

## Grounded by Gravity: A Well-Behaved Trade Model with Industry-Level Economies of Scale<sup>†</sup>

By KONSTANTIN KUCHERYAVYY, GARY LYN, AND ANDRÉS RODRÍGUEZ-CLARE\*

*We propose a model to study the role of industry-level external economies of scale in open economies. If the elasticity governing the strength of external economies is below the inverse of the trade elasticity in each industry, then specialization under frictionless trade is consistent with comparative advantage, the model is tractable even with trade frictions, and all countries gain from trade. External economies lower gains from trade except if the country specializes in industries with high scale economies, and they amplify the gains from further trade liberalization except if it leads to specialization in industries with low scale economies. (JEL D24, F11, F12, F13)*

For over a century trade economists have struggled to understand the positive and normative implications of localized industry-level external economies of scale—see Marshall (1890), Graham (1923), Chipman (1965), and Ethier (1982). Standard models yielded some discomfiting results, including “a bewildering variety of equilibria” (Krugman 1995)—even ones in which trade patterns do not conform to comparative advantage and lead to losses relative to autarky—that render them useless for quantitative analysis. The root of the problem is a circularity whereby the scale of an industry affects its productivity, while an industry’s productivity affects its scale through the impact on the pattern of trade and specialization.

Grossman and Rossi-Hansberg (2010) proposed a two-country Ricardian model with national industry-level external economies of scale that avoids this circularity by moving away from perfect competition and assuming Bertrand competition instead. Since firms in each industry understand the implications of their decisions

\*Kucheryavyy: University of Tokyo and HSE University (email: k.s.kucheryavyy@gmail.com); Lyn: Iowa State University (email: glyn@iastate.edu); Rodríguez-Clare: University of California at Berkeley (email: andres@econ.berkeley.edu). Simon Gilchrist was coeditor for this article. We thank James Anderson, Costas Arkolakis, Gaurab Aryal, Backhoon Song, Lorenzo Caliendo, Arnaud Costinot, Richard Cottle, Svetlana Demidova, Dave Donaldson, Jonathan Eaton, Michal Fabinger, Cecile Gaubert, Gene Grossman, Nail Kashaev, Tim Kehoe, Hideo Konishi, Sam Kortum, Beresford Parlett, Donald Richards, Steve Redding, Michael Tsatsomeris, Guang Yang, Xi Yang, and Stephen Yeaple for valuable discussions, and Kala Krishna for pointing us back to perfect competition. We thank Piyush Panigrahi, Mauricio Ulate, and Yipei Zhang for valuable research assistance. Konstantin Kucheryavyy acknowledges the financial support of JSPS Kakenhi Grant Numbers 17K13721, 20K13476, and 20H01487 as well as the financial support of the HSE University Basic Research Program. All errors are our own.

<sup>†</sup>Go to <https://doi.org/10.1257/mac.20190156> to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

on industry output and productivity, the equilibrium is unique and similar to that in Dornbusch, Fischer, and Samuelson (1977) when trade is frictionless. Unfortunately, as shown by Lyn and Rodríguez-Clare (2013a, b), the framework quickly becomes intractable under trade costs.

In this paper we present a new approach to study the consequences of external economies of scale in a trade context. We retain the assumption of perfect competition, as in the standard Ricardian model, but relax the implicit assumption in that model that the whole industry produces a single homogeneous good. Instead, we allow for product differentiation across countries within each industry, as in a multi-industry version of Armington (1969), and show that the model is well behaved if the strength of external economies is not too high. In particular, the equilibrium analysis remains tractable even under positive trade costs and satisfies the gravity equation, making the model useful to examine the quantitative importance of external economies of scale.<sup>1</sup>

Our multi-industry model with external economies of scale turns out to be isomorphic to a generalized version of the multi-industry Krugman (1980) model of product differentiation with internal economies of scale as well as a generalized version of the multi-industry Melitz (2003) model.<sup>2</sup> The common mathematical structure that characterizes the equilibrium in all models is governed by two industry-level elasticities: the elasticity of bilateral trade flows with respect to bilateral trade costs, commonly referred to as the *trade elasticity*; and the elasticity of productivity with respect to industry size, which we will refer to as the *scale elasticity*.<sup>3</sup> We show that the equilibrium properties of the model depend critically on whether the product of these two elasticities is higher or lower than one.<sup>4</sup> If this product is higher than one in some particular industry, then any pattern of cross-country specialization in which a nonempty set of countries are the sole suppliers in that industry is consistent with equilibrium. Otherwise, the pattern of specialization is always consistent with comparative advantage under frictionless trade, and the model remains tractable in the presence of trade costs.<sup>5</sup>

In the second half of the paper, we use our unified framework to study the implications of scale economies for the welfare effects of trade. We first establish that all countries gain from trade relative to autarky as long as the product of the trade and scale elasticities is weakly lower than one in all industries. This is so even if the scale elasticity differs across industries, an important finding in light of previous results with this type of model where countries could lose from opening up to trade.<sup>6</sup>

<sup>1</sup>Our analysis restricts to the case of external economies of scale operating inside each industry. An alternative case is the one in which some of the externalities operate across industries, as studied in Yatsynovich (2017)—for example, for the case with frictionless trade.

<sup>2</sup>Abdel-Rahman and Fujita (1990); Allen, Arkolakis, and Takahashi (2020); and Redding (2016) explore similar isomorphisms for spatial equilibrium models in the economic geography literature.

<sup>3</sup>The trade elasticity is defined as the elasticity of bilateral trade flows with respect to bilateral trade costs, holding wages and productivity constant—see Arkolakis, Costinot, and Rodríguez-Clare (2012).

<sup>4</sup>A similar threshold plays an important role in quantitative economic geography settings such as those described in Allen and Arkolakis (2014). See Kucheryavyy, Lyn, and Rodríguez-Clare (2021) for a detailed discussion.

<sup>5</sup>In Section III we formalize what we mean by saying that the pattern of specialization is “consistent with comparative advantage” and discuss results for the more general case with trade costs.

<sup>6</sup>See, for example, Ethier (1982).

We extend the “sufficient statistics approach” to the quantification of the gains from trade (i.e., the negative of the welfare change caused by a move to autarky) in Arkolakis, Costinot, and Rodríguez-Clare (2012, henceforth ACR), to multi-industry models with scale economies. The isomorphism that ACR establish across single-industry models now applies to multi-industry models as well: for the same industry-level trade and scale elasticities, the models deliver the same gains from trade and the same counterfactual implications given trade and production data. More generally, given trade and scale elasticities, we derive a general formula for the gains from trade in terms of the share of industry-level expenditure allocated to domestic goods, as well as the share of a country’s expenditure and revenue allocated to each industry. We use this formula to explore the way in which scale economies affect the gains from trade.

For the simple case in which the scale elasticity is the same across industries, we find that the gains from trade are lower with scale economies than without. This result may seem counterintuitive, but the reader should keep in mind that the gains from trade are defined as conditional on trade shares, so that—as in ACR and Costinot and Rodríguez-Clare (2014)—we can compare the gains from trade implied by different models that are consistent with the same data. Thus, the intuition that scale economies should lead to larger gains from trade through deeper industry-level specialization and larger trade flows is simply not operative here, although, as we explain below, this intuition is relevant for the gains from trade liberalization, defined as the welfare effects of a counterfactual decline in trade costs starting at the observed equilibrium.

So why do scale economies lead to lower gains from trade? The move back to autarky implies a reallocation of labor across industries that, in the presence of scale economies, leads to productivity gains in expanding industries and productivity losses in contracting industries. Since the industries that expand are those where the country has positive net imports, it must be that they have a high expenditure share or a low employment share. A higher expenditure share implies that a given productivity gain matters more for welfare, whereas a low employment share implies that a given absolute increase in employment leads to a higher proportional expansion and a higher productivity gain. As a result, a move back to autarky generates an expenditure-weighted average productivity improvement, implying lower welfare losses. A corollary of this reasoning is that the decline in the gains from trade from the presence of economies of scale is stronger for economies that exhibit a higher degree of industry-level specialization.

The previous results are specific to the case in which the scale elasticity is the same across industries. If the scale elasticity varies across industries, the implication of economies of scale depends not only on the degree of industry-level specialization, but also on its pattern. Everything else equal, countries that happen to specialize in industries with high scale elasticities will gain more from trade than countries that specialize in industries with low scale elasticities. Countries that specialize in industries with high scale elasticities may even gain more from trade in the presence of economies of scale compared to the standard model without them.

We next consider two simple cases for which we can analytically characterize the impact of scale economies on the gains from trade liberalization: first, a case of two

mirror-image countries, and second, a case with exogenous wages. In the first case we find that the gains from trade liberalization are higher with scale economies than without, a reflection of the magnified response of industry-level specialization and trade to the decline in trade costs in the presence of economies of scale. In the second case we find that countries lose from unilateral trade liberalization if the product of the trade and scale elasticities is above a threshold value that is a function of industry-level import and export shares, and we argue that this is a straightforward implication of the fact that trade liberalization leads to further specialization in an industry with relatively weak external economies of scale, a generalization of a key result in Venables (1987).

We complement our exploration of the effect of scale economies on the gains from trade and the gains from trade liberalization by applying our framework to data from the World Input Output Database (WIOD, Timmer et al. 2015) in 2008. A key ingredient for this quantitative analysis is a set of values for industry-level trade and scale elasticities. We take those from Bartelme et al. (2021), who set trade elasticities at the median value of recent estimates in the literature, and estimate scale elasticities using an approach motivated by the model we have developed in this paper.<sup>7</sup> The effects of scale economies on the gains from trade discussed above turn out to be quantitatively important. For example, they lead to an increase in the gains from trade for Germany (from 5.9 to 6.6 percent) and Korea (from 6.9 to 7.5 percent) but a decrease for Greece (from 5.9 to 4.3 percent) and, most prominently, for Russia (from 2.8 to 1.2 percent).

We also use the model to quantify the welfare implications of removing all tariffs observed in the data. We find that the presence of scale economies implies, on average, larger gains from the removal of all tariffs, with gains more than tripling for China and more than doubling for some countries. Still, gains from trade liberalization fall for countries that are pushed to specialize in industries with relatively low economies of scale. The most dramatic example of this force is for Russia and Mexico, which experience welfare losses that are twice and three times as large with economies of scale than without, respectively.

We end with a quantitative exploration of the effect of scale economies on industry specialization and trade flows. We find that the removal of economies of scale implies a decline in the degree of specialization and total trade, but the effects are small, implying that scale economies are much less important than Ricardian comparative advantage in driving industry-level specialization.

Our welfare analysis is related to that in Costinot and Rodríguez-Clare (2014), who compute both gains from trade and gains from trade liberalization for multi-industry economies under perfect and monopolistic competition. Compared to that paper, we further establish analytically that all countries gain from trade as long as the product of the trade and scale elasticities is weakly lower than one in all industries, we connect a country's decline in the gains from trade to its degree of industry specialization, we analyze how varying scale elasticities across industries interact with a country's

<sup>7</sup>Lashkaripour and Lugovskyy (2021) have also estimated sector-level trade and scale elasticities in an environment compatible with our model. Reassuringly, the estimates in both of these papers satisfy the condition that the product of the trade and scale elasticities be lower than one in all sectors.

interindustry trade pattern to affect its gains from trade, and we connect the results for the gains from trade liberalization to the insights in Venables (1987).

Our paper is also related to Somale (2021), who introduces sector-specific innovation into a multisector Eaton and Kortum (2002) model (via mechanisms from Eaton and Kortum 2001) to quantify its implications for welfare. Interestingly, although the model in Somale (2021) is dynamic, the balanced growth path is also characterized by the same system of equations as all the models that we consider in this paper and has related existence results. Somale (2021) also studies the quantitative importance of scale economies in determining industry-level specialization, but whereas he focuses on the variance of comparative advantage, we compare direct measures of trade and specialization between the data and those that would arise in a counterfactual world where everything is the same except that there are no economies of scale. Our work is also related to Caliendo et al. (2015), who compute the gains from trade liberalization in a multi-industry Melitz model extended to allow for intermediate goods as in Caliendo and Parro (2015). Our focus is on the theoretical properties of the model, including isomorphisms across different micro-foundations for economies of scale as well as understanding the properties of equilibria and the mechanisms through which economies of scale affect gains from trade and gains from trade liberalization.

Finally, our paper is related to Bartelme et al. (2021, henceforth BCDR), who quantify the gains from industrial policy in the multi-industry gravity model of trade with scale economies developed in this paper. A critical ingredient of BCDR's quantification is a set of industry-level scale elasticities, which they estimate by exploiting variation in industry size caused by an exogenous component of demand and using trade and production data for a cross-section of countries. We use these estimates in our quantification of the welfare effects from trade in Sections IVB and VB.

### I. An Armington Model with External Economies of Scale

In this section we present a multi-industry version of the Armington model of trade extended to allow for external economies of scale (EES) operating at the level of each industry. Formally, there are  $N$  countries indexed by  $n$ ,  $i$ , and  $l$  and  $K$  industries indexed by  $k$ . Each country produces a differentiated good in each industry. The only factor of production is labor, which is immobile across countries and perfectly mobile across industries within a country. We use  $\bar{L}_i$  and  $w_i$  to denote the inelastic labor supply and the wage level in country  $i$ , respectively.

Each country has a representative consumer with upper-tier Cobb-Douglas preferences with industry-level expenditure shares  $\beta_{i,k} \in (0, 1)$  satisfying  $\sum_{k=1}^K \beta_{i,k} = 1$  for all  $i$  and lower-tier CES preferences with elasticity of substitution  $\sigma_k > 1$  across goods from different countries within industry  $k$ . Thus, letting  $p_{ni,k}$  denote the price of good  $(i, k)$  in country  $n$ , the corresponding demand function is

$$(1) \quad q_{ni,k} = p_{ni,k}^{-\sigma_k} P_{n,k}^{\sigma_k - 1} \beta_{n,k} w_n \bar{L}_n,$$

where

$$(2) \quad P_{n,k} = \left( \sum_i P_{ni,k}^{1-\sigma_k} \right)^{\frac{1}{1-\sigma_k}}$$

is the price index for industry  $k$  in country  $n$ , and the aggregate price index is  $P_n = \tilde{\beta}_n \prod_{k=1}^K P_{n,k}^{\beta_{n,k}}$ , where  $\tilde{\beta}_n \equiv \prod_{k=1}^K \beta_{n,k}^{-\beta_{n,k}}$  is the standard Cobb-Douglas term.

The production technology exhibits constant or increasing returns to scale due to EES. In particular, labor productivity in good  $(i, k)$  is  $\tilde{A}_{i,k} \equiv A_{i,k} L_{i,k}^{\phi_k}$ , where  $A_{i,k}$  is an exogenous productivity parameter and  $L_{i,k}$  is total employment in industry  $k$  in country  $i$ . The parameter  $\phi_k$  governs the strength of EES in industry  $k$  (i.e., the elasticity of industry productivity with respect to industry size), which we refer to as the *scale elasticity*.

