific cost shifter, a_z , and an exporter-importer specific term, $\tau_{ji,k}$, that accounts for international differences in production and transportation costs—as shown in Appendix B, the above cost function and the corresponding analysis can be easily generalized to allow for the dependence of $\mathcal{C}(\cdot)$ on output quality. Finally, $w_i f_{i,k}$ accounts for the fixed cost of serving market i which is paid in terms of labor in the destination market. The cost function can be re-formulated in terms of effective output, \tilde{q} , as follows:

$$\tilde{\mathcal{C}}(\tilde{q}, \varphi; \tau_{ji,k}, f_{i,z}, w_j, w_i) = \tau_{ji,k} a_z w_j \tilde{q} \varphi^{\frac{\rho_z}{\rho_z-1}} + w_i f_{i,z}.$$

Under monopolistic competition, a firm operating with an *effective* variable cost, $\tilde{c}_z(\varphi; \tau_{ji,k}, w_j) \equiv \tau_{ji,k} a_z w_j \varphi^{\frac{\rho_z}{\rho_z-1}}$, chooses a market and segment-specific effective price $\tilde{p}_z(\tilde{c}_z, P_{i,z}, Q_{i,z})$ in order to maximize profits:

$$\pi(\tilde{c}_z, P_{i,z}, Q_{i,z}, L_i) = \max_{\tilde{p}_z} (\tilde{p}_z - \tilde{c}_z) \tilde{q}(\tilde{p}_z, P_{i,z}, Q_{i,z}) L_i,$$

where $\tilde{q}(\cdot) \times L_i$ is the total demand facing the firm in market i . Given that there is zero cross-cost passthrough between different varieties of a given firm, the first-order condition for the above problem entails that firms charge a monopoly markup that is inversely related to the underlying demand elasticity (Laffont and Tirole 1993). Considering the optimal markup, a firm will serve segment z of market i , if and only if $\pi(\tilde{c}_z, P_{i,z}, Q_{i,z}, L_i) \geq w_i f_{i,z}$.

A. Trade Equilibrium

Below, I characterize equilibrium outcomes at the firm and at the aggregate levels—in doing so, I borrow heavily from the analysis in Arkolakis et al. (2019).¹⁸ Subsequently, I present a formal summary of the trade equilibrium.

¹⁷By allowing firms to serve multiple segments, the present model accommodates firm heterogeneity in product scope. According to Hottman, Redding, and Weinstein (2016), product quality and product scope, together, explain more than four fifths of the variation in firm sales.

¹⁸While I closely follow Arkolakis et al. (2019) in deriving the gravity equation, the present model exhibits key differences with Arkolakis et al. (2019). Most importantly, aside from considering a more general demand system, Arkolakis et al. (2019) is a special case of the present model where ρ_z is uniform across segments and equal to 1/2. As a result, the pattern of specialization highlighted in Section IIC is absent in the Arkolakis et al. (2019) framework.

Firm-Level Outcomes.—To outline the firm-level outcomes, consider a firm serving segment z in market i , with effective cost $\tilde{c}_{i,z}$ and effective price $\tilde{p}_{i,z}$ —keep in mind that $\tilde{c}_{i,z} \equiv \tilde{c}_z(\varphi; \tau_{ji,k}, w_j)$ is a function of country of origin characteristics and firm-level quality. Following Assumption A2, the Marshallian demand elasticity facing such a firm can be expressed as a function of its effective price, $\tilde{p}_{i,z}$, relative to a market-level index, $P_{i,z} \equiv w_i/F_{i,z}$, as follows:

$$\varepsilon_z(\tilde{p}_{i,z}/P_{i,z}) \equiv -\partial \ln \mathcal{D}_z(\tilde{p}_{i,z}/P_{i,z}) / \partial \ln \tilde{p}_{i,z}.$$

Hence, defining $\nu_{i,z} \equiv P_{i,z}/\tilde{c}_{i,z}$, the optimal markup, $m_{i,z} \equiv \tilde{p}_{i,z}/\tilde{c}_{i,z} = p_{i,z}/c_{i,z}$ on any variety sold in segment z can be calculated as

$$m_{i,z} = \frac{\varepsilon_z(\tilde{p}_{i,z}/P_{i,z})}{\varepsilon_z(\tilde{p}_{i,z}/P_{i,z}) - 1} = \frac{\varepsilon_z(m_{i,z}/\nu_{i,z})}{\varepsilon_z(m_{i,z}/\nu_{i,z}) - 1}.$$

It is straightforward to verify that if $\varepsilon'_z(x) > 0$, there exists a unique markup function, $m_{i,z} \equiv \mu_{i,z}(\nu_{i,z})$, that implicitly solves the above equation. More importantly, $\mu_{i,z}(\cdot)$ does not depend on the firm's country of origin. Altogether, given the demand function specified by Assumption A2, and noting that $\tilde{p}_{i,z} = m_{i,z}\tilde{c}_{i,z}$ with $\tilde{c}_{i,z} = P_{i,z}/\nu_{i,z}$, the total sales of a firm serving segment $z \in \mathcal{Z}_k$ of market i can be calculated as

$$x(\nu_{i,z}, P_{i,z}, Q_{i,z}, L_i) = \frac{\mu_{i,z}(\nu_{i,z})}{\nu_{i,z}} P_{i,z} \mathcal{D}_z\left(\frac{\mu_{i,z}(\nu_{i,z})}{\nu_{i,z}}\right) Q_{i,z} L_i.$$

The above expression then implies the following segment-specific profit function:

$$\pi(\nu_{i,z}, P_{i,z}, Q_{i,z}, L_i) = \left[(\mu_{i,z}(\nu_{i,z}) - 1) / \mu_{i,z}(\nu_{i,z}) \right] x(\nu_{i,z}, P_{i,z}, Q_{i,z}, L_i).$$

The zero-profit cutoff condition for segment z in market i can, therefore, be expressed as follows:

$$\pi(\nu_{i,z}^*, P_{i,z}, Q_{i,z}, L_i) = w_i f_{i,z}.$$

The implicit solution to the above equation, $\nu_{i,z}^*$, implies a cut-off that depends solely on the characteristics of market i and is independent of the origin country.¹⁹ Alternatively, the demand structure specified by Assumption A2, is general enough to admit a choke price, $b_{i,z}$, such that $\pi_{i,z} = 0$ if $\tilde{c}_{i,z} \geq b_{i,z}$. So even in the case where $f_{i,z} = 0$, there exists a cut-off $\nu_{i,z}^* = P_{i,z}/b_{i,z}$ that is invariant to the characteristics of the origin country.

¹⁹The quality cutoff, $\varphi_{ji,z}^*$, implied by $\nu_{i,z}^*$ is, however, origin country specific.

Aggregate Outcomes.—Total sales of country j to market i in segment $z \in \mathcal{Z}_k$ can be calculated as the sum of all firm-level sales as follows:

$$X_{ji,z} = N_{j,k} \int_{\varphi} x \left(P_{i,z} / \tau_{ji,k} a_z w_j \varphi^{\frac{\rho_z}{\rho_z-1}}, P_{i,z}, Q_{i,z}, L_i \right) dG_{j,z}(\varphi).$$

Noting that (i) $dG_{j,z} = \gamma_k \mathcal{A}_{j,z} \varphi^{-\gamma_k-1}$, by Assumption A3, and (ii) $\nu_{i,z} = (P_{i,z} / \tau_{ji,k} a_z w_j) \varphi^{\frac{\rho_z}{1-\rho_z}}$, the above equation can be reformulated as

$$(3) \quad X_{ji,z} = \chi_{i,z} \mathcal{A}_{j,z} N_{j,k} (\tau_{ji,k} w_j)^{-\gamma_k \left(\frac{1}{\rho_z} - 1 \right)},$$

where $\chi_{i,z} \equiv \gamma_k a_z^{-\theta_z} Q_{i,z} P_{i,z} \int_{\nu_{i,z}^*}^{\infty} \frac{\mu_{i,z}(\nu)}{\nu} \mathcal{D}_z \left(\frac{\mu_{i,z}(\nu)}{\nu} \right) \nu^{-\gamma_k \frac{1-\rho_z}{\rho_z}} \frac{d\nu}{\nu}$.²⁰ Correspondingly, the aggregate segment-level profits collected from sales to market i is given by

$$\begin{aligned} \Pi_{ji,z} &= N_{j,k} \int_{\varphi} \pi \left(P_{i,z} / \tau_{ji,k} a_z w_j \varphi^{\frac{\rho_z}{\rho_z-1}}, P_{i,z}, Q_{i,z}, L_i \right) dG_{i,z}(\varphi) \\ &= \zeta_{i,z} \chi_{i,z} \mathcal{A}_{j,z} N_{j,k} (\tau_{ji,k} w_j)^{-\gamma_k \left(\frac{1}{\rho_z} - 1 \right)}, \end{aligned}$$

where

$$\zeta_{i,z} \equiv \frac{\int_{\nu_{i,z}^*}^{\infty} \frac{\mu_{i,z}(\nu) - 1}{\mu_{i,z}(\nu)} \left(\frac{\mu_{i,z}(\nu)}{\nu} \right) \mathcal{D}_z \left(\frac{\mu_{i,z}(\nu)}{\nu} \right) \nu^{-\gamma_k \frac{1-\rho_z}{\rho_z}} \frac{d\nu}{\nu}}{\int_{\nu_{i,z}^*}^{\infty} \left(\frac{\mu_{i,z}(\nu)}{\nu} \right) \mathcal{D}_z \left(\frac{\mu_{i,z}(\nu)}{\nu} \right) \nu^{-\gamma_k \frac{1-\rho_z}{\rho_z}} \frac{d\nu}{\nu}}.$$

The above equations imply that the share of country j 's profits, $\Pi_{j,k}$, from total sales, $\mathcal{R}_{j,k}$, in industry k is equal to $\bar{\zeta}_{j,k} \equiv \Pi_{j,k} / \mathcal{R}_{j,k} = (\sum_{i,z \in \mathcal{Z}_k} \zeta_{i,z} X_{ji,z}) / (\sum_{i,z \in \mathcal{Z}_k} X_{ji,z})$. Note that if we were also to assume that $\nu_{i,z}^*$ is uniform across markets à la Arkolakis et al. (2019), then $\zeta_{i,z}$ will be constant and uniform within segment z and profits will be a constant share of total revenue.

Summary of Equilibrium.—A trade equilibrium is an $\mathcal{N} \times \mathcal{K}$ vector of entrants $\mathbf{N} \equiv [N_{i,k}]$, an $\mathcal{N} \times 1$ vector of national wage rates, $\mathbf{w} \equiv [w_i]$, an $\mathcal{N} \times \mathcal{K}$ vector of industry-level employment, $\mathbf{L} \equiv [L_{i,k}]$, an $\mathcal{N} \times 1$ vector of income levels, $\mathbf{y} \equiv [y_i]$, and an $\mathcal{N} \times \mathcal{K} \times \mathcal{Z}$ vector of market-specific cut-offs, $\boldsymbol{\nu}^* \equiv [\nu_{i,z}^*]$, that satisfy (i) the free entry condition (FE), (ii) the industry and aggregate-level

²⁰The term $\theta_z \equiv \gamma_k(1/\rho_z - 1)$ denotes the trade elasticity, which I will elaborate on later. Also, in the integrals, the subscript for ν is dropped in the interest of brevity.

labor market clearing conditions (LMC), (iii) the representative consumers budget constraint, and (vi) the zero-profit cutoff condition (ZPC):²¹

$$(FE) \quad N_{j,k} = \bar{\zeta}_{j,k} L_{j,k} / f_k^e,$$

$$(\text{industry LMC}) \quad w_j L_{j,k} = \sum_{i=1}^{\mathcal{N}} \sum_{z \in \mathcal{Z}_k} X_{ji,z} + N_{ji,z} w_j f_{j,z} - N_{ij,z} w_i f_{i,z},$$

$$(\text{aggregate LMC}) \quad \sum_{k=1}^{\mathcal{K}} L_{j,k} = L_j,$$

$$(BC) \quad y_i L_i = w_i L_i = \sum_{n=1}^{\mathcal{N}} \sum_{k=1}^{\mathcal{K}} \sum_{z \in \mathcal{Z}_k} X_{in,z},$$

$$(ZPC) \quad \pi(\nu_{i,z}^*, P_{i,z}, Q_{i,z}, L_i) = w_i f_{i,z},$$

where $N_{ji,z} \equiv N_{j,k}(1 - \mathcal{G}_{ji,z}(\nu_{i,z}^*))$ denotes the mass of country j firms that can profitably serve market i in segment $z \in \mathcal{Z}_k$.²² In constructing the equilibrium, note that given the markup equation, the vector of wages, $\mathbf{w} = \mathbf{y}$, fully determines the income levels, the entire schedule of prices in each market, as well as $P_z(\tilde{\mathbf{p}}, y)$ and $Q_z(\tilde{\mathbf{p}}, y)$. Also, given that the Marshallian demand must satisfy the budget constraint (BC) of the representative consumer, trade is necessarily balanced, i.e., $\sum_n \sum_k \sum_{z \in \mathcal{Z}_k} X_{in,z} = \sum_j \sum_k \sum_{z \in \mathcal{Z}_k} X_{ji,z}$ for every country i .

Gravity Equation.—An attractive feature of the present framework is that country-level export flows within each segment are characterized by a gravity equation. Specifically, let $\alpha_{i,k}$ denote country i 's expenditure share on industry k , and $e_{i,z}$ denote country i 's (within-industry) expenditure share on segment z . Equation (3) then implies that total export flows from country j to i within segment z of industry k are given by

$$(4) \quad X_{ji,z} = \frac{\mathcal{A}_{j,z} N_{j,k} (\tau_{ji,k} w_j)^{\gamma_k (1 - \frac{1}{\rho_z})}}{\sum_{n=1}^{\mathcal{N}} \mathcal{A}_{n,z} N_{n,k} (\tau_{ni,k} w_n)^{\gamma_k (1 - \frac{1}{\rho_z})}} e_{i,z} \alpha_{i,k} Y_i,$$

where $Y_i \equiv w_i L_i$ denotes total income in country i . The *trade elasticity* underlying the above gravity equation, $\theta_z \equiv \gamma_k (1 - 1/\rho_z)$, depends on both (i) the industry-level degree of firm heterogeneity, γ_k , and (ii) the segment-level degree of quality intensity, ρ_z . Only in the case of exact isomorphism between quality and quantity (which amounts to setting $\rho_z = 1/2, \forall z$), the trade elasticity becomes identical to the Pareto shape parameter, γ_k , as in Chaney (2008).

²¹The existence and uniqueness of equilibrium in the present setup is nontrivial. But, for a given vector ν^* , the present model falls under the class of multi-industry gravity models studied in Kucheryavy, Lyn, and Rodríguez-Clare (2016); and satisfies the necessary condition for uniqueness and existence of equilibrium in these models—see Propositions 2–6 in Kucheryavy, Lyn, and Rodríguez-Clare (2016).

²² $d\mathcal{G}_{ji,z}(\nu) \equiv dG_{j,z}(\tau_{ji,k} a_z w_j \nu^{1/\rho_z - 1} / P_{i,z})$ denotes the PDF of ν for country j firms selling to market i in segment z .

B. A Special Case: CES without q - φ Isomorphism

To put the above model in perspective, it is useful to outline a special case of the model with familiar CES preferences. This special case is essentially a variation of Chaney (2008) that relaxes the exact isomorphism between quality and quantity. In this special case, the utility of the representative consumer has the following parametric structure:

$$W_i = \prod_{k=1}^K \prod_{z \in \mathcal{Z}_k} \left(\int_{\omega \in \Omega_{i,z}} \varphi_{\omega}^{\frac{1}{\epsilon_z}} q_{\omega}^{1-\frac{1}{\epsilon_z}} d\omega \right)^{\frac{\epsilon_z}{\epsilon_z-1} e_{i,z} \alpha_{i,k}},$$

where $\alpha_{i,k}$ denotes the share of market i 's expenditure on industry k , while $e_{i,z}$ denotes the (within-industry) share of expenditure on segment z , with $\sum_k \alpha_{i,k} = 1$ and $\sum_{z \in \mathcal{Z}_k} e_{i,z} = 1$. Based on the above utility, $\rho_z = 1/\epsilon_z$, and the demand function facing firm ω is $\tilde{q}_{\omega} = (\tilde{p}_{\omega}^{-\epsilon_z} / \sum_k \int_{\omega'} \tilde{p}_{\omega'}^{1-\epsilon_z} d\omega') \bar{e}_{i,z} y_i$, with the effective quantities and prices defined as $\tilde{q}_{\omega} \equiv q_{\omega} \varphi_{\omega}^{1/(\epsilon_z-1)}$ and $\tilde{p}_{\omega} \equiv p_{\omega} / \varphi_{\omega}^{1/(\epsilon_z-1)}$. One can verify, either directly or by using equation (4), that the gravity equation in this special case adopts the following formulation:

$$X_{ji,z} = \frac{\mathcal{A}_{j,z} N_{j,k} (\tau_{ji,k} w_j)^{\gamma_k (1-\epsilon_z)}}{\sum_{n=1}^N \mathcal{A}_{n,z} N_{n,k} (\tau_{ni,k} w_n)^{\gamma_k (1-\epsilon_z)}} e_{i,z} \alpha_{i,k} Y_i.$$

Based on the above gravity equation, the trade elasticity, $\theta_z = \gamma_k (1 - \epsilon_z)$, depends on both the firm-level demand elasticity, ϵ_z , which governs the degree of quality intensity, and the Pareto shape parameter, γ_k . In other words, the model predicts a strong link between the degree of firm-level and country-level market powers, which, as I will argue later, has sharp macro-level implications.

