Industry-Level Scale Economies: *from Micro-Estimation to Macro-Implications*[*]

Ahmad Lashkaripour
Indiana University

Volodymyr Lugovskyy
Indiana University

January 29, 2018

Abstract

What are the macroeconomic impacts of trade and industrial policy? A long tradition in economic analysis indicates that our answer to such questions depend upon (i) the industry-level degree of comparative advantage, which is reflected in the *trade elasticity*, and (ii) the industry-level degree of scale economies, which is reflected in the *scale elasticity*. In this paper we propose an empirical methodology to jointly recover these two key elasticities from firm-level trade data. Our key finding is that scale economies are weaker than assumed in standard theories, and vary considerably across industries. We use our estimates to shed new light on cross-national income differences, the gains from trade, and the empirical relevance of classic trade and industrial policy prescriptions.

1 Introduction

For centuries economist have identified "comparative advantage" and "scale economies" as the two central pillars of trade and industrial policy. The preoc-

[*]We are grateful to Adina Ardelean, Dominick Bartelme, Kerem Cosar, Arnaud Costinot, Alan Deardorff, Farid Farrokhi, Ana Cecilia Fieler, David Hummels, Konstantin Kucheryavyy, Gary Lyn, Ralph Ossa, Andrés Rodríguez-Clare, Peter Schott, Alexandre Skiba, Anson Soderbery, Jonathan Vogel and conference participants at the Fall 2016 Midwest Trade, UECE Lisbon Meetings, and 2017 NBER ITI Summer Institute, for helpful comments and suggestions. We thank Nicolas de Roux and Santiago Tabares for providing us with the data on the Colombian HS10 product codes changes over time. We are grateful to Fabio Gomez for research assistance. Lugovskyy thanks Indiana University SSRC for financial support. All remaining errors are our own.

cupation with these two forces harks back to a long tradition in economic analysis. Many trade-promoting policies are inspired by Ricardo's (1891) insight that due to technological dissimilarities, nations can gain from trade based on comparative advantage. Relatedly, there is a long strand of policy prescriptions inspired by Smith (1776) and Mill (1848) that promote import protection in scale-intensive industries as a means to capitalize on industry-level scale economies.

All the technical advances in recent years have produced quantifiable formulas that can evaluate the macro-level effects of "comparative advantage" and "scale economies" based on only two structural parameters: the *trade* and *scale* elasticities. To demonstrate this, consider the seminal model of Krugman (1980), which predicts that country i's aggregate TFP is given by:

$$TFP_i = A_i \text{ (import intensity)}^{\frac{1}{\theta}} \text{ (population size}_i)^{\psi}$$

Based on the above formula (among other things) aggregate TFP depends on two key factors. The first is the *import intensity* of country i elevated to the ***trade elasticity***, $\boldsymbol{\theta}$—a parameter that reflects the degree of technical heterogeneity across countries or, as some put it, the degree of comparative advantage.[1] The second factor is the *size* of country i's economy elevated to the ***scale elasticity***, $\boldsymbol{\psi}$—a parameter that reflects the strength of scale economies, which is the rate at which production costs diminish with scale. The above formula is quite revealing. It indicates that to understand cross-country income differences we need credible estimates for both θ and ψ.

The scale and trade elasticity parameters assume an even more subtle role in the implementation of industrial policy. For instance, in the multi-industry Krugman model, the change in aggregate TFP in response to a policy shock is given by the following formula (see Kucheryavyy et al. (2016)):[2]

$$\%\Delta TFP_i = \sum_s \beta_{i,s} \left(\frac{1}{\theta_s} \%\Delta \text{ import intensity}_{i,s} + \psi_s \%\Delta \text{ industry size}_{i,s} \right), \quad (1)$$

where s indexes an industry in country i. Based on the above decomposition, the macro-economic effects of industrial policy depend critically on both

[1] Eaton and Kortum (2002) popularized the idea that θ reflects the force of comparative advantage in a global economy. Prior to their seminal paper, many interpreted θ as the degree of dissimilarity between nations in terms of their technical ability in producing a unique set of differentiated varieties.

[2] The above expression is derived assuming a Cobb-Douglas utility aggregator across industries, with $\beta_{i,s}$ denoting the share of country i's expenditure on industry s.

the industry-level scale elasticity, ψ_s, and the industry-level trade elasticity, θ_s. Among other things, the above formula captures the old idea that countries can boost their aggregate TFP by promoting scale-intensive (high-ψ) industries (Rodrik (1988)). However, to evaluate the classic insights implicit in this formula, we need credible estimates for both ψ_s and θ_s.

Given the key role of the industry-level scale and trade elasticities in policy evaluation, one may expect an abundance of estimates for these two parameters. However, while an extensive literature is devoted to estimating the industry-level trade elasticity, θ_s,[3] we know surprisingly little about the empirical size of the scale elasticity, ψ_h. Even more surprising, in most applications the scale elasticity is arbitrarily set to either *zero* ($\psi_s = 0$) or to the inverse of the trade elasticity ($\psi_s = 1/\theta_s$). In fact, the lack of credible estimates for ψ_s have impeded the quantitive exploration of many old policy prescriptions, some of which date back to Alexander Hamilton and Friedrich List.

Against the backdrop of this historic disconnect between theory and evidence, we take a first stab at jointly recovering the scale and trade elasticity parameters from micro-level trade data.

If the scale elasticity was the *only* parameter of interest, we could have employed the *macro-level* approach of using industry-level input cost and output data to recover the industry-level scale elasticity. This approach obviously has many merits, including the fact that it can measure both external and internal economies of scale (Bartelme et al. (2017)). However, the macro-level approach has been often criticized for (i) facing measurement issues, (ii) being susceptible to aggregation bias, and (iii) having difficulty disentangling the long-run and short-run scale elasticities.[4] Beyond these limitations, the macro-level approach faces an even bigger challenge in the context of trade: it cannot separately identify the scale elasticity from the trade elasticity. Instead it can only recover the scale elasticity up to an externally chosen trade elasticity.

Considering these underlying issues we propose a novel methodology to *jointly* recover the industry-level scale and trade elasticity parameters from *micro-*

[3]See for example, Broda and Weinstein (2006); Simonovska and Waugh (2014); Caliendo and Parro (2014); Soderbery (2015), among others.

[4]The measurement issue is due to many factors of production being unobservable, non-rivalrous, or under-utilized (see Hall (1989); Morrison (1992); Basu and Fernald (1997); Tybout (2001)). See Basu and Fernald (1997) for a discussion of the aggregation bias underlying the macro-level approach. Finally, see Bresnahan (1989) and Chirinko and Fazzari (1994) for a discussion on the challenging nature of disentangling the long-run scale elasticity from the short-run elasticity.

level trade data. Our approach relies on (i) a structural import demand function that underlies nearly all quantitative spatial economy and trade models, and (ii) free entry. Under these two conditions we show that if firms are profit-maximizers, the scale elasticity, ψ_s, always equals an industry-wide measure of firm market power and the trade elasticity, θ_s, can be identified as the *aggregate* demand elasticity facing a national industry. This observation is robust to how the cost function is parameterized or how many inputs are employed into production.

Based on this observation, we jointly estimate the industry-level scale and trade elasticities by fitting our structural import demand function to data on the universe of Colombian import transactions from 2007 to 2013. The firm-level nature of our empirical strategy exposes us to an important identification challenge. Specifically, standard estimations of import demand are often conducted at the country-level, using tariffs as an exogenous instrument to identify the underlying parameters. This identification strategy is, however, not applicable to our firm-level estimation.

To achieve identification, we take inspiration from the *Bartik* instrument that has been popularized by labor economists. To this end, we compile a comprehensive database on monthly exchange rates. Then, we interact aggregate movements is monthly exchange rates with the monthly structure of firm-level exports to construct a *shift-share* instrument that measures exposure to exchange rate shocks at the firm-product-year level.

Using the above strategy, we conduct both a pooled estimation on the entire sample and an industry-level estimation to identify the industry-level trade and scale elasticities. Our pooled estimation indicates that $\alpha = \textit{scale elasticity} \times \textit{trade elasticity} = 0.6$ across all manufacturing industries. To give perspective, this number lies between two polar assumptions in the literature: perfectly competitive models (e.g., Armington and Eaton and Kortum (2002)) assume that $\alpha = 0$, whereas monopolistically competitive models (e.g., Krugman (1980); Melitz (2003)) assume that $\alpha = 1$. Our industry-level scale elasticity estimates, meanwhile, display a considerable amount of variation. More specifically, we estimate scale economies to be stronger in the Machinery and Transportation sectors, and weaker in the Agriculture, Mining and Mineral sectors.

We highlight the importance of our micro-level estimates in three different contexts. First, given our scale and trade elasticity estimates, we can produce more credible estimates for the gains from trade and isolate the exact contribution of scale economies to these welfare gains. To be specific, trade has two effects on aggregate welfare: a *direct* effect (that operates through the term $1/\theta_s \times \%\Delta$ import intensity$_{i,s}$ in Equation 1) and a *scale-driven* effect. The scale-driven effect corresponds to how trade alters the structure of production and specialization in a country. Intuitively, if trade induces countries to specialize in scale-intensive industries (where returns to specialization are higher), the resulting gains will be larger.

The trade literature has been quite successful at quantifying the direct gains from trade. However, we know surprisingly little about the empirical size of the scale-driven gains. To eliminate this gap, we use our micro-level elasticity estimates to compute the scale-driven gains from trade. We find that, accounting for scale-driven gains, countries such as Germany, Finland, and Sweden benefit relatively more from trade at the expense of Greece, Mexico, and Russia. The intuition is that industries such as machinery or transportation feature stronger scale economies and offer higher returns to specialization. Countries that have a comparative advantage in these industries, therefore, gain relatively more from trade-induced specialization.

Our second application emphasizes industrial policy. First, we show that the optimal rate of import protection increases systematically with the industry-level scale elasticity but is rather insensitive to the trade elasticity. These results sheds empirical light on the old idea that countries should protect scale-intensive industries.[5] Second, we find that optimal production subsidies also vary systematically with the scale elasticity, and cross-subsidies from low-α to high-α industries. While preliminary, these applications lay a foundation for empirical assessment of many classical protectionist policies.

Finally, we show that our estimated scale elasticity values can help resolve the *income-size* elasticity puzzle (Rose (2006); Ramondo et al. (2016)). The puzzle concerns the observation that standard quantitative trade models predict a strong and positive relationship between real per capita income and population size.[6] The factual relationship, however, is negative and statistically insignificant. Furthermore, the income-size elasticity predicted by standard models is so large that introducing domestic trade frictions only partially mitigates it (Ramondo et al. (2016)). We show that plugging our estimated scale elasticity into

[5]This finding corroborates the theoretical result in Beshkar and Lashkaripour (2017).

[6]The Eaton and Kortum (2002) model would also predict a counterfactually strong income-size elasticity given that the stock of non-rival ideas in any country is proportional to its population size (see Ramondo et al. (2016)).

this class of models eliminates this counterfactual prediction altogether.

Our main contribution to the existing literature is methodological. Many empirical studies, including Head and Ries (2001), Antweiler and Trefler (2002), Davis and Weinstein (2003), and Costinot et al. (2016), have confirmed the prevalence of scale economies in various settings. Some recent studies have estimated the scale elasticity as the elasticity at which aggregate export sales increase with the size of the industry-level labor input (Somale (2017); Bartelme et al. (2017)). This approach has the advantage of accounting for external economies of scale. By contrast, we adopt a more standard definition of the scale elasticity[7] and estimate it using firm-level rather than industry-level data. In addition to simultaneously identifying both the trade and scale elasticities, our *micro-level* approach has the advantage of requiring minimal parametric restrictions on the cost function or the number of inputs used in production.

Our results also contribute to a historical debate about the origins of intra-industry trade (Head and Ries (2001); Davis and Weinstein (2003)). The trade literature hosts two dominant but competing views on this topic. The first point of view is the so-called *biological view*, where nationally-differentiated products or technologies arise in response to peculiarities in local demand (Bhagwati (1982); Feenstra (1982)). Second, is the popular view that intra-industry trade is driven by increasing returns to scale (Krugman (1980)). Our estimation sheds fresh light on the relative importance of these competing forces across different industries. Overall, we estimate a greater role for economies of scale in most industries. Nonetheless, with the exception of a few industries, we find national differentiation to be an economically significant force.

While not the primary focus of this paper, our estimation of the trade elasticity exhibits two novel elements. First, to the best of our knowledge, our estimation is the first to identify the trade elasticity using firm-level variations. Second, unlike traditional approaches, we estimate the trade elasticity without imposing restrictions on the scale elasticity parameter. Quite encouragingly, our approach delivers trade elasticity estimates that resemble those in the existing literature, including, among others, Broda and Weinstein, 2006, Caliendo and Parro (2014), Feenstra et al. (2017), and Simonovska and Waugh (2014).[8]

At a broader level this paper contributes to a vast literature studying cross-country income differences. These differences are puzzlingly large, prompting

[7] We use the standard definition that $\psi = \partial \ln \text{Output} / \partial \ln \text{Total Cost}$—see Morrison (2012).

[8] Among the aforementioned studies, the analysis in Feenstra et al. (2017) is more closely

a vast literature that seeks possible explanations (e.g., Hall and Jones (1999); Acemoglu et al. (2001); Parente et al. (2002); Klenow and Rodriguez-Clare (2005); Caselli and Coleman (2006); Restuccia and Rogerson (2008); Jones (2011)). In theory, an important fraction of cross-country income differences may be driven by specialization across high-return (scale-intensive) and low-return industries. The micro-level scale elasticity estimates provided by this paper, pave the way for quantifying such macro-level implications.

2 Reduced-Form Evidence

More often than not, quantitive trade models featuring scale economies assume the following rather arbitrary normalization

$$\alpha \equiv \text{scale elasticity} \times \text{trade ealsticity} = 1$$

The above normalization underlies, for instance, the widely-used Melitz-Pareto and Krugman models. In the words of Benassy (1996), the above normalization creates a somewhat arbitrary link between the market power of firms, which is reflected in the scale elasticity, and the *international* 'taste for variety', which is reflected in the trade elasticity. Setting $\alpha = 1$ of course serves a practical purpose: it allow the researchers to back out the industry level scale elasticity from a wide variety of readily-available trade elasticity estimates.

Considering the above background, it may be fruitful that before delving into the main analysis we highlight the predictions that result from setting $\alpha = 1$. Following Kucheryavyy et al. (2016) and Ramondo et al. (2016), the Melitz-Pareto or Krugman models featuring $\alpha = 1$ predict that

i. The number of export varieties increases proportionally with the exporting country's population size;

ii. Export sales increase proportionally with the number of export varieties;

related to ours. Using sector-level trade and production data, they estimate the macro and micro elasticities of substitution in a nested CES utility function, where the upper-level nest distinguishes between domestic and foreign composites. Despite the apparent similarities our framework is somewhat orthogonal to theirs. Whereas they distinguish only between domestic and foreign varieties, our framework distinguishes between all varieties by country of origin. Such a flexible aggregation fits our goal better, as it imposes minimal parametric restrictions on the underlying scale elasticity.

iii. In a cross-section of countries, real income per worker is strongly increasing in population size.