Trade costs are of the standard iceberg type so that delivering a unit of the industry- $k$  good from country  $i$  to country  $n$  requires shipping  $\tau_{ni,k} \geq 1$  units of the good, with  $\tau_{ii,k} = 1$  for all  $i$  and all  $k$  and  $\tau_{nl,k} \leq \tau_{ni,k} \tau_{il,k}$  for all  $n, l, i$ , and  $k$  (triangular inequality).

The above assumptions on technology and trade costs imply that the unit cost of good  $(i, k)$  for delivery to market  $n$  is

$$(3) \quad c_{ni,k} = \frac{\tau_{ni,k} W_i}{A_{i,k} L_{i,k}^{\phi_k}}$$

Since firms take  $L_{i,k}$  as given, this is also the perceived marginal cost faced by firms producing good  $(i, k)$  for market  $n$ .

Assuming that all markets are perfectly competitive, and letting  $y_{ni,k}$  denote the quantity of good  $(i, k)$  produced for delivery to market  $n$ , profit maximization then implies the following complementary slackness condition for all  $n, i$ , and  $k$ :

$$(4) \quad p_{ni,k} \leq c_{ni,k} \perp y_{ni,k} \geq 0.$$

Here the sign  $\perp$  between two weak inequalities means that they are satisfied with complementary slackness—that is,  $x \leq x_0 \perp y \geq y_0$  means that  $x \leq x_0, y \geq y_0$ , and if  $x < x_0$  then  $y = y_0$ . In addition, total employment  $L_{i,k}$  must be consistent with the amounts produced for each market,

$$(5) \quad L_{i,k} = \sum_n \frac{\tau_{ni,k}}{A_{i,k} L_{i,k}^{\phi_k}} y_{ni,k}.$$

The standard way to define an equilibrium is a set of good and factor prices,  $\{p_{ni,k}\}$  and  $\{w_i\}$ , and a set of consumption and production choices,  $\{q_{ni,k}\}$  and  $\{y_{ni,k}\}$ , such that consumers maximize utility and firms maximize profits, (1)–(4) for all  $(n, i, k)$ , and all markets clear, which here means  $q_{ni,k} = y_{ni,k}$  for all  $(n, i, k)$ , and

$$(6) \quad \sum_k L_{i,k} = \bar{L}_i$$

for all  $i$ , with sector-level employment levels  $\{L_{i,k}\}$  satisfying condition (5). A challenge with this definition is that we cannot evaluate condition (4) at a corner with  $L_{i,k} = 0$  for some  $(i, k)$ . The reason is that in such a case equation (3) would imply that  $c_{ni,k}$  is not well defined, while the inverse demand associated with equation (1) would also imply that  $p_{ni,k}$  is not well defined, since  $q_{ni,k} = y_{ni,k} = 0$  when  $L_{i,k} = 0$ .

To proceed, we focus instead on the labor market counterpart of the equilibrium condition (4). Letting  $VMPL_{ni,k} \equiv p_{ni,k}A_{i,k}L_{i,k}^{\phi_k}/\tau_{ni,k}$  be the value of the marginal product of labor in the production of good  $(i, k)$  for market  $n$ , we define the equilibrium in the same way as above but with conditions (3) and (4) replaced with

$$(7) \quad VMPL_{ni,k} \leq w_i \perp y_{ni,k} \geq 0.$$

The benefit of this condition is that, in contrast to (4), at a corner with  $L_{i,k} = 0$ , it can be evaluated by the limit as  $L_{i,k} \rightarrow 0$ , since now  $VMPL_{ni,k}$  fully captures the interaction between cost and prices as we approach a corner, as we discuss further below.

To derive a mathematically more convenient characterization of equilibrium, note first that the Inada property of CES demand in (1) implies that at an equilibrium either  $y_{ni,k} > 0$  for all  $n$  or  $y_{ni,k} = 0$  for all  $n$ ,<sup>8</sup> so we can write (7) more compactly as

$$(8) \quad VMPL_{ni,k} \leq w_i \perp L_{i,k} \geq 0.$$

Next, to incorporate the limit approach to evaluate (8) at a corner, let us start by considering an equilibrium with  $L_{i,k} > 0$ . Letting  $X_{i,k} \equiv \sum_n p_{ni,k}y_{ni,k}$  denote the revenue from sales of good  $(i, k)$  across all destinations, such an equilibrium would satisfy  $w_i L_{i,k} = X_{i,k}$ ; hence, from (8) we have  $VMPL_{ni,k} = X_{i,k}/L_{i,k}$  for every  $n$ . Equations (1), (2), and (8) further imply  $X_{i,k} = \sum_n \lambda_{ni,k} \beta_{n,k} w_n \bar{L}_n$  with

$$(9) \quad \lambda_{ni,k} = \frac{A_{i,k}^{\sigma_k-1} L_{i,k}^{\alpha_k} (\tau_{ni,k} w_i)^{-(\sigma_k-1)}}{\sum_l A_{l,k}^{\sigma_k-1} L_{l,k}^{\alpha_k} (\tau_{nl,k} w_l)^{-(\sigma_k-1)}},$$

the industry-level trade shares, and  $\alpha_k \equiv (\sigma_k - 1)\phi_k$ . Letting

$$(10) \quad G_{i,k}(\mathbf{w}, \mathbf{L}_k) = w_i - \frac{1}{L_{i,k}} \sum_n \lambda_{ni,k}(\mathbf{w}, \mathbf{L}_k) \beta_{n,k} w_n \bar{L}_n$$

<sup>8</sup>Consider an equilibrium with  $y_{ni,k} > 0$  and  $y_{n'i,k} = 0$  for some  $(i, k)$  and  $n' \neq n$  for any finite trade costs. Since this is an equilibrium, then  $q_{ni,k} = y_{ni,k} > 0$  and  $q_{n'i,k} = y_{n'i,k} = 0$ , so (1) implies that  $p_{ni,k}$  is finite and  $p_{n'i,k}$  is infinite. But (7) and  $y_{ni,k} > 0$  and  $y_{n'i,k} = 0$  imply that  $VMPL_{ni,k} \leq VMPL_{n'i,k} = w_i$ , which implies that  $p_{n'i,k} \leq p_{ni,k} \tau_{n'i,k} / \tau_{ni,k}$ , leading to a contradiction.

denote profits per worker as a function of wages  $\mathbf{w} \equiv (w_1, \dots, w_N)$  and sector  $k$  employment  $\mathbf{L}_k \equiv (L_{1,k}, \dots, L_{N,k})$ , an equilibrium with  $L_{i,k} > 0$  then satisfies  $G_{i,k}(\mathbf{w}, \mathbf{L}_k) = 0$ . This suggests the following condition as an alternative to (8):

$$(11) \quad G_{i,k}(\mathbf{w}, \mathbf{L}_k) \geq 0 \perp L_{i,k} \geq 0,$$

with  $G_{i,k}(\mathbf{w}, \mathbf{L}_k)$  at  $L_{i,k} = 0$  defined by its limit as  $L_{i,k} \rightarrow 0$ . We now argue that if  $(\mathbf{w}, \mathbf{L})$  satisfies both (6) for all  $i$  and (11) for all  $i, k$ , then it constitutes an equilibrium.

If  $L_{i,k} > 0$ , then (11) implies  $G_{i,k}(\mathbf{w}, \mathbf{L}_k) = 0$ , and so  $VMPL_{ni,k} = w_i - G_{i,k} = w_i$  for all  $n$ , and prices are  $p_{ni,k} = \tau_{ni,k} w_i / (A_{i,k} L_{i,k}^{\phi_k})$ . One can then obtain quantities consumed  $q_{ni,k}$  from (1) and (2), and by construction of  $G_{i,k}$ , equation (5) is satisfied with  $y_{ni,k} = q_{ni,k}$ . Alternatively, if  $L_{i,k} = 0$ , then (11) implies that  $G_{i,k} \geq 0$ . Equations (1) and (2) combined with  $q_{ni,k} = 0$  imply that prices  $p_{ni,k}$  are infinite, so  $VMPL_{ni,k} \equiv p_{ni,k} A_{i,k} L_{i,k}^{\phi_k} / \tau_{ni,k}$  is not well defined; hence, we evaluate it by its limit as  $L_{i,k} \rightarrow 0$ . This implies that  $VMPL_{ni,k} = w_i - G_{i,k} \leq w_i$  for all  $n$ , and so condition (8) is satisfied.

For future purposes, we define an equilibrium labor allocation given wages  $\mathbf{w}$  as an  $\mathbf{L}$  such that  $(\mathbf{w}, \mathbf{L})$  satisfies (11). An equilibrium is then a  $\mathbf{w}$  such that there is an equilibrium labor allocation given  $\mathbf{w}$  that satisfies (6).

We illustrate the previous ideas by considering the simple case of two countries under frictionless trade and focusing on sector  $k = 1$ . Assuming for simplicity that  $A_{1,1} = A_{2,1}$ ,  $\bar{L}_1 = \bar{L}_2 = 1$ , and  $\beta_{1,1} = \beta_{2,1} = 1/2$ , an equilibrium labor allocation given wages  $w_1 = w_2 = 1$  satisfies (11) for  $k = 1$  with

$$G_{i,1}(\mathbf{L}_1) = 1 - \frac{L_{i,1}^{\alpha_1 - 1}}{L_{1,1}^{\alpha_1} + L_{2,1}^{\alpha_1}}.$$

From (11) for  $k = 1$  we also get  $\sum_i L_{i,1} G_{i,1}(\mathbf{L}_1) = 0$ , which implies  $L_{1,1} + L_{2,1} = 1$ , and so we can think of  $G_{1,1}$  as a function of  $L_1$  by setting  $L_{2,1} = 1 - L_{1,1}$ . This function is shown in Figure 1 for  $L_{1,1} \in [0, 1/2]$ . If  $\alpha_1 < 1$  (as in the left panel), then  $G_{1,1}$  becomes minus infinity as  $L_{1,1} \rightarrow 0$ , and so equation (11) cannot hold, implying that there are no equilibria with corner allocations; while if  $\alpha_1 > 1$ , then  $G_{1,1} = 1$  at  $L_{1,1} = 0$ , and a corner allocation can be part of an equilibrium. We discuss this more formally in Section III.

We now make three clarifying observations before proceeding with the rest of the analysis. First, as argued above, a corner equilibrium entails infinite prices for goods that are not produced, a standard feature in trade and spatial equilibrium models.<sup>9</sup> Second, our analysis lays out the equilibrium conditions using standard complementary slackness conditions to properly evaluate corner allocations and

<sup>9</sup>For example, in the Melitz (2003) trade model, varieties that are not sold in a destination have infinite prices. Similarly, in the Allen and Arkolakis (2014) economic geography model, an equilibrium with an uninhabited location entails infinite prices for that location's differentiated variety.

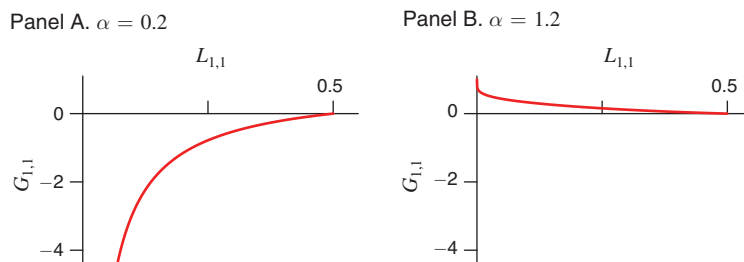


FIGURE 1.  $G_{1,1}$ , TWO COUNTRIES, TWO INDUSTRIES SYMMETRIC CASE

does not introduce any equilibrium selection mechanism. Third, reformulating the complementary slackness condition in terms of  $G_{i,k}$  allows us to evaluate a corner allocation by its limit  $L_{i,k} \rightarrow 0$ . Mathematically, we are working with the extended real line and defining all functions by their limits as a way to resolve indeterminacies of the kind  $0 \times (\pm\infty)$ ,  $\pm\infty/\pm\infty$ .<sup>10</sup>

In the next section we show isomorphisms between our model and other work-horse models in trade. A consequence is that our characterization of equilibrium in Section III will apply to these other settings as well.

## II. A Common Framework

In this section we discuss how generalized multisector versions of the models in Krugman (1980) and Melitz (2003) lead to the same equilibrium conditions as the model of the previous section—except, of course, that the trade and scale elasticities correspond to different structural parameters.<sup>11,12</sup> We also briefly discuss the implications of allowing for EES in the multisector Eaton and Kortum (2002) model, with a fuller treatment in Appendix A.A3.

As we discuss below, a multisector version of Krugman (1980) would imply that  $\alpha_k = 1$  for all  $k$ . To add flexibility and allow for  $\alpha_k \neq 1$ , we assume that the elasticity of substitution across varieties from different countries can differ from the elasticity of substitution across varieties from the same country. Specifically, there is a continuum of differentiated varieties within each industry, and preferences are

<sup>10</sup>In Section 2 of the online Appendix, we demonstrate that our approach is identical to that in the recent economic geography literature. In particular, we show that the well-known Allen and Arkolakis (2014) model in which each location produces an Armington differentiated good leads to a condition on parameters ruling out irregular equilibria (some locations are uninhabited corner allocations) that is similar to our condition for ruling out corner equilibria.

<sup>11</sup>In Section 3 of the online Appendix, we additionally show that adding heterogeneous labor supply to the Armington model à la Galle, Rodríguez-Clare, and Yi (2023) also leads to the same equilibrium conditions as in the Armington model of the previous section, except with  $\alpha_k$  now given by  $(\sigma_k - 1) \left( \phi_k - \frac{1}{\kappa - 1} \right) \left( \frac{\kappa - 1}{\kappa} \right)$ , where  $\kappa$  is a parameter regulating the degree of worker heterogeneity.

<sup>12</sup>In a similar spirit, Helpman and Krugman (1985) provide an integrated framework for international trade in the presence of different returns to scale and market structures. While their analysis is theoretical and limited to the case of frictionless trade, we focus on gravity models and the quantitative implications of scale economies for trade flows and welfare.

multitiered: Cobb-Douglas across industries with weights  $\beta_{i,k}$ , CES across country bundles within an industry with elasticity of substitution  $\eta_k$ , and CES across varieties within a country bundle with elasticity of substitution  $\sigma_k > 1$ .<sup>13</sup> Everything else is as in the standard Krugman model: labor is the only factor of production, and the total labor cost of producing quantity  $q$  of any variety in industry  $(i, k)$  is  $F_{i,k} + q/A_{i,k}$ , where labor productivity  $A_{i,k}$  and fixed cost  $F_{i,k}$  are exogenous but can vary across industry-country pairs.