C. The Structure of Within-Industry Specialization

The present framework offers a new perspective on the structure of within-industry specialization. As a first step in demonstrating this, I use the gravity equation to determine the composition of country-level exports across segments of an industry. Following equation (4), the relative exports of countries j and n to market i across segments z and $z' \in \mathcal{Z}_k$ are given by

$$(5) \quad \frac{X_{ji,z}/X_{ji,z'}}{X_{ni,z}/X_{ni,z'}} = \left(\frac{\mathcal{A}_{j,z}/\mathcal{A}_{j,z'}}{\mathcal{A}_{n,z}/\mathcal{A}_{n,z'}} \right) \left(\frac{\tau_{ji,k} w_j}{\tau_{ni,k} w_n} \right)^{\gamma_k \left(\frac{1}{\rho_z} - \frac{1}{\rho_{z'}} \right)}.$$

Based on equation (5), *revealed* comparative advantage across segments of an industry is governed by (i) *technical* comparative advantage through $\mathcal{A}_{j,z} \mathcal{A}_{n,z'} / \mathcal{A}_{j,z'} \mathcal{A}_{n,z}$, as well as (ii) differences in trade costs, τ , and labor cost, w . The latter margin yields a clear pattern of specialization, which is outlined below.

LEMMA 1 (ρ -Driven Specialization): *Absent technical comparative advantage (i.e., $\mathcal{A}_{j,z} = \mathcal{A}_{j,k}\mathcal{A}_z$), high- w and high- τ economies have a revealed comparative advantage in high- ρ (quality-intensive) segments of industries.*

Lemma 1 is a reduced-form result that follows directly from equation (5), relating a country's export composition to its wage level and geo-location, which are both observable. While Lemma 1 is agnostic to the source of cross-national wage differences, it requires the segment-neutrality of \mathcal{A} 's, a sufficient condition to rule out cases where low- w or low- τ economies have a technical comparative advantage (i.e., a higher relative \mathcal{A}) in high- ρ segments. To the extent that Lemma 1 outlines a pattern of specialization across industry segments with different trade elasticities, $\theta_z \equiv \gamma_k(1 - 1/\rho_z)$, it resembles the theory in Fieler (2011). The present model, however, identifies a novel source of variation in the trade elasticities that operates through ρ_z , and that is absent in Fieler (2011) or prior gravity models. Accordingly, specialization across *low-* and *high-trade elasticity* segments, in the present model, assumes a different interpretation compared to Fieler (2011).

The next step is to determine how the pattern of specialization outlined by Lemma 1 dictates the structure of export market power and markups. This transition is guided by an important feature of the model where even though markups are variable, the distribution of firm-level markups is invariant to the characteristics of the exporting country. To demonstrate this, let $\mathcal{M}_{ji,z}(m; w_j, \tau_{ji,k}) = \Pr\{\mu_z(\nu) \leq m \mid \nu \geq \nu_{i,z}^*\}$ denote the distribution of firm-level export markups from country j to i in segment z . Appendix C demonstrates that $\mathcal{M}_{ji,z}$ is invariant to the exporting country's characteristics, $\tau_{ji,k}$ and w_j , and is given by

$$\mathcal{M}_{ji,z}(m; w_j, \tau_{ji,k}) = \mathcal{M}_{i,z}(m) = 1 - \left(\nu_{i,z}^* \mu_{i,z}^{-1}(m)\right)^{-\gamma_k\left(\frac{1}{\rho_z}-1\right)}.$$

The distribution of export markups, however, varies systematically across segments and also with the characteristics of the importing country i . The *across-market* variation in markups is a reflection of pricing-to-market, as in Simonovska (2015). The following lemma summarizes the result concerning the invariance of markup distribution across exporters.

LEMMA 2 (Within-Segment Markup Distribution): *The distribution of export markups within segment z is invariant to the exporting country's characteristics.*

The intuition behind Lemma 2 is similar to that provided by Arkolakis et al. (2019), which is based on the countervailing effects of higher costs on markups. While lower transport costs, $\tau_{ji,k}$, or labor costs, w_j , induce firms from country j to charge higher markups in market i , they also induce lower quality firms from country j to export, and such firms charge lower markups. These two effects exactly offset one another in the Pareto case, leading to the invariance of the markup distribution. Importantly, Lemma 2 indicates that in the extensive class of models covered here (e.g., Melitz and Ottaviano 2008 and Simonovska 2015), variable markups do not create a systematic link between country-level characteristics and the

average export markup *within* segments. Instead, the industry-level heterogeneity in export markup levels is driven exclusively by the composition of exports across segments.

Lemma 2 points to two important corollaries. First, it entails that the distribution of demand elasticities facing firms exporting from country j to market i is also invariant to country j 's characteristics: $\mathcal{E}_{ji,z}(m; w_j, \tau_{ji,k}) = \mathcal{E}_{i,z}(\varepsilon)$. Given the invariance of the *markup* and *demand elasticity* distributions, I henceforth use $\bar{\varepsilon}_{i,z}$ and $\bar{m}_{i,z}$ to denote the average demand elasticity and markup level associated with segment z in market i :

$$\bar{\varepsilon}_{i,z} \equiv \int_{\varepsilon} \varepsilon d\mathcal{E}_{i,z}(\varepsilon),$$

$$\bar{m}_{i,z} \equiv \int_m m d\mathcal{M}_{i,z}(m).$$

Correspondingly, in each market, we can uniquely rank segments within industry k by their average firm-level markup or elasticity level. Such a ranking is also significant from a macro-level perspective given that, under free entry, $\bar{m}_{i,z}$ represents the degree of scale economies in segment z . To elaborate, consider the standard definition of the *scale elasticity*, which is the elasticity at which firm-level output increases with total input cost: $\psi \equiv \partial \ln q / \partial \ln C$. It is a well-known that if (i) firms are profit-maximizing and (ii) there is free entry ($AC = p$), this elasticity equals the firm-level markup (see Hanoch 1975 and Helpman 1984). In particular,

$$\text{scale elasticity: } \quad \psi \equiv \frac{AC}{MC} = \frac{p}{MC} = \frac{\varepsilon}{\varepsilon - 1} = m,$$

where MC and AC , respectively, denote the marginal cost and average cost at the firm level. Considering the above equation, high- \bar{m} or low- $\bar{\varepsilon}$ segments can be viewed as high-scale intensive. As one may expect, this connection has basic implications for the gains from trade and cross-country income differences—a point I will return to in Section IV.

The final step in establishing my theoretical proposition is to relate ρ -driven specialization (Lemma 1) to the markup and scale-intensity of a nation's exports. To this end, I impose the following testable assumption that links the segment-level quality intensity, ρ_z , to its average demand elasticity, $\bar{\varepsilon}_{i,z}$.

ASSUMPTION A4 (Relation between $\bar{\varepsilon}$ and ρ): *Firms on average face a lower demand elasticity, $\bar{\varepsilon}_{i,z}$, in high-quality-intensive (high- ρ_z) segments of an industry.*

The above assumption is already implicit in the CES case outlined in Section IIB, where $\rho_z = 1/\varepsilon_z$. Beyond CES, the prior literature has occasionally invoked Assumption A4, with Rodrik (1994), for instance, arguing that quality-intensive segments are less standardized and therefore less price sensitive. But more importantly,

Assumption A4 is empirically testable. In Section III, I employ rich micro-level data to establish that the empirical relationship between quality intensity and the demand elasticity aligns with Assumption A4.

Assumption A4 paired with Lemmas 1 and 2 imply my main proposition that high-income and geographically remote economies specialize in high-markup, high-scale-intensive segments of industries. Appendix C provides a formal proof of this claim, and the following provides a formal statement.

PROPOSITION 1: *Absent technical comparative advantage, high-wage and geographically remote economies specialize in high-scale intensive, high-markup segments of industries.*

The intuition behind Proposition 1 is straightforward. National industries that face high labor or transportation costs tend to export relatively more in quality-intensive (low trade elasticity) segments where aggregate sales are less sensitive to aggregate cost measures. Per Assumption A4, high quality-intensive segments are also those in which firms, as an individual unit, enjoy greater market power and set higher markups. Altogether, these two patterns yield the proposition stated above. To better understand Proposition 1, bear in mind that higher wage levels in the present model are a reflection of the economy being agglomerated with high-quality firms (which will be reflected in a higher \mathcal{A}). Proposition 1, merely states that an agglomeration of high-quality firms in an economy puts upward pressure on the local wage, inducing firms to specialize in high-market power, high-scale-intensive segments of industries.

Discussion.—Assumptions A1–A4 outline a set of sufficient conditions that deliver Proposition 1. Of these, Assumption A1 is less consequential. It relaxes the quality-quantity isomorphism in favor of a less-restrictive assumption to identify a new source of variation in the trade elasticities. Assumptions A2 and A3 are conventional assumptions that deliver the gravity equation—as noted earlier, both assumptions underlie an important class of constant and variable markup trade models. Assumption A4, however, is a major assumption that can be tested with micro-level data, and the following section is dedicated to this task.

Note, however, that even a weaker version of Assumption A4 will still yield Proposition 1. Specifically, as long as the segment-wide trade elasticity $\theta_z \equiv \gamma_k(1 - 1/\rho_z)$ is positively correlated with the average firm-level demand elasticity, $\bar{\epsilon}_{i,z}$, the pattern of revealed comparative advantage outlined by Lemma 1, will yield Proposition 1. The same applies even if the Pareto shape parameter, γ , is segment-specific.

Finally, the specialization pattern outlined by Proposition 1 can concurrently occur at various levels of aggregation. In the interest of consistency, the theory was cast here as one of within industry specialization. But one can alternatively assume that A1–A4 hold across (rather than within) industries to arrive at a similar proposition regarding across-industry specialization. In that case, high-wage and remote economies specialize in high-market power, high-scale-intensive industries.

D. An Alternative View on Export Price Levels

Proposition 1 offers a new perspective on the previously highlighted facts (i) and (ii): that export price levels increase with distance and the income per capita of the exporting country. As noted before, the dominant view in the literature is that these facts are driven by quality specialization. Against the backdrop of this dominant view, Proposition 1 highlights the markup channel as an alternative driver of the aforementioned facts, while Lemma 1 provides an alternative view on the origins of quality specialization.

To elaborate, note that the average free on board price of exports from country j to market i in industry k can be decomposed as follows:

$$\bar{p}_{ji,k} = \underbrace{\bar{\mu}(\tau_{ji,k}, w_j)}_{\text{markup-driven}} \times \underbrace{\bar{\mathbf{a}}(\tau_{ji,k}, w_j)w_j}_{\text{cost-driven}},$$

where $\bar{\mathbf{a}}(\tau_{ji,k}, w_j) \equiv \sum_{z \in \mathcal{Z}_k} a_z r_{ji,z}(\tau_{ji,k}, w_j)$ and $\bar{\mu}(\tau_{ji,k}, w_j) \equiv \sum_{z \in \mathcal{Z}_k} \bar{m}_{i,z} r_{ji,z}(\tau_{ji,k}, w_j)$, respectively, denote the average markup and unit labor cost embedded in country j 's industry-level exports to markets i , with $r_{ji,z} \equiv X_{ji,z} / \sum_{z \in \mathcal{Z}_k} X_{ji,z}$ being the share of segment z from total industry-level export sales. In the above expression, the markup-driven competent, $\bar{\mu}(\tau, w)$, has a clear connotation. The cost-driven term, $\bar{\mathbf{a}}(\tau, w)w$, meanwhile, primarily accounts for quality differences. In particular, $\bar{\mathbf{a}}(\tau, w)w$ may be higher due to (i) a higher wage, which reflects an agglomeration of high-quality firms in the economy, or (ii) a higher $\bar{\mathbf{a}}(\cdot)$, which reflects a higher share of exports in high-quality-intensive segments, provided that a_z is increasing in the degree of quality intensity, ρ_z .

From the lens of the present model, the markup-driven term, alone, can explain why exports from rich and distant suppliers exhibit higher price levels. In particular, Proposition 1 asserts that $\partial^2 r_{ji,z} / \partial \tau_{ji,k} \partial \bar{m}_{i,z} > 0$ and $\partial^2 r_{ji,z} / \partial w_j \partial \bar{m}_{i,z} > 0$, which along with the fact that $\sum_{z \in \mathcal{Z}_k} r_{ji,z} = 1$, imply the following:

$$(6) \quad \frac{\partial \bar{\mu}(\tau_{ji,k}, w_j)}{\partial \tau_{ji,k}} = \sum_{z \in \mathcal{Z}_k} \bar{m}_{i,z} \frac{\partial r_{ji,z}}{\partial \tau_{ji,k}} > 0,$$

$$\frac{\partial \bar{\mu}(\tau_{ji,k}, w_j)}{\partial w_j} = \sum_{z \in \mathcal{Z}_k} \bar{m}_{i,z} \frac{\partial r_{ji,z}}{\partial w_j} > 0.$$

Based on the above, the “Washington apples” effect, $\partial \bar{p}_{ji,k} / \partial \tau_{ji,k} > 0$, is partly driven by distant exporters sorting into high-markup segments; or the observation that $\partial \bar{p}_{ji,k} / \partial w_j > 0$ is reflective of high-income economies exporting relatively more goods pertaining to high-markup segments of industries. Even though this mechanism concerns export markups, it operates orthogonal to the well-known variable markup mechanism. As implied by Lemma 2, markups can be variable across firms, but the distribution of markups, *within segments*, is the same for all exporting countries. That being the case, variable markups can explain why Australian firms set higher prices in the US market than in China, but not why the

US imports higher price goods from Australia than from, say, Canada. The above mechanism, by contrast, explains this exact type of variation.

Finally, in the present model, the quality margin also contributes to the positive relationship between $\tau_{ji,k}$, w_j , and $\bar{p}_{ji,k}$, but through a mechanism that is distinct from those specified in the prior literature. Specifically, provided that high-quality-intensive (high- ρ) segments involve higher production cost, a_z , Lemma 1 implies that $\partial \bar{a}(\tau, w) / \partial \tau > 0$ and $\partial \bar{a}(\tau, w) / \partial w > 0$. That is, absent markup heterogeneity, a higher $\tau_{ji,k}$ or w_j increases the aggregate export price level, $\bar{p}_{ji,k}$, by inducing specialization in quality-intensive segments.

III. Micro-Level Evidence

Even if the full schedule of export markups were observable, testing Proposition 1 *directly* would remain challenging as it requires information on the full vector of deep technology parameters, $\mathbf{A} \equiv [\mathcal{A}_{i,z}]$. To elaborate, suppose the data suggests that high-wage and remote economies export relatively more in high-markup segments. It may very well be the case that they have a factor-driven or a historical scale-driven comparative advantage in these segments, which will be reflected in a higher \mathcal{A} . Alternatively, high-wage and remote economies may have a fundamental comparative disadvantage in high-markup segments, which will negate the mechanism highlighted in this paper.

Despite such complications, one can still *indirectly* test Proposition 1 by evaluating the sufficient conditions that lead to it. As noted earlier, of the assumptions leading to Proposition 1, Assumption A1 is not consequential; Assumptions A2 and A3 are standard assumptions that deliver the gravity equation (see Arkolakis et al. 2019); but Assumption A4 is a unique assumption that is empirically testable. Considering this, I first conduct a formal test of Assumption A4 in this section. I then present direct evidence that factual export patterns do in fact comply with Proposition 1. Finally, I discuss the limitations of my analysis and provide auxiliary evidence on across-segment specialization.

A. Data Description

The empirical analysis that follows uses import transactions data from the Colombian Customs Office for the 2007–2013 period.²³ The data covers the universe of all import transactions in Colombia, reporting the following information per transaction: date of transaction, Harmonized System 10-digit product category (HS10), importing and exporting firm's IDs,²⁴ free on board values of shipments in US dollars, quantity and the unit it is measured in, freight in US dollars, insurance

²³The data is obtained from Datamyne, a company that specializes in documenting import and export transactions in the Americas. For more detail, please see www.datamyne.com.

²⁴The identification of the Colombian importing firms is standardized by the national tax ID number. For the foreign exporting firms, the data provide the name of the firm, phone number, and address. The names of the firms are not standardized, and thus there are instances in which the name of a firm and its address are recorded differently (e.g., using abbreviations, capital and lower-case letters, dashes, etc.). To deal with this problem, I follow the guidelines in Lashkaripour and Lugovskyy (2019), which involves standardizing the spelling and the length of the names along with utilizing the data on firms' phone numbers.

in US dollars, import tariff and value-added tax rates, and country of origin. The online Appendix (Table 1) reports a summary of basic trade statistics for this dataset. A possible challenge when working with this dataset is that Colombia has been changing the HS10 classification for some products between 2007 and 2013. Fortunately, the Colombian Statistical Agency, DANE, has kept track of these changes, permitting the concordance of Colombian HS10 codes over time—the concordance procedure closely follows the guidelines in Lashkaripour and Lugovskyy (2019), with changes in HS10 codes affecting less than 0.1 percent of the observations between 2007 and 2013. My analysis also requires (i) data on monthly average exchange rates, which are taken from the Bank of Canada; (ii) data on national accounts that are taken from the Penn World Table version 9; and (iii) data on geo-distance, which are taken from the CEPII database (Mayer and Zignago 2011).²⁵

To map theory to data, I take a similar approach to Holmes and Stevens (2014), using the 4-digit Standard Industrial Classification (SIC) to define industries and the 10-digit HS10 product classification to specify segments within an industry. To fix, consider the following group of HS10 product segments: “PASSENGER CARS WITH 1000–1500CC ENGINES,” “PASSENGER CARS WITH LARGER THAN 3000CC ENGINES,” “LIMOUSINES,” “TRACTORS,” “GOLF CARTS,” and so on. Based on the Standard Industrial Classification, these segments are grouped together into SIC 3711, which corresponds to “MOTOR VEHICLES AND PASSENGER CAR BODIES.” The analysis that follows concerns specialization across these narrowly defined HS10 segments that are grouped into broadly defined SIC industries.