Prediction (i), unlike the other two predictions, stems from the *free entry* condition and has less to do with $\alpha = 1$—in fact, (i) is implied by wide range of non-CES frameworks as well. Predictions (ii) and (iii), however, are direct artifacts of setting $\alpha = 1$. Below, we contrast predictions (i)-(iii) against reduced form evidence, to note that setting $\alpha = 1$ may not be as innocuous as is often assumed.

2.1 Data Description

Our primary data source covers daily import transactions from the Colombian Customs Office for the 2007–2013 period.[9] The data include detailed information about each transaction, such as the Harmonized System 10-digit product category (HS10), importing and exporting firms,[10] f.o.b. (free on board) and c.i.f. (customs, insurance, and freight) values of shipments in US dollars, quantity, unit of measurement (of quantity), freight in US dollars, insurance in US dollars, value-added tax in US dollars, country of origin, and weight. The uniqueness of this data set is that it reports the identities of all foreign firms exporting to Colombia. This allows us to define a variety as a firm-product combination—In comparison, most papers focusing on international exports to a given location typically treat varieties as a more aggregate country-product combination. Table 1 reports a summary of basic trade statistics in our data.

When working with the above data set we face the challenge that, for some products, Colombia has been changing the HS10 classification between 2007 and 2013. Fortunately, the Colombian Statistical Agency, DANE, has kept track of these changes,[11] and we utilized this information to concord the Colombian HS10 codes over time. In the process, we followed the guidelines outlined by

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

[10] The 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.). We deal with this problem by standartizing the spelling and the length of the names along with utilizing the data on firms' phone numbers. The detailed description of cleaning the exporters' names is provided in the Technical Appendix.

[11] We thank Nicolas de Roux and Santiago Tabares for providing us with this information.

Table 1: *Summary Statistics of the Colombian Import Data.*

	Year						
Statistic	2007	2008	2009	2010	2011	2012	2013
F.O.B. value (billion dollars)	30.77	37.26	31.39	38.41	52.00	55.79	56.92
$\frac{\text{C.I.F. value}}{\text{F.O.B. value}}$	1.08	1.07	1.05	1.06	1.05	1.05	1.05
$\frac{\text{C.I.F. + tax value}}{\text{F.O.B. value}}$	1.28	1.21	1.14	1.19	1.15	1.18	1.15
No. of exporting countries	210	219	213	216	213	221	224
No. of imported varieties	483,286	480,363	457,000	509,524	594,918	633,008	649,561

Notes: Tax value includes import tariff and value-added tax (VAT). The number of varieties corresponds to the number of country-firm-product combination imported by Colombia in a given year.

Pierce and Schott (2012) for the concordance of the U.S. HS10 codes over time.[12] Overall, changes in HS10 codes between 2007 and 2013 affect a very small portion (less than 0.1%) of our dataset.

We also use aggregate statistics from other sources: *(i)* aggregate bilateral merchandise trade flows from the U.N. COMTRADE database (Comtrade (2010)); *(ii)* national account data from the World Bank database (World-Bank (2012)) and PENN WORLD TABLE version 9; *(iii)* monthly average exchange rates from the Bank of Canada;[13] and *(iv)* data on bilateral distance, common official language, and borders from Mayer and Zignago (2011). Below, we present three empirical patterns regarding the relationship between population size, the number of exported varieties, export values, and real per capita income.

2.2 Suggestive Evidence on Scale Effects

Utilizing the micro-level data described above we present three stylized facts relating to predictions (i)-(iii). We start with an observation that relates to prediction (i) about the number of export varieties.

Pattern 1. *The number of exported varieties to Colombia increases proportionally with population size.*

We establish this pattern using the transaction-level import data from Colom-

[12] To preserve the industry identifier of the product codes, and in contrast to Pierce and Schott (2012), we try to minimize the number of the synthetic codes. The concordance data and do files are provided in the data appendix.

[13] See *http://www.bankofcanada.ca/rates/exchange/monthly-average-lookup/*.

Table 2: *The Number of Exported Varieties vs. National Characteristics*

Dependent variable: Number of exported varieties (log)

	Regressor (log)					
Sample	Population (L)	GDPpc	Distance	Border	R^2	Obs.
All Products	1.01***	1.09***	-0.42***	1.42***	0.71	839
	(0.03)	(0.03)	(0.06)	(0.27)		
Manufacturing	1.02***	1.08***	-0.34***	1.50***	0.69	832
	(0.03)	(0.04)	(0.07)	(0.29)		

Notes: Estimation results of Equation (2). Robust standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. The estimation is conducted with year fixed effects.

bia. We define a variety as a country-firm-product combination imported by Colombia in a given year. Using this definition, we calculate the *nominal* number of varieties exported to Colombia from country i in year t, labeling it N_{it}.[14] We run the following regression to estimate the elasticity with which the number of export varieties increase with population size:

$$\ln N_{it} = \kappa \cdot \ln L_{it} + \beta \cdot \text{Controls}_{it} + \delta_t + \varepsilon_{it}. \quad (2)$$

The above regression includes standard controls such as GDP per capita, distance, and a common border, plus a set of year dummy variables, δ_t. The estimation results (displayed in Table 2) suggest that not only the number of export varieties increases with population size, but coefficient κ assumes a value of exactly "one." There are various ways to interpret this result. In the context of variety-based models (e.g., Krugman (1980) and Melitz (2003)) N_{it} reflects the number of differentiated firm-product varieties developed in country i. In the context of idea-based models (e.g.,Eaton and Kortum (2001)), N_{it} is reflective of the stock of technological knowledge in country i. Either way, the evidence aligns with the prediction that free-entry leads to the stock of varieties/knowledge to increase proportionally with population size.

Next, we turn to prediction (ii), which states that aggregate export sales also increase proportionally with the number of export varieties. Recall that –unlike prediction (i)– prediction (ii) is an artifact of setting $\alpha = 1$, so perhaps

[14] For example, in 2007, a product with an HS10 code 8428101000 *(Ascensores sin cabina ni contrapeso —Elevators without a cabin or counterweight),* was exported to Colombia by firm *"MITSUBISHICO"* from two countries of origin—namely, Japan and Thailand. We treat these as two distinct varieties: one exported to Colombia by Japan and one exported by Thailand.

not surprisingly it does not align well with the following observation.

Pattern 2. *Export sales to Colombia increase less-than-proportionally with the number of export varieties.*

To establish the above pattern, we estimate the elasticity of X_{it} (country i's export sales to Colombia in year t) with respect to N_{it} (the nominal number of varieties exported from country i to Colombia in year t):

$$\ln X_{it} = \mu \cdot \ln N_{it} + \delta_t + \varepsilon_{it}. \tag{3}$$

The above equation resembles a constant elasticity gravity estimation. Silva and Tenreyro (2006) point out that in the presence of heteroskedasticity, consistent estimates of the elasticity μ could be attained with a Poisson pseudo-maximum-likelihood (PPML) estimator. Complying with this well-established approach, we estimate Equation 3 with a PPML estimator (OLS estimates are also reported for comparison).

The estimation results displayed in Table 3 suggest that export sales increase less-than-proportionally with the number of exported varieties. Consequently, even though a larger population size increases the stock of varieties/knowledge proportionally, it has a less-than-proportional effect on export sales: $\frac{\partial \ln X}{\partial \ln L} = \frac{\partial \ln X}{\partial \ln N}\frac{\partial \ln N}{\partial \ln L} \approx \mu < 1$.[15] This observations is clearly inconsistent with setting $\alpha = 1$, which as noted entails that $\frac{\partial \ln X}{\partial \ln L} = 1$.

While quite revealing, Patterns 1 and 2 should be interpreted with some caution. We are analyzing the number of exported firm-product combinations, as opposed to the number of exporting firms (as in Fernandes et al. (2015)) or the number of exported product categories (as in Hummels and Klenow (2005)). This distinction is critical because (i) each firm typically exports multiple product lines, or (ii) each product category features multiple exporting firms. In fact, while sales per variety fall with total exports, sales per firm increase, which reflects the fact that firms in general export multiple product lines.[16] Similarly, in

[15]The idea-based interpretation of Pattern 2. can be stated as follows. If one perceives the number of distinct varieties as presenting the stock of technological knowledge, Pattern 2 points to diminishing returns to knowledge in *levels* (see Jones (1995) for parallels in growth theory). Among other things, pattern 2 may reflect the possibility that product varieties or technologies are relatively similar *within* countries. Under that interpretation, the *extreme love-of-variety* assumption would be analogous to the *independence of irrelevant alternatives* (IIA) assumption. Pattern 2, therefore, may suggest that the IIA assumption is too restrictive in the context of international trade.

[16]We estimate that the number of exporting firms increases with less-than-proportionally

***Table 3:** Export Sales vs. Number of Exported Varieties.*

Dependent: Total Export Sales to Colombia (log)

	All product		Manufacturing	
Estimator	PPML	OLS	PPML	OLS
No. of Exported Varieties (log)	0.82***	0.68***	0.82***	0.70***
	(0.03)	(0.03)	(0.02)	(0.02)
R^2	0.87	0.79	0.87	0.78
Observations	861	861	854	854

Notes: Estimation results of Equation (3). Robust standard errors in parentheses. *** denotes significance at 1%. The estimation is conducted with year fixed effects on a panel of 123 exporting countries over 7 years.

line with the findings of Hummels and Klenow (2005), sales per product category also increase with total national sales. Again, this observation reflects the fact that each additional product category exported by country i typically involves multiple firms.[17]

Finally, we turn to prediction (iii) about cross-country income difference. To this end, we document the following macro-level observation.

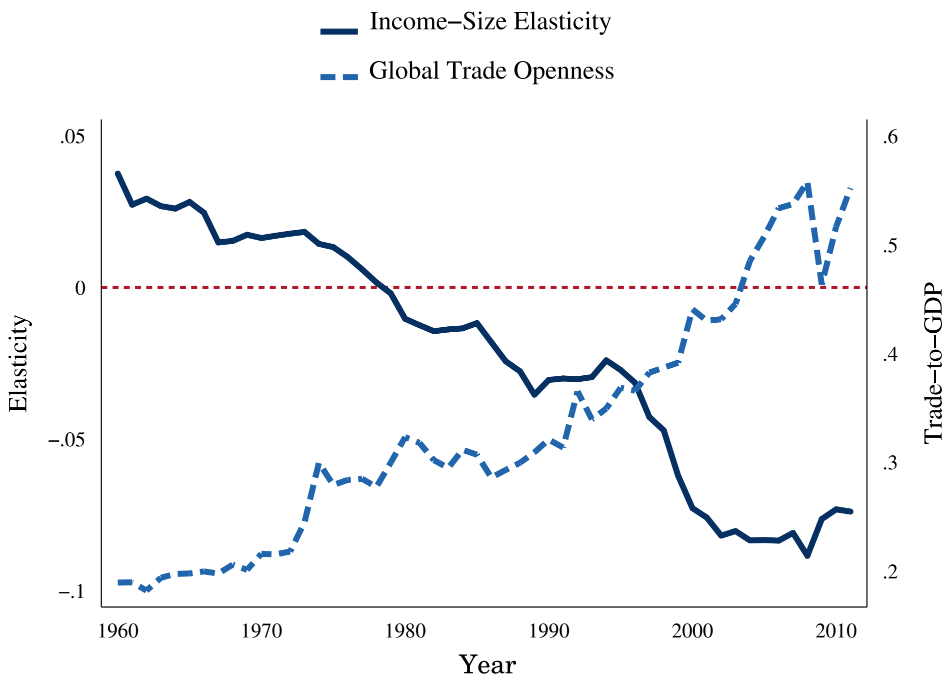
Pattern 3. *With the rise of international trade, the income-size elasticity has transformed from weakly positive in the 1960s to negative in the late 2000s.*

Figure 1 illustrates the above pattern. For each year from 1960 and 2008, we estimate the income-size elasticity (i.e., $\frac{\partial \ln \text{Real GDP p/c}}{\partial \ln \text{Population size}}$) using the cross section of countries from the PENN WORLD TABLES. Starting from 1960, the is income-size elasticity is weakly positive but becomes negative post 1980. As illustrated by the top panel in Figure 1, the decline of the *income-size* elasticity coincides with the rise of international trade. To an extent, this trend helps us understand why previous studies have struggled to find a systematic relationship between population size and national prosperity (e.g., Rose (2006)): both the magnitude and the sign of the income-size elasticity depends on the size of the trade barriers and can change over time. However, as noted by Ramondo et al. (2016),

with population size, whereas total sales increase more-than-proportionally to the number of exporting firms. This result resonates with the findings in Fernandes et al. (2015) plus Eaton et al. (2011).

[17] Altogether, we believe Patterns 1.A and 1.B lend themselves well to inferences regarding the import demand system. In particular, the co-existence of multiple varieties confirms the love-of-variety. However, the love-of-variety and the resulting scale economies may be more nuanced than is generally assumed.

Figure 1: *Trade Openness and the Income-Size Elasticity Over Time.*



Notes: the graph suggests that after 1980, larger economies have on average a lower real income per-capita, which contradicts the prediction (iii) implied by $\alpha = 1$.

the negative income-size elasticity cannot be easily reconciled with standard quantitative trade models. A trade model assuming $\alpha = 1$ predicts that international trade should weaken the effect of population size on the standard of living. However, if fitted to trade data in 2011 the model will nevertheless predict a positive relationship between population size and real per capita income (see Section 5). In other words, the scale effects implied by $\alpha = 1$ are too strong to be eliminated by the observed levels of economic integration.

3 Theoretical Framework

Now that we have illustrated the pitfalls of normalizing the scale elasticity, this section formally presents a methodology that can jointly recover the industry-level scale and trade elasticities from firm-level trade data.

Environment. There are multiple countries indexed by i and multiple industries or product categories indexed by h—one can, for instance, think of h as either HS6 industries or HS10 products categories. Each industry is served by a multitude of firms indexed by ω that compete in an imperfectly competitive setup.

The utility of the representative consumer in country i is given by

$$W_i = U_i(Q_{i,1}, ..., Q_{i,H})$$

where $Q_{i,h}$ denotes the industry k consumption bundle that aggregates across various firm varieties. Specifically speaking, $Q_{i,h} = Q_{i,h}(\boldsymbol{q}_{\omega i,h})$, where $\boldsymbol{q}_{\omega i,h}$ is a vector, with each element describing consumption of a given firm variety ω.

Production may use multiple factors of production, with the total cost of production for firm ω determined uniquely by it total production quantity, q, and economy-wide factor prices, $\mathbf{w}$. In particular, the total cost can be specified as

$$TC_{\omega,h}(q, \mathbf{w}) = VC_{\omega,h}(q, \mathbf{w}) + F_{\omega,h}(\mathbf{w}),$$

where $VC_{\omega,h}(q, \mathbf{w})$ denotes a variable cost that varies with output, and $F_{\omega,h}(\mathbf{w})$ denotes a fixed cost that is independent of output level. We impose no particular restriction on functions $VC_{\omega,h}(.)$ and $F_{\omega,h}(.)$ beyond them being continuous and twice differentiable. Given the above characterization, fixed costs are the main driver of scale economies. They account for *non-rivalrous* inputs such as R&D workers who develop *non-rival* ideas or blueprints.