In Appendix A we show that the expression for trade shares in this generalized Krugman model collapses to that in equation (9) but now with trade and scale elasticities given by  $\eta_k - 1$  and  $(\sigma_k - 1)^{-1}$ , respectively, and  $\alpha_k \equiv (\eta_k - 1)/(\sigma_k - 1)$ .<sup>14</sup> If  $\sigma_k = \eta_k$  for all  $k$ , then this is just the standard multi-industry Krugman model and  $\alpha_k = 1$ .<sup>15</sup> See rows 2 and 4 of Table 1.

Now consider the Melitz (2003) model with Pareto-distributed productivity and the same preferences as in the Krugman model above. After paying a fixed “entry” cost  $F_{i,k}$  in units of labor in country  $i$ , firms are able to produce a variety in industry  $(i, k)$  with labor productivity drawn from a Pareto distribution with shape parameter  $\theta_k > \sigma_k - 1$  and location parameter  $b_{i,k}$ . Firms from  $i$  can then pay a fixed “marketing” cost  $f_{n,k}$  in units of labor of  $n$  to serve that market.<sup>16,17</sup> As shown in Appendix A, the expression of trade shares in this model collapses to that in equation (9) but now with trade and scale elasticities given by  $\theta_k / \left[ 1 + \theta_k \left( \frac{1}{\eta_k - 1} - \frac{1}{\sigma_k - 1} \right) \right]$  and  $1/\theta_k$ , respectively. Again, if  $\sigma_k = \eta_k$  for all  $k$ , then this is just the standard multisector Melitz-Pareto model and  $\alpha_k = 1$  for all  $k$ . See rows 3 and 5 in Table 1.

In Appendix A we show that both the generalized multisector Krugman and Melitz models lead to an equilibrium system that is equivalent to the one for the model presented in Section I except for constants that play no role for equilibrium analysis and comparative statics.<sup>18</sup> We can then think of a common framework that nests the multi-industry Armington model with EES of Section I as well as the Krugman and Melitz models presented in the current section. We refer to this common framework in the remainder of the paper and use notation  $\varepsilon_k$  and  $\psi_k$  to denote the trade and scale

<sup>13</sup>Feenstra et al. (2018) also consider a multi-industry Melitz-Pareto model with possibly different elasticities of substitution across varieties from different countries and across varieties from the same country.

<sup>14</sup>In contrast to the Armington case, in this model one can directly evaluate the zero-profit condition at a corner. If  $L_{i,k} = 0$ , the unit cost of any firm in  $i$  selling a variety of good  $k$  in any market is finite, and the price would just be the CES markup  $\sigma/(\sigma - 1)$  times the unit cost. With  $\alpha_k < 1$ , the demand for varieties of good  $k$  from  $i$  would be infinite if  $L_{i,k} = 0$  in each market  $n$ , contradicting the zero-profit condition. The same reasoning applies to the Melitz-Pareto model below.

<sup>15</sup>It is also straightforward to incorporate external economies of scale into the multi-industry Krugman model presented above. For instance, letting  $A_{i,k} \equiv \tilde{A}_{i,k} L_{i,k}^{\phi_k}$  leads to scale and trade elasticities  $(\sigma_k - 1)^{-1} + \phi_k$  and  $\eta_k - 1$ , respectively.

<sup>16</sup>To simplify the analysis, we assume that the fixed marketing cost to serve destination  $n$  does not vary across origins  $i$ . Allowing these fixed costs to vary across country pairs would not change any of our main conclusions below.

<sup>17</sup>The assumption that fixed marketing costs are paid in units of labor of the destination country is critical for the result that this model collapses to the general structure introduced above. This is related to the discussion in ACR about how their macro-level restriction R3' obtains in the Melitz-Pareto model if and only if the fixed cost is paid in units of labor of the destination country.

<sup>18</sup>This is true in the Melitz model under the balanced-trade assumption. If a comparative statics exercise involves changing trade deficits, then the outcome of the Melitz model would not be equivalent to the outcome of the other two models. We ignore this issue in the analysis that follows and recognize that whenever our comparative static exercises involve changing trade deficits, the results apply only to the Armington and Krugman models.

TABLE 1—MAPPING TO DIFFERENT MODELS

Model	Trade elasticity, $\varepsilon_k$	Scale elasticity, $\psi_k$	$\alpha_k = \varepsilon_k \psi_k$
Multisector Armington with EES	$\sigma_k - 1$	$\phi_k$	$(\sigma_k - 1)\phi_k$
Multisector Krugman	$\sigma_k - 1$	$\frac{1}{\sigma_k - 1}$	1
Multisector Melitz-Pareto model	$\theta_k$	$\frac{1}{\theta_k}$	1
Generalized multisector Krugman	$\eta_k - 1$	$\frac{1}{\sigma_k - 1}$	$\frac{\eta_k - 1}{\sigma_k - 1}$
Generalized multisector Melitz-Pareto	$\frac{\theta_k}{1 + \theta_k(\frac{1}{\eta_k - 1} - \frac{1}{\sigma_k - 1})}$	$\frac{1}{\theta_k}$	$\frac{1}{1 + \theta_k(\frac{1}{\eta_k - 1} - \frac{1}{\sigma_k - 1})}$

elasticities,  $S_{i,k}$  to capture industry productivity, and  $\mu_{n,k}$  for the constant affecting the price index. With  $\alpha_k \equiv \varepsilon_k \psi_k$ , the expressions for trade shares and price indices are now

$$(12) \quad \lambda_{ni,k}(\mathbf{w}, \mathbf{L}_k) = \frac{S_{i,k} L_{i,k}^{\alpha_k} (w_i \tau_{ni,k})^{-\varepsilon_k}}{\sum_l S_{l,k} L_{l,k}^{\alpha_k} (w_l \tau_{nl,k})^{-\varepsilon_k}}$$

and

$$(13) \quad P_{n,k} = \mu_{n,k} \left[ \sum_l S_{l,k} L_{l,k}^{\alpha_k} (w_l \tau_{nl,k})^{-\varepsilon_k} \right]^{-1/\varepsilon_k}.$$

The Krugman and Melitz models provide different ways to micro-found external economies of scale. The source of economies of scale in the Krugman model is the love-of-variety effect: keeping everything else constant, an increase in  $L_{i,k}$  of one log point leads to a proportional increase in the number of varieties produced in industry  $(i, k)$ , which has an effect equivalent to an increase in productivity of that industry (in log points) of  $1/(\sigma_k - 1)$ . The source of economies of scale in the Melitz model is the selection effect: keeping everything else constant, an increase in  $L_{i,k}$  of one log point leads to a proportional increase in firm entry and an increase in productivity thresholds for industry  $(i, k)$  for selling in any market as well as an effect equivalent to an increase in productivity of that industry (in log points) of  $1/\theta_k$ .

We finish this section by briefly discussing the case in which sector-level EES applies to the multisector Eaton and Kortum (2002) model in Costinot, Donaldson, and Komunjer (2012), with details provided in Appendix A3. In the presence of EES, the multisector Eaton-Kortum (EK) model with EES exhibits exactly the same interior equilibria as the corresponding Armington model, with trade elasticities given by the shape parameter of the Fréchet distribution of productivity in sector  $k$  rather than  $\sigma_k - 1$  and scale elasticities given by the technology parameter  $\phi_k$  in both cases. The only substantive difference between these two perfectly competitive models is that condition (11) does not correctly capture the zero-profit condition that must hold at a corner with  $L_{i,k} = 0$  in the EK model with EES. In contrast to what happens in the Armington model, in the EK model prices are finite and well defined at a corner, and so it is possible to directly evaluate the zero-profit condition there. With  $L_{i,k} = 0$ , productivity in  $(i, k)$  is zero, and zero employment is then compatible with equilibrium as long as some other country has positive employment in sector  $k$  (so that prices are well defined). Still, both models are equivalent

outside of corners, and so—as discussed in the next section—if  $\alpha_k < 1$ , then an equilibrium with  $L_{i,k} = 0$  in the multisector EK model with EES has the peculiar feature that profits per worker in  $(i, k)$  are  $-w_i$  but positive and unbounded if  $L_{i,k}$  is positive and small. In contrast, if  $\alpha_k > 1$ , then corner equilibria would not feature this discontinuity. An equilibrium refinement that required profits per worker to be continuous at the equilibrium would get rid of corner equilibria if  $\alpha_k \leq 1$  but not otherwise, thereby restoring the isomorphism with the Armington model.

### III. Characterizing Equilibrium

We now turn to characterizing the equilibrium of the multi-industry gravity model with industry-level economies of scale. The analysis proceeds in two steps: we first characterize the equilibrium labor allocations given wages—formally defined as an  $\mathbf{L}$  such that  $(\mathbf{w}, \mathbf{L})$  satisfies (11)—and then we characterize wages that satisfy labor market clearing given the corresponding equilibrium labor allocations. This proves convenient because our stronger result is for the first step, where we have a sharp necessary and sufficient condition for uniqueness of labor allocations given wages—namely,  $\alpha_k < 1$  or  $\alpha_k = 1$  plus a regularity condition for trade costs—and this is also where we characterize intuitive parameter restrictions for whether the equilibrium labor allocation exhibits corners or not. In contrast, for the second step we are not able to show that the condition above is sufficient for there to be a unique wage vector that satisfies the labor-market-clearing conditions in all countries. We are only able to do this for the case of frictionless trade, two countries, or a small open economy in the presence of trade costs. Although clearly special, these cases are important because in the standard textbook treatment there is multiplicity of equilibria in any of these special cases. Moreover, in extensive simulations we have never found an instance of multiple equilibria for the more general case.<sup>19</sup>

*Two-Step Equilibrium Definition.*—The equilibrium labor allocations for some wage vector  $\mathbf{w}$  are given by  $\mathbf{L}$  that satisfy (11) for all  $(i, k)$ . Let  $\mathcal{L}(\mathbf{w})$  be the set of such equilibrium allocations. A wage vector  $\mathbf{w}$  is an equilibrium wage vector if there exists an element  $\mathbf{L} \in \mathcal{L}(\mathbf{w})$  such that  $\mathbf{L}$  also satisfies (6) for all  $i$ .

Note that given wages for each industry  $k$  we have a system of  $N$  nonlinear complementary slackness conditions in  $L_{i,k}$  for  $i = 1, \dots, N$  specified by (11). For the first step we exploit the fact that this system is independent across  $k$ .<sup>20</sup> We now introduce some additional notation and definitions.

*Interior, Corner, and Complete Specialization Allocations.*—An allocation  $\mathbf{L}_k$  is an interior allocation if  $L_{i,k} > 0$  for all  $i$ , an allocation  $\mathbf{L}_k$  is a corner allocation if

<sup>19</sup>We have systematically looked for examples of multiple equilibria by generating thousands of random economies with three and four countries in the presence of trade costs but have found none.

<sup>20</sup>The assumption of upper-tier Cobb-Douglas preferences plays an important role in allowing the system to be independent across industries. More generally, industry-level expenditure shares,  $\beta_{n,k}$ , in (10) need not be fixed and would be a function of prices across all sectors, thereby breaking the independence property. As is well known, with upper-tier CES preferences and a high enough elasticity of substitution, there would be multiple equilibria even in autarky.

$L_{i,k} = 0$  for at least one  $i$ , and an allocation  $\mathbf{L}_k$  is a complete specialization allocation if there is a unique  $i^*(k)$  such that  $L_{i,k} = 0$  for all  $i \neq i^*(k)$ .

*Industry-Level Equilibrium Labor Allocations.*—Given wage  $\mathbf{w}$ ,  $\mathcal{L}_k(\mathbf{w})$  denotes the set of equilibrium labor allocations in industry  $k$ ; i.e., for any  $\mathbf{L}_k \in \mathcal{L}_k(\mathbf{w})$ ,  $\mathbf{L}_k$  satisfies complementary slackness conditions (11) for industry  $k$ .

### A. Step 1: Equilibrium Labor Allocations

Given the previous definitions, we are now ready to state our first Proposition. The formal proof is provided in Appendix B.

**PROPOSITION 1:** *If either (a)  $0 \leq \alpha_k < 1$  or (b)  $\alpha_k = 1$  and the matrix  $\{\tau_{ni,k}^{-\varepsilon_k}\}_{n,i}$  is nonsingular, then the set  $\mathcal{L}_k(\mathbf{w})$  is a singleton; if  $\alpha_k > 1$ , then the set  $\mathcal{L}_k(\mathbf{w})$  contains multiple allocations, including (but not necessarily limited to) one for each complete specialization allocation. Moreover, the unique allocation in  $\mathcal{L}_k(\mathbf{w})$  is an interior allocation if  $0 \leq \alpha_k < 1$ , while it may be an interior or a corner allocation if  $\alpha_k = 1$ .*

We now provide some intuition for this result, focusing for concreteness on the Armington specification. We can think of a battle between two forces that are activated as the labor allocation to an industry vanishes. The first is consumers' preferences for goods differentiated by origin—a force for cross-country industry diversification with strength  $1/(\sigma_k - 1)$ —which implies that the marginal utility (and hence price) increases to infinity as consumption of a good from a particular country goes to zero. The second is external economies of scale—a force for cross-country industry specialization with strength  $\phi_k$ —which implies that productivity falls to zero; thus, cost rises to infinity as industry employment falls to zero. For  $\alpha_k \in (0, 1)$  we have  $1/(\sigma_k - 1) > \phi_k$ , and the force for diversification dominates the one for specialization, implying a unique and strictly positive equilibrium labor allocation across countries.

### B. Step 2: Equilibrium Wages

In Section 4.2 of the online Appendix, we show formally that an equilibrium always exists. In Section 4.4 of the online Appendix we also show that  $\alpha_k \leq 1$  for all  $k$  is a necessary condition for unique equilibrium wages, a direct consequence of Proposition 1. We next provide a Proposition that characterizes sufficient conditions for a unique equilibrium in the absence of trade costs and then provide a brief description of results for the more general case with trade costs. All proofs are provided in Section 4 of the online Appendix.

**PROPOSITION 2:** *If  $0 \leq \alpha_k \leq 1$  and trade is frictionless in all industries, then there is a unique wage equilibrium.*

Combined with Proposition 1, this result implies that the labor allocation is unique in industries with  $\alpha_k < 1$ . If there are two or more industries with  $\alpha_k = 1$ , then there can be a continuum of labor allocations in knife-edge cases when  $S_{i,k} w_i^{-\varepsilon_k - 1}$  is the same for two or more countries, just as in the standard Ricardian model. Regardless of whether labor allocations are unique or not, however, the pattern of specialization is always compatible with comparative advantage.<sup>21</sup>

In a setting with trade frictions, we show formally that the condition  $\alpha_k \leq 1$  for all  $k$  also guarantees that the equilibrium is unique in two special cases: that of a small open economy and that of two countries, with the latter requiring the additional regularity condition on trade costs in Proposition 1 for the case with  $\alpha_k = 1$ . Despite extensive simulation analysis, we have never encountered an instance of multiple equilibria even with more than two countries.<sup>22</sup> We take this as evidence that the model is well behaved under these parameter restrictions. For the formal propositions, proofs, and a more detailed discussion, we refer the reader to Section 4 of the online Appendix.