B. Estimating the Segment-Level Demand Elasticity: $\bar{\epsilon}_z$

For every variety ω, z, t exported to Colombia by firm ω , in HS10 segment z , in year t , I observe export quantity, $q_{\omega, z, t}$, as well as the free on board value plus transportation costs and tariff/tax charges, which can be used to construct the consumer price, $p_{\omega, z, t}$, associated with that variety. Given data on firm-level quantities and prices, I estimate the average demand elasticity facing firms in segment z using the following estimating equation:

$$(7) \quad \ln q_{\omega, z, t} = -\bar{\epsilon}_z \ln p_{\omega, z, t} + \delta_{\omega z} + \delta_{z t} + \xi_{\omega, z, t},$$

where $\delta_{\omega z}$ and $\delta_{z t}$ control for firm-product and product-year fixed effects, and $\xi_{\omega, z, t}$ represent non-price-demand shifters such as measurement errors or time-varying components of quality, $\varphi_{\omega, z, t}$.

To eliminate the firm-product fixed effects, I take first-differences from equation (7) to arrive at the following estimating equation:

$$(8) \quad \Delta \ln q_{\omega, z, t} = -\bar{\epsilon}_z \Delta \ln p_{\omega, z, t} + \Delta \delta_{z t} + \Delta \xi_{\omega, z, t}$$

²⁵ See <http://www.bankofcanada.ca/rates/exchange/monthly-average-lookup/>.

TABLE 2—SUMMARY OF THE ESTIMATED SEGMENT-LEVEL ELASTICITIES

	Mean	Median	First quartile	Thrid quartile	Number of sig. estimates
<i>Panel A. Segment-level demand elasticity: $\bar{\epsilon}_z$</i>					
OLS	1.13	0.90 (0.11)	0.61 (0.25)	1.23 (0.16)	4,391
2SLS	2.74	1.88 (1.22)	1.20 (0.68)	3.03 (2.61)	1,283
LIML	2.78	1.93 (1.26)	1.21 (0.17)	3.26 (1.89)	1,627
<i>Panel B. Segment-level trade elasticity: $\theta_z \equiv (1/\rho_z - 1)\gamma_k$</i>					
OLS	1.05	0.76 (0.14)	0.42 (0.28)	1.28 (0.36)	2,706
2SLS	2.24	1.69 (0.77)	1.05 (0.11)	2.75 (1.88)	1,827
LIML	2.58	1.89 (0.50)	1.16 (0.69)	3.21 (1.48)	1,701

The first-difference estimator exhibits a couple of attractive features for the purpose of my analysis: (i) it eliminates observations pertaining to one-time exporters, (ii) it is more efficient than the fixed effect estimator when $\xi_{\omega,kt}$ s are sequentially correlated, and (iii) it requires a weaker orthogonality condition, $\mathbb{E}[\Delta \ln p_{\omega,zt} | \Delta \xi_{\omega,zt}] = 0$, than the fixed effect estimator.

As a first step, I separately estimate equation (8) for 5,669 HS10 segments using the first-difference OLS estimator. For 4,391 of these HS10 segments, the estimation produces a positive elasticity, $\bar{\epsilon}_z$, (i.e., a negative coefficient on unit price) with a robust t -statistic that is greater than one.²⁶ A basic summary of the OLS-estimated demand elasticities is reported in the first row of Table 2.

IV Estimation.—A general concern when estimating equation (8) is that demand shocks, $\Delta \xi_{\omega,zt}$, may be correlated with longitudinal changes in price, $\Delta \ln p_{\omega,zt}$. Such a correlation, which may be driven by measurement errors or by quality upgrading, leads to a violation of the orthogonality condition, $\mathbb{E}[\Delta \ln p_{\omega,zt} | \Delta \xi_{\omega,zt}] \neq 0$. The standard approach to handling such a concern is to instrument for $\Delta \ln p_{\omega,zt}$ with a plausibly exogenous cost shifter that is orthogonal to $\Delta \xi_{\omega,zt}$. Changes in import tariff rates are often considered to be an appropriate instrument in these circumstances. But since equation (8) is being estimated at the firm level, import tariffs do not qualify as a strong instrument, as they exhibit little-to-no variation across firms from the same country.

To overcome the identification challenge, I use the methodology in Lashkaripour and Lugovskyy (2019), which is to instrument for $\Delta \ln p_{\omega,zt}$ using a plausibly exogenous measure of exposure to aggregate exchange rate shocks.²⁷

²⁶Throughout this section, I drop elasticity estimates that exhibit a robust t -statistic below or equal to one. My analysis, however, produces qualitatively similar results when I drop elasticities that exhibit a robust t -statistic below or equal to two.

²⁷The estimation conducted here differs from Lashkaripour and Lugovskyy (2019) in two basic aspects. First, Lashkaripour and Lugovskyy (2019) *jointly* estimate the country-level and firm-level demand elasticities,

This procedure involves compiling an external database on aggregate monthly exchange rates, and interacting the monthly variation in aggregate exchange rates with the monthly composition of firm-level exports in the *prior* two years. Doing so delivers the following *shift-share* instrument:

$$\mathcal{Z}_{\omega,zt} = \sum_{\ell=1,2} \sum_{mo=1}^{12} \Delta E_{jt}(mo) \times r_{\omega,zt-\ell}(mo),$$

with $\Delta E_{jt}(mo)$ denoting the change in country j 's exchange rate with Colombia in month mo of year t ; and $r_{\omega,zt-\ell}(mo)$ and denotes the prior share of firm ω 's month mo sales in year $t - \ell$. To elaborate, $\mathcal{Z}_{\omega,zt}$ measures the variety-specific *exposure to exchange rate shocks*. It builds on the fact that aggregate exchange rate movements may have differential effects on firms, depending on the monthly composition their export sales. In that regard, $\mathcal{Z}_{\omega,zt}$ resembles the widely used *Bartik* instrument. Perhaps most importantly, $\mathcal{Z}_{\omega,zt}$ captures a plausibly exogenous source of variation in firm-level costs, provided that aggregate exchange rates and historical export patterns are independent of concurrent demand shocks.

I conduct a battery of tests to diagnose the strength of my price instruments in Table 3. Given that errors may be heteroskedastic, I use the heteroskedasticity-robust Kleibergen and Paap (2006) test statistics to diagnose underidentification and weak instruments. Based on the Kleibergen-Paap LM statistic, I can reject the null hypothesis that the model is unidentified. To test if the instruments are only weakly correlated with unit price, I also report heteroskedasticity-robust Kleibergen and Paap (2006) Wald rk F -statistic. While critical values for the Cragg and Donald (1993) F -statistic have been provided by Stock and Yogo (2002) in the homoskedastic case, there is no tabulation of such values for the heteroskedastic case. So, as is common in the literature, I compare the heteroskedasticity-robust statistic to the Stock and Yogo (2002) critical values, which allows me to reject the null of weak instruments. Finally, Table 3 also reports the Anderson and Rubin (1949) Wald test, which strictly rejects the null that the coefficients on the excluded instruments are jointly zero when they are included in place of unit price, $\Delta \ln p_{\omega,zt}$, in the estimating equation (8).

The results attained under the 2SLS estimator are reported in the second row of Table 2. Encouragingly, the 2SLS estimates are lower than the OLS estimates, which suggest that the instruments are operating in the right direction. However, as expected, the standard errors corresponding to the 2SLS estimation are higher, so that for only 1,283 HS10 segments the estimated elasticity is positive and exhibits a robust t-statistic greater than one. Given that HS10 segments may feature a small sample of firm-level observations, I also estimate equation (8) using the Limited Information Maximum Likelihood (LIML) estimator, which is more robust than 2SLS in the presence of weak instruments and small sample bias (Stock and Yogo

by assuming a nested-CES demand system. Second, given the focus of that paper, the elasticities are estimated at WIOD industry level of aggregation. Here, I neither impose that the underlying demand is CES nor do I assume that the elasticities are uniform within industries. As a result, I can estimate reduced-form demand elasticities at the HS10 level of aggregation, in order to test Assumption A4. However, unlike Lashkaripour and Lugovskyy (2019), the estimated elasticities no longer assume a structural interpretation and the country-level trade elasticities have to be estimated separately from the firm-level demand elasticities.

TABLE 3—FIRST-STAGE DIAGNOSTICS: POOLED SAMPLE

	Estimating parameter	
	Demand elasticity $\bar{\epsilon}_z$	Trade elasticity θ_z
Kleibergen-Paap LM statistic (under-identification)	373.58	1,589.54
Kleibergen-Paap LM p -value	0.000	0.000
Kleibergen-Paap Wald rk F -statistic (weak instruments)	121.54	494.14
Anderson-Rubin Wald test F -statistic	235.72	2,223.92
Anderson-Rubin Wald test p -value	0.000	0.000
Observations	1,341,480	500,064

2002). The results corresponding to the LIML estimation are summarized in the third row of Table 2 and are qualitatively similar to the 2SLS estimates.

C. Estimating the Segment-Level Quality Intensity: ρ_z

Using the gravity equation one can estimate a segment-specific trade elasticity, $\theta_z = \gamma_k(1/\rho_z - 1)$, which determines ρ_z up to an industry-wide Pareto shape parameter, γ_k . There is an extensive literature on estimating the trade elasticity, and what follows builds heavily on this literature. To express the estimating equation, I take logs from equation (3); drop the importer subscript i as I am dealing with data from a unique market; and write the equation in stochastic form. Altogether, country j 's sales-per-firm to the Colombian market in HS10 segment z , in year t , can be stated as

$$(9) \quad \ln X_{j,z,t}/N_{j,kt} = -\gamma_k \left(\frac{1}{\rho_z} - 1 \right) \ln \tau_{j,z,t} w_{j,t} + \delta_{z,t} + \xi_{j,z,t}, \quad z \in \mathcal{Z}_k,$$

where $\delta_{z,t} = \ln \chi_{z,t}$ can be treated as an HS10-year fixed effect and $\xi_{j,z,t}$ is an unobserved error term that, among other things, encompasses measurement errors or longitudinal variations in national product quality, $\mathcal{A}_{j,z,t}$. Based on Head and Mayer's (2014) survey of the literature, there are multiple approaches to estimating the above equation. The first approach, which is more suited to cross-sectional data, is to control for w_j using exporter fixed effects and identify the trade elasticity using the variation in transport costs (e.g., Hummels 2001, Hertel et al. 2007) or tariff rates (e.g., Romalis 2007, Caliendo and Parro 2014). An alternative approach for panel data is to control for $\tau_{j,z,t} w_{j,t}$ using the average unit price of country-level exports, $\bar{p}_{j,z,t} = \bar{m}_{z,t} a_{z,t} \tau_{j,z,t} w_{j,t}$ (e.g., Erkel-Rousse and Mirza 2002, Broda and Weinstein 2006); and to instrument for $\bar{p}_{j,z,t}$ in order to address the possible endogeneity concern. Given the panel structure of my data, I adopt the latter approach.

To estimate equation (9), all firm-level variables are aggregated up to construct country-level variables. This procedure leaves me with 500,064 observations with the unit of observation being *exporting country–HS10 segment–year*. For each observation, I have the average sales per firm $X_{j,z,t}/N_{j,kt}$ and the average

price, $\bar{p}_{j,zt} = \bar{m}_{zt} a_{zt} \tau_{j,zt} w_{j,t}$, associated with these sales. Using these variables, I first estimate equation (9) using an OLS estimator, with a summary of the estimation results reported in the fourth row of Table 2.²⁸ The results indicate that for 2,706 HS10 segments, the estimated trade elasticity is positive and exhibits a robust *t*-statistic greater than one. Perhaps expectedly, the median trade elasticity is also lower than the median demand elasticity facing individual firms.

As with the firm-level estimation, one may be concerned that the orthogonality condition, $\mathbb{E}[\bar{p}_{j,zt} | \xi_{j,zt}] = 0$, is violated due to measurement error or longitudinal changes in the national quality distribution, $A_{j,zt}$. To address these concerns, I also estimate equation (9) using a 2SLS estimator, where I instrument for $\bar{p}_{j,zt}$ with the exporter-specific tariff and value-added tax rates. Here, the identifying assumption, as in Caliendo and Parro (2014), is that import tax rates are plausibly exogenous to $\xi_{j,zt}$. Correspondingly, the second column in Table 3 reports a set of first-stage diagnosis tests, which reject the null hypothesis of weak instruments or the model being unidentified.

The 2SLS estimates of $\theta_z = \gamma_k(1/\rho_z - 1)$ are summarized in the fifth row of Table 2. The 2SLS estimates are encouragingly lower than the OLS estimates but exhibit higher standard errors. Overall, the 2SLS-estimated trade elasticities are positive and exhibit a robust *t*-statistic greater than one for only 1,827 HS10 segments. As before, I also estimate equation (8) using the LIML estimator, which is more robust than the 2SLS estimator in the presence of weak instruments and small sample bias. The results, which are summarized in the last row of Table 2, are qualitatively similar to the 2SLS case. Generally speaking, my trade elasticity estimates are slightly lower than those produced by Caliendo and Parro (2014) and Simonovska and Waugh (2011). These differences may stem from the fact that (i) I explicitly control for the number of firms, $N_{j,kt}$, instead of treating countries as one integrated entity; and (ii) I estimate the trade elasticity at a more refined level of aggregation, which circumvents some of the aggregation biases highlighted in Imbs and Mejean (2015).

Before employing the estimated elasticities to test my theory, let me briefly discuss the plausibility of my estimates. In the online Appendix (Table 2), I report the average estimated demand elasticity and quality intensity for a range of broadly defined industries. The quality intensity is simply measured as $\rho_z = 1/(1 + \theta_z/\gamma_k)$, where γ_k is assigned a value of 2.46 based on the estimates of Eaton et al. (2010). Based on my estimates, the “Machinery” and “Electrical and Optical Equipment” sectors are comprised of the most quality-intensive HS10 segments, while the “Food” and “Paper” sectors are comprised of the least quality-intensive HS10 segments. These results sit well with evidence from consumer surveys that durables are more quality intensive than nondurables (Fornell et al. 1996). Similarly, the “Machinery” and “Electrical and Optical Equipment” sectors are comprised of the least price-elastic segments, while the “Food” and “Agriculture and Mining” sectors include the most price-elastic HS10 segments. These rankings align well with the classification in Campa and Goldberg (1995).

²⁸ Stated formally, the estimating equation is $\ln X_{j,zt}/N_{j,kt} = -\gamma_k(\frac{1}{\rho_z} - 1) \ln \bar{p}_{j,zt} + \tilde{\delta}_{zt} + \xi_{j,zt}$.

TABLE 4—THE RELATIONSHIP BETWEEN QUALITY-INTENSITY AND DEMAND ELASTICITY (ASSUMPTION A4)

Regressor	Dependent: trade elasticity			Dependent: quality intensity		
	OLS	2SLS	LIML	OLS	2SLS	LIML
Demand elasticity, $\log \bar{\varepsilon}_z$	0.346 (0.029)	0.250 (0.051)	0.177 (0.046)	-0.102 (0.007)	-0.115 (0.023)	-0.084 (0.021)
Observations	1,851	435	516	1,851	435	516
Number of SIC Industry FE	95	52	63	95	52	63
R^2	0.45	0.28	0.20	0.39	0.27	0.20

D. Testing Assumption A4: The Link between ρ_z and $\bar{\varepsilon}_z$

The segment-level trade elasticity, $\theta_z = (1/\rho_z - 1)\gamma_k$, identifies the segment-level quality intensity up to an industry-wide Pareto shape parameter, γ_k . Hence, to test Assumption A4, I can use the estimated HS10-level trade and demand elasticities (θ_z and $\bar{\varepsilon}_z$), to run the following regression:²⁹

$$(10) \quad \ln(1/\rho_z - 1)\gamma_k = \beta \cdot \ln \bar{\varepsilon}_z + \delta_k + \xi_z, \quad z \in \mathcal{Z}_k$$

where δ_k controls for SIC industry fixed effects. I estimate the above relationship using the elasticities attained under all estimators—namely, OLS, 2SLS, and LIML. As reported in Table 4, in all incidences the segment-level trade elasticity is positively correlated with the average demand elasticity, $\bar{\varepsilon}_z$, facing individual firms in that segment, with the relationship being statistically significant at the 99 percent confidence level. This positive relationship assures that high-quality-intensive (high- ρ_z) segments of the industry exhibit a lower demand elasticity, which is the assertion of Assumption A4.

When running regression (10), the value of γ_k is irrelevant as it is absorbed by the industry fixed effect. However, I can alternatively test Assumption A4 by calculating $\rho_z = 1/(1 + \theta_z/\gamma_k)$ based on $\gamma_k = 2.46$ from Eaton, Kortum, and Kramarz (2011), and running the following regression instead:

$$\ln \rho_z = \beta \cdot \ln \bar{\varepsilon}_z + \delta_k + \xi_z, \quad z \in \mathcal{Z}_k.$$

The estimation results corresponding to the above regression are reported in the right panel of Table 4. Expectedly, this approach delivers a similar set of results, pointing to a strong, negative relationship between quality-intensity and demand elasticity across segments, which again corroborates Assumption A4.

As a clarifying point, note that the number of observations in Table 4 varies across columns because each of the OLS, 2SLS, and LIML estimators produces a different set of plausible and statistically significant trade and demand elasticity estimates.