The Scale Elasticity Within the above setup, we define the long-run *"scale elasticity"* as the elasticity at which output increases with production inputs. That is,

$$\psi_h \equiv \frac{\partial \ln q_h}{\partial \ln TC_h} - 1,$$

where $q_h \equiv \sum_\omega q_{\omega,h}$ and $TC_h \equiv \sum_\omega TC_{\omega,h}$. The above definition is relatively standard in the presences of multiple factors of production. It gauges the increase in inputs by looking at the change in total cost holding factor prices fixed. Put differently, ΔTC gives an aggregate of changes in all the different factors weighted by the initial factor prices. The above definition highlights three different possibilities:

i. Constant returns to scale (CRS) when $\psi_h = 0$,

ii. Increasing returns to scale (IRS) when $\psi_h > 0$, and

iii. decreasing returns to scale (DRS) when $\psi_h < 0$.

The *direct approach* to estimating ψ_h is to use industry-level data on input costs, TC_h, and output, q_h, to evaluate $\partial \ln q_h / \partial \ln TC_h$. Identification in that case will rely on exogenous demand shifters that are orthogonal to cost shocks. This approach obviously has many merits and can potentially identify both *external* and *internal* economies of scale (see Ciccone and Hall (1996) and Henderson (2003), as well as Bartelme et al. (2017) for an application in trade). The direct approach, however, has been historically christianized along three lines. First, credibly measuring all production inputs is quite challenging. Many factors of production are unobservable, non-rivalrous, or under-utilized, so accounting for them can be quite difficult (Hall (1989); Morrison (1992); Basu and Fernald (1997)). Second, as noted by Basu and Fernald (1997), estimating the scale elasticity using industry-level rather than firm-level data can lead to lead to substantive aggregation bias. Last, and perhaps most importantly from a policy perspective, the direct approach faces difficultly disentangling the long-run elasticity from the short-run elasticity (see Bresnahan (1989) and Chirinko and Fazzari (1994)). Given these challenges, it is perhaps not surprising that many direct estimates of the scale elasticity point to either constant or decreasing returns to scale (Burnside (1996); Basu and Fernald (1997)).

One can also apply the direct approach to trade data, and recover ψ_h based on the elasticity at which industry-level export sales increase with the size of the industry-level labor force (see Bartelme et al. (2017)). As with the above literature, this strategy has various merits but in addition to the challenges highlighted above, it faces another challenge that arises due to the structure of trade data. When applied to trade data, the direct approach cannot separately identify the scale elasticity from the trade elasticities—both of which are key to policy analysis. Instead, the direct approach can identify the scale elasticity only up to an externally chosen trade elasticity.

To overcome these underlying challenges, we propose an alternative approach that simultaneously recovers the industry-level scale *and* trade elasticities from firm-level trade data. In addition to jointly identifying two key elasticities, our approach has three basic merits (i) it imposes no particular parametric restriction on the cost function, (ii) by construction, it always identifies the long-run scale elasticity, and (iii) it is robust to the presence of many factors of production. Below, we thoroughly describe the theoretical underpinnings of our approach.

3.1 Recovering the Long-Run Scale Elasticity

Described in a nutshell, our approach recovers the industry-level scale elasticity, ψ_h, from an industry-wide measure of *firm* market power. To do so, we inevitably need to impose structure on (i) the demand function facing the firms and (ii) firm entry into industries. Encouragingly, we can rely exclusively on a set of restrictions that are consistent with an important class of quantitative trade and geography models. That being the case, the elasticities we recover can be consistently plugged in to these models to conduct macro-level policy analysis.

First, we assume the following parametric demand structure, which underlies nearly all workhorse quantitative trade models.

A1. *The within-industry demand structure is nested CES. That is, firm ω from country j in industry h faces the following demand function in market i*

$$q_{\omega i,h} = \varphi_{\omega i,h} \left(\frac{p_{\omega i,h}}{P_{ji,h}} \right)^{-(\vartheta_h+1)} \left(\frac{P_{ji,h}}{P_{i,h}} \right)^{-(\theta_h+1)} Q_{i,h},$$

where $p_{\omega i,h} = \tau_{ji,h} p_{\omega,h}$ denotes the consumer price that is composed of the factory gate price, $p_{\omega,h}$, and transport cost, $\tau_{ji,h}$. Moreover, $\varphi_{\omega i,h}$ is a variety-specific demand shifter; $P_{ji,h} = \gamma \left(\sum_{\omega' \in \Omega_{i,h}} \varphi_{\omega' i,h} p_{\omega' i,h}^{-\vartheta_h} \right)^{-1/\vartheta_h}$ denotes the aggregate price index of country j's export sales to market i, and $P_{i,h} = \tilde{\gamma} \left(\sum_j P_{ji,h}^{-\theta_h} \right)^{-1/\theta_h}$ denotes the overall price index of industry h in market i.

The nested-CES demand system characterized by A1 can arise either from a nested-CES utility function, in which case ϑ_h and θ_h respectively reflect the degrees of sub-national and cross-national product differentiation. Or alternatively it can arise from Ricardian specialization within industries, in which case ϑ_h and θ_h respectively reflect the degrees of sub-national and cross-national technology/knowledge differentiation (see Appendix B). Considering this, A1 underlies a large class of quantitative trade models. The Armington and Eaton and Kortum (2002) models for instance feature a special case of A1 where $\vartheta_h \longrightarrow \infty$. Krugman (1980) and Melitz (2003) feature another special case where $\theta_h = \vartheta_h$. Kucheryavyy et al. (2016), meanwhile, adopt A1 in its more general formulation to study the general equilibrium effects of scale economies.

Let $MR_{\omega,h} \equiv \sum_i \partial p_{\omega,h} q_{\omega i,h} / \partial q_{\omega i,h}$ denote the marginal revenue of firm ω.[18]

[18] The implicit assumption here is that transport costs are invariant to scale at the firm-level,

Assumption A1 simply asserts that the *price-to-marginal revenue* ratio is constant across firms serving industry h and equal to

$$\frac{p_{\omega,h}}{MR_{\omega,h}} = 1 + \frac{1}{\vartheta_h} \tag{4}$$

In that regards, A1 leads us to an industry-level measure of market power that is uniform across firms within industry h. In addition to A1, we need a second assumption regarding the structure of firm entry to relate the measure of firm-level market power to the scale elasticity, ψ_h. Here, we impose the following *long-run* assumption that is motivated by evidence in Hall (1989) and Rotemberg and Woodford (1999) that there are no significant pure profits in the United States.

A2. (**Free Entry**) *for each firm in industry h price equals average cost:* $p_{\omega,h} = TC_{\omega,h}/q_{\omega,h}$

A2 simply links the industry-wide degree of firm market power to the scale elasticity, ψ_h. To illustrate this, recall that

$$\psi_h = \frac{TC_h/q_h}{\partial TC_h/\partial q_h} - 1.$$

That is, by definition, the scale elasticity equals to the industry-level ratio of average to marginal cost. Also, note that for every firm serving industry h, (i) profit maximization entails that marginal cost equals marginal revenue: $MR = \frac{\partial TC}{\partial q}$, and (ii) A2 ensures that, in the long-run, price equals average cost: $p = TC/q$. Given (i) and (ii), then $\frac{TC_{\omega,h}/q_{\omega,h}}{\partial TC_{\omega,h}/\partial q_{\omega,h}} = \frac{p_{\omega,h}}{MR_{\omega,h}}$ for every firm in industry h, which considering Equation 4 implies that $\frac{TC_{\omega,h}/q_{\omega,h}}{\partial TC_{\omega,h}/\partial q_{\omega,h}} = 1 + 1/\vartheta_h$ for all firms in industry h. Considering that this relationship holds for all firms in industry h, the industry-level scale elasticity is then given by:

$$\psi_h = \frac{1}{\vartheta_h}$$

The above equation basically states that (conditional on A1 and A2) the industry-level scale elasticity equals the industry-wide degree of firm market power. That being the case, we can recover the industry-wide scale elasticity by estimating parameter ϑ_h per industry. Moreover, conditional on the set of firms (i.e., conditional on selection effects), the other structural demand parameter,

i.e., $\partial \tau_{ji,h}/\partial q_{\omega i,h} \approx 0$.

θ_h, determines the aggregate industry-level trade elasticity. Namely,

$$\frac{\partial \ln X_{ji,h}}{\partial \ln \tau_{ji,h}} = \sum \theta_k \frac{\partial \ln p_{\omega i,h}}{\partial \ln \tau_{ji,h}} + (\vartheta_h - \theta_h) \frac{\partial \ln P_{ji,h}}{\partial \ln \tau_{ji,h}} = \theta_h$$

In the presence of selection effects the trade elasticity will also depend on the firm productivity distribution, which we elaborate on in Appendix C.

Considering, the above framework the trade elasticity, θ_h, reflects the degree of international "taste for variety"—or in the case of idea-based models, it reflects the degree of international comparative advantage. The scale elasticity, ψ_h, meanwhile, reflects the industry-level degree of firm market power. This observation perhaps brings to light the arbitrary nature of setting, $\theta_h \psi_h = 1$, which as noted earlier is common-place in the literature. As Benassy (1996) notice early on, such a normalization creates an arbitrary link between the market power of firms, which is reflected in ψ_h, and the *international* 'taste for variety', which is reflected in θ_h. As we show next, since ψ_h and θ_h govern distinct macro-economic channels, setting $\theta_h \psi_h = 1$ can also cast problems for policy evaluation.

3.2 The Macro-Level Implications of ψ_h

To highlight the role of the scale elasticity in macro-level policy analysis, we place it into a general equilibrium context. It is needless to say that, while identifying the scale elasticity from market power relies on rather weak parametric assumptions, the implied macro-level effects will be more-or-less sensitive to the underlying general equilibrium model. A natural choice of model here is a multi-sector version of Krugman (1980), within which we show how the scale elasticity regulates both (i) the macro-level effect of industrial policy and (ii) the choice of optimal policy across industries.

The Macro-Level Effects of Policy Shocks To analyze the macro-level effects of industrial policy, we consider a multi-industry Krugman model featuring a Cobb-Douglas utility aggregator across industries, i.e, $Q_i = \prod_h Q_{i,h}^{\beta_{i,h}}$, and a within-industry demand system that is characterized by A1. Each industry operates based on monopolistic competition and free entry (A2). On the supply side, the model involves labor as the sole factor of production. Firms are symmetric within country i, and operate using on the following industry-specific total cost function:

$$TC_{i,h}(q) = \frac{w_i}{z_{i,h}} q + w_i f_h^e,$$

where w_i denotes the national wage rate in country i, $z_{i,h}$ corresponds to marginal productivity, and f_h^e denotes the fixed entry cost. Following A1, and noting that $\psi_h = 1/\vartheta_h$, each country j firm serving market i in industry h, generates the following sales:

$$p_{ji,h} q_{ji,h} = \frac{\varphi_{ji,h}}{N_{j,h}} \left(\frac{N_{j,h}^{\psi_k} p_{j,h}}{P_{i,h}} \right)^{-\theta_h} E_{i,h}, \tag{5}$$

where $E_{i,h} = \beta_{i,h} w_i L_i$ denotes total expenditure on industry h by market i. The free entry condition implies that

$$N_{j,h} = \frac{\psi_h f_h^e}{1 + \psi_h} L_{j,h} \tag{6}$$

where $L_{j,h}$ denotes the size of the country j's labor force in industry h. Plugging Equation 6 into Equation 5, the *aggregate* sales of country j to market i in industry h are be given by

$$X_{ji,h} = A_{ji,h}\, L_{j,h}^{\psi_h \theta_h}\, w_j^{-\theta_h}\, P_{i,h}^{\theta_h}\, E_{i,h},$$

where $A_{ji,h}$ is a function of (and only of) structural parameters that are invariant to trade policy shocks. Rearranging the above formula for the case of $X_{ii,h}$, leads us to the following expression:

$$Q_{i,h} = \frac{w_i}{P_{i,h}} = A_{ii,h}\, L_{j,h}^{\psi_h}\, \lambda_{ii,h}^{-1/\theta_h},$$

where $\lambda_{ii,h} \equiv X_{ii,h}/E_{i,h}$ denotes country i's domestic expenditure share in industry h. Therefore, noting that that $W(\boldsymbol{Q_i}) = \prod_h Q_{i,h}^{\beta_{i,h}}$, the total welfare or TFP of country i can be stated as

$$\ln W_i = \tilde{A}_i \prod_h L_{i,h}^{\beta_{i,h} \psi_h} \prod_h \lambda_{ii,h}^{-\frac{\beta_{i,h}}{\theta_h}}.$$

Finally, given the above formula, the change in aggregate TFP in response to a policy shock is given by

$$d \ln W_{i,h} = \sum_h \beta_{i,h} \psi_h d \ln L_{i,h} - \frac{\beta_{i,h}}{\theta_h} d \ln \lambda_{ii,h}$$

The above equation states that changes in TFP in response to policy, will depend on the change in (i) industry-level trade shares, $d \ln \lambda_{ii,h}$, and (ii) industry-level employment, $d \ln L_{i,h}$. The former reflects the *pure* gains from specialization and is governed by the trade elasticity, θ_h, while the latter reflects the scale-driven driven gains from specialization and is governed by the scale elasticity, ψ_h. Considering the above decomposition, setting $\psi_h \alpha_h = 1$ is not only arbitrary but also quite problematic vis-à-vis policy analysis.

Optimal Import Policy. Not only do the macroeconomic effects of industrial policy depend on the sector-level scale elasticities, but so does the optimal choice of policy. We demonstrate this using the case of the optimal import policy. As is common in the trade policy literature, let us temporarily abstract from income effects by assuming a quasi-linear utility aggregator across sectors. Our goal is to characterize the optimal industry-level tariff, $t_{ji,h}$, which is imposed by country i on country j's exports in sector h. In the presence of such a tariff policy, the aggregate import demand function can be stated as

$$X_{ji,h} = A_{ji,h} L_{j,h}^{\psi_h \theta_h} \left(1 + t_{ji,h}\right)^{-\theta_h} w_j^{-\theta_h} P_{i,h}^{\theta_h} E_{i,h}.$$

Additionally, total expenditure in country will include both labor income and tax revenues: $E_i = w_i L_i + \sum_{j,h} \frac{t_{ji,h}}{1+t_{ji,h}} X_{ji,h}$. As we show in Appendix ???, the optimal tariff in this setting will depend directly on the industry-level scale elasticity as follows:

$$t^*_{ji,h} = \frac{\psi_h}{\psi_h - 1} r_{ji,h}.$$

Perhaps surprisingly, and in line with the finding in Beshkar and Lashkaripour (2017), the optimal tariff does not vary with the industry-level trade elasticity. The above result has deep roots in the history of economics thought. It brings to light classic arguments by Alexander Hamilton and Friedrich List (among others) that countries should protect scale-intensive (high-ψ) industries. However, as noted earlier, empirical exploration of these old ideas have been impeded by lack of credible estimates for the long-run scale elasticity.

4 Estimation

In this section we use the universe of Colombian import transactions from 2007–2013 to estimate the structural parameters $\psi_h = 1/\vartheta_h$ and θ_h. Our estimation relies on the import demand function described by A1. Since we are focusing on one importer, we hereafter drop the importer's subscript i, and add a year subscripts t to account for the time dimension of our data. With this slight change of notation, the export sales of firm ω located in country j to Colombia in industry or product category h are given by

$$x_{\omega,ht} = p_{\omega,ht} q_{\omega,ht} = \varphi_{\omega,ht} \left(\frac{p_{\omega,ht}}{P_{j,ht}} \right)^{-\vartheta_h} \left(\frac{P_{j,ht}}{P_{ht}} \right)^{-\theta_h} E_{ht}, \tag{7}$$

Note that subscript h has been used thus far to references industries. But in our empirical analysis, h will designate the most disaggregated industry/product category in our dataset, which is an HS10 product category. Variety "ω, ht" therefore denotes an imported variety sourced from firm ω (located in country of origin j) within HS10 product category h, in year t.