#### IV. Scale Economies and the Gains from Trade Relative to Autarky

In this section we explore the implications of scale economies for the gains from trade, defined as in ACR as the negative of the percentage change in real income as we move from the observed equilibrium to autarky. We first do so at a theoretical level and then use the World Input-Output Dataset (WIOD) to quantify the effects. In Section V we turn to the related but distinct question of how scale economies affect the gains from trade liberalization, defined as the welfare effects of a counterfactual decline in tariffs or trade costs starting at the observed equilibrium.

##### A. Theoretical Analysis

In principle, countries that specialize in industries with weak economies of scale could even lose from trade—the premise of Frank Graham’s argument for protection. It turns out, however, that this cannot happen if  $0 \leq \alpha_k \leq 1$  for all  $k$ . A simple argument establishes this result for an industry-country pair  $(i, k)$  with  $L_{i,k} > 0$ . Using equations (12) and (13) and setting  $w_i = 1$  by choice of numeraire, the industry-level price index is  $P_{i,k} = \mu_{i,k} S_{i,k}^{-1/\varepsilon_k} L_{i,k}^{-\psi_k} \lambda_{ii,k}^{1/\varepsilon_k}$ . Since autarky employment

<sup>21</sup>To see this, consider the case with  $\alpha_k = \alpha$  and  $\varepsilon_k = \varepsilon$ . If  $\alpha < 1$ , then  $\ln\left(\frac{L_{i,k}/L_{j,k}}{L_{i,s}/L_{j,s}}\right) = \frac{1}{1-\alpha} \ln\left(\frac{S_{i,k}/S_{j,k}}{S_{i,s}/S_{j,s}}\right)$ , a generalization of the result in Costinot, Donaldson, and Komunjer (2012) to the case with  $\alpha > 0$ : economies of scale simply amplify the effect of comparative advantage on the pattern of specialization by the factor  $1/(1-\alpha)$ . If  $\alpha = 1$ , then  $S_{i,k}/S_{i,s} > S_{j,k}/S_{j,s}$  implies that  $L_{i,s}L_{j,k} = 0$  so that for any two countries and any two industries, both countries cannot have positive employment in their industries of comparative disadvantage. We prove this by showing that if  $L_{i,s} > 0$ , then  $L_{j,k} = 0$ , with the converse proven similarly. In particular, if  $L_{i,s} > 0$ , then  $S_{i,s}w_i^{-\varepsilon-1} \geq S_{j,s}w_j^{-\varepsilon-1}$  for all  $l$ , and so we have  $S_{i,s}w_i^{-\varepsilon-1} \geq S_{j,s}w_j^{-\varepsilon-1} > S_{j,k}w_j^{-\varepsilon-1}S_{i,s}/S_{i,k}$ ; hence,  $S_{i,k}w_i^{-\varepsilon-1} > S_{j,k}w_j^{-\varepsilon-1}$ , implying that  $L_{i,k} = 0$ .

<sup>22</sup>There are two papers that study the question of uniqueness of equilibrium in the multi-industry Krugman model: Hanson and Xiang (2004) consider the case of two countries and a continuum of industries, while Behrens et al. (2009) consider the case of many countries, one industry, and exogenous wages. They both show uniqueness under the assumption that there are no corner allocations. We extend their result to a more general environment, introduce the key condition that the product of the trade and scale elasticities is weakly lower than one in all industries, and allow for corner allocations.

levels are  $L_{i,k}^A = \beta_{i,k} \bar{L}_i$ , autarky price indices are then  $P_{i,k}^A = \mu_{i,k} S_{i,k}^{-1/\varepsilon_k} (\beta_{i,k} \bar{L}_i)^{-\psi_k}$ . We can then write

$$(14) \quad P_{i,k} = P_{i,k}^A (\lambda_{ii,k})^{1/\varepsilon_k} \left( \frac{L_{i,k}}{\beta_{i,k} \bar{L}_i} \right)^{-\psi_k}.$$

From (14) it is now evident that in the presence of scale economies there are two potentially countervailing forces acting on price indices: the standard gains from trade captured by  $(\lambda_{ii,k})^{1/\varepsilon_k}$  and the potentially negative scale effect arising from a contraction in employment relative to autarky, which is captured by  $\left( \frac{L_{i,k}}{\beta_{i,k} \bar{L}_i} \right)^{-\psi_k}$ .

Note, however, that in an equilibrium with  $L_{i,k} > 0$  we must have  $L_{i,k} > \lambda_{ii,k} \beta_{i,k} \bar{L}_i$ , since the RHS is just total employment associated with domestic sales. This implies that  $P_{i,k} < P_{i,k}^A (\lambda_{ii,k})^{1/\varepsilon_k - \psi_k}$ ; hence,  $\psi_k \leq 1/\varepsilon_k$  (or  $\alpha_k \equiv \varepsilon_k \psi_k \leq 1$ ) implies that  $P_{i,k} < P_{i,k}^A$ . In Appendix C.C1 we show that this strict inequality also holds in the case in which  $L_{i,k} = 0$ , thus establishing the following result:

**PROPOSITION 3:** *If  $0 \leq \alpha_k \leq 1$  for all  $k$ , then all countries gain from trade.*

Intuitively, the strength of the standard gains from trade is regulated by the inverse of the trade elasticity,  $1/\varepsilon_k$ , while the strength of the potentially negative scale effects is regulated by  $\psi_k$ . For  $\alpha_k \leq 1$  the standard gains from trade always neutralize any potentially opposing scale effects that could lead to higher prices with trade than in autarky. On the contrary, if  $\alpha_k > 1$ , then one could have higher prices in some industries with trade than without, leading to the possibility of losses from trade.

Proposition 3 can be seen as a generalization of Proposition 1 in Venables (1987), which states that in a Krugman (1980) model with an “outside good,” all countries gain from trade. Formally, the model in Venables (1987) is isomorphic to ours when we consider two countries and two industries—one having no trade costs, no scale economies, and an infinite trade elasticity (the “outside good”) and the other having trade costs, scale economies, and a finite trade elasticity, with  $\alpha_k = 1$ . Proposition 3 shows that this generalizes to a case without an outside good (i.e., with endogenous wages), with multiple industries and microfoundations different than Krugman, and arbitrary scale economies as long as  $\alpha_k \leq 1$  for all  $k$ .

To further explore the implications of scale economies for the magnitude of the gains from trade, we assume that the equilibrium is interior so that all trade shares and labor allocations are strictly positive. This allows us to derive an expression for the gains from trade as a function of industry-level data and the trade and scale elasticities that extend the multisector expressions in ACR.<sup>23</sup>

Equations (12) and (13), together with  $P_n = \tilde{\beta}_n \prod_{k=1}^K P_{n,k}^{\beta_{n,k}}$ , imply that

$$W_n \equiv w_n/P_n = \tilde{\beta}_n^{-1} \prod_k \left( \mu_{n,k}^{-1} S_{n,k} L_{n,k}^{\varepsilon_k \psi_k} \lambda_{nn,k}^{-1} \right)^{\beta_{n,k}/\varepsilon_k}.$$

<sup>23</sup> Given Proposition 1, the assumption that labor allocations are strictly positive is not restrictive for the case with  $0 \leq \alpha_k < 1$  for all  $k$ .

Using hat notation,  $\hat{x} = x'/x$ , a foreign shock (i.e., a shock that does not affect the exogenous variables in country  $n$ ) induces a change in welfare in country  $n$  equal to

$$(15) \quad \hat{W}_n = \prod_k \hat{\lambda}_{nm,k}^{-\beta_{n,k}/\varepsilon_k} \cdot \prod_k \hat{L}_{n,k}^{\beta_{n,k}\psi_k}.$$

The first term on the RHS of this expression is the standard multi-industry formula for gains from trade (with upper-tier Cobb-Douglas preferences), while the second term is an adjustment for scale economies.

To better understand this expression, we can use the fact that the welfare effect of an infinitesimally small change in wages and prices is  $d\ln W_n \equiv_{-\varepsilon_k} d\ln w_n - \sum_k \beta_{n,k} d\ln P_{n,k}$ . Log differentiating  $\lambda_{nm,k} = (\mu_{n,k} S_{n,k}^{-1/\varepsilon_k} L_{n,k}^{-\psi_k} w_n/P_{n,k})$  with respect to  $L_{n,k}$ ,  $w_n$ , and  $P_{n,k}$  and substituting into the previous equation yields

$$(16) \quad d\ln W_n = -\sum_k \beta_{n,k} \frac{d\ln \lambda_{nm,k}}{\varepsilon_k} + \sum_k \beta_{n,k} \psi_k d\ln L_{n,k}.$$

The first term on the right-hand side captures the welfare effect of an infinitesimally small foreign shock taking home productivity as given, while the second term captures the welfare effect of that shock through home productivity changes caused by changing industry employment levels in the presence of scale effects. Integrating the first term over some discrete shock yields  $\prod_k \hat{\lambda}_{nm,k}^{-\beta_{n,k}/\varepsilon_k}$ , while integrating the second term yields  $\prod_k \hat{L}_{n,k}^{\beta_{n,k}\psi_k}$ .<sup>24</sup>

Following ACR, we define the gains from trade as the negative of the percentage change in real income as we move from the observed equilibrium to autarky,

$$GT_n \equiv \frac{W_n - W_n^A}{W_n}.$$

We compute  $GT_n$  by applying (15) and noting that for the move back to autarky we have  $\hat{\lambda}_{nm,k} = 1/\lambda_{nm,k}$  and  $\hat{L}_{n,k} = \beta_{n,k}/r_{n,k}$ , where  $r_{n,k} \equiv L_{n,k}/\bar{L}_n$  denotes the industry revenue (or employment) shares in the observed equilibrium. Using  $e_{n,k} \equiv X_{n,k}/X_n$  for observed industry expenditure shares (of course,  $e_{n,k} = \beta_{n,k}$  in the model), this leads to a formula for the gains from trade that depends only on the

<sup>24</sup> It is interesting to compare the formula in (16) to the one that would obtain if instead of external economies of scale we had wedges affecting the allocation of labor across sectors, as in Krueger and Summers (1986) and Baqaee and Farhi (2020). The corresponding equation would be

$$d\ln W_n = -\sum_k \beta_{n,k} \frac{d\ln \lambda_{nm,k}}{\varepsilon_k} + \sum_k \frac{L_{n,k}}{\bar{L}_n} \frac{w_{n,k}}{\bar{w}_n} d\ln L_{n,k},$$

where  $\bar{w}_n \equiv \sum_k w_{n,k} L_{n,k}/\bar{L}_n$  is the average wage. Here,  $L_{n,k}/\bar{L}_n$  and  $w_{n,k}/\bar{w}_n$  play the role of  $\beta_{n,k}$  and  $\psi_{n,k}$  in (16). Whereas  $w_{n,k} = w_n$  for all  $k$  implies  $d\ln W_n = -\sum_k \beta_{n,k} \frac{d\ln \lambda_{nm,k}}{\varepsilon_k}$  in the case of wedges, this would not be true if  $\psi_k = \psi$  for all  $k$  in (16), except if  $\beta_{n,k} = L_{n,k}/L_n$ , which will generally not hold with trade. Another difference is that while we can be sure that all countries gain from trade in the case of scale economies if the necessary condition for uniqueness holds, no such clear result emerges under wedges. Finally, note that this is just comparing first-order effects; higher-order effects are different with economies of scale relative to the case with wedges.

country’s observables  $\lambda_{nm,k}$ ,  $e_{n,k}$ , and  $r_{n,k}$  as well as the trade and scale elasticities,  $\varepsilon_k$  and  $\psi_k$ ,

$$(17) \quad GT_n = 1 - \Delta_n \prod_k \lambda_{nm,k}^{e_{n,k}/\varepsilon_k},$$

where  $\Delta_n \equiv \prod_k (e_{n,k}/r_{n,k})^{e_{n,k}\psi_k}$ . The expression for the gains from trade in the standard perfectly competitive model with no scale economies obtains from (17) by setting  $\psi_k = 0$  for all  $k$ , which leads to  $\Delta_n = 1$ . The effect of scale economies on the gains from trade then depends on whether  $\Delta_n$  is higher or lower than one.

Consider first the case in which the scale elasticity is the same across industries ( $\psi_k = \psi$  for all  $k$ ) and note that  $\Delta_n^{1/\psi} = \exp D_{KL}(\mathbf{e}_n \parallel \mathbf{r}_n)$ , where  $\mathbf{r}_n \equiv (r_{n,1}, \dots, r_{n,K})$ ,  $\mathbf{e}_n \equiv (e_{n,1}, \dots, e_{n,K})$ , and

$$(18) \quad D_{KL}(\mathbf{e}_n \parallel \mathbf{r}_n) \equiv \sum_k e_{n,k} \ln(e_{n,k}/r_{n,k})$$

is the Kullback-Leibler divergence of  $\mathbf{r}_n$  from  $\mathbf{e}_n$ .<sup>25</sup> We can think of  $D_{KL}(\mathbf{e}_n \parallel \mathbf{r}_n)$  as a measure of industry specialization in country  $n$ —in autarky we would have  $\mathbf{r}_n = \mathbf{e}_n$  and  $D_{KL}(\mathbf{e}_n \parallel \mathbf{r}_n) = 0$ , while if  $\mathbf{r}_n \neq \mathbf{e}_n$ , then  $D_{KL}(\mathbf{e}_n \parallel \mathbf{r}_n) > 0$ . This implies that  $\Delta_n > 1$  (except if  $\mathbf{r}_n = \mathbf{e}_n$ , in which case  $\Delta_n = 1$ ) so that given trade shares, scale economies actually reduce the gains from trade, with a larger decline for higher values of  $\psi$  and for countries that exhibit higher levels of specialization.<sup>26</sup>

We can gain intuition about this result by going back to equation (16) and noting that if  $\psi_k = \psi$  for all  $k$  then the second term on the RHS of that equation can be written as  $\psi \sum_k e_{n,k} \frac{dL_{n,k}}{L_{n,k}}$ . A move back to autarky implies the expansion of industries with net imports and hence a high expenditure share or low employment. In either case, the expenditure-weighted productivity gain in expanding industries will be, on average, higher than the expenditure-weighted productivity loss in contracting industries, and hence  $\psi \sum_k e_{n,k} \frac{dL_{n,k}}{L_{n,k}} > 0$ . After integration, this leads to  $\Delta_n > 1$ .

Readers may be surprised by the result that the gains from trade are lower with economies of scale (i.e.,  $0 < \alpha < 1$ ) than without (i.e.,  $\alpha = 0$ ). As we mentioned above, this is because we are following the ACR approach of taking trade flows as data that is held fixed as we compare different models. The intuition that scale economies should lead to larger gains from trade through deeper industry-level specialization and larger trade flows is not operative here, although as we illustrate in Section V, this intuition is relevant for the gains from trade liberalization.