²⁹ Assumption A4 only requires that ρ_z and $\bar{\varepsilon}_z$ be positively related. The parametric form of the relationship is inconsequential. Nonetheless, the online Appendix (Table 4) also reports the relationship between ρ_z and $\bar{\varepsilon}_z$ in levels (rather than in logs), which yields similar outcomes.

For 1,851 HS10 segments (spanning 95 SIC industries) the OLS estimation delivers a statistically significant estimate of both $\bar{\varepsilon}_z$ and θ_z , which can be used to estimate equation (10). The 2SLS and LIML estimators, while conceivably more reliable, produce statistically significant trade and demand elasticities for a fewer number of SIC industries and HS10 segments.³⁰

E. The Composition of Industry-Level Exports

The previous section presented empirical support for Assumption A4, as the key condition that prompts across-segment specialization in Proposition 1. Now, I turn to *directly* testing the outcomes predicted by Proposition 1 and by Lemma 1. To test Lemma 1, I use my trade elasticity estimates to calculate the degree of quality intensity, ρ_z , per HS10 segment in order to construct the following measure:

$$\bar{\rho}_{j,kt} = \frac{\sum_{z \in \mathcal{Z}_k} \rho_z X_{i,zt}}{\sum_{z \in \mathcal{Z}_k} X_{i,zt}}.$$

To be more specific, $\bar{\rho}_{j,kt}$ measures the average quality intensity of country j 's exports to Colombia in SIC industry k . Lemma 1 predicts that $\bar{\rho}_{j,kt}$ is increasing in country j 's income per worker, $w_{j,t}$, and its distance to Colombia, $Dist_j$; which I can test by running the following regression:

$$\ln \bar{\rho}_{j,kt} = \beta_1 \ln w_{j,t} + \beta_2 \ln Dist_j + \beta_3 \ln Y_{j,t} + \delta_{kt} + \xi_{j,kt}.$$

In the above regression, δ_{kt} accounts for industry-year fixed effects, and $Y_{j,t}$ controls for country j 's total GDP. The results corresponding to this regression are reported in the left panel of Table 5, and confirm the prediction that high-wage and distant economies export relatively more in high-quality-intensive segments of each industry.

Similarly, to test Proposition 1, I use my estimated HS10-level demand elasticities to construct the following measure:

$$\overline{\mathcal{M}\mathcal{P}}_{j,kt} = \frac{\sum_{z \in \mathcal{Z}_k} X_{i,zt} / \bar{\varepsilon}_z}{\sum_{z \in \mathcal{Z}_k} X_{i,zt}},$$

which reflects the average market power of firms' exporting from country j to Colombia in SIC industry k . Proposition 1 predicts that $\overline{\mathcal{M}\mathcal{P}}_{j,kt}$ is increasing in country j 's income per worker, $w_{j,t}$, and its distance to Colombia, $Dist_j$, which I can test by running the following regression:

$$\ln \overline{\mathcal{M}\mathcal{P}}_{j,kt} = \beta_1 \ln w_{j,t} + \beta_2 \ln Dist_j + \beta_3 \ln Y_{j,t} + \delta_{kt} + \xi_{j,kt}.$$

³⁰Industries that produce significant 2SLS and LIML estimates are typically those with significantly more observations. This is important to note as one may be concerned with the issue of endogenous selection.

TABLE 5—THE COMPOSITION OF COUNTRY-LEVEL EXPORTS TO COLOMBIA

Regressor	Average quality intensity			Average market power		
	OLS	2SLS	LIML	OLS	2SLS	LIML
$\ln w_j$	0.005 (0.001)	0.001 (0.002)	0.004 (0.002)	0.013 (0.002)	0.021 (0.004)	0.031 (0.004)
$\ln Dist_j$	0.014 (0.001)	0.023 (0.002)	0.027 (0.002)	0.017 (0.003)	0.029 (0.004)	0.028 (0.005)
$\ln Y_j$	-0.005 (0.001)	-0.005 (0.001)	-0.006 (0.001)	-0.010 (0.001)	-0.019 (0.002)	-0.028 (0.003)
Observations	16,207	15,582	15,353	23,095	14,896	16,555
R^2	0.59	0.59	0.56	0.61	0.55	0.50

As before, δ_{kt} accounts for industry-year fixed effects and $Y_{j,t}$ controls for county j 's total GDP. The results corresponding to this regression are reported in the right panel of Table 5, and again decisively confirm the prediction that high-wage and distant economies export relatively more in high-market power segments of each industry. While these results are encouraging, one should exercise caution when interpreting them. Specifically, as noted earlier, the findings presented in Table 5 may be an artifact of factor-driven specialization. That is, it can be the case that distant and high-wage economies have a factor-driven comparative advantage in low- θ and low- ε segments, which reinforces the mechanism highlighted in this paper.

F. Discussion and Auxiliary Evidence

Here, I will briefly discuss the extensions and limitations of my analysis. Before doing so, let me take stock. This section estimated two key elasticities for a range of HS10 segments: (i) the trade elasticity θ_z , which reflects the collective market power of a country in segment z ; and (ii) the firm-level demand elasticity, $\bar{\varepsilon}_z$, which reflects the average market power of individual firms in that segment. Following Chaney (2008), it has been widely believed that the link between θ_z and $\bar{\varepsilon}_z$ is theoretically unclear. The results in this section, however, established a positive *empirical* link between θ_z and $\bar{\varepsilon}_z$. This relationship plus trade flows being governed by a gravity equation suffice to prompt specialization based on the mechanism introduced in this paper.

The assumption that γ_k does not vary across segments is not consequential to the conclusions in this section. More specifically, the across segment variation in the trade elasticity, θ_z , may very well reflect across-segment heterogeneity in γ . But this does not cast a problem for my analysis, as Proposition 1 holds so long as θ_z is positively correlated with $\bar{\varepsilon}_z$ across segments and irrespective of what triggers the variation in θ_z .

Looking closer at the evidence, the theory developed here can also explain *across-industry* and *within-segment* specialization. Specifically, as shown in the online Appendix (Table 3), the negative relationship between ρ_z and $\bar{\varepsilon}_z$ prevails not only within industries but also across industries. Per Proposition 1, such a relationship will prompt specialization across low- and high-markup (or low- and

TABLE 6—EVIDENCE ON WITHIN-SEGMENT SPECIALIZATION

Regressor	Estimator		
	LIML	2SLS	OLS
$\Delta \ln p_{\omega,zt}(-\varepsilon_c)$	-5.509 (0.733)	-3.787 (0.355)	-0.702 (0.005)
$\Delta \ln p_{\omega,zt} \times \ln Dist_j(-\varepsilon_\tau)$	4.456 (1.246)	3.001 (0.591)	-0.109 (0.009)
$\Delta \ln p_{\omega,zt} \times \ln w_j(-\varepsilon_w)$	16.586 (2.018)	9.804 (0.834)	0.562 (0.012)
Kleibergen-Paap LM statistic	235.4
Kleibergen-Paap LM p -value	0.000
Kleibergen Paap Wald rk F -statistic	25.32
Anderson-Rubin Wald F -statistic	82.22
Anderson-Rubin Wald p -value	0.000
Observations	1,279,875		

high-scale-intensive) industries. Moreover, a similar pattern of specialization may occur within HS10 segments. In fact, I find basic evidence that high-wage and distant economies specialize in high-markup niches of HS10 segments. The evidence is based on the following demand estimation, which admits elasticity heterogeneity across suppliers within an HS10 segment:

$$\Delta \ln q_{\omega,zt} = -(\varepsilon_c + \varepsilon_\tau \cdot \ln Dist_j + \varepsilon_w \cdot \ln w_j) \Delta \ln p_{\omega,zt} + \Delta \delta_{zt} + \Delta \xi_{\omega,zt}, \quad \omega \in \Omega_{j,zt}.$$

In the above equation, $\Delta \delta_{zt}$, as before, controls for HS10-year fixed effects, while $Dist_j$ and w_j denote the normalized distance and income per worker of country j , which applies to each firm ω exporting from that country. I use the same methodology, explained earlier, to estimate the above equation. The estimation results presented in Table 6 indicate that firm-level exports from high-wage and distant economies face significantly lower demand elasticities, which is indicative of specialization in high-market power niches within HS10 segments.

The obvious limitation of the present analysis is its reliance on the monopolistic competition assumption. This assumption is widely used but difficult to test. Importantly, the link between the average firm-level demand elasticity and the average markup should be amended if the monopolistic competition assumption is violated. That being said, in line with Hottman, Redding, and Weinstein (2016), I find that in most product categories in my sample, the number of exporting firms is large enough to render the monopolistic competition assumption plausible. The other shortcoming, which is due to a lack of better data, is the inability to test Lemma 1 and Proposition 1, while controlling for factor-driven specialization at the firm level.

IV. Macro-Level Implications

Within-industry specialization based on Proposition 1 determines the distribution of economic activity across industry segments that differ in their (i) degree of

market power, (ii) degree of returns to scale, and (iii) trade elasticity. The latter two characteristics determine the extent to which countries can gain from international specialization or trade. So, how national resources are divided along these margins can profoundly influence macroeconomic outcomes. Below, I fit my theoretical model to aggregate trade data to quantify these macro-level implications.

A. Quantitative Strategy

To set the stage for a transparent presentation of macro-level implications, I first describe the quantitative strategy by which the model is fitted to data. The strategy is closely related to a structural gravity estimation à la Anderson and van Wincoop (2003).³¹ That is, I identify the structural parameters of the model by fitting it to the matrix of aggregate bilateral trade values. However, unlike the standard structural gravity estimation, I allow for industries to be comprised of multiple segments with different trade elasticities.

I estimate the model using two distinct datasets. First, the 2008 version of the World Input Output Database (WIOD), which reports industry-level trade and production data for 32 major economies across 34 industries (15 of which are tradable). Estimating the model with this dataset will allow me to quantify the gains from within-industry specialization. Second, I also estimate the model using country-level trade and production data, which covers 100 economies in the year 2000. Estimating the model with this second dataset allows me to study cross-country TFP differences for a wider range of economies. Appendix D provides a more detailed description of both datasets.

Parametric Restrictions.—The theoretical model presented in Section II only required preferences to be additively separable across industries. But to map the model to data, I need to impose more specific structure on the cross-industry utility aggregator. Considering this, I impose the following restriction, which allows me to estimate the model separately for each industry.

R1. The cross-industry utility aggregator is Cobb-Douglas. That is,
$$W_i = \prod_{k=1}^K \mathcal{U}_{i,k}^{\alpha_{i,k}}, \text{ where } \sum_k \alpha_{i,k} = 1.$$

The above restriction is a standard staple in the quantitative trade literature (see Costinot and Rodríguez-Clare 2014). Since the industry-level expenditure shares, $\alpha \equiv [\alpha_{i,k}]$, are observable, Restriction R1 reduces the estimation problem to that of matching an $\mathcal{N} \times (\mathcal{N} - 1)$ matrix of trade values, $\mathbf{X}_k \equiv [X_{ji,k}]$, separately for each industry k .

To characterize the industry-level trade values, I further impose the following restriction on the demand structure across segments within an industry.

³¹ See Head and Mayer (2014) for a review of the literature on structural gravity estimation.

R2. Each industry k is comprised of two segments, $\mathcal{Z}_k = \{\mathcal{H}, \mathcal{L}\}$; with $e_{i,z}$ denoting the constant share of expenditure on segment z in country i .

Unlike α_k , the segment-level expenditure shares, $\mathbf{e}_k \equiv [e_{i,z}]$, are not observable. So, I treat them as free-moving parameters that are estimated. Given R1 and R2, the industry-level bilateral trade values, $X_{ji,k}$, can be calculated as

$$X_{ji,k} = \sum_{z \in \mathcal{Z}_k} \lambda_{ji,k} e_{i,z} \alpha_{i,k} w_i L_i$$

where $\lambda_{ji,z}$ denotes the share of country i 's expenditure on country j varieties in segment z of industry k . Following the gravity equation (4), derived in Section II, $\lambda_{ji,z}$ can be expressed as follows:

$$(11) \quad \lambda_{ji,z} = \frac{\mathcal{T}_{j,k} (w_j \tau_{ji,k})^{-\theta_z}}{\sum_{n=1}^{\mathcal{N}} \mathcal{T}_{n,k} (w_n \tau_{ni,k})^{-\theta_z}}, \quad z \in \mathcal{Z}_k.$$

where $\theta_z = \gamma_k(\epsilon_z - 1)$ denotes the segment-level trade elasticity, and $\mathcal{T}_{j,k} \equiv \mathcal{A}_{j,k} M_{j,k}$. The implicit assumption underlying the above equation is that technology parameters are segment-neutral, i.e., $\mathcal{A}_{j,z} = \mathcal{A}_{j,k} \mathcal{A}_z$. Correspondingly, \mathcal{A}_z drops out of the gravity equation and $\mathcal{T}_{j,k} \equiv \mathcal{A}_{j,k} M_{j,k}$ can be treated as an *industry* \times *exporter* fixed effect that is estimated. Considering the above, $\boldsymbol{\lambda}_z \equiv [\lambda_{ji,z}]$ is fully determined as a function of the segment-level trade elasticity, θ_z , country-level wages, $\mathbf{w} \equiv [w_j]$, the vector of exporter fixed effects, $\boldsymbol{\mathcal{T}}_k \equiv [\mathcal{T}_{j,k}]$, and the industry-wide matrix of variable trade costs $\boldsymbol{\tau}_k \equiv [\tau_{ji,k}]$.

If one were to treat the elements of $\boldsymbol{\tau}_k$ as free-moving parameters, the estimation will be overidentified, as the matrix of trade flows, $\mathbf{X}_k = [X_{ji,k}]$, can be perfectly matched by freely choosing the elements of $\boldsymbol{\tau}_k$. So, in line with the literature, I handle the overidentification problem by imposing the following parametric restriction on the bilateral trade costs.

R3. The iceberg trade costs assume the following parametrization:

$$\tau_{ji,k} = \beta_k \cdot \beta_{border,k} \cdot \beta_{lang,k} \cdot \beta_{agreement,k} (\text{Dist}_{ji})^{\beta_{dist,k}},$$

where Dist_{ji} denotes the distance (in thousands of kilometers) between countries j and i . The terms β_{border} , β_{lang} , $\beta_{agreement}$ are, respectively, “one” if countries j and i do not have a common border, common language, and trade agreement, and an estimated parameter otherwise.

An analogue of the above parameterization, which imposes symmetry in trade costs, underlies many applications including Anderson and van Wincoop (2003),

Fieler (2011), and Caliendo and Parro (2014).³² Based on R3, the matrix of trade costs in industry k is fully characterized by a vector of parameters, $\beta_k \equiv \{\beta_k, \beta_{border,k}, \beta_{lang,k}, \beta_{agreement,k}, \beta_{dist,k}\}$, plus data on distance, trade agreements, common-language, and borders, which I henceforth refer to as, \mathbf{D} . Correspondingly, given equation (11), the matrix of trade shares in segment $z \in \mathcal{Z}_k$ can be fully calculated as function of \mathbf{w} and \mathbf{D} , and estimated parameters \mathcal{T}_k , β_k , and θ_z ; namely, $\lambda_z = \lambda(\theta_z, \beta_k, \mathcal{T}_k; \mathbf{w}, \mathbf{D})$.

Subsequently, the entire matrix of industry-level trade flows can be computed as a function of free-moving parameters $\theta_{\mathcal{H}}$, $\theta_{\mathcal{L}}$, \mathcal{T}_k , β_k , and \mathbf{e}_k , plus data on wage, population size, geo-political characteristics, and industry-level expenditure shares (namely, \mathbf{w} , \mathbf{L} , \mathbf{D} , α_k) as follows:

$$(12) \quad X_{ji,k} = X_{ji,k}(\theta_{\mathcal{H}}, \theta_{\mathcal{L}}, \mathbf{e}_k, \mathcal{T}_k, \beta_k; \alpha_k, \mathbf{w}, \mathbf{L}, \mathbf{D}) \\ = \sum_{z \in \mathcal{Z}_k} \frac{\mathcal{T}_{j,k}(w_j \tau_{ji,k}(\beta_k; \mathbf{D}))^{-\theta_z}}{\sum_{n=1}^{\mathcal{N}} \mathcal{T}_{n,k}(w_n \tau_{ni,k}(\beta_k; \mathbf{D}))^{-\theta_z}} e_{i,z} \alpha_{i,k} w_i L_i.$$

The above estimation is subject to a basic identification challenge, which is well known in the structural gravity literature. That is, the effect of both $\theta_{\mathcal{H}}$ and $\theta_{\mathcal{L}}$ on trade values cannot be separately identified from that of β_k . The standard solution to this problem is to set the trade elasticity to a value, which is externally estimated with auxiliary data. This solution is, however, not fully applicable here, given that the trade elasticity is estimable only up to the same level of aggregation at which multilateral trade flows are reported. To handle this issue, I impose the following restriction that trade elasticities estimated with industry-level data are a trade-weighted average of the segment-level elasticities.

R4. If $\hat{\theta}_k$ is a trade elasticity estimated using industry-level trade data, then

$$\hat{\theta}_k = \frac{\sum_{j,i} \sum_{z \in \mathcal{Z}_k} \theta_z X_{ji,z}}{\sum_{j,i} \sum_{z \in \mathcal{Z}_k} X_{ji,z}},$$

where θ_z is the unobservable segment-level trade elasticity.