Given that $\psi_h = 1/\vartheta_h$, we rearrange Equation 7 to attain the following formulation (see Appendix B.3 for a detailed derivation):

$$x_{\omega,ht} = \chi_{ht} \, \tilde{\varphi}_{\omega,ht} \, p_{\omega,ht}^{-\theta_h} \left(\lambda_{\omega|j,ht} \right)^{1-\psi_h\theta_h}, \tag{8}$$

Where in the above equation $\chi_{ht} \equiv P_{ht}^{\theta_h} E_{ht}$, $\tilde{\varphi}_{\omega i,h} \equiv \varphi_{\omega i,h}^{\psi_h\theta_h}$, and $\lambda_{\omega|j,ht}$ denotes the within-national market share, which we will formally define later. Taking logs from the above Equation, and decomposing $\ln \tilde{\varphi}_{\omega,ht} = \delta_{\omega,h} + \varepsilon_{\omega,ht}$ into a systematic component, $\delta_{\omega,h}$, and an idiosyncratic component, $\varepsilon_{\omega,ht}$, yields the following stochastic log-linear import demand function:

$$\ln x_{\omega,ht} = -\theta_h \ln p_{\omega,ht} + (1 - \psi_h\theta_h) \ln \lambda_{\omega|j,ht} + \delta_{ht} + \delta_{\omega,h} + \varepsilon_{\omega,ht}, \tag{9}$$

In the above equation $\delta_{ht} \equiv \ln \chi_{ht}$ denotes a product-year fixed effect that is composed of the product-wide price index and total expenditure, $\delta_{\omega,h}$ is a firm-product fixed effect that (among other things) reflects firm-level quality, and $\varepsilon_{\omega,ht}$ is an unobservable demand shifter that reflects idiosyncratic variations in quality, measurement errors, or non-technological demand shifters specific to variety ω, ht.

In the estimating Equation 9, price and sales levels ($p_{\omega,ht}$ and $x_{\omega,ht}$) are directly observable per import variety, and the the within-national market share, $\lambda_{\omega|j,ht}$, can be calculated as follows. For every imported variety ω, ht, sourced from country j we observe total sales, $x_{\omega,ht}$, so we can calculate $\lambda_{\omega|j,ht}$ as

$$\lambda_{\omega|j,ht} = \frac{x_{\omega,ht}}{\sum_{\omega' \in \Omega_{j,ht}} x_{\omega',ht}},$$

where $\Omega_{j,ht}$ denotes the set of all varieties exported from country j to Colombia in product category h, in year t. Overall, our estimating equation closely resembles the nested demand function analyzed in Berry (1994). However, given the structure of our data, we adopt a distinct identification strategy.

Before presenting our estimation strategy, a discussion of what identifies $\alpha = \psi\theta$ is in order. Parameter $\alpha = \psi\theta$ essentially reflects the degree of national product/technology differentiation. Broadly speaking, a variety is (i) either imported from a thick market like China, in which case it is one of the many imported Chinese varieties (hence a low $\lambda_{\omega|j,ht}$), or (ii) it is imported from a thin market like Taiwan, in which case it is one of the few imported varieties from that country (hence a high $\lambda_{\omega|j,ht}$). Conditional on prices, if varieties originating from thick markets generate lower sales, our import demand function identifies this as a case where $1 > \alpha > 0$. That is to say, consumers are discounting low-$\lambda_{\omega|j,ht}$ varieties due to national product differentiation. By contrast, if consumers do not discount varieties based on country of origin, the import demand function identifies this as $\alpha = 1$.

We first estimate Equation 9 under a pooled specification where α_h and ψ_h are assumed to be uniform across product categories. To this end, we employ a first-difference estimator that eliminates the firm-product fixed effect, $\delta_{\omega,h}$, and drops observations pertaining to one-time exporters. We deem the first-difference approach appropriate given the possibility that $\varepsilon_{\omega,ht}$'s are sequentially correlated. Stated in terms of first-differences, our estimating equation takes the following form

$$\Delta \ln x_{\omega,ht} = -\theta \Delta \ln p_{\omega,ht} + (1-\alpha)\, \Delta \ln \lambda_{\omega|j,ht} + \Delta\delta_{ht} + \tilde{\varepsilon}_{\omega,ht}, \qquad (10)$$

where $\tilde{\varepsilon}_{\omega,ht} \equiv \Delta\varepsilon_{\omega,ht}$ represents a variety-specific demand shock, and $\Delta\delta_{ht}$ is a product-year fixed effect that controls for (i) product-wide inflation and (ii) growth in national expenditure on product h. The identification challenge here is that price and within-national market share are endogenous and may respond to the demand shock, $\tilde{\varepsilon}_{\omega,ht}$.[19]

[19] Note that the within-national market share, $\lambda_{\omega|j,ht}$, could be correlated with $\varepsilon_{\omega,ht}$ due to

While the first-difference transformation is a partial remedy to the identification problem, we employ an instrumental variable strategy to recover the price and within-national market share coefficients. To this end, we construct a variety-specific (or local) cost shifter that is uncorrelated with the variety-specific demand shock, $\tilde{\varepsilon}_{\omega,ht}$. Our identification strategy capitalizes on the monthly frequency of import transactions in our data. We compile an external database on aggregate monthly exchange rates. Then, we interact the monthly variation in aggregate exchange rates with the monthly composition of firm-level exports to construct a *shift-share* instrument, $z_{\omega,ht}$:

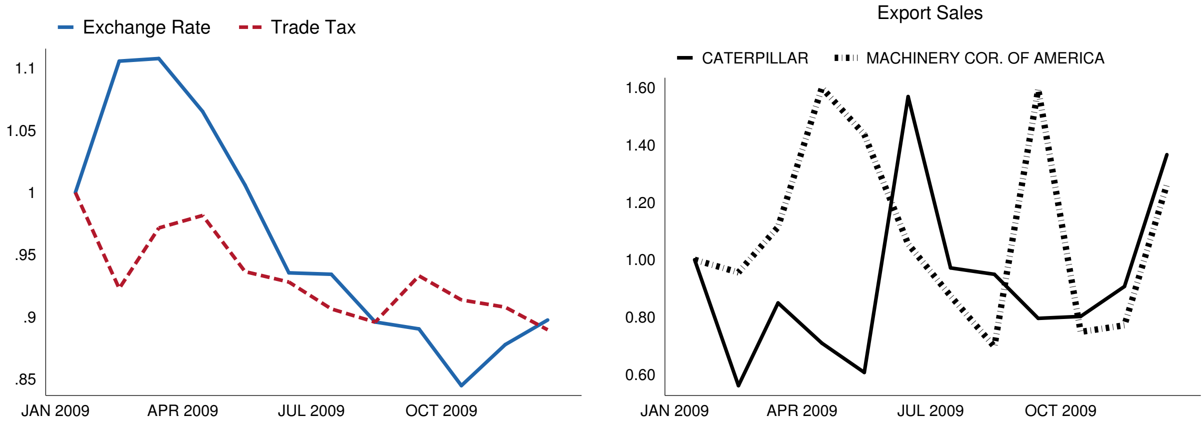
$$z_{\omega,ht} = \sum_{m=1}^{12} \frac{x_{\omega,hmt-1}}{x_{\omega,ht-1}} \Delta E_{jmt},$$

In the above expression, ΔE_{jmt} denotes the change in country j's exchange rate with Colombia in month m, year t; and $x_{\omega,hmt-1}$ represents month m sales of firm ω (from country j, within product category h) in the prior year, $t-1$. Hence, for variety ω, ht, the term $z_{\omega,ht}$ measures the variety-specific (or local) *exposure to exchange rate shocks* in year t. Put differently, depending on the monthly composition of sales to Colombia, aggregate exchange rate movements have differential effects on firms. These differential exposures to exchange rate shocks are picked up by $z_{\omega,ht}$—compare this with the widely-used *Bartik* instrument, which asserts that different regions are affected differentially by national labor market shocks depending on the industry composition of the region's production.

Figure 2 illustrates the workings of our instrument in more detail. It corresponds to U.S. exports in product category HS8431490000 (PARTS AND ATTACHMENTS OTHER FOR DERRIKS ETC.)—a product category that features one of the most frequently imported varieties: *machine parts* from "CATERPILLAR." The left panel of Figure 2 displays how both the exchange rate and the *average* import tax rate paid by U.S. based firms varied considerably on a monthly basis in 2009. The right panel plots the monthly variation in the export sales of the two largest U.S. based firms within category HS8431490000 (namely, "CATERPILLAR" and "MACHINERY CORP. OF AMERICA"). Given that the monthly

measurement errors in export sales. Our identification strategy takes care of this alternative source of endogeneity as long as our instruments are not correlated with variety-specific measurement errors. Similarly, unit prices could be contaminated with measurement errors, as they are averaged across transactions and consumers. This type of measurement error, however, is fairly innocuous given the log-linear structure of our demand function (see Berry (1994)),

Figure 2: *Monthly variations in national exchange rate, firm-specific exports, and import tax within a selected product-country-year category.*



Notes: The left panel plots monthly variations in exchange rate and (value-weighted) average import tax for US-based firms within product category HS8431490000-year 2009. The right panel plots monthly movements in export sales for the two biggest US firms in product HS8431490000 in year 2009— namely, Caterpillar and Machinery Corp. of America.

composition of exports from "CATERPILLAR" and "MACHINERY CORP. OF AMERICA" are markedly different, the two firms are affected differently by aggregate movements in the monthly exchange rate.

In addition to our measure of exchange rate shock exposure, we adopt a set of standard instruments borrowed from the existing literature. In particular, we instrument for the c.i.f. price with variety-specific import tax rates (which include tariff and the Columbian value added tax). Several studies, most notably Caliendo and Parro (2014), treat import taxes as an exogenous cost shifter to identify the trade elasticity. Following Khandelwal (2010), we construct two additional instruments for annual variations in the within-national market share, $\Delta \ln \lambda_{\omega|j,ht}$: *(i)* annual changes in the total number of country j firms serving the Colombian market in product category h, and *(ii)* changes in the total number of HS10 product categories actively served by firm ω in year t. These count measures will be correlated with $\Delta \ln \lambda_{\omega|jht}$ and uncorrelated with $\tilde{\varepsilon}_{\omega,ht}$ if variety entry and exit occur prior to, or independent of, the demand shock realization of competing varieties. This assumption is rather standard in the literature on estimating discrete choice demands curves (see Berry et al. (1995) or Khandelwal (2010)).

For obvious reasons, we are interested in estimating the trade and scale elasticities on a sector-specific basis. This involves estimating the import demand function separately for each sector s to identify a sector-level trade elasticity, θ_s, and a sector-level $\alpha_s = \psi_s \theta_s$. Each of the 14 sectors in our sample is comprised of multiple HS10 product categories. We thus pool all HS10 products belonging to sector s (i.e., $h \in H_s$) and estimate the following import demand function:[20]

$$\Delta \ln x_{\omega,ht} = -\theta_s \Delta \ln p_{\omega,ht} + (1-\alpha_s)\Delta \ln \lambda_{\omega|j,ht} + \Delta \Psi_{ht} + \tilde{\varepsilon}_{\omega,ht}\,, \qquad (11)$$

Note that, in principle, we can also estimate the import demand function separately for each HS10 product category. We opt, however, for elasticities estimated at the sector level for a practical reason. Our goal is to incorporate our estimates into the gains from trade formula, which requires data on trade, revenue, and expenditure shares. Since this data is available only at the sector-level, we only need sector-level elasticities to evaluate the gains from trade.

Results. The results of the pooled estimation (Equation (10)) are reported in Table 4. The estimation implies a pooled scale elasticity of $\psi = \alpha/\theta \approx 0.3$ across all manufacturing sectors. Given that $\alpha \approx 0.6 < 1$, the estimation essentially rejects the independence of irrelevant alternatives (IIA).[21] Or put differently, it rejects the arbitrary link between firm market power and the taste for international variety that is implied by $\alpha = 1$. Previously, some trade economists have referred to α as the degree of national product/technology differentiation. Under that definition, we find that national differentiation is prevalent, but not as extreme as assumed in competitive models like Armington and Eaton and Kortum (2002) where $\alpha = 0$.

The sector-level elasticity estimates, corresponding to Equation (11) are reported in Table 5. Perhaps expectedly, the sector-level elasticities display a considerable amount of variation. The estimated scale elasticity, $\psi = \alpha/\theta$, is highest in the Electrical & Optical Equipment ($\psi = 1.4$) and Transportation

[20] Sector groupings are based on the ISIC Rev. 3 classification to match the WIOD sectors—see https://wits.worldbank.org/product_concordance.html for concordance between HS10 codes and two-digit ISIC sectors.

[21] The *independence of irrelevant alternatives* (IIA) is rejected because product varieties or technologies are less differentiated intra-nationally than inter-nationally. While the IIA assumption has garnered considerable attention in the industrial organization literature, the trade literature has only recently tested this assumption against data. Redding and Weinstein (2016) estimate an international demand system that relaxes the IIA assumption by accommodating heterogeneous taste across consumers. Adao et al. (2015) estimate a trade model that (unlike standard CES models) permits varieties from certain countries to be closer substitutes. Our results contribute to this emerging literature by highlighting another aspect of the trade data that is at odds with the IIA assumption.

Table 4: *Import Demand Estimation*

	Manufacturing		Non-Manufacturing	
Variable (log)	IV	OLS	IV	OLS
Price, $-\theta$	-2.055***	0.070***	-3.058***	0.031***
	(0.119)	(0.001)	(0.300)	(0.003)
Within-national share, $1-\alpha$	0.367***	0.885***	0.186***	0.858***
	(0.013)	(0.001)	(0.031)	(0.008)
Weak Identification Test	177.99		28.07	
Under-Identification P-value	0.00		0.00	
N of Product-Year Groups	21,416	21,416	8,903	8,903
Observations	1,136,775	1,136,775	205,634	205,634
R-squared	...	0.82	...	0.76

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Estimation results of Equation (10). Standard errors in brackets are robust to clustering within product-year. The estimation is conducted with HS10 product-year fixed effects. The reported R^2 in the OLS specifications correspond to within-group goodness of fit. Weak identification test statistics is the F statistics from the Kleibergen-Paap Wald test for weak identification of all instrumented variables. The p-value of the under-identification test of instrumented variables is based on the Kleibergen-Paap LM test. The test for over-identification is not reported due to the pitfalls of the standard over-identification Sargan-Hansen J test in the multi-dimensional large datasets pointed by Angrist et al. (1996).

($\psi = 0.6$) sectors, and lowest in Agricultural & Mining ($\psi = 0.19$), Pulp & Paper ($\psi = 0.18$), and Mineral ($\psi = 0.22$) sectors. Furthermore, with the exception of Agriculture & Mining, Wood, and Minerals, we cannot reject the prevalence of scale economies.

While not the main focus of the paper, our trade elasticity estimates, θ_s, also display some novel properties. To our knowledge, our estimation is the first to identify sector-level trade elasticities using firm-level data. In general, our estimated trade elasticities are slightly lower than those estimated in Caliendo and Parro (2014) and Simonovska and Waugh (2014). Aside from the firm-level nature of our estimation, these differences may be driven by the fact that *(i)* instead of controlling for f.o.b. prices with exporter fixed effects, we directly use data on f.o.b. price levels, and *(ii)* we estimate a less-parametric import demand function that does not restrict the scale elasticity by imposing $\alpha = 0$.