<sup>25</sup> In the information theory literature, this measure is also called *relative entropy* of  $\mathbf{e}_n$  with respect to  $\mathbf{r}_n$ .

<sup>26</sup> The opposite result would hold if instead of economies of scale we had diseconomies of scale. For example, in a setting with  $\psi = 0$  and worker-level heterogeneity, Galle, Rodríguez-Clare, and Yi (2023) show that  $GT_n = 1 - \prod_k \lambda_{nm,k}^{e_{n,k}/\varepsilon_k} (e_{n,k}/r_{n,k})^{-e_{n,k}/\kappa}$ , where  $\kappa$  is a parameter that determines the degree of heterogeneity. The argument above now implies that the gains from trade are higher than in the case with no scale economies, which obtains here in the limit as  $\kappa \rightarrow \infty$  and corresponds to the case in which workers are homogeneous. We discuss this more formally in Section 3 of the online Appendix.

Turning now to the more general case in which  $\psi_k$  varies across  $k$ ,  $\Delta_n$  can be rewritten as

$$(19) \quad \Delta_n = \exp\{\bar{\psi}[DS_n - PS_n]\},$$

where  $DS_n \equiv D_{KL}(\mathbf{e}_n \parallel \mathbf{r}_n)$  and

$$(20) \quad PS_n \equiv \sum_k \frac{\psi_k - \bar{\psi}}{\bar{\psi}} \ln\left(\frac{r_{n,k}}{e_{n,k}}\right)^{e_{n,k}},$$

with  $\bar{\psi} \equiv (1/K)\sum_k \psi_k$ . Expression (19) makes it clear that in general there are two (possibly competing) forces: the first measures the degree of specialization ( $DS$ ) in country  $n$  while the second measures the pattern of specialization ( $PS$ ), i.e., the tendency of country  $n$  to specialize in industries with either higher or lower than average scale economies. Since  $DS$  always pushes toward lower gains relative to the case with no scale economies, the overall effect of scale economies on gains from trade depends on the direction and magnitude of  $PS$ . Countries that tend to specialize in industries with lower than average scale economies—so that  $PS$  is negative—gain less from trade with scale economies than without. However, in countries that tend to specialize in industries with higher than average scale economies—so that  $PS$  is positive—the effect of scale economies on the gains from trade is ambiguous. If a country's  $PS$  is strong enough to overcome its  $DS$ , then such a country could have higher gains with than without scale economies.

### B. Quantitative Analysis

We now explore the quantitative implications of the previous observations by computing the gains from trade using actual data combined with equation (17). A critical input for the analysis is a set of values for the sector-level trade and scale elasticities. We take those from BCDR, who, in turn, take trade elasticities from the median value of recent estimates in the literature and estimate scale elasticities using an approach that is motivated by the model we have developed in this paper.<sup>27,28</sup> These elasticities satisfy the condition  $\alpha_k \leq 1$  for all  $k$  and are listed in Table C2 of Appendix C.C2. We follow Costinot and Rodríguez-Clare (2014) and compute measures of  $\lambda_{nn,k}$ ,  $e_{n,k}$ , and  $r_{n,k}$  employing data on 31 sectors and 33 countries (plus the rest of the world aggregate) from the World Input Output Database (WIOD, Timmer et al. 2015) in 2008.<sup>29</sup>

We present the results of this exercise in Table 2. Column 1 reports the gains from trade when setting all scale economies to zero, while column 2 sets them to the

<sup>27</sup>BCDR apply this procedure for manufacturing, mining, and agriculture sectors only. For the other sectors they set trade elasticities at four and scale elasticities at zero. We follow the same approach here.

<sup>28</sup>In their main specification, BCDR assume that upper-tier preferences are CES and estimate this upper-tier elasticity to be 1.5. This is then used in the construction of the instrumental variables for the estimation of scale elasticities. Although, formally, this leads to a difference relative to our Cobb-Douglas model, the estimates obtained by BCDR when they restrict to Cobb-Douglas are virtually identical to their baseline estimates.

<sup>29</sup>Equation (17) ignores trade deficits. As discussed in Costinot and Rodríguez-Clare (2014), this implies that our results in this subsection capture the change in real income rather than the change in real expenditure caused by shutting down trade and closing any trade deficits that exist in the data.

TABLE 2—GAINS FROM TRADE, DEGREE, AND PATTERN OF SPECIALIZATION

Country	Gains from trade, %			<i>PS</i> , %	Country	Gains from trade, %			<i>PS</i> , %
	No scale	Scale	<i>DS</i> , %			No scale	Scale	<i>DS</i> , %	
	(1)	(2)	(3)	(4)		(1)	(2)	(3)	(4)
AUS	2.79	2.13	3.57	-8.74	IRL	10.95	11.29	6.90	13.95
AUT	8.40	8.48	1.54	3.21	ITA	3.81	3.96	2.60	5.47
BEL	14.51	14.82	7.24	13.83	JPN	2.61	2.91	4.05	9.78
BRA	1.62	1.56	0.38	-0.70	KOR	6.94	7.45	13.25	23.31
CAN	4.78	4.36	2.41	-5.69	MEX	3.77	3.04	1.75	-12.29
CHN	2.86	3.01	1.59	4.46	NLD	10.07	10.49	1.98	10.74
CZE	7.42	7.56	1.43	4.29	POL	5.67	5.52	0.86	-2.04
DEU	5.91	6.55	3.27	15.90	PRT	6.29	5.77	3.16	-6.95
DNK	9.27	9.00	8.96	3.54	ROM	5.70	4.83	2.04	-14.83
ESP	4.05	3.97	2.26	0.77	RUS	2.84	1.23	7.65	-22.59
FIN	5.48	5.88	3.63	11.40	SVK	10.72	10.80	3.81	5.48
FRA	4.00	4.15	1.97	4.76	SVN	11.06	10.37	11.22	-3.00
GBR	4.43	4.20	0.77	-3.58	SWE	6.68	6.99	1.90	8.10
GRC	5.85	4.30	10.97	-19.04	TUR	3.82	3.77	1.77	0.80
HUN	10.87	10.53	2.68	-4.17	TWN	8.81	9.04	11.24	15.92
IDN	3.23	3.06	1.27	-1.96	USA	1.97	1.76	0.72	-3.31
IND	3.40	3.35	1.32	0.28	ROW	6.49	5.74	4.99	-9.55
					Average	6.09	5.94	3.98	1.10

Notes: Gains from trade in columns 1 and 2 are calculated according to (17). For columns 1 and 2, values of  $\varepsilon_k$  for manufacturing industries are set based on values provided by BCDR, and for nonmanufacturing industries  $\varepsilon_k = 4$ . For column 1  $\psi_k = 0$  for all  $k$ . For column 2  $\psi_k$  for manufacturing industries are set based on estimates from BCDR, and for nonmanufacturing industries  $\psi_k = 0$ . BCDR-based values of  $\varepsilon_k$  and  $\psi_k$  are provided in Appendix Table C2. *DS* in column 3 are calculated according to expression (18) and do not depend on  $\psi_k$  or  $\varepsilon_k$ . *PS* in column 4 are calculated according to expression (20) and correspond to the parameterization in column 2. *DS* and *PS* are expressed in percentages. Average values in the last row are calculated based on the full set of countries.

values estimated in BCDR. Column 3 reports the degree of industry specialization, *DS*, and column 4 reports the pattern of specialization, *PS*. From the theoretical analysis above, we know that, given the average scale elasticity  $\bar{\psi} \equiv (1/K)\sum_k \psi_k$ , *DS* and *PS* are sufficient statistics for the effect of economies of scale on the gains from trade (relative to the ACR formula):  $(GT - GT_{ACR})/(1 - GT_{ACR}) \approx -\bar{\psi}(DS - PS)$ .<sup>30</sup>

Consistent with Proposition 3, gains from trade are positive for all countries even when there are scale economies. On average, the gains from trade with and without scale economies are very similar, but there are significant differences for certain countries. For example, allowing for economies of scale leads to an increase in the gains from trade for Germany (from 5.9 to 6.6 percent) and Korea (from 6.9 to 7.5 percent) but a decrease for Greece (from 5.9 to 4.3 percent) and, most prominently, for Russia (from 2.8 to 1.2 percent). Scale economies lead to lower gains from trade for Greece and Russia because these countries happen to be specialized in industries with relatively weak economies of scale as revealed by a negative *PS*. In contrast, the pattern of specialization in Germany and Korea is such that *PS* is positive and dominates the negative effect of *DS*, thus leading to higher gains from trade with economies of scale.

<sup>30</sup>This is an approximation. We start with  $GT - GT_{ACR} = (1 - \Delta)(1 - GT_{ACR})$ . Combined with  $1 - \Delta = 1 - \exp[\bar{\psi}(DS - PS)] \approx -\bar{\psi}(DS - PS)$ , the result in the text follows immediately.

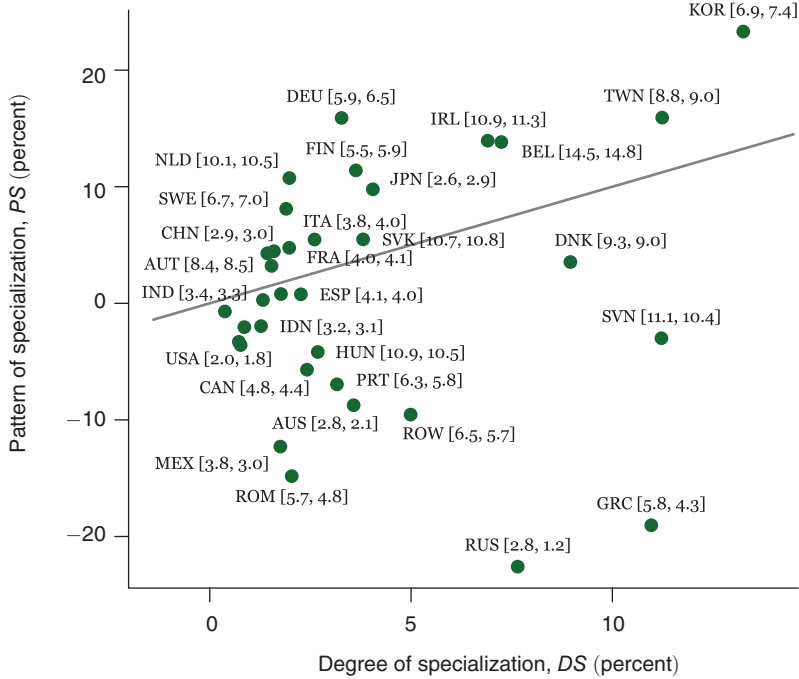


FIGURE 2. DEGREE OF SPECIALIZATION, PATTERN OF SPECIALIZATION, AND GAINS FROM TRADE IN THE PRESENCE OF ECONOMIES OF SCALE

Notes:  $DS$  and  $PS$  are calculated according to expressions (18) and (20), correspondingly. The numbers in parentheses are, respectively, the standard ACR gains,  $GT^{ACR}$ , and the gains from trade in the presence of economies of scale,  $GT^{scale}$ , both calculated according to (17).  $PS$ ,  $GT^{ACR}$ , and  $GT^{scale}$  are calculated by setting  $\varepsilon_k$  for manufacturing industries based on estimates from BCDR and setting  $\varepsilon_k = 0$  for nonmanufacturing industries. For  $GT^{ACR}$   $\psi_k = 0$  for all  $k$ . For  $PS$  and  $GT^{scale}$   $\psi_k$  for manufacturing industries are set based on estimates from BCDR, and  $\psi_k = 0$  for nonmanufacturing industries. The straight line is the 45-degree line. To avoid cluttering, some country labels are dropped.

The role of  $DS$  and  $PS$  in affecting the gains from trade is illustrated in Figure 2. The figure is a scatter diagram of  $DS$  (on the horizontal axis) against  $PS$  (on the vertical axis) for all countries in our sample, with most points also indicating the name of the country and the gains from trade without and with economies of scale, as in columns 1 and 2 of Table 2. More than half of the countries lie below the 45-degree line, indicating that  $PS < DS$ , so that economies of scale decrease the gains from trade. However, in addition to Germany and South Korea, there are a few countries (for example, Belgium, China, Finland, Japan, and Sweden) with  $PS > DS$ , implying higher gains from trade with economies of scale than without, thanks to a pattern of specialization tilted in favor of sectors with relatively high scale elasticities.

### V. Scale Economies and the Gains from Trade Liberalization

Whereas the previous section was devoted to the gains from trade using autarky as the counterfactual, we now study the gains from a decline in trade barriers, focusing on how these gains are affected by scale economies. We again split the section in two parts: theory and quantitative analysis.

### A. Theoretical Analysis

We consider two simple cases for which we can derive analytical results for the gains from a decline in trade costs. Both cases consist of two countries and two industries under conditions implying that relative wages are not affected by the trade shock.

*Mirror-Image Countries.*—With two mirror-image countries the wage is equalized, and we can normalize both wages to one. For ease of exposition we index countries by  $i = H, F$ , where  $H$  and  $F$  represent Home and Foreign, respectively. Let  $\bar{L} = 2$ ,  $\beta_{i,k} = 1/2$  for all  $(i, k)$ , and let  $S_{H,1} = S_{F,2} = 2$  and  $S_{H,2} = S_{F,1} = 1$ . Hence, Home has the comparative advantage in industry 1 and Foreign in industry 2. We assume that  $\varepsilon_k = \varepsilon$  and  $\psi_k = \psi$  for  $k = 1, 2$ .

To establish a link with the results of Section IV, we first illustrate that the gains from trade are decreasing in  $\psi$ . We then show that the conclusion is reversed once we consider a trade liberalization exercise in which trade shares respond endogenously as we lower trade costs. There we find that the gains from trade liberalization are increasing in  $\psi$ .

Home's gains from trade are simply

$$GT_H = 1 - \left( \frac{1/2}{r_{H,1}} \cdot \frac{1/2}{1 - r_{H,1}} \right)^{\psi/2} \left( \lambda_{HH,1}^{1/2\varepsilon} \cdot \lambda_{HH,2}^{1/2\varepsilon} \right).$$

The term  $\left( \frac{1/2}{r_{H,1}} \cdot \frac{1/2}{1 - r_{H,1}} \right)^{\psi/2}$  corresponds to  $\Delta_H$  in equation (17) and is higher than one as long as there is industry-level specialization.<sup>31</sup> Thus, given trade shares, gains are lower with scale effects ( $\psi > 0$ ) than without ( $\psi = 0$ ). It is also easy to see that these gains are decreasing in  $\psi$ .