To elaborate on Restriction R4, consider my main dataset that reports trade flows at the level of WIOD industries. Caliendo and Parro (2014) merge the WIOD data with industry-level tariff data to estimate the trade elasticity, $\hat{\theta}_k$, for each WIOD industry. R4 allows me to relate their estimated elasticity, $\hat{\theta}_k$, to segment-level elasticities, $\theta_{\mathcal{H}}$ and $\theta_{\mathcal{L}}$, for each industry. In other words, per R4, I can take the externally

³² An alternative approach is to allow for reduced-form asymmetries in trade costs as in Waugh (2010). However, as noted by Waugh (2010), such asymmetries “may be reduced-form representations of equilibrium responses to the fundamentals faced by agents.” As I will argue shortly, and in-line with the quoted remark, even under R2 the present model features asymmetries in *real* trade costs that arise due to equilibrium responses.

estimated $\hat{\theta}_k$ and estimate a spread, $\rho_k \equiv \theta_{\mathcal{L}}/\theta_{\mathcal{H}}$, to uniquely determine both $\theta_{\mathcal{H}}$ and $\theta_{\mathcal{L}}$ as follows:

$$(13) \quad \theta_{\mathcal{H}}(\rho_k, \mathbf{X}_{\mathcal{H}}, \mathbf{X}_{\mathcal{L}}; \hat{\theta}_k) = \left(\frac{\sum_{j,i} \sum_{z=\mathcal{H},\mathcal{L}} X_{ji,z}}{\sum_{j,i} X_{ji,\mathcal{H}} + \rho^{-1} \sum_{j,i} X_{ji,\mathcal{L}}} \right) \hat{\theta}_k,$$

$$\theta_{\mathcal{L}}(\rho_k, \mathbf{X}_{\mathcal{H}}, \mathbf{X}_{\mathcal{L}}; \hat{\theta}_k) = \left(\frac{\sum_{j,i} \sum_{z=\mathcal{H},\mathcal{L}} X_{ji,z}}{\rho \sum_{j,i} X_{ji,\mathcal{H}} + \sum_{j,i} X_{ji,\mathcal{L}}} \right) \hat{\theta}_k.$$

Importantly, unlike the trade elasticity levels, the spread $\rho_k \equiv \theta_{\mathcal{L}}/\theta_{\mathcal{H}}$ can be separately identified from β_k .³³

The Estimation Problem.—Given parametric restrictions R1–R4; data on wage and population size, \mathbf{w} and \mathbf{L} , geo-political variables, \mathbf{D} , industry-level expenditure shares, α_k , and revenue shares, $\mathbf{r}_k = [r_{j,k}]$; we can estimate the vector of parameters, $\Theta_k = \{\rho_k, \mathbf{e}_k, \mathcal{T}_k, \beta_k\}$, separately for each industry to match the factual matrix of industry-level trade values, $\mathbf{X}_k^D = [X_{ji,k}^D]$. The estimation problem can be formally stated as follows:

$$\min_{\Theta_k} \sum_{j=1}^N \sum_{i \neq j} \left(\ln X_{ji,k}^D - \ln X_{ji,k}(\Theta_k; \alpha_k, \mathbf{w}, \mathbf{L}, \mathbf{D}) \right)^2$$

subject to

$$X_{ji,k}(\cdot) = \sum_{z=\mathcal{H},\mathcal{L}} \lambda_{ji,z} e_{i,z} \alpha_{i,k} w_i L_i \quad \forall i, j,$$

$$\lambda_{ji,z} = \mathcal{T}_{j,k} (w_j \tau_{ji,k})^{-\theta_z} / \left(\sum_{n=1}^N \mathcal{T}_{n,k} (w_n \tau_{ni,k})^{-\theta_z} \right) \quad \forall i, j; \quad z = \mathcal{H}, \mathcal{L},$$

$$\tau_{ji,k} = \beta_k \cdot \beta_{border,k} \cdot \beta_{lang,k} \cdot \beta_{agreement,k} (\text{Dist}_{ji})^{\beta_{dist,k}} \quad \forall i, j,$$

$$\theta_z = \theta_z(\rho_k, \mathbf{X}_{\mathcal{H}}, \mathbf{X}_{\mathcal{L}}; \hat{\theta}_k), \quad z = \mathcal{H}, \mathcal{L},$$

$$(LMC) \quad \sum_{i=1}^N X_{ji,k} = r_{j,k} w_j L_j \quad \forall j,$$

where the last constraint corresponds to the equilibrium labor market-clearing condition (LMC), and $\theta_k(\cdot)$ is described by (13). The above estimator is a basic extension of the structural gravity estimation in Anderson and van Wincoop (2003). More specifically, the structural gravity estimation in Anderson and van Wincoop (2003) corresponds to a restricted version of the above estimator that imposes $\rho_k \equiv \theta_{\mathcal{L}}/\theta_{\mathcal{H}} = 1$. Following Anderson and van Wincoop (2003), the nonlinear least square (NLLS) estimator, specified above, is unbiased if $\varepsilon \equiv X^D - X$

³³ See Fieler (2011) for a thorough discussion on how the trade elasticity spread, ρ_k , is separately identified from the trade cost parameters, β_k .

is uncorrelated with the derivative of $X_{ji,k}(\Theta_k; \alpha_k, \mathbf{w}, \mathbf{L}, \mathbf{D})$ with respect to $\alpha_k, \mathbf{w}, \mathbf{L}$, and \mathbf{D} , which is the case if ε represents measurement errors.

As noted earlier, I run the above estimation using two different datasets. First, I estimate the model with country-level trade data for a sample of 100 countries. The country-level estimation treats each economy as one integrated industry. Second, I estimate the model separately for 15 *traded* WIOD industries using industry-level bilateral trade and production data for 32 major economies. The former estimation is used to study the implications of the model for international TFP differences. The latter estimation is used to quantify the gains from within-industry specialization; Appendix D provides further details about each estimation.

Finally, to conduct the above estimation, I borrow trade elasticity estimates (namely, $\hat{\theta}_k$) from the literature. For the industry-level estimation conducted on the WIOD sample, I take the elasticities from Caliendo and Parro (2014), with details provided in Table D2. For the country-level estimation, where each country is treated as one integrated industry, I set $\hat{\theta} = 4$ as in Anderson and van Wincoop (2003). This choice according to the meta-analysis in Head and Mayer (2014) approximates the median elasticity estimated in the existing literature.

In-Sample Fit.—Before moving forward, a brief discussion regarding the in-sample fit of the model is in order (a more thorough discussion is also provided in Appendix D). I center the discussion here around the model estimated using country-level data, noting that similar arguments apply to the industry-level estimation. The country-level estimation (where $\mathcal{N} = 100$) fits the model to $100 \times (100 - 1) = 9,900$ data points on bilateral trade values.³⁴ Figure 1 displays the in-sample fit of the estimated model. Table D1 (in the Appendix) also reports a summary of the main estimated parameters and compares them those attained under the standard structural gravity estimation à la Anderson and van Wincoop (2003). A notable contrast between the two estimations is that the main model exhibits an $R^2 = 0.65$, which is significantly higher than that of the standard structural gravity estimation.³⁵

As discussed in Appendix D, the superior in-sample fit of the main model is due to its ability to match (i) the low volume of trade between similar trading partners (South-South and North-North trade) as well as (ii) the high-volume of trade between dissimilar partners (North-South trade). The standard gravity estimation, by comparison, has difficulty matching the differential levels of similar-similar and dissimilar-dissimilar trade (see Figure D1 in the Appendix). This feature of the present model is comparable to that of Fieler (2011).

³⁴The number of estimated parameters in the country-level case is 206, which includes 100 exporter fixed effects (labeled \mathcal{T}), 100 country-level expenditure shares on segment \mathcal{H} (labeled \mathbf{e}), 6 trade cost parameters (labeled β), plus an across-segment trade elasticity spread (labeled ρ).

³⁵Another way of enhancing the in-sample fit of the structural gravity model is to relax the symmetric trade cost assumption. In that case, trade is no longer bilaterally balanced and the exporter and importer fixed effects need not to be equal. In fact, with asymmetric trade costs, a reduced-form gravity estimation that uses a PPML estimator, becomes isomorphic to the structural gravity estimation (see Head and Mayer 2014).

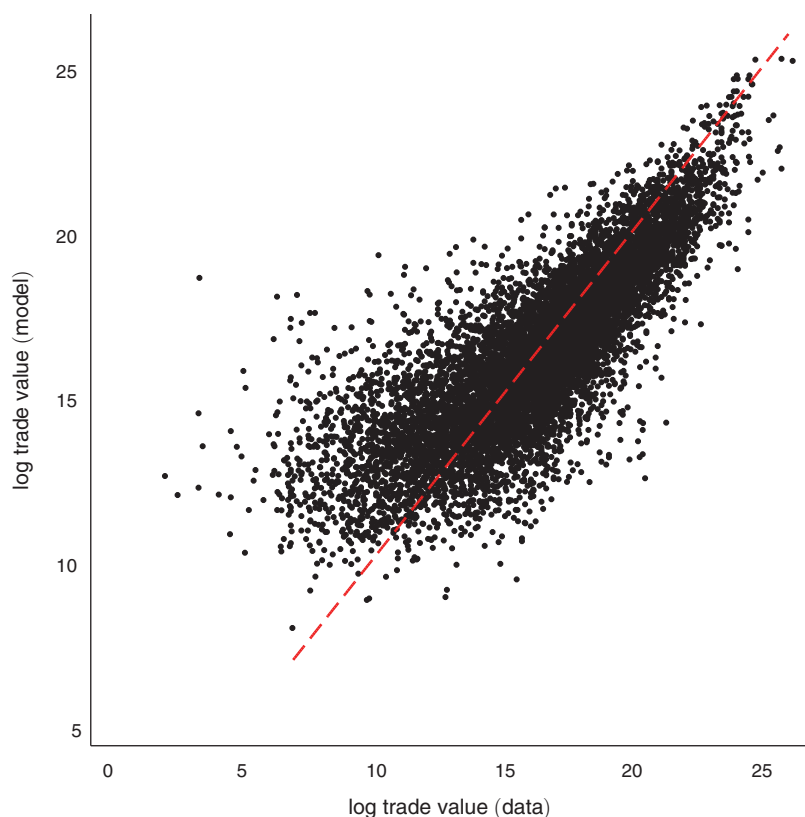


FIGURE 1. FIT OF THE MODEL ESTIMATED WITH COUNTRY-LEVEL DATA

Out-of-Sample Prediction with respect to Markups.—The initial motivation behind my theory were two facts: that export price and markup levels increase with the exporting economy’s (i) income per capita, and (ii) distance to global markets. Encouragingly, the estimated model reproduces these *out-of-sample* facts relatively well. To demonstrate this, I suppose the underlying demand system is CES, so that markups are constant within each segment. Based on this assumption, I can easily compute the model-implied average export markup for each country j in the sample, labeling it \hat{m}_j . After calculating the export markup levels, I run the following regression,

$$\ln \hat{m}_j = \underset{(0.005)}{0.07} \ln GDPcap_j + \underset{(0.006)}{0.32} \ln \overline{Dist}_{ji} + \epsilon_j,$$

which implies that the predicted export markup level increases significantly with both income per capita, $GDPcap_j$, and remoteness, \overline{Dist}_{ji} . In comparison, the standard gravity model that does not admit low- and high-markup segments within industries, predicts a uniform markup across all exporters.

B. The Gains from Within-Industry Specialization

In this section, I use the model estimated with industry-level data to compute the gains from within-industry specialization. As shown in Appendix C, the overall gains from trade for country i can be calculated using the following formula:³⁶

$$(14) \quad GT_i = 1 - \prod_{k=1}^{\mathcal{K}} \prod_{z \in \mathcal{Z}_k} (\mathcal{S}_{i,k} \lambda_{ii,z})^{e_{i,z} \alpha_{i,k} / \theta_z},$$

where $\lambda_{ii,z}$ denotes the share of country i 's expenditure on domestic varieties in segment z , and $\mathcal{S}_{i,k} \equiv \alpha_{i,k} / r_{i,k} \left[\left(\sum_{z \in \mathcal{Z}_k} e_{i,z} / \epsilon_z \right) / \left(\sum_{z \in \mathcal{Z}_k} r_{i,z} / \epsilon_z \right) \right]$ is a term that accounts for the scale-driven gains—as before, $r_{i,k}$ is used to denote the share of country i 's output generated in industry k and, correspondingly, $r_{i,z}$ denotes the (within-industry) fraction of output generated in segment $z \in \mathcal{Z}_k$. The above equation is a simple extension of the celebrated Arkolakis, Costinot, and Rodríguez-Clare (2012—henceforth, ACR) formula. The challenge in evaluating equation (14) is that it requires data on segment-level variables, $\lambda_{ii,z}$, $r_{i,z}$, and θ_z . The estimation outlined in the previous section identifies these exact variables from industry-level trade and production data as well as industry-level trade elasticity estimates (from Caliendo and Parro 2014). Since among the estimated parameter values the segment-level elasticities, $\theta_{\mathcal{H}}$ and $\theta_{\mathcal{L}}$, are especially key to equation (14), Table D2 explicitly reports them for each of the WIOD industries.

Equation (14) determines the overall gains from trade, which include the traditional ACR-type gains plus the gains from within-industry specialization. Hence, to determine the *pure* gains from within-industry specialization, I need to purge GT_i from the ACR-driven gains. To pin down the ACR-driven gains, one can shut down heterogeneity within industries to eliminate within-industry specialization across low- and high- θ segments. Doing so, the gains from trade will be driven solely by traditional forces, and following Costinot and Rodríguez-Clare (2014) will be given by

$$GT_i^{ACR} = 1 - \prod_{k=1}^{\mathcal{K}} (\mathcal{S}_{i,k}^{ACR} \lambda_{ii,k})^{\alpha_{i,k} / \theta_k},$$

with $\mathcal{S}_{i,k}^{ACR} \equiv \alpha_{i,k} / r_{i,k}$ accounting for scale effects in the traditional ACR class of models. To evaluate the above equation, I take an identical approach to that of Costinot and Rodríguez-Clare (2014), which is to use industry-level data on $\lambda_{ii,k}$, $r_{i,k}$, and $\alpha_{i,k}$ from the WIOD, plus trade elasticity estimates, θ_k , from Caliendo and Parro (2014). After computing GT_i and GT_i^{ACR} , I calculate the pure gains from within-industry specialization for every country i as follows:

$$\text{Gains from specialization} = GT_i - GT_i^{ACR}.$$

Table 7 reports the gains from specialization, along with the overall and ACR-driven gains from trade. In summary, the main model featuring within-industry specialization

³⁶The above formula implicitly assumes that the within-segment demand is CES and markups are constant. Nonetheless, as it will become clear, variable markups are inconsequential to the insights presented hereafter.

TABLE 7—THE GAINS FROM WITHIN-INDUSTRY SPECIALIZATION

Country	GDP (US = 1)	Population (US = 1)	Remoteness (US = 1)	Gains from trade		Gains from specialization (%)
				ACR (%)	Main model (%)	
AUS	0.07	0.07	1.47	3.7	16.7	12.9
AUT	0.03	0.03	0.41	30.5	33.7	3.1
BEL	0.04	0.04	0.57	32.4	57.6	25.3
BRA	0.11	0.63	0.74	4.3	9.5	5.2
CAN	0.10	0.11	0.98	15.3	25.2	9.9
CHN	0.31	4.36	0.59	4.0	8.3	4.3
CZE	0.02	0.03	0.40	21.2	19.0	-2.2
DEU	0.26	0.27	0.53	17.6	32.7	15.1
DNK	0.02	0.02	0.52	24.8	25.1	0.3
ESP	0.11	0.15	0.54	9.5	24.3	14.8
FIN	0.02	0.02	0.47	10.5	10.3	-0.3
FRA	0.20	0.21	0.49	11.1	23.4	12.3
GBR	0.19	0.20	0.62	11.7	21.3	9.6
GRC	0.02	0.04	0.53	4.7	9.1	4.4
HUN	0.01	0.03	0.46	31.2	23.9	-7.3
IDN	0.03	0.77	0.72	4.0	6.4	2.4
IND	0.08	3.86	0.64	4.3	16.7	12.4
IRL	0.02	0.01	0.53	14.2	14.7	0.5
ITA	0.16	0.19	0.45	9.2	21.6	12.4
JPN	0.33	0.42	0.53	3.7	11.4	7.7
KOR	0.07	0.16	0.63	8.6	14.3	5.7
MEX	0.07	0.38	0.80	12.1	15.6	3.5
NLD	0.06	0.05	0.73	23.1	43.4	20.3
POL	0.04	0.13	0.40	19.7	28.0	8.3
PRT	0.02	0.03	0.48	20.6	15.6	-5.0
ROM	0.01	0.07	0.36	12.7	12.5	-0.2
RUS	0.11	0.47	0.65	0.9	19.4	18.5
SVK	0.01	0.02	0.52	23.6	16.2	-7.4
SVN	0.00	0.01	0.55	39.3	24.5	-14.8
SWE	0.03	0.03	0.47	14.5	19.4	4.9
TUR	0.05	0.23	0.51	13.3	23.8	10.6
TWN	0.03	0.08	0.98	9.9	14.6	4.7
USA	1.00	1.00	1.00	3.8	8.6	4.8
RoW	0.61	8.00	2.49	7.3	14.8	7.5

predicts gains from trade that are, on average, 46 percent larger than those implied by the ACR formula (20.4 percent versus 14 percent increase in real GDP). Based on these numbers, one can deduce that within-industry specialization alone increases the real GDP by around 6.4 percent for the average economy.

The gains from specialization are, however, quite heterogeneous across countries. High-wage or remote economies like Germany, France, and Australia gain significantly more as they specialize in high-scale-intensive segments, where returns to specialization are higher. In comparison, the pure gains from specialization are considerably smaller (and even negative) for developing economies like Mexico and Indonesia, as they specialize in low-scale-intensive segments where returns to specialization are relatively low.