4.1 Discussion

To summarize our results, we estimate that the actual force of scale economies, $\alpha = 0.6$, lies somewhere in between the two polar assumptions: competitive

Table 5: Import Demand Estimation by Sector

Sector	ISIC4 codes	Estimated Parameter θ_s	$1-\alpha_s$	ψ_s	Obs.	Weak Ident. Test
Agriculture & Mining	100-1499	4.584 (1.357)	0.137 (0.127)	0.188 (0.184)	11,715	3.50
Food	1500-1699	2.036 (0.928)	0.137 (0.046)	0.423 (0.240)	19,914	2.69
Textiles, Leather & Footwear	1700-1999	2.418 (0.214)	0.328 (0.022)	0.278 (0.031)	129,913	79.39
Wood	2000-2099	2.376 (1.269)	0.197 (0.201)	0.338 (0.389)	4,509	1.53
Paper	2100-2099	4.765 (2.157)	0.140 (0.132)	0.181 (0.189)	36,215	1.44
Petroleum	2300-2399	0.328 (0.176)	0.351 (0.108)	1.979 (1.227)	4,046	3.65
Chemicals	2400-2499	2.389 (0.216)	0.140 (0.024)	0.360 (0.070)	127,259	47.24
Rubber & Plastic	2500-2599	3.020 (0.571)	0.275 (0.075)	0.240 (0.080)	109,470	9.57
Minerals	2600-2699	3.912 (0.970)	0.135 (0.117)	0.221 (0.199)	24,572	5.11
Basic & Fabricated Metals	2700-2899	2.250 (0.351)	0.408 (0.039)	0.263 (0.048)	156,644	23.86
Machinery	2900-3099	2.471 (0.391)	0.222 (0.056)	0.315 (0.093)	247,482	27.68
Electrical & Optical Equipment	3100-3399	0.394 (0.143)	0.462 (0.020)	1.367 (0.502)	233,520	24.31
Transport Equipment	3400-3599	0.463 (0.235)	0.734 (0.024)	0.575 (0.293)	82,045	6.34
N.E.C. & Recycling	3600-3799	4.946 (0.549)	-0.014 (0.076)	0.205 (1.089)	153,493	25.44

Notes. Estimation results of Equation (11). Standard errors in parentheses. The estimation is conducted with HS10 product-year fixed effects. All standard errors are simultaneously clustered by product-year and by product-'country of origin.' The weak identification test statistics is the F statistics from the Kleibergen-Paap Wald test for weak identification of all instrumented variables. The test for over-identification is not reported due to the pitfalls of the standard over-identification Sargan-Hansen J test in the multi-dimensional large datasets pointed by Angrist et al. (1996).

Table 6: *Scale Elasticity in Mainstream Trade Models ($\alpha \equiv$ scale elast $\times$ trade elast).*

Estimation	Canonical Trade Models	
	Armington, Eaton-Kortum 2002	Krugman, Melitz, Eaton-Kortum 2001
$\alpha \approx 0.6$	$\alpha = 0$	$\alpha = 1$

trade models where $\alpha = 0$ (e.g., Eaton and Kortum (2002); Anderson (1979)) and imperfectly competitive models where $\alpha = 1$ (e.g., Krugman (1980); Eaton and Kortum (2001); Melitz (2003))—see Table 6. That is to say, scale economies, while prevalent, are weaker than generally assumed in "New" trade theories. We demonstrate in Appendix C that the same conclusion, regarding the strength of scale economies, applies to heterogeneous firm models. As we show in the following section, this finding revises many of the macroeconomic predictions produced by an important class of quantitative trade models.

Note that our identification approach rests on the nested CES import demand structure that underlies two distinct class of trade models: (i) *idea-based* models such as Eaton and Kortum (2002) and (ii) *variety-based* models such as Krugman (1980) and Melitz (2003). Our identification strategy, therefore, assumes different interpretations depending on the underlying micro-foundation. From the lens of *variety-based* models, our estimation simply identifies the extent to which an increase in industry-level inputs increases the stock of non-rival varieties/blueprints, and hence the real output-to-input ratio. Alternatively, from the lens of *idea-based* models, our estimation identifies the extent to which an increase in industry-level inputs increases the stock of quality-enhancing knowledge, and hence the real output-to-input ratio.

An attractive feature of our empirical strategy is that it simultaneously recovers the trade and scale elasticities without imposing any general equilibrium structure on firm entry or cost functions. In fact, our estimation recovers these elasticities, conditional on an observable vector of firm-level prices and sales. More importantly, by construction, our approach identifies the long-run scale elasticity that is more relevant to certain policy applications. Despite these merits, our approach has an basic limitation. When analyzing industrial policy, our estimates will deliver macro-level implications that are strictly model-specific. For example, the estimated ψ_s and θ_s's enter the aggregate welfare function differently in models *with* and *without* firm selection effect (see AppendixC).

Finally, our analysis provides an empirical basis for discriminating between two classic but competing views of intra-industry trade. The first one is the so-called *biological view,* where nationally differentiated products or technologies arise in response to peculiarities in local demand (Bhagwati (1982); Feenstra (1982)). This view corresponds to an α that is closer to *zero*. The second one is the popular view that intra-industry trade is driven by increasing returns to scale (Krugman (1980)). This latter view is more relevant the closer α is to *one*. Considering our results in Table 5, we estimate a greater role for aggregate returns to scale in most sectors. Nonetheless, with the exception of a few sectors, national differentiation is an economically significant force. These findings complement previous attempts to discriminate between these two views with more aggregated data (e.g., Head and Ries (2001)).

5 Macro-Level Policy Analysis

This section incorporates our sector-level elasticity estimates into a general equilibrium, multi-country model of the global economy. Our analysis focuses on three main general equilibrium applications. First, we use our micro-level estimates to identify the scale-driven gains from trade. Second, we study the implications of sector-level scale economies for the optimal design of industrial and trade policy. Third, we argue that our micro-level elasticity estimates can help resolve the cross-national income-size elasticity puzzle.

5.1 Sectoral Specialization and the Gains From Trade

Our first application computes the gains from trade based on the estimated scale and trade elasticities. Theoretically, it is well-understood that the gains from trade are highly sensitive to the sectoral variation in scale elasticities (Kucheryavyy et al. (2016); Costinot and Rodríguez-Clare (2014)). However, we know relatively little about the actual size of the scale-driven gains. Standard quantitative trade models often normalize the scale-driven gains by setting $\alpha = \psi\theta$ to either 0 or 1.

Here, using our micro-level estimates, we take a first stab at identifying the actual contribution of scale economies to the gains from trade. Following Proposition 2, we can write the gains from trade formula in terms of sector-level elasticities and trade statistics as follows:

$$\hat{W}_i = \prod_s \left(\frac{\beta_{i,s}}{r_{i,s}}\right)^{-\beta_{i,s}\psi_s} \prod_s \lambda_{ii,s}^{-\frac{\beta_{i,s}}{\theta_s}},$$

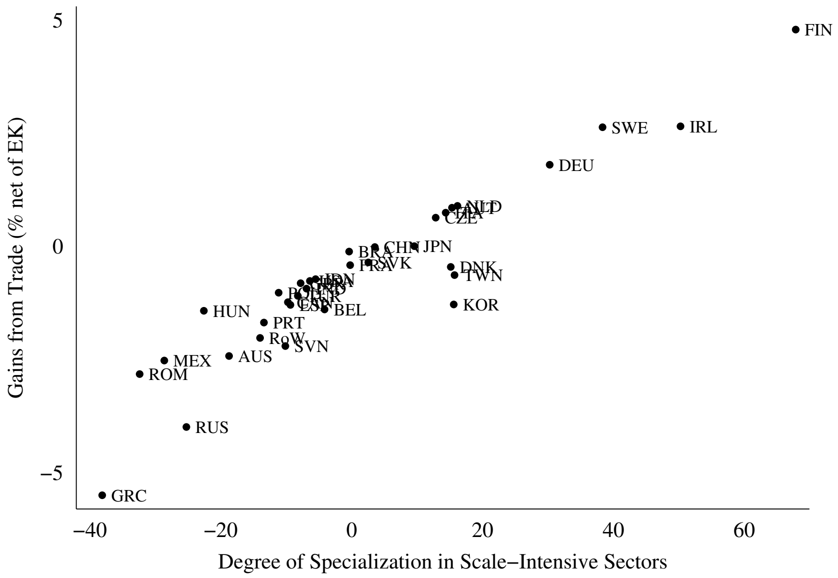
where s denotes a sector for which we have data on $\beta_{i,s}$, $r_{i,s}$ and $\lambda_{ii,s}$ from the world input-output database (WIOD) covering 32 sectors and 34 countries. In the previous section, we estimated θ_s and ψ_s for 15 traded sectors in the WIOD sample. Following Costinot and Rodríguez-Clare (2014), we normalize θ_s to 100 and the scale elasticity, ψ_s, to *zero* for the non-traded sectors. The computed gains from trade are reported in Table 9. Additionally, Table 9 reports the gains implied by two polar but standard frameworks: *(i)* the perfectly competitive EK-Armington framework (in which $\alpha_s = 0$ for all manufacturing sectors), and *(ii)* the standard monopolistic competition framework à la Krugman (1980) (in which $\alpha_s = 1$ for all manufacturing sectors).

Whereas the overall size of the gains from trade is fairly similar across all specifications, the cross-national allocation of the gains differs systematically. In particular, the scale-driven component of the gains from trade, $\prod_s (\beta_{i,s}/r_{i,s})^{-\beta_{i,s}\psi_s}$, favors *(i)* countries with higher degrees of specialization and *(ii)* countries that specialize in high-ψ sectors with stronger returns to scale. To elaborate on this claim, we use our micro-level elasticity estimates to compute two national indexes identified by Kucheryavyy et al. (2016). The first one is an index that captures the overall degree of specialization: $\sum_s \beta_s \log \frac{\beta_{i,s}}{r_{i,s}}$. The second index (namely, *scale intensity* of national output) measures the degree of specialization in scale-intensive (i.e., high-ψ) industries:

$$\text{Scale-Intensity}_i = \sum_{s=1}^{32} \left(\frac{\psi_s - \bar{\psi}}{\bar{\psi}}\right) \beta_s \log \frac{\beta_{i,s}}{r_{i,s}},$$

where $\overline{\psi} = \frac{1}{32}\sum_{s=1}^{32} \psi_s$ reflects the scale elasticity of the average sector. The above index adopts a higher value if more revenue is generated in sectors with strong returns to scale. To demonstrate this, Figure 3 plots the *scale-driven* gains from trade against the scale intensity of a country's output. Clearly, countries with more a scale-intensive output structure gain relatively more from trade. For example, accounting for factual scale elasticities, western European countries like Finland, Germany, and Sweden experience relatively larger gains from trade given their specialization in scale-intensive sectors. Greece, Mexico, Romania, and Russia, meanwhile, experience smaller gains given their specialization in

Figure 3: *Sectoral specialization and the scale-driven gains from trade*



sectors with weaker returns to scale.

The intuition behind the above findings is straightforward. International trade induces production specialization, leading to the expansion of comparatively advantaged sectors in each country. Low-ψ sectors such as Agriculture & Mining as well as Minerals are subject to weak scale economies, offering smaller returns to specialization and expansion. Greece, Mexico, and Russia have a comparative advantage in these low-return sectors, which effectively contracts their gains from trade. By contrast, high-ψ sectors such as Transport Equipment and Machinery feature strong scale economies, and offer greater returns to specialization. Finland, Germany and Sweden have a comparative advantage in these high-return sectors and, therefore, gain relatively more from trade-induced specialization.

5.2 Scale Economies and Industrial Policy

Much of the historical interest in scale economies was triggered by industrial policy considerations—these considerations can be traced back to the work of Alexander Hamilton and Friedrich List. There is a general consensus that industrial policy should promote sectors that offer higher returns to scale and specialization. While these arguments are well-established conceptually, their application has been hindered by a lack of empirical estimates for sector-level scale economies. Our estimated elasticities fill this gap and can be potentially used to guide industrial policy.

Here, we highlight the implications of our micro-level estimates for the design of two specific policies. In doing so, we fit the general equilibrium multi-sector model outlined in Section 3, to sectoral production and trade data from the WIOD. Our application involves the US and an aggregate of the rest of the world (ROW), 14 traded sectors and an aggregated non-traded sector. Using the calibrated model we analyze two unilateral policies conducted by the US government (i) production subsidies and (ii) import taxes.

Figure 4 displays the computed optimal US import tariffs under the estimated scale and trade elasticities and contrast them to the optimal tariff rates implied by the standard EK model. The graph reveals two general patterns. First, introducing sector-wide scale effects increases the sectoral heterogeneity in optimal tariffs. In particular, the multi-sector EK model without scale economies implies uniform tariffs across sectors, despite variation in the underlying trade elasticities. By comparison, under the estimated scale elasticities, optimal tariffs are non-uniform and display a considerable amount of sectoral variation.[22] This result confirms the widespread belief that scale economies should be a key consideration when in the design of trade policy.

Our second observation is that in the presence of sector-level scale economies, optimal tariffs adopt systematically lower values. While this finding resonates with earlier assertions in Markusen and Wigle (1989), the size of the effect is incredibly sensitive to the underlying scale elasticity. Previous arguments that highlight the effect of scale economies on optimal tariff levels are based on extreme scale elasticity normalizations (in spirit of Krugman (1980)). Our application, therefore, offers a credible view on the actual importance of these effects.

Using our calibrated model, we also compute the optimal sector-level production subsidies for the US economy. These marginal cost subsidies apply to all domestically produced varieties in a given sector, and are financed with a lump sum income tax. Figure 5 displays the optimal US production subsidies across various traded sectors. Evidently, the optimal subsidy depends on α=scale elasticity $\times$ trade elasticity, cross-subsidizing from low-α sectors to high-α sectors. These findings provide a first glimpse into the empirical relevance of classical industry policy recommendations, like those in Graham (1923).

[22] This result is in line with the theoretical finding of Beshkar and Lashkaripour (2017).