Next we study the gains from trade liberalization, allowing for endogenous responses of trade shares to trade costs.<sup>32</sup> We use the autarky economy (i.e., the economy with  $\tau = \infty$ ) as the baseline case and define gains from trade liberalization as welfare changes resulting from the move from autarky to an economy with a finite level of trade costs. We set  $\varepsilon = 5$  and, consistently with the exercises of Section IV,  $\psi \in \{0, 0.14, 0.2\}$ , which corresponds to  $\alpha \in \{0, 0.7, 1\}$ . In all these cases  $L_{i,k} = 1$  for all  $(i, k)$  under autarky. As  $\tau$  falls from  $\infty$ , country  $H$  specializes in industry 1 and country  $F$  specializes in industry 2, but the extent of specialization will be stronger with  $\psi = 0.2$  than  $\psi = 0.14$  and with  $\psi = 0.14$  than  $\psi = 0$ , as illustrated in Figure 3, panel A. Figure 3, panel B shows the implications for the gains from trade liberalization for each of these three cases. We see that the gains from trade liberalization increase with  $\psi$ . The intuition is simple: countries gain by specializing according to comparative advantage, and the concentration of

<sup>31</sup> The term  $\frac{1/2}{r_{H,1}} \cdot \frac{1/2}{1 - r_{H,1}}$  is minimized at  $r_{H,1} = 1/2$ , and specialization according to comparative advantage implies  $r_{H,1} > 1/2$ ; hence, we must have  $\left( \frac{1/2}{r_{H,1}} \cdot \frac{1/2}{1 - r_{H,1}} \right)^{\psi/2} > 1$ .

<sup>32</sup> In the terminology of ACR, this corresponds to an “ex ante analysis,” whereas the results for the gains from trade above correspond to an “ex post analysis.”

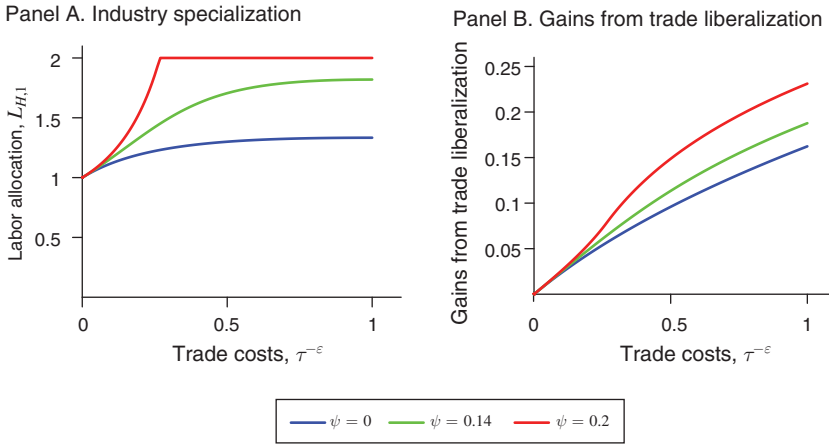


FIGURE 3. EFFECT OF ECONOMIES OF SCALE

production also allows for a greater exploitation of scale economies, which, in turn, generates additional efficiency gains.<sup>33</sup>

*Freely Traded Outside Good.*—The result established in Proposition 3—that if the product of the trade and scale elasticities is weakly lower than one in all industries, countries always gain from trade—does not necessarily imply that there are always gains from further trade liberalization. In fact, our model nests the one in Venables (1987), and so we know that a unilateral decline in inward trade costs may decrease welfare. To see this more explicitly, consider a case with two countries and two industries, with  $\varepsilon_1 = \infty > \varepsilon_2$ ,  $\psi_1 = 0 < \psi_2 \leq 1/\varepsilon_2$  (so that  $\alpha_2 \leq 1$ ), and with no trade costs in industry 1 (i.e.,  $\tau_{12,1} = \tau_{21,1} = 1$ ). If we start with an interior equilibrium (i.e.,  $L_{i,k} > 0$  for  $i = 1, 2$  and  $k = 1, 2$ ), then wages are pinned down by (exogenous) productivities in industry 1 (the outside good), and—suppressing the industry subindex—the labor allocation in industry 2 is given by  $(L_1, L_2)$  that solves

$$(21) \quad w_i L_i = \sum_n S_i L_i^\alpha (w_i \tau_{ni})^{-\varepsilon} P_n^\varepsilon \beta_n w_n \bar{L}_n$$

for  $i = 1, 2$ , with  $P_n^{-\varepsilon} = \sum_j S_j L_j^\alpha (w_j \tau_{nj})^{-\varepsilon}$ . The case considered by Venables (1987) entails  $\alpha = 1$ , in which case the previous system can be rewritten as a system in  $(P_1, P_2)$ ,

$$w_i = \sum_n S_i (w_i \tau_{ni})^{-\varepsilon} P_n^\varepsilon \beta_n w_n \bar{L}_n.$$

<sup>33</sup>To understand this further, note that the gains from trade liberalization can be seen as the increase in welfare  $W_H = 1/(1 - GT_H) = \left(\frac{1/2}{r_{H,1}} \cdot \frac{1/2}{1 - r_{H,1}}\right)^{-\psi/2} \left(\lambda_{HH,1}^{-1/2\varepsilon} \cdot \lambda_{HH,2}^{-1/2\varepsilon}\right)$  as  $\tau$  falls. The decline in  $\tau$  leads to deeper industry-level specialization, as captured by a higher  $r_{H,1}$ , and this decreases  $\left(\frac{1/2}{r_{H,1}} \cdot \frac{1/2}{1 - r_{H,1}}\right)^{-\psi/2}$  and decreases  $W_H$ . But there is also a change in trade shares, and this increases  $\lambda_{HH,1}^{-1/2\varepsilon} \cdot \lambda_{HH,2}^{-1/2\varepsilon}$ , which more than offsets the previous effect.

In this case it is easy to see that a decline in  $\tau_{12}$  leads to an increase in  $P_1$  and a decrease in  $P_2$ , exactly as in Venables (1987). Of course, if  $\alpha = 0$ , then  $P_n^{-\varepsilon} = \sum_j S_j^\alpha (w_j \tau_{nj})^{-\varepsilon}$ , and so  $P_1$  would decrease while there would be no change in  $P_2$ , implying further gains from unilateral trade liberalization. More generally, we can prove the following proposition (see Section 5.1 of the online Appendix for the details):

**PROPOSITION 4:** *Assume there are two countries and two industries, with industry 1 playing the role of an outside good (i.e.,  $\varepsilon_1 = \infty, \psi_1 = 0$  and  $\tau_{12,1} = \tau_{21,1} = 1$ ) and industry 2 having scale economies with  $\alpha > 0$ . Assume that the initial equilibrium is interior (i.e.,  $L_{i,k} > 0$  for  $i = 1, 2$  and  $k = 1, 2$ ). There exists a threshold  $\bar{\alpha}^{n,\tau} \in (0, 1)$ —depending on import and export shares—such that country  $n$  loses from a small unilateral trade liberalization in industry 2 if and only if  $\alpha \in (\bar{\alpha}^{n,\tau}, 1]$ .<sup>34</sup>*

This proposition generalizes the result of immiserizing inward trade liberalization in Venables (1987) in two ways. First, the result holds outside of the Krugman model. What is needed is that the scale economies vary across sectors, but the source of such economies (for example, love of variety or external economies of scale) is irrelevant. Second, the results are a manifestation of the more general idea that a shock that pushes a country to specialize in an industry with weak economies of scale (here, the outside good) may lower the gains from trade. Using the notion of the pattern of specialization ( $PS$ ) introduced in the previous section, we can restate this as implying that if a shock leads to a decline in a country's  $PS$ , then the gains from trade liberalization may be lower with economies of scale than without.

### B. Quantitative Analysis

The results of the previous subsection imply that scale economies tend to increase the gains from trade liberalization except if they lead to a decline in the pattern of specialization. We illustrate the relevance of this finding by quantifying the welfare effects of removing all observed tariffs across the countries and sectors in the WIOD for the year 2008. We use data on ad valorem tariffs for the year 2008 from the United Nations Statistical Division, Trade Analysis and Information System (UNCTAD-TRAINS, n.d.) as described in detail in Appendix C2. As in Section IV, we take the trade and scale elasticities from BCDR. To solve for counterfactual equilibria, we use the exact hat algebra approach popularized by Dekle, Eaton, and Kortum (2008) but extended here to allow for complementary slackness conditions, as described in Section 5.2 of the online Appendix.

The results of this exercise are presented in Table 3. Columns 1 and 2 report the welfare gains for the cases without and with economies of scale. Welfare gains in these columns are calculated explicitly as the change in real income, which is

<sup>34</sup>In Section 5.1 of the online Appendix, we also show that there exists a threshold  $\bar{\alpha}^{n,S} \in (0, 1)$  such that country  $n$  loses from a small foreign productivity improvement in industry 2 if and only if  $\alpha \in (\bar{\alpha}^{n,S}, 1]$ .

TABLE 3—EFFECTS OF MULTILATERAL ELIMINATION OF TARIFFS

Country	Gains from elimination of tariffs, %		ACR GTL diff., p.p.	$\hat{D}S$ , %	$\hat{P}S$ , %	Country	Gains from elimination of tariffs, %		ACR GTL diff., p.p.	$\hat{D}S$ , %	$\hat{P}S$ , %
	No scale	Scale					No scale	Scale			
	(1)	(2)					(1)	(2)			
AUS	0.19	0.34	0.51	4.30	-2.43	IRL	0.38	0.52	0.40	3.39	-1.42
AUT	0.15	0.24	0.17	1.05	-0.37	ITA	0.11	0.16	0.02	0.12	0.74
BEL	0.14	0.25	0.21	1.03	-0.59	JPN	0.33	0.42	0.80	6.71	-6.15
BRA	0.06	0.02	0.15	1.78	-1.73	KOR	0.63	0.83	0.48	5.95	1.04
CAN	0.15	0.13	0.55	3.80	-6.73	MEX	-0.03	-0.09	0.18	1.86	-2.37
CHN	0.12	0.38	0.39	2.22	-0.04	NLD	0.22	0.34	0.13	1.24	1.05
CZE	0.10	0.13	0.05	0.56	0.24	POL	0.11	0.19	0.09	0.62	0.47
DEU	0.22	0.30	0.09	0.59	0.50	PRT	0.08	0.11	0.08	0.25	-0.67
DNK	0.22	0.46	0.48	4.01	-0.26	ROM	0.05	0.12	0.04	0.28	0.85
ESP	0.06	0.11	0.03	0.09	0.42	RUS	-0.16	-0.32	2.87	23.21	-31.53
FIN	0.28	0.36	0.18	1.69	-0.14	SVK	0.14	0.24	0.18	2.08	0.64
FRA	0.13	0.20	0.03	0.15	0.84	SVN	0.15	0.32	0.25	2.04	0.75
GBR	0.10	0.21	0.11	0.64	0.52	SWE	0.16	0.31	0.14	1.11	1.27
GRC	0.09	0.24	0.42	1.72	-3.30	TUR	0.12	0.23	0.11	0.89	0.96
HUN	0.08	0.13	-0.04	-0.21	1.39	TWN	1.05	0.72	1.30	14.07	-15.44
IDN	0.19	0.32	0.17	1.08	0.45	USA	0.07	0.11	0.26	2.19	-2.01
IND	0.02	0.01	0.23	1.09	-3.22	ROW	0.12	0.10	0.68	4.44	-8.23
						Average	0.17	0.24	0.35	2.82	-2.19

Notes: Columns 1 and 2 present results on welfare gains of a separate counterfactual exercise that removes all tariffs observed in the data. For columns 1 and 2, values of  $\varepsilon_k$  for manufacturing industries are set based on values provided by BCDR, and for nonmanufacturing industries,  $\varepsilon_k = 4$ . For column 1  $\psi_k = 0$  for all  $k$ . For column 2  $\psi_k$  for manufacturing industries are set based on estimates from BCDR, and for nonmanufacturing industries,  $\psi_k = 0$ . BCDR-based values of  $\varepsilon_k$  and  $\psi_k$  are provided in Appendix Table C2. Column 3 shows the difference in percentage points of ACR welfare gains between the cases in columns 2 and 1, correspondingly.  $\hat{P}S$  in column 4 are calculated for the exercises in column 2. Average values in the last row are calculated based on the full set of countries.

given by the sum of labor income and tariff revenues deflated by the price index. Comparing columns 1 and 2 reveals that multilateral liberalization tends to generate larger gains with economies of scale than without: average gains increase from 0.17 to 0.24 percent. However, there are large differences, with some countries more than doubling the gains (for example, China, the United Kingdom, and Greece) and others exhibiting significantly lower gains (for example, Brazil, Mexico, and Russia).

There are two main forces behind these results. The first one is that, as discussed in Section VA, the presence of economies of scale tends to amplify the effect of trade liberalization on specialization and trade, leading to larger gains from trade. We capture this force by computing the difference in the gains from trade liberalization with and without economies of scale using the ex post ACR welfare formula,  $\prod_k \hat{\lambda}_{nn,k}^{-\beta_{nn,k}/\varepsilon_k} - 1$ . This ignores tariff revenues and the role of scale economies on domestic productivity and isolates the effect of changes in domestic trade shares on welfare. The results are shown in column 3. On average, economies of scale lead to an increase in the gains from trade liberalization by 0.35 percentage points.

The second force is that, because of scale economies, the expansion and contraction of industries induced by trade liberalization affects productivity and hence

welfare. As explained in Section IVA, we capture this by the change in  $DS$  and  $PS$ . In fact, a very good approximation of the difference in the gains from trade liberalization in columns 1 and 2 is given by the value in column 3 minus the average scale elasticity  $\bar{\psi}$  times the difference of columns 4 and 5:  $c2 - c1 \approx c3 - \bar{\psi}(c4 - c5)$ , with  $\bar{\psi} = 0.05$ . Thus, for example, the gains from trade liberalization in Brazil are lower once we allow for scale economies in spite of a large positive value in column 3—which, by itself, would imply that those gains would be 0.15 percent higher—because of the large increase in  $DS$  and decline in  $PS$ , with  $0.15 - 0.05(1.78 + 1.73) = -0.03$ , which is close to the difference between columns 1 and 2. The same forces explain lower gains from trade liberalization in Canada, India, Mexico, Russia, and Taiwan.

## VI. Scale Economies and Trade Flows

In this section we quantify the role of scale economies in determining industry-level specialization and trade flows. In particular, we ask how these variables would change if we shut down scale economies but left all other exogenous variables unchanged.