C. International Specialization and TFP Differences

Beyond the gains from trade, the present model sheds new light on the profound gap in the standard of living across rich and poor countries. As noted by

Jones (2011), “by the end of the twentieth century, per capita income in the United States was more than 50 times higher than per capita income in Ethiopia and Tanzania. Dispersion across the ninety-fifth–fifth percentiles of countries was more than a factor of 32.” There is an abundance of theories that seek to explain these persistent gaps.³⁷ The general theme of such theories is to identify multiplier effects that amplify cross-country income/TFP differences.

The theory of international specialization presented here identifies a possibly new multiplier effect. To demonstrate this, consider the special case of constant markups presented in Section IIB. Let $L_{i,k}$ denote the number of workers engaged in industry k ; free entry entails that country i will host a mass $\ell_{i,k} \equiv \left(\sum_{z \in \mathcal{Z}_k} r_{i,z}/\epsilon_z\right) L_{i,k}/f_e^k$ of local firms in that industry. Consequently, the real income per worker in country i will be given by (see Appendix C)³⁸

$$(15) \quad W_i = A_i \underbrace{\left(\prod_{k=1}^{\mathcal{K}} \prod_{z \in \mathcal{Z}_k} \ell_{i,k}^{\alpha_{i,k} e_{i,z}/\theta_z}\right)}_{\text{scale effects}} \underbrace{\left(\prod_{k=1}^{\mathcal{K}} \prod_{z \in \mathcal{Z}_k} \lambda_{ii,z}^{-\alpha_{i,k} e_{i,z}/\theta_z}\right)}_{\text{trade openness}},$$

where A_i , the fundamental TFP of country i , reflects its absolute advantage in producing high-quality output; $\ell_{i,k}$ reflects the effective scale of country i in industry k ; and $\theta_z = \gamma_k/(\psi_z - 1)$ is the segment-level trade elasticity, which is inversely related to the segment-level scale elasticity $\psi_z = \epsilon_z/(\epsilon_z - 1)$ (as defined in Section II). Aside from the fundamental TFP, A_i , the above equation states that real income depends on the scale-driven gains from specialization (first parentheses) and trade openness (second parenthesis).

Given Proposition 1, the scale-driven term operates as a TFP multiplier. To elaborate, note that all else equal, high- A countries pay higher equilibrium wages, which is easily verifiable using the balanced trade condition. Given Proposition 1, the higher wage induces high- A countries to specialize in low- ϵ_z segments, which leads to a higher $\ell_{i,k}$. Relatedly, high- A economies also specialize in industries that exhibit on average a higher scale elasticity. Together, these effects multiply the TFP differences, whereby the scale-driven term, $\prod_{k=1}^{\mathcal{K}} \prod_{z \in \mathcal{Z}_k} \ell_{i,k}^{\alpha_{i,k} e_{i,z}/\theta_z}$, is systematically larger for high- A_i economies.

To calculate the contribution of *specialization-driven* scale effects to income differences, note that the scale-driven component in equation (15) can be decomposed into (i) a standard size-driven component, and (ii) a specialization-driven component as follows:

$$\prod_{k=1}^{\mathcal{K}} \prod_{z \in \mathcal{Z}_k} \ell_{i,k}^{\alpha_{i,k} e_{i,z}/\theta_z} = \underbrace{\left(\prod_{k=1}^{\mathcal{K}} L_{i,k}^{\alpha_{i,k}/\theta_k}\right)}_{\text{size-driven}} \underbrace{\left[\left(\prod_{k=1}^{\mathcal{K}} \prod_{z \in \mathcal{Z}_k} \ell_{i,k}^{\alpha_{i,k} e_{i,z}/\theta_z}\right) / \left(\prod_{k=1}^{\mathcal{K}} L_{i,k}^{\alpha_{i,k}/\theta_k}\right)\right]}_{\text{specialization-driven}},$$

³⁷ See Klenow and Rodríguez-Clare (1997); Prescott (1998); Hall and Jones (1999); Parente and Prescott (1999); Acemoglu, Johnson, and Robinson (2000); Parente and Prescott (2000); Caselli (2005); and Jones (2011).

³⁸ In the above expression $A_i \equiv \varsigma \prod_{k=1}^{\mathcal{K}} \prod_{z \in \mathcal{Z}_k} \mathcal{A}_{i,z}^{e_{i,z}/\gamma_k(1-\sigma_z)}$, where ς encapsulates a combination of constant parameters. Also, the equation assumes firms are symmetric, but can be trivially extended to account for firm heterogeneity and selection effects—see Kucheryavy, Lyn, and Rodríguez-Clare (2016).

where θ_k is the industry-level trade elasticity that is itself a weighted average of the segment-level trade elasticities per R4. Note that the size-driven component is relatively well-known, as it is also present in the standard ACR class of models. But the specialization-driven component is unique to the present model, and absent (or equal to one for all countries) in the ACR class of models.

Considering the above decomposition, we can appeal to the estimated model to calculate the counterfactual real income of each country i , net of specialization-driven scale effects. Doing so using the model fitted to country-level trade data for 100 economies, implies that specialization across low- and high-scale-intensive segments explains around 37.3 percent of the cross-national heterogeneity in real income levels, as measured by $\text{var}(\log W_i)$.³⁹

V. Concluding Remarks

This paper presented an alternative view of within-industry specialization; the thesis of which is that, even in the absence of Ricardian or factor-driven comparative advantage, high-wage and remote economies have a tendency to specialize in high-market power segments of industries. I argued that this patterns of specialization (i) explains the cross-national and spatial variation in export price levels, and (ii) delivers macro-level implications that are distinct from traditional theories of within-industry quality specialization. Some of these macro-level implications were discussed thoroughly in this paper, but several others merit further attention.

First, the present theory can shed light on a recently circulating thesis that market power has been rising in the US economy (De Loecker and Eeckhout 2017). Many factors can be responsible for this rise. From the perspective of the present theory, this rise can reflect a trade-induced hyper-specialization of US-based firms in high-market power segments of industries. An empirical assessment of this hypothesis presents a promising avenue for future research.

Second, the present paper raises some old questions about industrial policy in an open economy. First, in the present framework, the labor allocation in high-market power segments is inefficiently low. This creates a scope for policy intervention, which is absent in the standard heterogeneous firm models of trade and geography. Additionally, the differential degrees of scale economies across segments creates a scope for industrial targeting, whereby countries can improve their terms-of-trade by heavily protecting the high-returns to scale segments of each industry.

APPENDIX A. DESCRIPTION OF DATA USED IN SECTION I

The data used to calculate export unit prices is from the COMTRADE-BACI database compiled by the CEPII. This is the most comprehensive publicly-available data that permits the construction of export unit prices. The data originates from the United Nations Comtrade database. The baci variation of the data uses a harmonization methodology to reconcile mirror flows, which provides more complete

³⁹These results are related to Bento (2014), who shows that accounting for heterogeneity within industries increases the elasticity of welfare with respect to entry costs.

geographical coverage than if only a single direction of Comtrade statistics were to be used—see Gaulier and Zignago (2010) for a description the Baci approach to harmonization. The COMTRADE-BACI database provides physical quantities (only in units of weight) and values of annual bilateral trade flows by 6-digit HS product categories from 2003 to 2015. For 224 countries Head, Mayer, and Ries (2010) report matching data for GDP, distance, and population size. The country-level production data is taken from the TRADEPROD database compiled by the CEPII. This dataset is built on the Trade, Production and Protection database made available by the World Bank (Nicita and Olarreaga 2007). It covers 26 industrial sectors in the ISIC Revision 2 from 1980 to 2006—see De Sousa, Mayer, and Zignago (2012) for more details.

APPENDIX B. GENERAL COST FUNCTION

In the interest of exposition, the main model abstracted from intranational cost heterogeneity across firms within a segment. However, given the weak isomorphism between quality and productivity, the model can be trivially amended to also accommodate cost heterogeneity across firms. To this end, suppose firms are heterogeneous in their unit labor cost, $a_z(\varphi)$, which increases with output quality at elasticity, κ —i.e., $\partial \ln a_z(\varphi) / \partial \ln \varphi = \kappa$. Considering this specification, the *effective* marginal cost facing a firm becomes

$$\tilde{c}_z(\varphi; \tau_{ji,k}, w_j) = \tau_{ji,k} a_z w_j \varphi^{\kappa + \frac{\rho_z}{\rho_z - 1}}.$$

Expectedly, under $\kappa = 0$, the above expression reduces to the one presented in Section II. The above marginal cost function can be reformulated in terms of an augmented degree of quality intensity, $\tilde{\rho}_z \equiv 1 - (1 - \rho_z) / (1 - \kappa(1 - \rho_z))$, as follows:

$$\tilde{c}_z(\varphi; \tau_{ji,k}, w_j) = \tau_{ji,k} w_j \varphi^{\frac{\tilde{\rho}_z}{\tilde{\rho}_z - 1}}.$$

If $1 + \kappa < 1 / (1 - \rho_z)$, then (i) $\tilde{\rho}_z < 1$, and (ii) $\tilde{\rho}_z$ is strictly increasing in ρ_z . That being the case, all the results presented in the paper follow qualitatively in this more general setup, but have to be reformulated in terms of $\tilde{\rho}_z$ instead of ρ_z . For example, the gravity equation (4) will assume the following formulation:

$$X_{ji,z} = \frac{A_{j,z} N_{j,k} (\tau_{ji,k} w_j)^{\gamma_k \left(1 - \frac{1}{\tilde{\rho}_z}\right)}}{\sum_{n=1}^N A_{n,z} N_{n,k} (\tau_{ni,k} w_n)^{\gamma_k \left(1 - \frac{1}{\tilde{\rho}_z}\right)}} e_{i,z} \alpha_{i,k} Y_i,$$

which immediately implies Lemma 1. That is, high- w and high- τ economies have a revealed comparative advantage in high- $\tilde{\rho}$ segments, which are also the high- ρ segments, by construction. The assumption that $1 + \kappa < 1 / (1 - \rho_z)$ ensures that a higher output quality does not increase the production cost so much that it puts the high-quality firms at a disadvantage—a very similar assumption underlies Baldwin and Harrigan's (2011) quality-sorting model.

APPENDIX C. PROOFS AND DERIVATIONS

A. Proof of Lemma 2

The distribution of export markups from country j to market i can be defined as a function of the bilateral trade costs, $\tau_{ji,k}$, and country j 's wage rate, w_j , as follows:

$$\mathcal{M}_{i,z}(m; w_j, \tau_{ji,k}) = \Pr\{\mu(\nu) \leq m \mid \nu \geq \nu_{i,z}^*\}.$$

Note that given the definition of $\nu \equiv P/\tilde{c}$, the market specific cut-off $\nu_{i,z}^*$ can be expressed as

$$(P_{i,z}/\tau_{ji,k}w_j)(\varphi_{ji,z}^*)^{\frac{\rho_z}{1-\rho_z}} = \nu_{i,z}^*,$$

where recall that $\varphi_{ji,z}^*$ denotes the lowest quality under which a firm can profitably export from country j to i . Considering (i) the above equation for $\nu_{i,z}^*$ and (ii) that for each firm exporting from country j , $\nu = \nu_{i,z}^* [\varphi/\varphi_{ji,z}^*]^{\rho_z/(1-\rho_z)}$, the markup distribution can be stated as

$$\mathcal{M}_{i,z}(m; w_j, \tau_{ji,k}) = \frac{\Pr\left(\mu_{i,z}\left(\nu_{i,z}^* \left[\varphi/\varphi_{ji,z}^*\right]^{\frac{\rho_z}{1-\rho_z}}\right) < m, \varphi < \varphi_{ji,z}^*\right)}{\Pr(\varphi < \varphi_{ji,z}^*)}.$$

Given Assumption A3, $dG_{j,z}(\varphi) = \gamma_k \mathcal{A}_{j,z} \varphi^{-\gamma_k-1}$, which simplifies the above expression as follows:

$$\mathcal{M}_{ji,z}(m; w_j, \tau_{ji,k}) = \frac{\int_{\varphi_{ji,z}^*}^{\left(\nu_{i,z}^* \mu_{i,z}^{-1}(m)\right)^{\frac{1-\rho_z}{\rho_z}} \varphi_{ji,z}^*} dG_{j,z}(\varphi)}{\int_{\varphi_{ji,z}^*}^{\infty} dG_{j,z}(\varphi)} = 1 - \left(\nu_{i,z}^* \mu_{i,z}^{-1}(m)\right)^{\frac{1-\rho_z}{\rho_z} \gamma_k},$$

which clearly does not depend on the characteristics of the exporting country, j . ■

B. Proof of Proposition 1

To prove Proposition 1, it suffices to show that high- w and high- τ economies specialize in high-trade elasticity segments, where the trade elasticity is $\theta_z = \gamma_k(1 - 1/\rho_z)$. Once this claim is established, Proposition 1 follows immediately from Assumption A4. First, note that the segment-neutrality of technology parameters entails that $\mathcal{A}_{i,z} = \mathcal{A}_{i,k} \mathcal{A}_z$ for $z \in \mathcal{Z}_k$. Plugging this condition into equation (3), the total output of country j in segment $z \in \mathcal{Z}_k$ can be expressed as follows:

$$\mathcal{R}_z(\mathcal{A}_j, w_j, \tau_{j,k}) = \mathcal{A}_{j,k} N_{j,k} w_j^{\gamma_k(\frac{1}{\rho_z}-1)} \sum_{i=1}^N \tau_{ji,k}^{\gamma_k(\frac{1}{\rho_z}-1)} \mathcal{Y}_{i,z}$$

where the term $\mathcal{Y}_{i,z} \equiv \mathcal{A}_z \chi_{i,z} Q_{i,z} Y_i$ is introduced to simplify the notation, and $\tau_{j,k} = \{\tau_{ji,k}\}$. Note that, based on the above equation, national output per

segment $z \in \mathcal{Z}_k$, is uniquely determined by $\mathcal{A}_{j,k}$, w_j , $\tau_{j,k}$, where w_j itself is an implicit function of $\mathcal{A}_{j,k}$ and $\tau_{j,k}$. Correspondingly, the share of segment $z \in \mathcal{Z}_k$ in total industry-level output can be expressed solely as a function of w_j and $\tau_{j,k}$:

$$r_z(w_j, \tau_{j,k}) \equiv \frac{R_z}{\sum_{z' \in \mathcal{Z}_k} R_{z'}} = \frac{w_j^{-\theta_z} \sum_{i=1}^N \tau_{ji,k}^{-\theta_z} \mathcal{Y}_{i,z}}{\sum_{z' \in \mathcal{Z}_k} w_j^{-\theta_{z'}} \sum_{i=1}^N \tau_{ji,k}^{-\theta_{z'}} \mathcal{Y}_{i,z'}}.$$

That is, the effect of $\mathcal{A}_{j,k}$ on r_z is entirely passed through the economy-wide wage rate. Now, consider segments \mathcal{H} and $\mathcal{L} \in \mathcal{Z}_k$ that differ in their trade elasticity $\theta_{\mathcal{H}} < \theta_{\mathcal{L}}$. An α -fold increase in country j 's wage (holding $\tau_{j,k}$ fixed) leads to an $\alpha^{\theta_{\mathcal{L}} - \theta_{\mathcal{H}}}$ -fold increase in $r_{\mathcal{H}}/r_{\mathcal{L}}$, where $\alpha^{\theta_{\mathcal{L}} - \theta_{\mathcal{H}}} > 1 \Leftrightarrow \alpha > 1$. In particular,

$$(C1) \quad \frac{r_{\mathcal{H}}(\alpha w_j, \tau_{j,k})}{r_{\mathcal{L}}(\alpha w_j, \tau_{j,k})} = \alpha^{\theta_{\mathcal{L}} - \theta_{\mathcal{H}}} \frac{r_{\mathcal{H}}(w_j, \tau_{j,k})}{r_{\mathcal{L}}(w_j, \tau_{j,k})} > \frac{r_{\mathcal{H}}(w_j, \alpha \tau_{j,k})}{r_{\mathcal{L}}(w_j, \alpha \tau_{j,k})}.$$

Similarly, an α -fold increase in country j 's remoteness (holding w_j fixed⁴⁰) leads to an $\alpha^{\theta_{\mathcal{L}} - \theta_{\mathcal{H}}}$ -fold increase in $r_{\mathcal{H}}/r_{\mathcal{L}}$, where $\alpha^{\theta_{\mathcal{L}} - \theta_{\mathcal{H}}} > 1 \Leftrightarrow \alpha > 1$. In particular,

$$(C2) \quad \frac{r_{\mathcal{H}}(w_j, \alpha \tau_{j,k})}{r_{\mathcal{L}}(w_j, \alpha \tau_{j,k})} = \alpha^{\theta_{\mathcal{L}} - \theta_{\mathcal{H}}} \frac{r_{\mathcal{H}}(w_j, \tau_{j,k})}{r_{\mathcal{L}}(w_j, \tau_{j,k})} > \frac{r_{\mathcal{H}}(w_j, \alpha \tau_{j,k})}{r_{\mathcal{L}}(w_j, \alpha \tau_{j,k})}.$$

Together, inequalities (C1) and (C2) imply that high-wage and remote economies specialize in low- θ segments, which per Assumption A4 are also the high-market power segments of the industry.