Figure 4: Optimal Import Policy: Benchmarks versus Eaton and Kortum (2002)

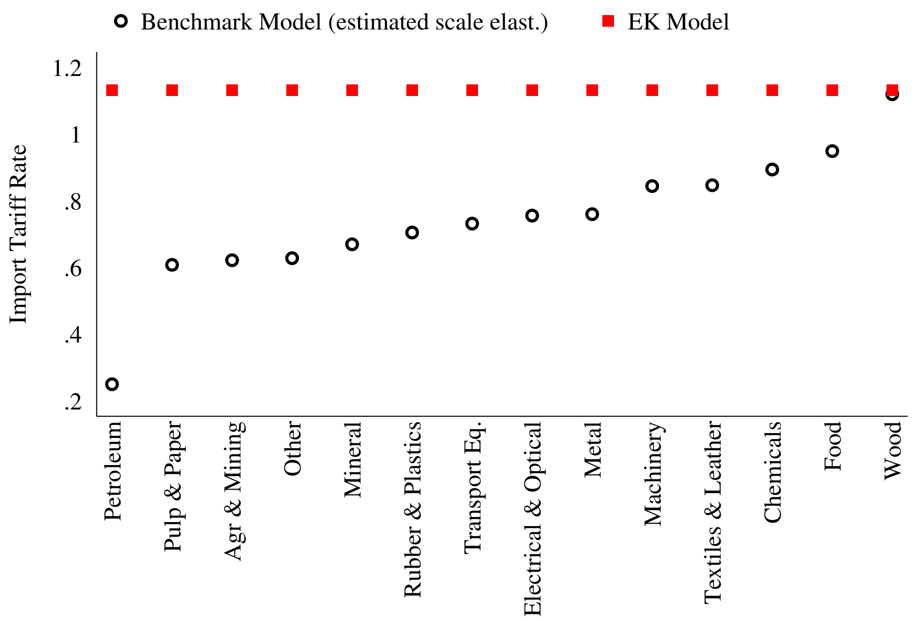
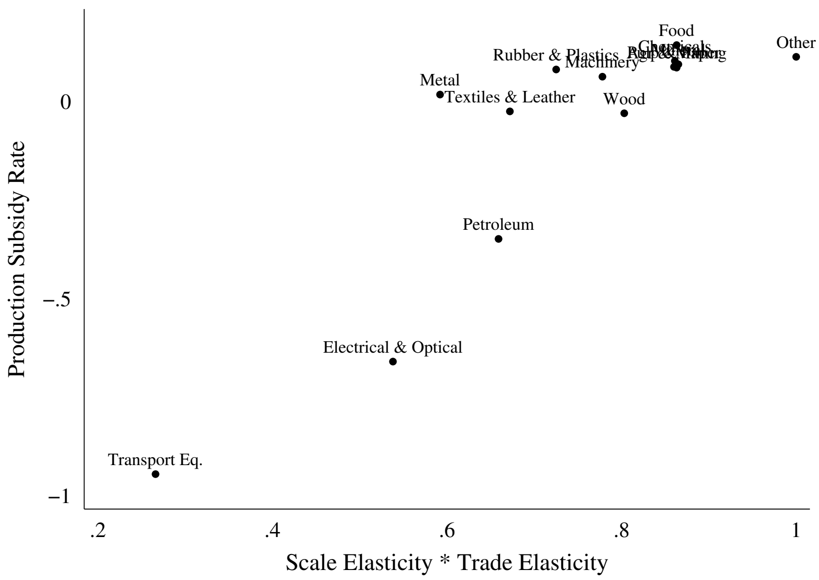


Figure 5: Optimal production subsidy vs. sector-level scale effects



5.3 Population Size and Welfare

In this section, we ask whether our estimates can help resolve the *income-size* elasticity puzzle. This puzzle, as noted by Ramondo et al. (2016), concerns the fact that a large class of quantitative trade models, including Krugman (1980), Eaton and Kortum (2001), and Melitz (2003), predict an income-size elasticity (i.e., the elasticity at which real per capita income increases with population size) that is counterfactually high. One straightforward remedy for this counterfactual prediction is introducing domestic trade frictions into these models. This treatment, however, is only a partial remedy. As shown by Ramondo et al. (2016), even after controlling for direct measures of internal trade frictions, the predicted income-size elasticity remains counterfactually strong.

To understand the income-size elasticity puzzle, consider a single-sector Krugman model that is governed by the demand structure and general equilibrium conditions outlined in Section 3. The model delivers the following expression relating country i's real income per worker or TFP ($W_i = w_i/P_i$) to its structural efficiency, A_i, population size, L_i, trade-to-GDP ratio, λ_{ii}, and a measure of internal trade frictions, τ_{ii}:

$$W_i = \gamma\, A_i\, L_i^{\psi}\, \lambda_{ii}^{-\frac{1}{\theta}}\, \tau_{ii}^{-1}. \tag{12}$$

The standard Krugman model assumes extreme love-of-variety (or extreme scale economies), which implies $\psi = 1/\theta$ and precludes internal trade frictions, which results in $\tau_{ii} = 1$. Given these two assumptions, we can compute the real income per worker predicted by the standard Krugman model and contrast it to actual data for a cross-section of countries.

For this exercise, we use data on the trade-to-GDP ratio, real GDP per worker, and population size for 116 countries from the PENN WORLD TABLES in the year 2011. Given our micro-estimated trade elasticity, θ, and plugging $\tau_{ii} = 1$ as well as $\psi = 1/\theta$ into Equation 12, we can compute the real income per worker predicted by the Krugman model. Figure 6 (top panel) reports these predicted values and contrasts them to factual values. Clearly, there is a sizable discrepancy between the income-size elasticity predicted by the standard Krugman model (0.36, standard error 0.03) and the factual elasticity (-0.04, standard error 0.06). To gain intuition, note that small countries import a higher share of their GDP (i.e., posses a lower λ_{ii}) which partially mitigates their size disadvantage. However, even after accounting for observable levels of trade openness, the scale economies underlying the Krugman model are so strong that they lead to a counterfactually high income-size elasticity.

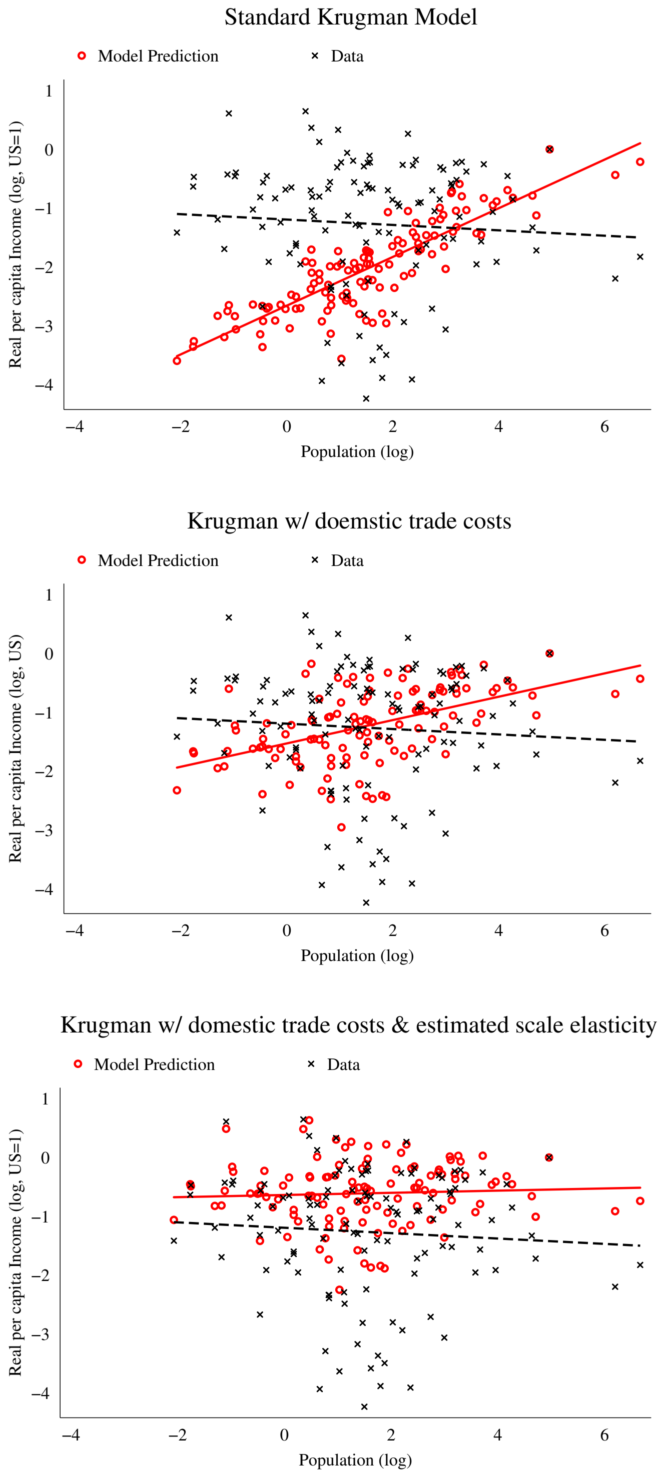
One solution to the income-size elasticity puzzle is introducing internal trade frictions into the Krugman model (i.e., relaxing the $\tau_{ii} = 1$ assumption). Ramondo et al. (2016) performs this task using direct measures of domestic trade frictions. Their calibration is suggestive of $\tau_{ii} \propto L_i^{0.17}$. Plugging this implicit relationship into Equation 12 and using data on population size and trade openness, we compute the model-predicted real income per worker and contrast it with actual data in Figure 6 (middle panel). Expectedly, accounting for internal frictions shrinks the the income-size elasticity. However, as pointed out by Ramondo et al. (2016), the income-size elasticity still remains puzzlingly large.

We ask if our micro-estimated scale elasticity can help resolve the remaining income-size elasticity puzzle. To this end, in Equation 12, we set the scale elasticity to $\psi = \alpha/\theta$ where α is set to 0.6 as implied by our micro-level estimation. Then, using data on population size and trade-to-GDP ratios, we compute the real income per capita predicted by a model that features both domestic trade frictions *and* adjusted scale economies. Figure 6 plots these predicted values, indicating that this adjustment indeed resolves the income-size elasticity puzzle. In particular, the income-size elasticity predicted by the Krugman model with adjusted scale economies is statistically insignificant (0.02, standard error 0.03), aligning very closely with the factual elasticity.

Economic versus political integration. At a broader level, the above application relates to a vibrant literature on optimal country size. An important goal of the aforementioned literature is characterizing the advantages of economic integration (free trade across political borders) versus political integration (elimination of political borders)—see Alesina et al. (2000) or Alesina and Spolaore (2005). Our micro-level estimation can shed new light on this rather old discussion. Specifically, given that scale economies are weaker than conventionally assumed, *do traditional analyses overstate the benefits of political integration?*

To address the above question, consider a symmetric world with $N > 1$ countries, each with population L. Let κ denote the elasticity at which internal trade frictions increase with population size ($\tau_{ii} \propto L^{\kappa}$). Considering Equation 12 and given that $\lambda_{ii} = \left(1 + (N-1)\,\tau^{-\theta}\right)^{-1}$ where τ denotes the *home-bias* or the *border effect*; real per capita income under economic integration will be $W_E = \gamma\left[1 + (N-1)\,\tau^{-\theta}\right]^{\frac{1}{\theta}} L^{\psi-\kappa}$. Under political integration, the world is re-

Figure 6: The income-size elasticity



duced to one country with a population $N \times L$ and real per capita income will equal $W_P = \gamma (N \times L)^{\psi - \kappa}$. Hence, the real income level achieved through political integration relative to economic integration can be stated as:

$$\frac{W_P}{W_E} = \left[\frac{N^{\alpha - \kappa\theta}}{1 + (N-1)\,\tau^{-\theta}} \right]^{\frac{1}{\theta}}$$

To put the above expression into perspective, consider a world consisting of 50 regions ($N = 50$). Based on our estimates of $\theta \approx 2.5$ and $\alpha \approx 0.6$, plus the $\kappa \approx 0.17$ implied by the Ramondo et al. (2016) analysis, then $\alpha - \kappa\theta \approx 0.17$. In this hypothetical world, the border effect should be around $\tau \approx 5$ for political integration (i.e., eliminating regional borders) to be superior to trade across political borders. By contrast, in the standard Krugman model where $\alpha = 1$, political integration is superior as long as the border effect is greater than $\tau \approx 2.1$. This back-of-the-envelop calculation demonstrates how the standard Krugman model overstates the benefits of size expansion through political integration. The intuition being that products and ideas developed within the same political borders are relatively similar. Hence, despite its benefits, political integration will diminish product diversity to some degree. Such an effect is overlooked under the standard love-of-variety specification. It is needless to say that the above arguments are extremely stylized to deliver any concrete resolution. At most our arguments are merely suggestive, whereas a comprehensive assessment of political versus economic integration should account for a long list of relevant factors overseen by our application.

6 Concluding Remarks

For centuries economists have theorized about the importance of scale economies in policy and welfare analysis. But perhaps surprisingly, we know relatively little about the empirical relevance of these classic theories. Beholding this gap, we took a preliminary step toward identifying the force of industry-level scale economies. To this end, we first utilized micro-level data to estimate industry-level scale elasticities. We then plugged our estimates in a general equilibrium quantitative model of trade and geography to study macro-level welfare and policy implications.

While we highlighted three macro-level implications, our micro-estimates have an even broader reach. Two implications, which were left out in the interest of space, merit particularly close attention. First, our scale elasticity estimates can help disentangle the relative contribution of scale economies and Ricardian comparative advantage to industry-level specialization. This is an old question, which from an empirical perspective, we know surprisingly little about.

Second, our estimates can shed fresh light on the puzzlingly large income gap between rich and poor countries. Economists have always hypothesized that a fraction of this income gap is driven by rich countries specializing in scale-intensive, high-return industries. Empirical assessment of these hypotheses, however, has been impeded by a lack of estimates for industry-level scale elasticities. Our micro-level estimates pave the way for an empirical exploration in this direction.

References

Acemoglu, D., S. Johnson, and J. A. Robinson (2001). The colonial origins of comparative development: An empirical investigation. *The American Economic Review 91*(5), 1369–1401.

Adao, R., A. Costinot, and D. Donaldson (2015). Nonparametric counterfactual predictions in neoclassical models of international trade. Technical report, National Bureau of Economic Research.

Alesina, A. and E. Spolaore (2005). *The size of nations*. Mit Press.

Alesina, A., E. Spolaore, and R. Wacziarg (2000). Economic integration and political disintegration. *American Economic Review 90*(5), 1276–1296.

Anderson, J. E. (1979). A theoretical foundation for the gravity equation. *The American Economic Review*, 106–116.

Angrist, J., G. Imbens, and D. Rubin (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association 91*, 444–455.

Antweiler, W. and D. Trefler (2002). Increasing returns and all that: a view from trade. *The American Economic Review 92*(1), 93–119.

Bartelme, D., A. Costinot, D. Donaldson, and A. Rodriguez-Clare (2017). External economies of scale and industrial policy: A view from trade. *Working Paper*.

Basu, S. and J. G. Fernald (1997). Returns to scale in us production: Estimates and implications. *Journal of political economy 105*(2), 249–283.

Benassy, J.-P. (1996). Taste for variety and optimum production patterns in monopolistic competition. *Economics Letters 52*(1), 41–47.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242–262.

Beshkar, M. and A. Lashkaripour (2017). Interdependence of trade policies in general equilibrium.

Bhagwati, J. N. (1982). Shifting comparative advantage, protectionist demands, and policy response. In *Import competition and response*, pp. 151–196. University of Chicago Press.

Bresnahan, T. F. (1989). Empirical studies of industries with market power. *Handbook of industrial organization 2*, 1011–1057.

Broda, C. and D. E. Weinstein (2006). Globalization and the gains from variety. *The Quarterly Journal of Economics 121*(2), 541–585.

Burnside, C. (1996). Production function regressions, returns to scale, and externalities. *Journal of monetary Economics 37*(2), 177–201.

Caliendo, L. and F. Parro (2014). Estimates of the trade and welfare effects of nafta. *The Review of Economic Studies*, rdu035.

Caselli, F. and W. J. Coleman (2006). The world technology frontier. *The American Economic Review 96*(3), 499–522.

Chirinko, R. S. and S. M. Fazzari (1994). Economic fluctuations, market power, and returns to scale: Evidence from firm-level data. *Journal of Applied Econometrics 9*(1), 47–69.