We rely on the fact that if  $\mathbf{L}$  is an equilibrium of the actual economy with scale economies, then it is also an equilibrium of the economy with no scale economies given by

$$(22) \quad w_i L_{i,k} = \sum_{n=1}^N \frac{T_{i,k}(w_i \tau_{ni,k})^{-\varepsilon_k}}{\sum_{l=1}^N T_{l,k}(w_l \tau_{nl,k})^{-\varepsilon_k}} \beta_{n,k} (w_n \bar{L}_n + D_n)$$

and

$$(23) \quad \sum_{k=1}^K L_{i,k} = \bar{L}_i,$$

where  $T_{i,k} \equiv S_{i,k} L_{i,k}^{\alpha_k}$  and where  $D_n$  are trade deficits satisfying  $\sum_n D_n = 0$ . Thus, if we want to know the counterfactual allocation for the economy with  $\alpha_k = 0$  for all  $k$  but everything else equal, we can use the exact hat algebra approach in the economy with no scale effects subjected to a shock to productivities  $T_{i,k}$  given by  $\hat{T}_{i,k} = S'_{i,k} / (S_{i,k} L_{i,k}^{\alpha_k}) = \hat{S}_{i,k} L_{i,k}^{-\alpha_k}$ .<sup>35</sup> To focus on the interaction between specialization and scale economies, we assume that  $\hat{S}_{i,k}$  is such that if country  $i$  was in autarky, then the shock would have no effect on productivity. Since in autarky  $L_{i,k} = \beta_{i,k} \bar{L}_i$ , this requires  $\hat{S}_{i,k} = (\beta_{i,k} \bar{L}_i)^{\alpha_k}$ . Using  $L_{i,k} = r_{i,k} \bar{L}_i$  and measuring  $\beta_{i,k}$  by  $e_{i,k}$ , this implies that  $\hat{T}_{i,k} = (e_{i,k}/r_{i,k})^{\alpha_k}$ . Using (22) and (23) we can derive a system in wage changes. The solution for  $\hat{w}_i$  can then be used to get the implied changes in labor allocations  $\hat{L}_{i,k}$  and, by extension, the implied changes in trade flows  $\hat{X}_{ni,k}$ . The details of this derivation are in Section 5.3 of the online Appendix.<sup>36</sup>

<sup>35</sup>We can ignore corner solutions because the data has no zeros at the industry level (i.e.,  $r_{i,k} > 0$  for all  $i, k$ ), and the shock that we consider moves us away from corners.

<sup>36</sup>In order to solve for  $\hat{w}_i$ , we only need to work with the model for  $\alpha_k = 0$  for all  $k$  (as in the standard Eaton and Kortum 2002 model), and so for this case, the algorithm that solves for  $\hat{w}_i$  is just a straightforward extension of the Alvarez and Lucas (2007) algorithm to multiple sectors.

TABLE 4—SCALE AND TRADE FLOWS

Country	$\hat{DS}, \%$ (1)	$\hat{EX}, \%$ (2)	Country	$\hat{DS}, \%$ (1)	$\hat{EX}, \%$ (2)	Country	$\hat{DS}, \%$ (1)	$\hat{EX}, \%$ (2)
AUS	-1.49	-3.84	GBR	-0.28	-0.46	PRT	-0.65	1.15
AUT	-0.38	-0.13	GRC	-2.85	-2.61	ROM	-0.64	0.98
BEL	-0.76	-1.17	HUN	-1.12	-2.39	RUS	-3.50	-9.37
BRA	-0.14	-3.23	IDN	-0.53	-2.03	SVK	-0.96	-1.89
CAN	-0.83	0.73	IND	-0.13	-2.22	SVN	-9.85	-0.46
CHN	-0.39	-3.73	IRL	-4.55	-4.40	SWE	-0.70	-2.06
CZE	-0.69	-1.96	ITA	-0.42	-2.65	TUR	-0.66	-6.81
DEU	-0.70	-2.79	JPN	-0.53	-5.57	TWN	-2.50	-8.16
DNK	-2.08	-0.37	KOR	-1.21	-5.65	USA	-0.19	-1.60
ESP	-0.23	-0.64	MEX	-0.79	-2.39	ROW	-1.03	-2.15
FIN	-1.23	-4.83	NLD	-1.04	-2.53			
FRA	-0.17	-1.77	POL	-0.42	-1.07			
Average							-1.28	-2.59

Notes: Columns 1 and 2 present the results of an exercise in which it is assumed that the data is generated by the model with  $\varepsilon_k$  and  $\psi_k$  for manufacturing industries based on estimates from BCDR (provided in Appendix Table C2), and  $\varepsilon_k = 4$  and  $\psi_k = 0$  for nonmanufacturing industries. In the exercise  $\psi_k$  for all  $k$  are set to zero, and all productivities are adjusted so that in autarky the changes in  $\psi_k$  would have no effect. All other parameters are held unchanged.  $\hat{DS}$  and  $\hat{EX}$  are the implied changes in the degree of specialization and total exports in terms of the world GDP. Average values in the last row are calculated based on the full set of countries.

Table 4 presents results of this exercise. For this exercise, we again take the values of trade and scale elasticities from BCDR. Columns 1 and 2 report the implied change in  $DS$  and in total exports for each country in the model. As expected, the removal of economies of scale implies a decline in the degree of specialization and total trade, but the effects are small. Without economies of scale amplifying Ricardian comparative advantage, industry-level specialization would be, on average, 1.3 percent smaller than what we observe in the data, implying that scale economies are much less important than Ricardian comparative advantage in driving industry-level specialization. In addition, without economies of scale, total world exports would be about 3 percent lower than what we see in the data (this number is close to the simple average decline in exports of 2.6 percent provided in Table 4).

## VII. Concluding Remarks

For over a century since Alfred Marshall's initial exposition, economists have been intrigued by the implications of industry-level external economies of scale for trading economies. Despite such interest, however, the discomfort with the plethora of equilibria and counterintuitive implications in early work implied that these externalities were mostly ignored in the recent trade literature. By leveraging the workhorse Armington (1969) framework, we show how one can tractably model external economies of scale together with Ricardian forces and put it to work within a quantitative setting. The resulting model has exactly the same mathematical structure as generalized versions of the multi-industry Krugman and Melitz-Pareto models, and so our results apply to these well-known models, as well. The condition for the model to be well behaved is simple and intuitive: the scale elasticity must be weakly lower than the inverse of the trade elasticity for all industries.

The model is rich in welfare implications. Most importantly, if the the product of the trade and scale elasticities is weakly lower than one in all industries, then all countries gain from trade. Economies of scale tend to make gains from trade lower and gains from trade liberalization higher relative to models without scale economies, with results more positive for countries that specialize in industries with stronger than average scale economies.

Motivated by the multi-industry gravity model with scale economies developed in this paper, BCDR have proposed a way to estimate industry-level scale elasticities and to quantify the consequences for industrial policy. Their estimates indicate that scale economies are positive but lower than the inverse of the trade elasticity, with substantial variation across manufacturing industries. We have explored the implications of these estimated elasticities for the gains from trade relative to autarky and for the gains from removal of all tariffs observed in the data for the countries in the WIOD. We find that the presence of economies of scale implies, on average, slightly lower gains from trade relative to autarky and significantly larger gains from the removal of all tariffs, which more than double for several countries while falling for others.

We also found that while economies of scale do play a role in explaining trade flows and industry-level specialization in the data, they pale in comparison to the importance of Ricardian comparative advantage: shutting down economies of scale would imply a decline in industry-level specialization of about 1.3 percent, with an implied fall in world exports of 3 percent.

#### APPENDIX A. A COMMON FRAMEWORK

In this appendix we derive expressions for price indices and trade shares for the multi-industry versions of Krugman (1980), Melitz (2003), and Eaton and Kortum (2002) models. In all models below variable trade costs  $\tau_{ni,k}$  are as in the baseline Armington model in Section I.

##### A1. A Krugman Model with Two-Tier CES Preferences

There is a continuum of differentiated varieties within each industry. Preferences are multitiered: Cobb-Douglas across industries with weights  $\beta_{i,k}$ , CES across country bundles within an industry with elasticity  $\eta_k$ , and CES across varieties within a country bundle with elasticity of substitution  $\sigma_k > 1$ . The corresponding demand function for a representative variety in  $n$  from  $i$  in industry  $k$  is

$$q_{ni,k} = p_{ni,k}^{-\sigma_k} (P_{ni,k})^{\sigma_k - 1} \left( \frac{P_{ni,k}}{P_{n,k}} \right)^{-(\eta_k - 1)} \beta_k X_n,$$

where  $p_{ni,k}$  is the price of the variety,  $P_{ni,k} = M_{i,k}^{1/(1-\sigma_k)} p_{ni,k}$  is the price index in country  $n$  of country  $i$  varieties of industry  $k$ ,  $M_{i,k}$  is the measure of varieties produced (or equivalently the measure of firms) in  $(i, k)$ , and  $P_{n,k} = \left( \sum_i P_{ni,k}^{1-\eta_k} \right)^{1/(1-\eta_k)}$  is the industry price index in  $n$ .

Let  $A_{i,k}$  be the exogenous productivity in  $(i,k)$  that is common across firms in that industry, and let  $F_{i,k}$  denote the fixed cost (in terms of labor) associated with the production of any variety in  $(i,k)$ . Monopolistic competition implies the price index in country  $n$  of country  $i$  varieties of industry  $k$  can be rewritten as

$$P_{ni,k} = M_{i,k}^{1/(1-\sigma_k)} (\bar{\sigma}_k w_i \tau_{ni,k} / A_{i,k}),$$

where  $\bar{\sigma}_k \equiv \sigma_k / (\sigma_k - 1)$  is the markup. Denoting the revenue of the representative firm from  $k$  in  $i$  by  $R_{i,k} \equiv \sum_n p_{ni,k} q_{ni,k}$ , the corresponding profit (gross of fixed costs) is then given by  $R_{i,k} / \sigma_k$ . Free entry then implies complementary slackness condition

$$(A1) \quad \frac{R_{i,k}}{\sigma_k} - F_{i,k} w_i \leq 0 \perp M_{i,k} \geq 0.$$

Analogous to the Armington case with EES, total employment  $L_{i,k}$  must also be consistent with the amounts produced for each market plus the fixed cost of entry,

$$L_{i,k} = \sum_n q_{ni,k} \tau_{ni,k} / A_{i,k} + M_{i,k} F_{i,k},$$

and labor markets must clear,  $\sum_k L_{i,k} = L_i$ .

Note that we can express the gross profits of the representative firm as  $R_{i,k} / \sigma_k = M_{i,k}^{-1} \sum_n \lambda_{ni,k} X_{n,k} / \sigma_k$  where

$$\lambda_{ni,k} = \left( \frac{P_{ni,k}}{P_{n,k}} \right)^{-(\eta_k - 1)} = \frac{A_{i,k}^{\eta_k - 1} M_{i,k}^{(\eta_k - 1) / (\sigma_k - 1)} (w_i \tau_{ni,k})^{1 - \eta_k}}{\sum_l A_{l,k}^{\eta_k - 1} M_{l,k}^{(\eta_k - 1) / (\sigma_k - 1)} (w_l \tau_{nl,k})^{1 - \eta_k}} \beta_{n,k} w_n \bar{L}_n$$

are industry-level trade shares. Substituting this expression for gross profits into (A1) and rearranging yields

$$M_{i,k}^{\frac{\eta_k - 1}{\sigma_k - 1} - 1} \frac{1}{\sigma_k} \sum_n \frac{A_{i,k}^{\eta_k - 1} (w_i \tau_{ni,k})^{1 - \eta_k}}{P_{n,k}^{-(\eta_k - 1)}} \beta_{n,k} w_n \bar{L}_n - F_{i,k} w_i \leq 0 \perp M_{i,k} \geq 0.$$

Here it becomes immediately evident that at a corner equilibrium,  $M_{i,k} = 0$ , the condition is not well defined for  $(\eta_k - 1) / (\sigma_k - 1) < 1$ , again motivating the need to evaluate the limit of this expression. In particular, we define  $R_{i,k} / \sigma_k - F_{i,k} w_i$  in the case of  $L_{i,k} = 0$  by the limit  $L_{i,k} \rightarrow 0$  for any fixed vector of wages  $\mathbf{w} = (w_1, \dots, w_N)$ .

To complete the link to the common framework, we now solve for the equilibrium measure of varieties  $M_{i,k}$  as a function of industry employment  $L_{i,k}$  and then use the result to derive an expression for trade shares in terms of labor allocations. Letting  $\Pi_{i,k}$  be total profits net of fixed costs in industry  $(i,k)$ , we then have  $\Pi_{i,k} = \sum_n \lambda_{ni,k} X_{n,k} / \sigma_k - w_i M_{i,k} F_{i,k}$ . If  $L_{i,k} > 0$ , then free entry implies zero profits, so total revenues must equal total wage payments in industry  $(i,k)$ ,  $\sum_n \lambda_{ni,k} X_{n,k} = w_i L_{i,k}$ . Combined with  $\Pi_{i,k} = 0$ , we then have

$$(A2) \quad M_{i,k} = L_{i,k} / (\sigma_k F_{i,k}).$$

Trade shares are then

$$(A3) \quad \lambda_{ni,k} = \frac{A_{i,k}^{\eta_k-1} F_{i,k}^{-\frac{\eta_k-1}{\sigma_k-1}} L_{i,k}^{\frac{\eta_k-1}{\sigma_k-1}} (w_i \tau_{ni,k})^{-(\eta_k-1)}}{\sum_l A_{l,k}^{\eta_k-1} F_{l,k}^{-\frac{\eta_k-1}{\sigma_k-1}} L_{l,k}^{\frac{\eta_k-1}{\sigma_k-1}} (w_l \tau_{nl,k})^{-(\eta_k-1)}}$$

with price indices given by

$$P_{n,k} = \mu_k^{Krug} \left[ \sum_l A_{l,k}^{\eta_k-1} F_{l,k}^{-\frac{\eta_k-1}{\sigma_k-1}} L_{l,k}^{\frac{\eta_k-1}{\sigma_k-1}} (w_l \tau_{nl,k})^{-(\eta_k-1)} \right]^{-1/(\eta_k-1)},$$

where  $\mu_k^{Krug} = \sigma_k^{-\frac{1}{\sigma_k-1}} \bar{\sigma}_k$ . Using the (A2) and (A3) together with the limit definition of  $R_{i,k}/\sigma_k - F_{i,k} w_i$  at  $L_{i,k} = 0$ , we can rewrite complementary slackness condition (A1) as

$$w_i - \frac{1}{L_{i,k}} \sum_n \lambda_{ni,k} \beta_k X_n \geq 0 \perp L_{i,k} \geq 0,$$

which, together with labor market clearing, fully characterizes equilibria in this Krugman setting. It is then immediately evident that the expressions for trade shares and industry price indexes in this Krugman model collapse to those in equations (12) and (13) by setting  $\mu_{n,k} = \mu_k^{Krug}$ ,  $S_{i,k} = A_{i,k}^{\eta_k-1} F_{i,k}^{-\frac{\eta_k-1}{\sigma_k-1}}$ ,  $\psi_k = (\sigma_k - 1)^{-1}$ , and  $\varepsilon_k = (\eta_k - 1)$ . Note also that if we set  $\sigma_k = \eta_k$  for all  $k$ , then this is just the standard multi-industry Krugman model, while if  $\sigma_k \rightarrow \infty$ , then  $(\eta_k - 1)/(\sigma_k - 1) \rightarrow 0$  and we obtain the multi-industry Armington model without external economies of scale. See rows 2 and 4 of Table 1.