Welfare Formulas.—Here, I derive formulas that characterize the gains from trade in the estimated model. To derive the formula for GT_i , I follow the same steps as in Costinot and Rodríguez-Clare (2014). From the gravity equation (4), I can deduce the following two expressions for $\lambda_{ii,z}$ and $P_{i,z}$,

$$\lambda_{ii,z} = \mathcal{A}_{i,z} N_{i,k} (\tau_{ii,k} w_j)^{-\theta_z} / \sum_j \mathcal{A}_{j,z} N_{j,k} (\tau_{ji,k} w_j)^{-\theta_z},$$

$$P_{i,z} = \varsigma_{i,z} \left(\sum_j \mathcal{A}_{j,z} N_{j,k} (\tau_{ji,k} w_j)^{\theta_z} \right)^{1/\theta_z},$$

where $\varsigma_{i,z}$ is composed of market-specific constants that are invariant to trade costs. Combining the above two equations implies that $C_{i,z} = w_i/P_i = \varsigma_{i,z} (\lambda_{ii,z}/\mathcal{A}_{i,z} N_{i,k})^{-1/\theta_z}$, which combined with R1 and R2 yields the following formula for aggregate welfare:

$$(C3) \quad W_i = \prod_k \prod_{z \in \mathcal{Z}_k} C_{i,z}^{\alpha_i k e_{i,z}} = \prod_k \prod_{z \in \mathcal{Z}_k} \varsigma_{i,z} (\lambda_{ii,z}/\mathcal{A}_{i,z} N_{i,k})^{-\alpha_i k e_{i,z}/\theta_z}.$$

⁴⁰Holding w_j fixed entails an increase in \mathcal{A}_j to counter the increase in trade costs.

Using the familiar hat notation to denote the change in variables when moving from the factual to the counter-factual equilibrium, I can arrive at the following expression for changes in welfare:

$$\widehat{W}_i = \prod_k \prod_{z \in \mathcal{Z}_k} \left(\widehat{\lambda}_{ii,z} / \widehat{N}_{i,k} \right)^{-\alpha_{i,k} e_{i,z} / \theta_z}.$$

To characterize $\widehat{N}_{i,k}$ in the above equation, I can appeal to the free entry condition (FE), which can be stated as follows in the constant markup case:

$$w_i N_{i,k} f_k^e = \sum_{z \in \mathcal{Z}_k} \frac{\mathcal{R}_{i,z}}{\epsilon_z},$$

with $\mathcal{R}_{i,z} = \sum_{n=1}^{\mathcal{N}} X_{in,z}$ denoting the total revenue generated by economy i in segment $z \in \mathcal{Z}_k$. Rearranging the above equation and noting that $\mathcal{R}_i = w_i L_i$, we can arrive at the following equation describing the mass of firms,

$$N_{i,k} = \frac{1}{f_k^e} \left(\sum \frac{\mathcal{R}_{i,z} / \mathcal{R}_{i,k}}{\epsilon_z} \right) \frac{\mathcal{R}_{i,k}}{\mathcal{R}_i} \frac{\mathcal{R}_i}{w_i} = \left(\sum_{z \in \mathcal{Z}_k} \frac{r_{i,z}}{\epsilon_z} \right) r_{i,k} L_i,$$

where, by definition, $r_{i,z} \equiv \mathcal{R}_{i,z} / \mathcal{R}_{i,k}$ and $r_{i,k} \equiv \mathcal{R}_{i,k} / \mathcal{R}_i$. The above equation immediately implies that $\widehat{N}_{i,k} = \left(\sum_{z \in \mathcal{Z}_k} r_{i,z} / \epsilon_z \right) \widehat{r}_{i,k}$, which when plugged back into the formula for \widehat{W}_i , delivers the following:

$$\widehat{W}_i = \prod_{k=1}^{\mathcal{K}} \prod_{z \in \mathcal{Z}_k} \left(\widehat{\lambda}_{ii,z} / \left[\sum_{z \in \mathcal{Z}_k} \frac{r_{i,z}}{\epsilon_z} \right] \widehat{r}_{i,k} \right)^{-\alpha_{i,k} e_{i,z} / \theta_z}.$$

Noting that in autarky (i) $\lambda_{ii,z} = 1$; (ii) $r_{i,k} = \alpha_{i,k}$, and (iii) $\sum_{z \in \mathcal{Z}_k} r_{i,z} / \epsilon_z = \sum_{z \in \mathcal{Z}_k} e_{i,z} / \epsilon_z$, the above equation implies the following formula for the gains from trade,

$$GT_i = 1 - 1 / \widehat{W}_i = \prod_{k=1}^{\mathcal{K}} \prod_{z \in \mathcal{Z}_k} (\mathcal{S}_{i,k} \lambda_{ii,z})^{\alpha_{i,k} e_{i,z} / \theta_z},$$

where $\mathcal{S}_{i,k} \equiv \alpha_{i,k} / r_{i,k} \left[(\sum_{z \in \mathcal{Z}_k} e_{i,z} / \epsilon_z) / (\sum_{z \in \mathcal{Z}_k} r_{i,z} / \epsilon_z) \right]$. Similarly, the expression for aggregate welfare (equation (15)) can be derived by appealing to equation (C3), as follows,

$$\begin{aligned} W_i &= \prod_{k=1}^{\mathcal{K}} \prod_{z \in \mathcal{Z}_k} \varsigma_{i,z} (\lambda_{ii,z} / \mathcal{A}_{i,z} N_{i,k})^{-\alpha_{i,k} e_{i,z} / \theta_z} \\ &= \left(\varsigma_i \prod_{k=1}^{\mathcal{K}} \prod_{z \in \mathcal{Z}_k} \mathcal{A}_{i,z}^{\alpha_{i,k} e_{i,z} / \theta_z} \right) \left(\prod_{k=1}^{\mathcal{K}} \prod_{z \in \mathcal{Z}_k} \ell_{i,k}^{\alpha_{i,k} e_{i,z} / \theta_z} \right) \left(\prod_{k=1}^{\mathcal{K}} \prod_{z \in \mathcal{Z}_k} \lambda_{ii,z}^{-\alpha_{i,k} e_{i,z} / \theta_z} \right), \end{aligned}$$

where $\varsigma_i \equiv \prod_k \prod_{z \in \mathcal{Z}_k} \varsigma_{i,z}$ and $\ell_{i,k} \equiv (\sum_{z \in \mathcal{Z}_k} r_{i,z} / \epsilon_z) L_{i,k}$. Replacing $\theta_z = \gamma_k (\epsilon_z - 1)$ and noting (from Section II) that $\psi_z = \epsilon_z / (\epsilon_z - 1)$, we will arrive at the welfare formula specified by equation (15).

APPENDIX D. GRAVITY ESTIMATION DETAILS

This Appendix provides details regarding the data employed and parameter values attained under the country- and industry-level estimations in Section IV.

A. Country-Level Estimation

The economy-wide estimation treats each economy as one integrated industry, i.e., $\mathcal{K} = 1$.⁴¹ So I present the results here without the industry subscript, k . To conduct the estimation, I take country-level trade data from the U.N. COMTRADE database (Comtrade 2010) for a sample of the 100 largest economies (in terms of real GDP) in year 2000—this sample accounts for more than 95 percent of world trade in that year. I merge this data with (i) data on population size and GDP from the World Bank (World Bank 2012), and data on geo-distance, common language, borders, and trade agreements from Mayer and Zignago (2011).

In the country-level implementation, I am estimating 206 structural parameters by fitting the model to $(100 - 1) \times 100 = 9,900$ data points on bilateral trade values, namely $\mathbf{X} = [X_{ji}]$. The estimated parameters are the following:

- (i) One hundred exporter fixed effects, $\mathcal{T} = [\mathcal{T}_j]$;
- (ii) One hundred imported fixed effects, $\mathbf{e} = [e_{i,\mathcal{H}}]$, that characterize the expenditure shares on segment \mathcal{H} ;⁴²
- (iii) Five parameters, vector β , that characterize the iceberg trade costs; and
- (iv) One parameter, $\rho = \theta_{\mathcal{H}}/\theta_{\mathcal{L}}$, describing the spread between the trade elasticity in segments \mathcal{H} and \mathcal{L} .

Table D1 presents the estimation results. In the interest of space, I only report the main parameters, β and ρ . The same table also compares the results to those attained under the standard structural gravity estimation with no heterogeneity in trade elasticities à la Anderson and van Wincoop (2003). Evidently, the main model exhibits a significantly higher predictive power than the standard, uniform elasticity model—the main model exhibits an $R^2 = 0.65$, whereas the standard model exhibits an $R^2 = 0.46$.

The present model's superior fit stems from its ability to match the following feature of bilateral trade data: rich-poor trading partners engage in significantly more two-way trade (i.e., have a higher X_{ji}) than rich-rich or poor-poor trading partners. To illustrate this point formally, let North denote the richest 20 countries in my sample and South denote the other 80 countries. Figure D1 illustrates that in year 2000 North-South trade was conducted more intensively than both North-North trade. As illustrated in Figure D1, the standard model

⁴¹This treatment of the economy resembles the approach taken by Fieler (2011).

⁴²Once \mathbf{e} is known, the expenditure share on segment \mathcal{L} is automatically given by $[e_{i,\mathcal{L}}] = \mathbf{1} - \mathbf{e}$.

TABLE D1—ESTIMATION RESULTS WITH COUNTRY-LEVEL DATA

Parameters	Standard gravity model $\theta_{\mathcal{H}} = \theta_{\mathcal{L}}$	Main model $\theta_{\mathcal{H}} < \theta_{\mathcal{L}}$
$\rho \equiv \frac{\theta_{\mathcal{H}}}{\theta_{\mathcal{L}}}$...	0.54 (0.004)
β_{const}	2.83 (0.020)	4.73 (0.065)
β_{dist}	0.43 (0.003)	0.36 (0.006)
β_{border}	0.53 (0.006)	0.87 (0.015)
β_{lang}	0.52 (0.005)	0.84 (0.011)
$\beta_{agreement}$	0.98 (0.010)	0.93 (0.020)
R^2	0.46	0.65
Observations		9,900

Note: Standard errors are reported in parentheses.

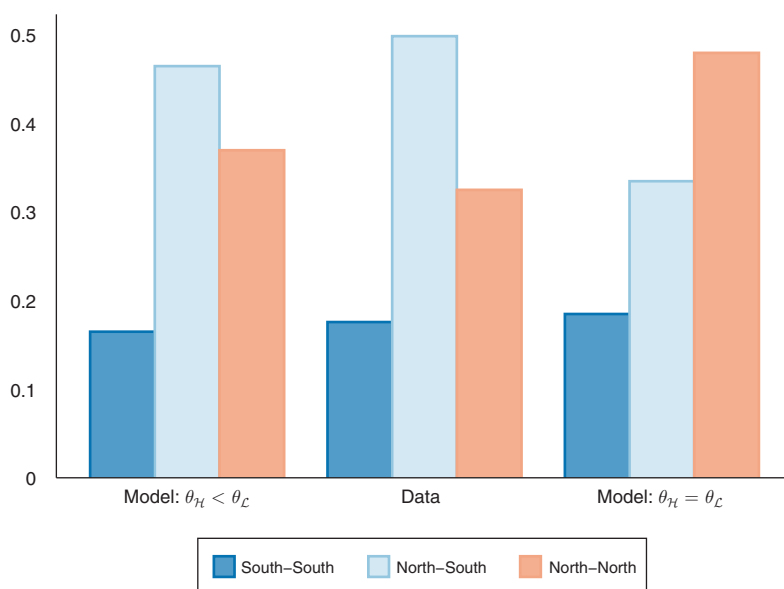


FIGURE D1. THE COMPOSITION OF AGGREGATE TRADE FLOWS: MODEL VERSUS DATA

Note: The North corresponds to the 20 richest countries in the sample.

estimated subject to $\theta_{\mathcal{H}}/\theta_{\mathcal{L}} = 1$ has difficulty matching this pattern. The reason being that, under $\theta_{\mathcal{H}}/\theta_{\mathcal{L}} = 1$, the model admits no systematic pattern of *intra-industry* specialization that makes North-South trade relatively more attractive than North-North trade. The estimated model that treats $\theta_{\mathcal{H}}/\theta_{\mathcal{L}} < 1$ as free-moving parameter, however, admits North-South specialization across low- and high- θ segments. This feature of the model renders North-South trade relatively more attractive. Importantly, the extent of North-South specialization

depends on $\theta_{\mathcal{H}}/\theta_{\mathcal{L}}$ and $\mathbf{e} = [e_{i,\mathcal{H}}]$, which in turn guides the identification of these parameters.

B. Industry-Level Estimation

The industry-level estimation uses data from the World Input-Output Database (WIOD, Timmer et al. 2012) in 2008. The original WIOD database covers 35 industries and 40 countries, which account for more than 85 percent of world GDP, plus an aggregate of the rest of the world. The countries in the sample include all 27 members of the European Union plus 13 other major economies, namely, Australia, Brazil, Canada, China, India, Indonesia, Japan, Mexico, Russia, South Korea, Taiwan, Turkey, and the United States. The 35 industries in WIOD database are listed in Table D2, which include 15 tradable industries plus 20 service-related industries. Following Costinot and Rodríguez-Clare (2014), I contract the WIOD database into a sample of 32 major economies (listed in Table 7) plus an aggregate of the rest of the world. For each industry k , the trimmed WIOD data reports the industry-level vector of expenditure shares, $\boldsymbol{\alpha}_k \equiv [\alpha_{i,k}]$, revenue shares, $\mathbf{r}_k \equiv [r_{i,k}]$, the full matrix of industry-level trade flows, $\mathbf{X}_k \equiv [X_{ji,k}]$, as well as country-level data on income, $Y_i = w_i L_i$, and population size L_i . As with the country-level data, I complement the WIOD with data on geo-distance, common language, borders, and trade agreements from Mayer and Zignago (2011). Following Costinot and Rodríguez-Clare (2014), the industry-level trade elasticity estimates are taken from Caliendo and Parro (2014) and are reported in Table D2.

Here, for each of the 15 (nonservice) industries in the sample, I am separately estimating 70 structural parameters by fitting the model to $(32 - 1) \times 32 = 992$ data points on industry-level bilateral trade values, namely $\mathbf{X}_k = [X_{ji,k}]$.⁴³ The estimated parameters are:

- (i) Thirty-two exporter \times industry fixed effects, $\mathcal{T}_k = [\mathcal{T}_{j,k}]$;
- (ii) Thirty-two importer \times industry fixed effects, $\mathbf{e}_k = [e_{i,\mathcal{H}}]$, that characterize the expenditure shares on segment \mathcal{H} in industry k ;⁴⁴
- (iii) Five parameters, vector $\boldsymbol{\beta}_k$, that characterize the iceberg trade costs in industry k ; and
- (iv) One parameter, $\rho_k = \theta_{\mathcal{H}}/\theta_{\mathcal{L}}$, describing the spread between the segment-level trade elasticities in industry k .

In all incidences, the industry-level estimation exhibits an R^2 of 0.50 or greater. The last two columns in Table D2 report the segment-level trade elasticities estimated for each of the 15 nonservice industries in the WIOD sample. For some

⁴³ So, in total, I am estimating $15 \times 70 = 1,050$ parameters across all industries in the sample.

⁴⁴ Once \mathbf{e}_k is known, the expenditure share on segment \mathcal{L} is automatically given by $[e_{i,\mathcal{L}}] = \mathbf{1} - \mathbf{e}_k$.

TABLE D2—LIST OF WIOD INDUSTRIES IN THE INDUSTRY-LEVEL ESTIMATION

WIOD sector	Sector's description	Industry-level θ (Caliendo-Parro)	θ_H	θ_L
1	Agriculture, hunting, forestry, and fishing	8.11	6.28	9.43
2	Mining and quarrying	15.72	11.53	17.12
3	Food, beverages, and tobacco	2.55	0.89	3.71
4	Textiles and textile products; leather and footwear	5.56	3.93	7.06
5	Wood and products of wood and cork	10.83	9.10	12.00
6	Pulp, paper, printing, and publishing	9.07	6.03	9.92
7	Coke, refined petroleum, and nuclear fuel	51.08	49.38	55.10
8	Chemicals and chemical products	4.75	3.80	5.82
9	Rubber and plastics	1.66	0.62	3.35
10	Other non-metallic mineral	2.76	0.91	3.57
11	Basic metals and fabricated metal	7.99	6.55	9.03
12	Machinery, nec.	1.52	0.56	4.90
13	Electrical and optical equipment	10.60	8.78	12.48
14	Transport equipment	0.37	0.16	0.75
15	Manufacturing, nec.; recycling	5.00	3.29	6.76
16	Electricity, gas, and water supply	5
	Construction			
	Sale, maintenance, and repair of motor vehicles and motorcycles; retail sale of fuel			
	Wholesale trade and commission trade, except of motor vehicles and motorcycles			
	Retail trade, except of motor vehicles and motorcycles; repair of household goods			
	Hotels and restaurants			
	Inland transport			
	Water transport			
	Air transport			
	Other supporting and auxiliary transport activities; activities of travel agencies			
	Post and telecommunications			
	Financial intermediation			
	Real estate activities			
	Renting of M&Eq and other business activities			
	Education			
	Health and social work			
	Public admin. and defence; compulsory social security			
	Other community, social, and personal services			
	Private households with employed persons			

industries, such as “Petroleum” or “Fabricated Metals,” θ_H and θ_L are not significantly different, whereas in the “Machinery” and “Food” industries the estimation identifies a significant degree of within-industry θ heterogeneity.

REFERENCES

- Acemoglu, Daron, Simon Johnson, and James A. Robinson.** 2000. “The Colonial Origins of Comparative Development: An Empirical Investigation.” NBER Working Paper 7771.
- Alessandria, George, and Joseph P. Kaboski.** 2011. “Pricing-to-Market and the Failure of Absolute PPP.” *American Economic Journal: Macroeconomics* 3 (1): 91–127.
- Anderson, James E., and Eric van Wincoop.** 2003. “Gravity with Gravitas: A Solution to the Border Puzzle.” *American Economic Review* 93 (1): 170–92.