Ciccone, A. and R. E. Hall (1996). Productivity and the density of economic activity. *The American Economic Review*, 54–70.

Comtrade, U. (2010). United nations commodity trade statistics database. *URL: http://comtrade. un. org*.

Costinot, A., D. Donaldson, M. Kyle, and H. Williams (2016). The more we die, the more we sell? a simple test of the home-market effect. Technical report, National Bureau of Economic Research.

Costinot, A. and A. Rodríguez-Clare (2014). Trade theory with numbers: Quantifying the consequences of globalization. *Handbook of International Economics 4*, 197.

Davis, D. R. and D. E. Weinstein (2003). Market access, economic geography and comparative advantage: an empirical test. *Journal of International Economics 59*(1), 1–23.

Eaton, J. and S. Kortum (2001). Technology, trade, and growth: A unified framework. *European economic review 45*(4), 742–755.

Eaton, J. and S. Kortum (2002). Technology, geography, and trade. *Econometrica 70*(5), 1741–1779.

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

Feenstra, R. (1982). Appendix. product creation and trade patterns: a theoretical note on the biological model of trade in similar products. *Import competition and response. Chicago: University of Chicago*.

Feenstra, R. C., P. A. Luck, M. Obstfeld, and K. N. Russ (2017). In search of the armington elasticity. *The Review of Economics and Statistics* (Accepted).

Fernandes, A. M., P. J. Klenow, S. Meleshchuk, and M. D. Pierola (2015). The intensive margin in trade: Moving beyond pareto.

Graham, F. D. (1923). Some aspects of protection further considered. *The Quarterly Journal of Economics 37*(2), 199–227.

Hall, R. E. (1989). Invariance properties of solow's productivity residual. Technical report, National Bureau of Economic Research.

Hall, R. E. and C. I. Jones (1999). Why do some countries produce so much more output per worker than others? *The quarterly journal of economics 114*(1), 83–116.

Head, K. and J. Ries (2001). Increasing returns versus national product differentiation as an explanation for the pattern of us-canada trade. *American Economic Review*, 858–876.

Henderson, J. V. (2003). Marshall's scale economies. *Journal of urban economics 53*(1), 1–28.

Hummels, D. and P. J. Klenow (2005). The variety and quality of a nation's exports. *The American Economic Review 95*(3).

Jones, C. I. (1995). R & d-based models of economic growth. *Journal of political Economy 103*(4), 759–784.

Jones, C. I. (2011). Intermediate goods and weak links in the theory of economic development. *American Economic Journal: Macroeconomics 3*(2), 1–28.

Khandelwal, A. (2010). The long and short (of) quality ladders. *The Review of Economic Studies 77*(4), 1450–1476.

Klenow, P. J. and A. Rodriguez-Clare (2005). Externalities and growth. *Handbook of economic growth 1*, 817–861.

Kortum, S. S. (1997). Research, patenting, and technological change. *Econometrica: Journal of the Econometric Society*, 1389–1419.

Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. *The American Economic Review 70*(5), 950–959.

Kucheryavyy, K., G. Lyn, and A. Rodríguez-Clare (2016). Grounded by gravity: A well-behaved trade model with industry-level economies of scale. Technical report, National Bureau of Economic Research.

Manski, C. F., D. McFadden, et al. (1981). *Structural analysis of discrete data with econometric applications*. Mit Press Cambridge, MA.

Markusen, J. R. and R. M. Wigle (1989). Nash equilibrium tariffs for the united states and canada: The roles of country size, scale economies, and capital mobility. *Journal of political Economy 97*(2), 368–386.

Mayer, T. and S. Zignago (2011). Notes on cepiis distances measures: The geodist database.

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

Mill, J. S. (1848). *Principles of political economy with some of their applications to social philosophy, by John Stuart Mill*. JW Parker.

Morrison, C. J. (1992). Unraveling the productivity growth slowdown in the united states, canada and japan: The effects of subequilibrium, scale economies and markups. *The Review of Economics and Statistics*, 381–393.

Morrison, C. J. (2012). *A microeconomic approach to the measurement of economic performance: Productivity growth, capacity utilization, and related performance indicators*. Springer Science & Business Media.

Parente, S. L., E. C. Prescott, et al. (2002). Barriers to riches. *MIT Press Books 1*.

Pierce, J. R. and P. K. Schott (2012). Concording u.s. harmonized system codes over time. *Journal of Official Statistics 28*(1), 53–68.

Ramondo, N., A. Rodríguez-Clare, and M. Saborío-Rodríguez (2016, October). Trade, domestic frictions, and scale effects. *American Economic Review 106*(10), 3159–84.

Redding, S. J. and D. E. Weinstein (2016). A unified approach to estimating demand and welfare. *Working Paper*.

Restuccia, D. and R. Rogerson (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic dynamics 11*(4), 707–720.

Ricardo, D. (1891). *Principles of political economy and taxation*. G. Bell.

Rodrik, D. (1988). Imperfect competition, scale economies, and trade policy in developing countries. In *Trade policy issues and empirical analysis*, pp. 109–144. University of Chicago Press.

Rose, A. K. (2006). Size really doesn't matter: In search of a national scale effect. *Journal of the Japanese and international Economies 20*(4), 482–507.

Rotemberg, J. J. and M. Woodford (1999). The cyclical behavior of prices and costs. *Handbook of macroeconomics 1*, 1051–1135.

Silva, J. S. and S. Tenreyro (2006). The log of gravity. *The Review of Economics and statistics 88*(4), 641–658.

Simonovska, I. and M. E. Waugh (2014). The elasticity of trade: Estimates and evidence. *Journal of international Economics 92*(1), 34–50.

Smith, A. (1776). An inquiry into the nature and causes of the wealth of nations.

Soderbery, A. (2015). Estimating import supply and demand elasticities: Analysis and implications. *Journal of International Economics 96*(1), 1–17.

Somale, M. A. (2017). Comparative advantage in innovation and production.

Tybout, J. R. (2001). Plant-and firm-level evidence on" new" trade theories. Technical report, National Bureau of Economic Research.

World-Bank (2012). *World Development Indicators 2012*. World Bank-free PDF.

A The Industry-Level Entry Elasticity

Section 2 presented evidence suggesting that the entry elasticity, $\kappa \equiv \frac{\partial N}{\partial L}$, is approximately "one." Also, note that, in the presence of *free entry* in variety-based models or *endogenous knowledge accumulation* in idea-based model, $\kappa = 1$ is the only theory-consistent value. Below, we examine (i) how κ varies across sectors, and (ii) if $\kappa = 1$ is consistent with alternative datasets. To this end, as our first analysis, we use sector-level employment data from the 2012 World Input Output Database (WIOD) to evaluate Fact 1.A (from Section 2) on an sector-level basis. Given data on the sector-level size of the labor force, $L_{i,s}$, we run the following regression separately for each *manufacturing* sector s:

$$\ln N_{i,st} = \kappa_s \cdot \ln L_{i,st} + \beta_s \cdot \text{Controls}_{it} + \delta_t + \varepsilon_{i,st}.$$

The estimation results reported in Table 7 are quite encouraging for the existing literature. They suggest that κ_s is rather uniform across all manufacturing sector, always assuming a value close to 1. Perhaps this outcome should not be surprising, given the strong *aggregate* explanatory power of the CES framework.

In our second analysis we ask whether the unit entry elasticity, $\kappa = 1$, is a regularity that spans beyond the Colombia sample. To seek an answer, we adopt the widely-used EXPORTER DYNAMICS DATABASE (EDD) described in Fernandes et al. (2015). The publicly-available version of the EDD covers the universe of export transactions provided by customs agencies from 60 countries for the 1997–2013 period. For each country, the data reports the number of firms per HS6 product category, allowing us to calculate the total number of export varieties (i.e., HS6-firm-year combinations) for each country in years 1997 to 2013. For 481 country-year pairs, we have matching data for population size, GDP per capita, and trade openness from the Penn World Table version 9. We run the following regression on this data, with GDP per capita and trade openness as additional controls:

$$\ln N_{i,t} = \kappa_s \cdot \ln L_{i,t} + \beta_s \cdot \text{Controls}_{it} + \delta_t + \varepsilon_{i,t}.$$

The estimation results reported in Table 8 are consistent with an entry elasticity of 1, and comply with our benchmark analysis conducted on the Colombia sample.

Table 7: *Industry-level entry effects in manufacturing.*

		Dependent Variable			
Sector	ISIC4 codes	L (industry-Level)	GDPcap	Dist	Obs.
Textiles, Leather & Footwear	1700-1999	1.097	1.337	0.514	146
		(0.035)	(0.061)	(0.079)	
Wood	2000-2099	1.011	0.544	-0.0908	88
		(0.048)	(0.090)	(0.116)	
Paper	2100-2099	0.909	0.615	-0.286	146
		(0.054)	(0.084)	(0.136)	
Petroleum	2300-2399	1.015	1.304	-0.844	85
		(0.050)	(0.039)	(0.116)	
Chemicals	2400-2499	0.766	0.662	-0.426	157
		(0.025)	(0.033)	(0.054)	
Rubber & Plastic	2500-2599	1.097	0.821	-0.848	162
		(0.029)	(0.041)	(0.057)	
Minerals	2600-2699	1.095	0.865	0.0171	147
		(0.033)	(0.042)	(0.056)	
Basic & Fabricated Metals	2700-2899	1.245	0.675	-0.736	156
		(0.032)	(0.032)	(0.050)	
Machinery	2900-3099	0.994	0.675	-0.896	161
		(0.017)	(0.026)	(0.040)	
Electrical & Optical Equipment	3100-3399	1.007	0.508	-0.577	179
		(0.020)	(0.026)	(0.038)	
Transport Equipment	3400-3599	0.947	0.566	0.0025	145
		(0.052)	(0.051)	(0.072)	
N.E.C. & Recycling	3600-3799	1.127	1.129	-0.197	170
		(0.042)	(0.058)	(0.068)	

Notes: the dependent variable is (log) number of export varieties (country-HS10 product-year combinations).

Table 8: *Entry effects in the* EXPORTER DYNAMICS DATABASE.

	Inependent Variable (log)			
Dependent	Population size	GDPcap	Trade Openness	Obs.
No. of export variaties (log)	1.0279	.00013	1.9404	403
	(.0579)	(.00000)	(.4126)	

Notes: the estimation includes year fixed effects, and each data point in the sample is weighted by its population size.

B Different Models Same Import Demand Function

In this appendix we show that the import demand function characterized by assumption A1, can be derived from two different class of model with different underlying micro-foundations.

B.1 Variety-based model

The Nested-CES import demand function may be derived trivially from a nested CES utility function, which nests the utility functions in Krugman (1980) and Armington as a special case. Specifically, suppose that preferences within product category h are described by a nested CES utility function that aggregates across various firm-level varieties in that product category. In particular,

$$W_{i,h} = \max_{q_{\omega i,h}} \left[\sum_{j=1}^{N} \left(\sum_{\omega \in \Omega_{ji,h}} \varphi_{\omega i,h}^{\frac{1}{1+\vartheta_h}} q_{\omega i,h}^{\frac{\vartheta_h}{1+\vartheta_h}} \right)^{\frac{1+\vartheta_h}{\vartheta_h} \frac{\theta_h}{1+\theta_h}} \right]^{\frac{1+\theta_h}{\theta_h}}$$

$$\text{s.t.} \quad \sum_{j=1}^{N} \sum_{\omega \in \Omega_{ji,h}} p_{\omega i,h} q_{\omega i,h} = E_{i,h} \ ,$$

where $p_{\omega i,h}$, $q_{\omega i,h}$, and $\phi_{\omega,h}$ respectively denote the price, quantity, and quality associated with variety $\omega i, h$ supplied by firm ω to market i in product category h; $\Omega_{ji,h}$ denotes the set of all varieties exported from country j to market i in product category h; and finally, $E_{i,h}$ denotes total spending on product category h. Also, $\theta_h + 1$ and $\vartheta_h + 1$ denote to the *cross-national* and *sub-national* elasticities of substitution. Utility maximization implies that the share of income spent on variety ω, ih from country j (i.e., $\omega \in \Omega_{ji,h}$) is given by:

$$\lambda_{\omega i,h} = \lambda_{\omega|ji,h} \times \lambda_{ji,h}, \tag{13}$$

where $\lambda_{\omega|ji,h}$ denotes the within-national market share of variety $\omega i, h$ (that is, the share of income spent on variety ω, ih conditional on buying from country j), and is $\lambda_{ji,h}$ is the overall share of country i's spending on all product h varieties originating from country j. These two components of the demand share are described by:

$$\lambda_{\omega|ji,h} = \frac{\varphi_{\omega i,h} p_{\omega i,h}^{-\vartheta_h}}{\sum_{\omega' \in \Omega_{ji,h}} \varphi_{\omega' i,h} p_{\omega' i,h}^{-\vartheta_h}} = \varphi_{\omega i,h} \left(\frac{p_{\omega i,h}}{P_{ji,h}} \right)^{-\vartheta_h}, \tag{14}$$

and,

$$\lambda_{ji,h} = \frac{\left(\sum_{\omega \in \Omega_{ji,h}} \varphi_{\omega i,h} p_{\omega' i,h}^{-\vartheta_h}\right)^{\frac{\theta_h}{\vartheta_h}}}{\sum_{k=1}^{N}\left(\sum_{\omega' \in \Omega_{ki,h}} \varphi_{\omega' i,h} p_{\omega' i,h}^{-\vartheta_h}\right)^{\frac{\theta_h}{\vartheta_h}}} = \left(\frac{P_{ji,h}}{P_{i,h}}\right)^{-\theta_h}, \tag{15}$$

where $P_{ji,h} \equiv \left[\sum_{\omega' \in \Omega_{ji,h}} \left(p_{\omega' i,h}/\phi_{\omega' i,h}\right)^{-1/\psi_h}\right]^{-\psi_h}$ denotes the price index of country j's exports to market i in product category h, and $P_{i,h} \equiv \left[\sum_{k=1}^{N}\left(P_{ki,h}\right)^{-\theta_h}\right]^{-\frac{1}{\theta_h}}$ is the overall price index of product h in country i, which aggregates over all national price indexes in product category h. Denoting $Q_{i,h} = E_{i,h}/P_{i,h}$, demand for variety $\omega i,h$ from country of origin j in market i is given by:

$$q_{\omega i,h} = \varphi_{\omega i,h}\left(\frac{p_{\omega i,h}}{P_{ji,h}}\right)^{-\vartheta-1}\left(\frac{P_{ji,h}}{P_{i,h}}\right)^{-\theta_h-1} Q_{i,h}.$$

B.2 Idea-based model: *Nested EK*

The nested CES import demand function specified by A1 can also arise from within-product specialization in an idea-based model a la Eaton and Kortum (2002). The idea-based framework departs from the variety-based framework in assuming that each product category is comprised of a continuum of homogenous goods. Suppliers have differentiated productivities across the goods in the continuum. To be more specific, let ν index a homogeneous good pertaining to product category h. The sub-utility of the representative consumer in country i with respect to product category h is a log-linear aggregator across the continuum of goods in that category:

$$Q_{i,h} = \int_0^1 \ln \tilde{q}_{i,h}(\nu) d\nu$$

Country j hosts various firms, each indexed by ω with $\Omega_{ji,h}$ denotes the set of all firms in country j actively serving market i. Firm ω supplies good ν to market i at the following *quality-adjusted* price:

$$\tilde{p}_{\omega i,h}(\nu) = \frac{p_{\omega i,h}}{z_\omega(\nu)}$$

Whereas the nominal price, $p_{\omega i,h}$, applies to all goods supplied by firm ω in product category h, the quality component, $z_\omega(\nu)$, varies systematically. For any given good ν, firm-specific qualities are drawn independently from the following nested Fréchet joint distribution:

$$F_h(\boldsymbol{z}(\boldsymbol{\nu})) = \exp\left[-\sum_{k=1}^{N}\left(\sum_{\omega\in\Omega_{ki,h}}\varphi_{\omega,h}z_{\omega}(\nu)^{-\vartheta_h}\right)^{\frac{\theta_h}{\vartheta_h}}\right],$$

The above distribution generalizes the basic Fréchet distribution in EK. In particular, it relaxes the restriction that productivities are perfectly correlated across firms within the same country. Instead, the above distribution allows for productivity differentiation within countries and for the degrees of within and cross-national productivity differentiations (ϑ_h and θ_h respectively) to diverge—a special case of the distribution where $\vartheta_h \longrightarrow \infty$ corresponds to the EK specification. Note that the above distribution has deep theoretical roots. The Fisher-–Tippett–Gnedenko theorem states that if ideas are drawn from a (normalized) distribution, in the limit the distribution of the best draw takes the form of a general extreme value (GEV) distribution, which includes the above Fréchet distribution as a special case. A special application of this result can be found in Kortum (1997) who develops an idea-based growth model where the limit distribution of productivities is Fréchet, with $\varphi_{\omega,h}$ reflecting the stock of technological knowledge accumulated by firms ω in category h.