- Anderson, T.W., and Herman Rubin. 1949. "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations." *Annals of Mathematical Statistics* 20 (1): 46–63.
- Arkolakis, Costas, Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare. 2019. "The Elusive Pro-competitive Effects of Trade." *Review of Economic Studies* 86 (1): 46–80.
- Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare. 2012. "New Trade Models, Same Old Gains?" *American Economic Review* 102 (1): 94–130.
- Baldwin, Richard, and James Harrigan. 2011. "Zeros, Quality, and Space: Trade Theory and Trade Evidence." *American Economic Journal: Microeconomics* 3 (2): 60–88.
- Baldwin, Richard E., and Tadashi Ito. 2008. "Quality Competition versus Price Competition Goods: An Empirical Classification." NBER Working Paper 14305.
- Bento, Pedro. 2014. "Niche Firms, Mass Markets, and Income across Countries: Accounting for the Impact of Entry Costs." *Journal of Economic Dynamics and Control* 48: 147–58.
- Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63 (4): 841–90.
- Bertoletti, Paolo, and Federico Etro. 2017. "Monopolistic Competition When Income Matters." *Economic Journal* 127 (603): 1217–43.
- Bertoletti, Paolo, Federico Etro, and Ina Simonovska. 2018. "International Trade with Indirect Additivity." *American Economic Journal: Microeconomics* 10 (2): 1–57.
- Bils, Mark, and Peter J. Klenow. 2001. "Quantifying Quality Growth." *American Economic Review* 91 (4): 1006–30.
- Brandt, Loren, Johannes Van Biesebroeck, Luhang Wang, and Yifan Zhang. 2017. "WTO Accession and Performance of Chinese Manufacturing Firms." *American Economic Review* 107 (9): 2784–2820.
- Broda, Christian, and David E. Weinstein. 2006. "Globalization and the Gains from Variety." *Quarterly Journal of Economics* 121 (2): 541–85.
- Brooks, Wyatt, and Pau Pujolas. 2016. "Nonlinear Gravity." <https://www3.nd.edu/~wbrooks/NonlinearGravity.pdf>.
- Caliendo, L., and F. Parro. 2014. "Estimates of the Trade and Welfare Effects of NAFTA." *Review of Economic Studies* 82 (1): 1–44.
- Campa, Jose, and Linda S. Goldberg. 1995. "Investment in Manufacturing, Exchange Rates and External Exposure." *Journal of International Economics* 38 (3–4): 297–320.
- Caselli, Francesco. 2005. "Accounting for Cross-Country Income Differences." In *Handbook of Economic Growth*, Vol. 1A, edited by Philippe Aghion and Steven N. Durlauf, 679–741. Amsterdam: North-Holland.
- Chaney, Thomas. 2008. "Distorted Gravity: The Intensive and Extensive Margins of International Trade." *American Economic Review* 98 (4): 1707–21.
- Choi, Yo Chul, David Hummels, and Chong Xiang. 2009. "Explaining Import Quality: The Role of the Income Distribution." *Journal of International Economics* 78 (2): 293–303.
- Coibion, Olivier, Liran Einav, and Juan Carlos Hallak. 2007. "Equilibrium Demand Elasticities across Quality Segments." *International Journal of Industrial Organization* 25 (1): 13–30.
- Costinot, Arnaud, and Andrés Rodríguez-Clare. 2014. "Chapter 4—Trade Theory with Numbers: Quantifying the Consequences of Globalization." In *Handbook of International Economics*, Vol. 4, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, 197–261. Amsterdam: North-Holland.
- Cragg, John G., and Stephen G. Donald. 1993. "Testing Identifiability and Specification in Instrumental Variable Models." *Econometric Theory* 9 (2): 222–40.
- Crozet, Matthieu, Keith Head, and Thierry Mayer. 2012. "Quality Sorting and Trade: Firm-level Evidence for French Wine." *Review of Economic Studies* 79 (2): 609–44.
- de Blas, Beatriz, and Kathryn N. Russ. 2015. "Understanding Markups in the Open Economy." *American Economic Journal: Macroeconomics* 7 (2): 157–80.
- De Loecker, Jan, and Jan Eeckhout. 2017. "The Rise of Market Power and the Macroeconomic Implications." NBER Working Paper 23687.
- De Loecker, Jan, and Jan Eeckhout. 2018. "Global Market Power." NBER Working Paper 24768.
- De Loecker, Jan, Catherine Fuss, and Johannes Van Biesebroeck. 2014. "International Competition and Firm Performance: Evidence from Belgium." National Bank of Belgium Working Paper 269.
- De Loecker, Jan, Pinelopi K. Goldberg, Amit K. Khandelwal, and Nina Pavcnik. 2016. "Prices, Markups, and Trade Reform." *Econometrica* 84: 445–510.
- De Sousa, José, Thierry Mayer, and Soledad Zignago. 2012. "Market Access in Global and Regional Trade." *Regional Science and Urban Economics* 42 (6): 1037–52.
- Díez, Federico J., Dabuek Leigh, and Suchanan Tambunlertchai. 2018. "Global Market Power and Its Macroeconomic Implications." IMF Working Paper 18/137.

- Dingel, Jonathan I.** 2016. "The Determinants of Quality Specialization." *Review of Economic Studies* 84 (4): 1551–82.
- Eaton, Jonathan, Marcela Eslava, C.J. Krizan, Maurice Kugler, and James Tybout.** 2010. "A Search and Learning Model of Export Dynamics." <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.189.2412&rep=rep1&type=pdf>.
- Eaton, Jonathan, Samuel Kortum, and Francis Kramarz.** 2011. "An Anatomy of International Trade: Evidence from French Firms." *Econometrica* 79 (5): 1453–98.
- Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu.** 2015. "Competition, Markups, and the Gains from International Trade." *American Economic Review* 105 (10): 3183–3221.
- Erkel-Rousse, Hélène, and Daniel Mirza.** 2002. "Import Price Elasticities: Reconsidering the Evidence." *Canadian Journal of Economics* 35 (2): 282–306.
- Fajgelbaum, Pablo, Gene M. Grossman, and Elhanan Helpman.** 2011. "Income Distribution, Product Quality, and International Trade." *Journal of Political Economy* 119 (4): 721–65.
- Fally, Thibault.** 2018. "Integrability and Generalized Separability." NBER Working Paper 25025.
- Fally, Thibault.** 2019. "Generalized Separability and the Gains from Trade." *Economics Letters* 178: 8–12.
- Fan, Haichao, Yao Amber Li, and Stephen R. Yeaple.** 2015. "Trade Liberalization, Quality, and Export Prices." *Review of Economics and Statistics* 97 (5): 1033–51.
- Feenstra, Robert C., and John Romalis.** 2014. "International Prices and Endogenous Quality." *Quarterly Journal of Economics* 129 (2): 477–527.
- Fieler, Ana Cecilia.** 2011. "Nonhomotheticity and Bilateral Trade: Evidence and a Quantitative Explanation." *Econometrica* 79 (4): 1069–1101.
- Flach, Lisandra.** 2017. "A Note on the Isomorphism of Heterogeneous Firms Models." *Economics Letters* 155: 24–27.
- Fornell, Claes, Michael D. Johnson, Eugene W. Anderson, Jaesung Cha, and Barbara Everitt Bryant.** 1996. "The American Customer Satisfaction Index: Nature, Purpose, and Findings." *Journal of Marketing* 60 (4): 7–18.
- Gaulier, Guillaume, and Soledad Zignago.** 2010. "Baci: International Trade Database at the Product-Level. The 1994–2007 Version." CEPII Research Center Working Paper 2010-23.
- Goldberg, Pinelopi Koujianou.** 1995. "Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry." *Econometrica* 63 (4): 891–951.
- Grossman, Gene M., and Elhanan Helpman.** 1991. "Quality Ladders in the Theory of Growth." *Review of Economic Studies* 58 (1): 43–61.
- Hall, Robert E., and Charles I. Jones.** 1999. "Why Do Some Countries Produce So Much More Output per Worker than Others?" *Quarterly Journal of Economics* 114 (1): 83–116.
- Hallak, Juan Carlos.** 2006. "Product Quality and the Direction of Trade." *Journal of International Economics* 68 (1): 238–65.
- Hallak, Juan Carlos, and Jagadeesh Sivadasan.** 2013. "Product and Process Productivity: Implications for Quality Choice and Conditional Exporter Premia." *Journal of International Economics* 91 (1): 53–67.
- Hanoch, Giora.** 1975. "The Elasticity of Scale and the Shape of Average Costs." *American Economic Review* 65 (3): 492–97.
- Hausman, Jerry, Gregory Leonard, and J. Douglas Zona.** 1994. "Competitive Analysis with Differentiated Products." *Annales d'Economie et de Statistique* 34: 159–80.
- Head, Keith, and Thierry Mayer.** 2014. "Gravity Equations: Workhorse, Toolkit, and Cookbook." In *Handbook of International Economics*, Vol. 4, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, 131–95. Amsterdam: North-Holland.
- Head, Keith, Thierry Mayer, and John Ries.** 2010. "The Erosion of Colonial Trade Linkages after Independence." *Journal of International Economics* 81 (1): 1–14.
- Helpman, Elhanan.** 1984. "Chapter 7—Increasing Returns, Imperfect Markets, and Trade Theory." In *Handbook of International Economics*, Vol. 1, edited by Ronald W. Jones and Peter D. Kenen, 325–65. Amsterdam: North-Holland.
- Hertel, Thomas, David Hummels, Maros Ivanic, and Roman Keeney.** 2007. "How Confident Can We Be of CGE-based Assessments of Free Trade Agreements?" *Economic Modelling* 24 (4): 611–35.
- Holmes, Thomas J., Wen-Tai Hsu, and Sanghoon Lee.** 2014. "Allocative Efficiency, Mark-ups, and the Welfare Gains from Trade." *Journal of International Economics* 94 (2): 195–206.
- Holmes, Thomas J., and John J. Stevens.** 2014. "An Alternative Theory of the Plant Size Distribution, with Geography and Intra- and International Trade." *Journal of Political Economy* 122 (2): 369–421.
- Hottman, Colin J., Stephen J. Redding, and David E. Weinstein.** 2016. "Quantifying the Sources of Firm Heterogeneity." *Quarterly Journal of Economics* 131 (3): 1291–1364.

- Hummels, David.** 2001. "Toward a Geography of Trade Costs." <http://www.krannert.purdue.edu/faculty/hummelsd/research/toward/tgtc.pdf>.
- Hummels, David, and Peter J. Klenow.** 2005. "The Variety and Quality of a Nation's Exports." *American Economic Review* 95 (3): 704–23.
- Hummels, David, and Volodymyr Lugovskyy.** 2009. "International Pricing in a Generalized Model of Ideal Variety." *Journal of Money, Credit and Banking* 41 (S1): 3–33.
- Hummels, David, and Alexandre Skiba.** 2004. "Shipping the Good Apples Out? An Empirical Confirmation of the Alchian-Allen Conjecture." *Journal of Political Economy* 112 (6): 1384–1402.
- Imbs, Jean, and Isabelle Mejean.** 2015. "Elasticity Optimism." *American Economic Journal: Macroeconomics* 7 (3): 43–83.
- Johnson, Robert C.** 2012. "Trade and Prices with Heterogeneous Firms." *Journal of International Economics* 86 (1): 43–56.
- Jones, Charles I.** 2011. "Intermediate Goods and Weak Links in the Theory of Economic Development." *American Economic Journal: Macroeconomics* 3 (2): 1–28.
- Khandelwal, Amit.** 2010. "The Long and Short (of) Quality Ladders." *Review of Economic Studies* 77 (4): 1450–76.
- Kimball, Miles S.** 1995. "The Quantitative Analytics of the Basic Neomonetarist Model." NBER Working Paper 5046.
- Kleibergen, Frank, and Richard Paap.** 2006. "Generalized Reduced Rank Tests Using the Singular Value Decomposition." *Journal of Econometrics* 133 (1): 97–126.
- Klenow, Peter J., and Andrés Rodríguez-Clare.** 1997. "The Neoclassical Revival in Growth Economics: Has It Gone Too Far?" *NBER Macroeconomics Annual* 12: 73–103.
- Klenow, Peter J., and Jonathan L. Willis.** 2016. "Real Rigidities and Nominal Price Changes." *Economica* 83 (331): 443–72.
- Kraay, Aart, and Jaume Ventura.** 2007. "Comparative Advantage and the Cross-Section of Business Cycles." *Journal of the European Economic Association* 5 (6): 1300–33.
- Kucheryavyi, Konstantin, Gary Lyn, and Andrés Rodríguez-Clare.** 2016. "Grounded by Gravity: A Well-Behaved Trade Model with Industry-Level Economies of Scale." NBER Working Paper 22484.
- Laffont, Jean-Jacques, and Jean Tirole.** 1993. *A Theory of Incentives in Procurement and Regulation*. Cambridge, MA: MIT Press.
- Lashkaripour, Ahmad, and Volodymyr Lugovskyy.** 2019. "Scale Economies and the Structure of Trade and Industrial Policy." pages.iu.edu/~alashkar/Lashkaripour_Lugovskyy.pdf.
- Levchenko, Andrei A., and Jing Zhang.** 2014. "Ricardian Productivity Differences and the Gains from Trade." *European Economic Review* 65: 45–65.
- Lugovskyy, Volodymyr, and Alexandre Skiba.** 2015. "How Geography Affects Quality." *Journal of Development Economics* 115: 156–80.
- Matsuyama, Kiminori, and Philip Ushchev.** 2017. "Beyond CES: Three Alternative Cases of Flexible Homothetic Demand Systems." Buffett Institute Global Poverty Research Lab Working Paper 17-109.
- Mayer, Thierry, and Soledad Zignago.** 2011. "Notes on CEPII's Distances Measures: The GeoDist Database." CEPII Working Paper 2011-25.
- McCalman, Philip, and Alan Spearot.** 2013. "Why Trucks Jump: Offshoring and Product Characteristics." *Journal of International Economics* 91 (1): 82–95.
- Melitz, Marc J., and Gianmarco I.P. Ottaviano.** 2008. "Market Size, Trade, and Productivity." *Review of Economic Studies* 75 (1): 295–316.
- Melitz, Marc J., and Stephen J. Redding.** 2014. "Chapter 1—Heterogeneous Firms and Trade." In *Handbook of International Economics*, Vol. 4, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, 1–54. Amsterdam: North-Holland.
- Nicita, Alessandro, and Marcelo Olarreaga.** 2007. "Trade, Production, and Protection Database, 1976–2004." *World Bank Economic Review* 21 (1): 165–71.
- Ossa, Ralph.** 2015. "Why Trade Matters After All." *Journal of International Economics* 97 (2): 266–77.
- Ottaviano, Gianmarco, Takatoshi Tabuchi, and Jacques-François Thisse.** 2002. "Agglomeration and Trade Revisited." *International Economic Review* 43 (2): 409–35.
- Parente, Stephen L., and Edward C. Prescott.** 1999. "Monopoly Rights: A Barrier to Riches." *American Economic Review* 89 (5): 1216–33.
- Parente, Stephen L., and Edward C. Prescott.** 2000. *Barriers to Riches*. Cambridge, MA: MIT Press.
- Prescott, Edward C.** 1998. "Lawrence R. Klein Lecture 1997: Needed: A Theory of Total Factor Productivity." *International Economic Review* 39 (3): 525–51.
- Rodrik, Dani.** 1994. "Industrial Organization and Product Quality: Evidence from South Korean and Taiwanese Exports." In *Empirical Studies of Strategic Trade Policy*, edited by Paul Krugman and Alasdair Smith, 195–210. Chicago: University of Chicago Press.

- Romalís, John.** 2007. "NAFTA's and CUSFTA's Impact on International Trade." *Review of Economics and Statistics* 89 (3): 416–35.
- Schott, Peter K.** 2004. "Across-Product versus Within-Product Specialization in International Trade." *Quarterly Journal of Economics* 119 (2): 647–78.
- Simonovska, Ina.** 2015. "Income Differences and Prices of Tradables: Insights from an Online Retailer." *Review of Economic Studies* 82 (4): 1612–56.
- Simonovska, Ina, and Michael E. Waugh.** 2011. "The Elasticity of Trade: Estimates and Evidence." NBER Working Paper 16796.
- Spearot, Alan C.** 2013. "Variable Demand Elasticities and Tariff Liberalization." *Journal of International Economics* 89 (1): 26–41.
- Stock, James H., and Motohiro Yogo.** 2002. "Testing for Weak Instruments in Linear IV Regression." NBER Working Paper 284.
- Sutton, John.** 2007. "Quality, Trade and the Moving Window: The Globalisation Process." *Economic Journal* 117 (524): F469–98.
- Sutton, John, and Daniel Treffer.** 2016. "Capabilities, Wealth, and Trade." *Journal of Political Economy* 124 (3): 826–78.
- Timmer, Marcel, Abdul A. Erumban, Reitze Gouma, Bart Los, Umed Temurshoev, Gaaitzen J. de Vries, Iñaki Arto, et al.** 2012. "The World Input-Output Database (WIOD): Contents, Sources and Methods." World Input-Output Database (WIOD) Working Paper 10.
- UN Comtrade Database** 2010. United Nations Commodity Trade Statistics Database. <http://comtrade.un.org>.
- Verhoogen, Eric A.** 2008. "Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector." *Quarterly Journal of Economics* 123 (2): 489–530.
- Waugh, Michael E.** 2010. "International Trade and Income Differences." *American Economic Review* 100 (5): 2093–2124.
- World Bank.** 2012. *World Development Indicators 2012*. Washington, DC: World Bank.