Given the vector of effective prices, the representative consumer in county i (who is endowed with income Y_i) maximizes their real consumption, $\tilde{q}_{i,h}(\nu) = \frac{\beta_{i,h}Y_i}{\tilde{p}_{i,h}(\nu)}$, by sourcing good ν from the cheapest provider. This amounts to solving the following discrete choice problem for each good ν:

$$\min_{\omega}\frac{p_{\omega i,h}}{z_{\omega}(\nu)} \sim \max_{\omega} - \ln\frac{p_{\omega i,h}}{z_{\omega}(\nu)} = -\ln p_{\omega i,h} + \ln z_{\omega}(\nu).$$

The determine the share of goods for which firm ω is the most competitive supplier, we can invoke the theorem of "general extreme value." Specifically, define $G(c)$ as follows

$$G_h(\mathbf{p}) = \sum_{k=1}^{N}\left(\sum_{\omega\in\Omega_{ki,h}}\varphi_{\omega,h}\exp(-\vartheta_h \ln p_{\omega i,h})\right)^{\frac{\theta_h}{\vartheta_h}} = \sum_{k=1}^{N}\left(\sum_{\omega\in\Omega_{ki,h}}\varphi_{\omega,h}p_{\omega i,h}^{-\vartheta_h}\right)^{\frac{\theta_h}{\vartheta_h}},$$

Note that $G_h(.)$ is a continuous and differentiable function of vector $\mathbf{c}$(where) and has the following properties:

i. $G_h(.) \geq 0$;

ii. $G_h(.)$ is a homogeneous function of rank θ_h: $G_h(\rho\mathbf{p}) = \rho^{\theta_h} G_h(\mathbf{p})$;

iii. $\lim_{c_{\omega i,h}\to\infty} G_h(\mathbf{p}) = \infty$, $\forall\omega$;

iv. the k'th partial derivative of $G_h(.)$ with respect to a generic combination of k variables $c_{\omega i,h}$, is non-negative if k is odd and non-positive if k is even.

Manski et al. (1981) show that if $G_h(.)$ satisfies the above conditions, and $z_\omega(\nu)$'s are drawn from

$$F_h(\boldsymbol{z}) = \exp\left(-G_h(e^{-\ln z})\right) = \exp\left(-\sum_{k=1}^{N}\left(\sum_{\omega\in\Omega_{ki,h}} \varphi_{\omega,h} z_\omega^{-\vartheta_h}\right)^{\frac{\theta_h}{\vartheta_h}}\right),$$

which is the exact same distribution specified above. In that case, the probability of choosing variety ω from country of origin j is

$$\lambda_{\omega i,h} = \frac{\left(\frac{p_{\omega i,h}}{\theta_h}\right)\frac{\partial G_h}{\partial p_{\omega i,h}}}{G_h(\mathbf{c})} = \frac{\varphi_{\omega,h} p_{\omega i,h} p_{\omega i,h}^{\vartheta_h - 1}\left(\sum_{\omega'\in\Omega_{ji,h}} \varphi_{\omega',h} p_{\omega' i,h}^{\vartheta_h}\right)^{\frac{\theta_h}{\vartheta_h}-1}}{\sum_{k=1}^{N}\left(\sum_{\omega'\in\Omega_{ki,h}} \varphi_{\omega',h} p_{\omega' i,h}^{\vartheta_h}\right)^{\frac{\theta_h}{\vartheta_h}}}$$
$$= \frac{\varphi_{\omega,h} p_{\omega i,h}^{\vartheta_h}}{\sum_{\omega'\in\Omega_{ji,h}} \varphi_{\omega',h} p_{\omega' i,h}^{\vartheta_h}} \cdot \frac{\left(\sum_{\omega'\in\Omega_{ji,h}} \varphi_{\omega',h} p_{\omega' i,h}^{\vartheta_h}\right)^{\frac{\theta_h}{\vartheta_h}}}{\sum_{k=1}^{N}\left(\sum_{\omega'\in\Omega_{ki,h}} \varphi_{\omega',h} p_{\omega' i,h}^{\vartheta_h}\right)^{\frac{\theta_h}{\vartheta_h}}}.$$

Rearranging the above equation yields the following expression:

$$\lambda_{\omega i,h} = \varphi_{\omega,h}\left(\frac{p_{\omega i,h}}{P_{ji,h}}\right)^{-\vartheta_h}\left(\frac{P_{ji,h}}{P_{i,h}}\right)^{-\theta_h},$$

where $P_{ji,h} \equiv \left[\sum_{\omega'\in\Omega_{ji,h}} \varphi_{\omega,h} p_{\omega' i,h}^{-\vartheta_h}\right]^{-1/\vartheta_h}$ and $P_{i,h} \equiv \left[\sum P_{ji,h}^{-\theta_h}\right]^{-\frac{1}{\theta_h}}$. Given the share of goods sourced from firm ω, total sales of firm $\omega \in \Omega_{ji,h}$ to market i, in product category h can be calculated as:

$$x_{\omega i,h} = p_{\omega i,h} q_{\omega i,h} = p_{\omega i,h}\lambda_{\omega i,h}\frac{X_{i,h}}{p_{\omega i,h}} = \lambda_{\omega i,h} E_{i,h}$$
$$= \varphi_{\omega,i}\left(\frac{p_{\omega i,h}}{P_{ji,h}}\right)^{-\vartheta_h}\left(\frac{P_{ji,h}}{P_{i,h}}\right)^{-\theta_h} E_{i,h}.$$

which is identical to the nested-CES function specified by A1—namely, $q_{\omega i,h} =$

$\varphi_{\omega,h}\left(\frac{p_{\omega i,h}}{P_{ji,h}}\right)^{-\vartheta-1}\left(\frac{P_{ji,h}}{P_{i,h}}\right)^{-\theta_h-1}Q_{i,h}.$

B.3 Deriving the Log-Linear Import Demand Function.

Below, we derive the estimating Equation 9 from the following demand function specified by A1:

$$x_{\omega i,h}=\varphi_{\omega i,h}\left(\frac{p_{\omega i,h}}{P_{ji,h}}\right)^{-\frac{\theta_h}{\alpha_h}}\left(\frac{P_{ji,h}}{P_{i,h}}\right)^{-\theta_h}E_{i,h},$$

where α_h is defined such that $\vartheta_h=\frac{\theta_h}{\alpha_h}$. We can rearrange the above equation as

$$\begin{aligned} x_{\omega i,h} &= \varphi_{\omega i,h}\, p_{\omega i,h}^{-\theta_h}\, p_{\omega i,h}^{\theta_h-\frac{\theta_h}{\alpha_h}}\, P_{ji,h}^{\frac{\theta_h}{\alpha_h}-\theta_h}\, P_{i,h}^{\theta_h}\, E_{i,h} \\ &= \varphi_{\omega i,h}^{\alpha_h}\, p_{\omega i,h}^{-\theta_h}\left\{\varphi_{\omega i,h}\left(\frac{p_{\omega i,h}}{P_{ji,h}}\right)^{-\frac{\theta_h}{\alpha_h}}\right\}^{1-\alpha_h} P_{i,h}^{\theta_h}E_{i,h}. \end{aligned}$$

Noting that $\lambda_{\omega|ji,h}=\varphi_{\omega i,h}\left(\frac{p_{\omega i,h}}{P_{ji,h}}\right)^{-\frac{\theta_h}{\alpha_h}}$ and taking logs, the above equation will deliver our estimating equation:

$$\ln x_{\omega i,h}=-\theta_h\ln p_{\omega i,h}+(1-\alpha_h)\ln\lambda_{\omega|ji,h}+\underbrace{\theta_h\ln P_{i,h}E_{i,h}}_{\delta_{i,h}}+\tilde{\varphi}_{\omega i,h},$$

where $\tilde{\varphi}_{\omega i,h}\equiv\alpha_h\varphi_{\omega i,h}$.

C Scale Elasticity in the Melitz-Pareto Model

In the presence of firm selection effects, the elasticity of welfare with respect to industry-level employment will depend on the scale elasticity and the firm productivity distribution. In the knife-edge case where the Melitz (2003) features a Pareto productivity distribution, we can still use our estimates to evaluate the macro-level effects of policy shocks. To demonstrate this, note that in the Melitz-Pareto model changes in welfare in response to a policy shock still follow the following characterization (see Kucheryavyy et al. (2016)):

$$d \ln W_i = \sum_{h \in H} \beta_{i,h} \left(\psi_h d \ln L_{i,h} - \frac{1}{\varepsilon_h} d \ln \lambda_{ii,h} \right),$$

where the trade elasticity, ε_h, can be estimated as the elasticity of export sales with respect to tariffs:[23]

$$\varepsilon_h = \frac{\partial \ln X_{ji,h}}{\partial \ln t_{ji,}}$$

Based on the above, one can identify ε_h using the triple-difference methodology in Caliendo and Parro (2014). However, even after knowing ε_h, the scale elasticity will still depend critically on our estimates of θ_h and α_h. In particular,[24]

$$\psi_h = \frac{1}{\varepsilon_h} - \frac{1}{\theta_h} (1 - \alpha_h) .$$

Contrast the above formula with the common practice of arbitrarily setting the scale elasticity to $\frac{1}{\varepsilon_h}$, by assuming $\alpha = 1$. According to our estimation results, this standard assumption is counterfactual, overstates the scale elasticity, ψ, and ignores the sectoral variation in ψ.

D Technical Appendix: Cleaning the data on the identity/name of exporting firms

Utilizing the information on the identity of the foreign exporting firm is a critical part of our empirical exercise. Unfortunately, the names of the exporting firms in our dataset are not standardized. As a result, there are instances when the same firm is recorded differently due to using or not using the abbreviations, capital and lower-case letters, spaces, dots, other special characters, etc. To standardize the names of the exporting firms, we used the following procedure.[25]

1. We deleted all observations with the missing exporting names and/or zero trade values.

2. We capitalized firms names and their contact information (which is either

[23] The trade elasticity fits the above definition as long as tariffs are applied to cost rather than revenue.

[24] This expression follows directly from Table 1 in Kucheryavyy et al. (2016).

[25] The corresponding Stata code is in the cleanFirmsNames.do.

email or phone number of the firm).

3. We eliminated abbreviation “LLC,” spaces, paraenthesis, and other special characters (. , ; / @ ‘ } - & “) from the firms names.

4. We eliminated all characters specified in 3. above and a few others (# : FAX) from the contact information.

5. We dropped observations without contact information (such as, "NOTIENE", "NOREPORTA", "NOREGISTRA," etc.), with non-existent phone numbers (e.g., “0000000000”, “1234567890”, “1”), and with six phone numbers which are used for multiple firms with different names (3218151311, 3218151297, 6676266, 44443866, 3058712865, 3055935515).

6. Next, we kept only up to first 12 characters in the firm’s name and up to first 12 characters in the firm’s contact information (which is either email or phone number). In our empirics, we treat all transaction with the same updated name and contact information as coming from the same firm.

7. We also analyzed all observations with the same contact information, but slightly different name spelling. We only focused on the cases in which there are up to three different variants of the firm name. For these cases, we calculated the Levenshtein distance in the names, which is the smallest number of edits required to match one name to another. We treat all export observations as coming from the same firm if the contact phone number (or email) is the same and the Levenshtein distance is four or less.

E Gains from Trade: *Detailed Estimates*

Table 9: *The Gains From Trade.*

	Gains from Trade			Specialization	
	Benchmark	EK-Armington	Krugman	Degree	Scale-Intesity
Country	(1)	(2)	(3)	(4)	(5)
AUS	10.0%	12.4%	5.8%	4.58	-18.75
AUT	27.8%	26.9%	28.6%	1.97	15.32
BEL	33.5%	35.0%	32.2%	9.31	-4.15
BRA	6.0%	6.1%	4.4%	0.49	-0.40
CAN	20.5%	21.7%	17.8%	3.16	-9.76
CHN	13.9%	13.9%	16.0%	2.04	3.53
CZE	34.4%	33.8%	35.7%	1.84	12.83
DEU	21.6%	19.8%	23.5%	4.20	30.25
DNK	25.6%	26.1%	26.1%	11.56	15.13
ESP	11.5%	12.8%	9.6%	2.90	-9.37
FIN	22.1%	17.3%	27.1%	4.64	67.85
FRA	13.7%	14.1%	13.4%	2.52	-0.24
GBR	14.9%	15.8%	14.0%	1.01	-7.81
GRC	11.5%	17.0%	6.6%	14.03	-38.13
HUN	50.1%	51.6%	50.5%	3.43	-22.60
IDN	14.3%	15.0%	12.1%	1.63	-5.53
IND	6.2%	7.1%	5.3%	1.69	-6.91
IRL	24.0%	21.4%	27.3%	8.83	50.24
ITA	10.8%	10.1%	11.8%	3.34	14.36
JPN	6.2%	6.2%	7.6%	5.20	9.56
KOR	16.7%	18.0%	19.2%	16.97	15.59
MEX	21.9%	24.5%	20.3%	2.24	-28.65
NLD	23.5%	22.6%	24.0%	2.54	16.16
POL	22.3%	23.3%	20.9%	1.12	-11.17
PRT	18.5%	20.2%	17.1%	4.04	-13.42
ROM	23.1%	25.9%	20.2%	2.62	-32.41
RUS	6.3%	10.3%	3.2%	9.80	-25.27
SVK	43.2%	43.5%	43.0%	4.87	2.53
SVN	38.8%	41.0%	38.9%	14.35	-10.15
SWE	24.2%	21.6%	26.3%	2.43	38.33
TUR	11.9%	13.0%	10.4%	2.28	-8.24
TWN	40.8%	41.5%	44.4%	14.40	15.71
USA	7.0%	7.8%	6.1%	1.00	-6.40
RoW	20.8%	22.8%	17.3%	6.42	-14.01