Finally, given the expressions for the price indices and trade shares derived above, the equilibrium conditions are given by the same expressions as in the Armington model with EES: goods-market-clearing conditions (10)–(11) and labor-market-clearing conditions (6).

### A2. A Melitz-Pareto Model with Two-Tier Preferences

We now present a model à la Melitz (2003) with Pareto-distributed productivity and with preferences exactly the same as the as in the Krugman model above.

After paying a fixed “entry” cost  $F_{i,k}$  in units of labor in country  $i$ , firms are able to produce a variety in industry  $(i, k)$  with labor productivity drawn from a Pareto distribution with shape parameter  $\theta_k > \sigma_k - 1$  and location parameter  $b_{i,k}$ . Firms from  $i$  can then pay a fixed “marketing” cost  $f_{n,k}$  in units of labor of  $n$  to serve that market.

Let  $\Omega_{ni,k}$  denote the set of varieties that  $i$  sells to  $n$  in industry  $k$ . The price index of these goods is  $P_{ni,k} \equiv \left( \int_{\omega \in \Omega_{ni,k}} p_{ni,k}(\omega)^{1-\sigma_k} d\omega \right)^{\frac{1}{1-\sigma_k}}$ . Let  $M_{i,k}$  denote total entry in

industry  $(i, k)$  and  $\varphi_{ni,k}^*$  denote the cutoff productivity such that  $i$  exports to  $n$  all goods with productivity higher than  $\varphi_{ni,k}^*$ . We have

$$\begin{aligned} P_{ni,k}^{1-\sigma_k} &= M_{i,k} \int_{\varphi_{ni,k}^*}^{\infty} [p_{ni,k}(\varphi)]^{1-\sigma_k} dG_{i,k}(\varphi) \\ &= \theta_k b_{i,k}^{\theta_k} M_{i,k} [\bar{\sigma}_k w_i \tau_{ni,k}]^{1-\sigma_k} \int_{\varphi_{ni,k}^*}^{\infty} \varphi^{\sigma_k - \theta_k - 2} d\varphi \\ &= \frac{\theta_k b_{i,k}^{\theta_k}}{\theta_k - \sigma_k + 1} M_{i,k} [\bar{\sigma}_k w_i \tau_{ni,k}]^{1-\sigma_k} (\varphi_{ni,k}^*)^{\sigma_k - \theta_k - 1}, \end{aligned}$$

where  $\bar{\sigma}_k \equiv \sigma_k / (\sigma_k - 1)$  is the markup.

Letting  $M_{ni,k}$  be the measure of firms from  $i$  in industry  $k$  that serve market  $n$  with  $M_{i,k} = \sum_n M_{ni,k}$ , the condition that determines the cutoff  $\varphi_{ni,k}^*$  implies complementary slackness condition

$$(A4) \quad \frac{1}{\bar{\sigma}_k} \left( \frac{\bar{\sigma}_k w_i \tau_{ni,k}}{\varphi_{ni,k}^*} \right)^{1-\sigma_k} P_{ni,k}^{\sigma_k - 1} \left( \frac{P_{ni,k}}{P_{n,k}} \right)^{1-\eta_k} X_{n,k} - f_{n,k} w_n \leq 0 \perp M_{ni,k} \geq 0.$$

The condition that determines the cutoff  $\varphi_{ni,k}^*$  is

$$\frac{1}{\bar{\sigma}_k} \left( \frac{\sigma_k}{\sigma_k - 1} \cdot \frac{w_i \tau_{ni,k}}{\varphi_{ni,k}^*} \right)^{1-\sigma_k} P_{ni,k}^{\sigma_k - 1} \left( \frac{P_{ni,k}}{P_{n,k}} \right)^{1-\eta_k} X_{n,k} = w_n f_{n,k}.$$

This implies that<sup>37</sup>

$$\varphi_{ni,k}^* = \frac{\sigma_k}{\sigma_k - 1} \cdot \frac{w_i \tau_{ni,k}}{P_{ni,k}} \left( \frac{\sigma_k w_n f_{n,k}}{X_{n,k}} \right)^{\frac{1}{\sigma_k - 1}} \left( \frac{P_{ni,k}}{P_{n,k}} \right)^{\frac{1-\eta_k}{1-\sigma_k}}.$$

Using the well-known result that marketing costs associated with sales  $X_{ni,k}$  are  $\nu_k X_{ni,k}$  with  $\nu_k = \frac{\theta_k - (\sigma_k - 1)}{\sigma_k \theta_k}$ , we get that the profits associated with sales  $X_{ni,k}$  (net of production and marketing costs) are  $(1/\sigma_k - \nu_k) X_{ni,k}$ . Letting  $L_{i,k}$  represent total labor used for production and entry by industry  $(i, k)$  (so that  $w_i L_{i,k} = (1 - \nu_k) \sum_n X_{ni,k}$ ), then the free-entry condition is  $\frac{1/\sigma_k - \nu_k}{1 - \nu_k} w_i L_{i,k} = w_i F_{i,k} M_{i,k}$ , which implies that  $M_{i,k} = \frac{1}{1 + \theta_k} \cdot \frac{L_{i,k}}{F_{i,k}}$ <sup>38</sup>

Plugging expressions for  $\varphi_{ni,k}^*$  and  $M_{i,k}$  into the expression for  $P_{ni,k}$  above using  $X_{n,k} = \beta_{n,k} w_n \bar{L}_n$  and solving for  $P_{ni,k}$  yields

$$P_{ni,k}^{1-\eta_k} = \left[ \mu_{n,k}^{Mel} \right]^{-\theta_k \xi_k} b_{i,k}^{\theta_k \xi_k} F_{i,k}^{-\xi_k} L_{i,k}^{\xi_k} (w_i \tau_{ni,k})^{-\theta_k \xi_k} P_{n,k}^{\frac{1-\eta_k}{\sigma_k - 1} (\sigma_k - \theta_k - 1) \xi_k},$$

<sup>37</sup> This expression is valid as long as  $\varphi_{ni,k}^* \geq b_{i,k}$ . We assume that this inequality holds for all  $n, i$ , and  $k$ .

<sup>38</sup> Note that  $M_{i,k}$  is the measure of entering firms, i.e., both operating and nonoperating. Letting  $M_{ni,k}$  be the measure of firms from  $i$  in industry  $k$  that serve market  $n$ , it is straightforward to show that  $M_{ni,k} = M_{i,k} (b_{i,k} / \varphi_{ni,k}^*)^{\theta_k}$ .

where  $\xi_k \equiv 1/\left[1 + \theta_k\left(\frac{1}{\eta_k - 1} - \frac{1}{\sigma_k - 1}\right)\right]$ , and

$$\mu_{n,k}^{Mel} \equiv \frac{\sigma_k}{\sigma_k - 1} \left[ \frac{\theta_k}{\theta_k - \sigma_k + 1} \cdot \frac{1}{1 + \theta_k} \right]^{-\frac{1}{\theta_k}} \left( \frac{\sigma_k f_{n,k}}{\beta_{n,k} \bar{L}_n} \right)^{\frac{1}{\sigma_k - 1} - \frac{1}{\theta_k}}.$$

The expression for trade shares can then be derived by using  $P_{n,k}^{1-\eta_k} = \sum_i P_{ni,k}^{1-\eta_k}$  together with  $\lambda_{ni,k} = \left(\frac{P_{ni,k}}{P_{n,k}}\right)^{1-\eta_k}$ . This gives

$$\lambda_{ni,k} = \frac{b_{i,k}^{\theta_k \xi_k} F_{i,k}^{-\xi_k} L_{i,k}^{\xi_k} (w_i \tau_{ni,k})^{-\theta_k \xi_k}}{\sum_l b_{l,k}^{\theta_k \xi_k} F_{l,k}^{-\xi_k} L_{l,k}^{\xi_k} (w_l \tau_{nl,k})^{-\theta_k \xi_k}}.$$

Finally, combining  $P_{n,k}^{1-\eta_k} = \sum_i P_{ni,k}^{1-\eta_k}$  with the result above for  $P_{ni,k}$  yields price indices

$$P_{n,k} = \mu_{n,k}^{Mel} \left[ \sum_l b_{l,k}^{\theta_k \xi_k} F_{l,k}^{-\xi_k} L_{l,k}^{\xi_k} (w_l \tau_{nl,k})^{-\theta_k \xi_k} \right]^{-\frac{1}{\theta_k \xi_k}}.$$

Note that if we set  $\sigma_k = \eta_k$  for all  $k$ , then  $\xi_k = 1$ , and this model is just a multi-industry version of the Melitz-Pareto model in Arkolakis et al. (2008).

We get expressions (12) and (13) for trade shares and price indices by setting  $\mu_{n,k} = \mu_{n,k}^{Mel}$ ,  $S_{i,k} = b_{i,k}^{\theta_k \xi_k} F_{i,k}^{-\xi_k}$ ,  $\psi_k = 1/\theta_k$ , and  $\varepsilon_k = \theta_k \xi_k$ . The goods-market-clearing conditions are then given by (10)–(11). The labor-market-clearing conditions need to take into account that labor is used not only for production and entry (which are all included in  $L_{i,k}$ ) but also for paying marketing costs (which are not part of  $L_{i,k}$  according to our definition above):  $\sum_k L_{i,k} = \left[1 - \sum_k \beta_{i,k} \nu_k\right] \bar{L}_i$ . This expression differs from the labor-market-clearing condition (6) in the Armington model by the constant  $\sum_k \beta_{i,k} \nu_k$ . This difference is inconsequential for equilibrium analysis and comparative statics; therefore, in the analysis pertaining to the general model, we ignore this term.

The above derivations for the Melitz model rely on the balanced-trade assumption. Without this assumption, the multiplicative term for the price index and the term on the right-hand side of the labor-market-clearing condition would depend on trade deficits. This has implications for the comparative statics exercises that involve changing trade deficits. At the same time, the derivations for the Armington model with external economies of scale and for the Krugman model do not depend on the balanced-trade assumption.

### A3. An Eaton-Kortum Model with External Economies of Scale

Consider the multi-industry Eaton and Kortum (2002) model as in Costinot, Donaldson, and Komunjer (2012), extended to feature EES. Each industry is composed of a countable number of goods or varieties  $\omega \in \Omega \equiv \{1, \dots, +\infty\}$ .

Preferences are Cobb-Douglas across industries with weights  $\beta_{i,k}$  and CES across varieties within an industry with elasticity of substitution  $\sigma_k$ . Labor productivity for good  $\omega$  in industry  $(i, k)$  is  $z_{i,k}(\omega)L_{i,k}^{\phi_k}$ , where  $z_{i,k}(\omega)$  is an exogenous productivity parameter,  $L_{i,k}$  is the total labor allocated to industry  $(i, k)$ , and  $L_{i,k}^{\phi_k}$  captures EES in industry  $k$ . Exogenous productivity parameters  $z_{i,k}(\omega)$  are independently drawn from a Fréchet distribution with shape parameter  $\theta_k$  and scale parameter  $T_{i,k}$ .

Let  $p_{n,k}(\omega)$  be the price of industry- $k$  good  $\omega$  in country  $n$ , and let  $l_{ni,k}(\omega)$  represent the quantity of labor used in country  $i$  to produce industry- $k$  good  $\omega$  sold to country  $n$ . Given wages  $w_i$ , the equilibrium conditions for industry  $k$  are

$$(A5) \quad w_i \geq \frac{p_{n,k}(\omega)z_{i,k}(\omega)}{\tau_{ni,k}}L_{i,k}^{\phi_k} \perp l_{ni,k}(\omega) \geq 0 \text{ for all } \omega, i, n;$$

$$(A6) \quad \frac{p_{n,k}(\omega)^{1-\sigma_k}}{\int_0^1 p_{n,k}(\omega)^{1-\sigma_k} d\omega} \beta_{n,k} w_n \bar{L}_n = \sum_i w_i l_{ni,k}(\omega) \text{ for all } \omega, n;$$

$$(A7) \quad \sum_{n,\omega} l_{ni,k}(\omega) d\omega = L_{i,k} \text{ for all } i.$$

The first condition is the standard complementary slackness condition for the allocation of labor across the production of different goods and destinations, the second condition is that demand equals supply (in value) for good  $\omega$  by country  $n$ , and the third condition just follows from the definition of  $l_{ni,k}(\omega)$  and  $L_{i,k}$ . Combining these three sets of conditions for each industry  $k$  with the labor-market-clearing condition  $\sum_k L_{i,k} = \bar{L}_i$  gives us the full set of equilibrium conditions.

Observe that as long as there is at least one country  $i$  with positive labor allocations in industry  $k$ , conditions (A5)–(A7) imply that prices  $p_{i,k}(\omega)$  are finite for all  $\omega$  because country  $i$  can produce any good  $\omega$  at a finite cost and export it to any other country. This, in turn, implies that zero labor allocation in industry  $k$  in any other country is trivially consistent with conditions (A5) and (A7) and can thus potentially be an equilibrium.

Let  $\mathcal{N}_k \subseteq \{1, \dots, N\}$  denote a set of countries with positive labor allocations in industry  $k$ . Formally,  $L_{i,k} > 0$  if and only if  $i \in \mathcal{N}_k$ . Consider any collection of nonempty sets  $\mathcal{N}_k$  for  $k = 1, \dots, K$ .<sup>39</sup> Given  $\mathcal{N}_k$  for  $k = 1, \dots, K$ , we can apply the same argument as in Costinot, Donaldson, and Komunjer (2012) and use conditions (A5)–(A7) to derive the aggregate goods-market-clearing conditions (in value) in each industry  $k$ :

$$(A8) \quad G_{i,k}(\mathbf{w}, \mathbf{L}_k) = 0, \quad i \in \mathcal{N}_k$$

<sup>39</sup>There are  $(2^N - 1)^K$  such collections of sets.

where

$$(A9) \quad G_{i,k}(\mathbf{w}, \mathbf{L}_k) \equiv w_i - \frac{1}{L_{i,k}} \sum_n \lambda_{ni,k}(\mathbf{w}, \mathbf{L}_k) \beta_{n,k} w_n \bar{L}_n, \quad i \in \mathcal{N}_k,$$

and

$$(A10) \quad \lambda_{ni,k}(\mathbf{w}, \mathbf{L}_k) = \frac{T_{i,k} L_{i,k}^{\alpha_k} (\tau_{ni,k} w_i)^{-\theta_k}}{\sum_{l \in \mathcal{N}_k} T_{l,k} L_{l,k}^{\alpha_k} (\tau_{nl,k} w_l)^{-\theta_k}}, \quad n = 1, \dots, N, i \in \mathcal{N}_k$$

where  $\alpha_k = \phi_k \theta_k$ . The equilibrium in the Eaton-Kortum model is then characterized by a collection of nonempty sets  $\mathcal{N}_k$ , conditions (A8), and the labor-market-clearing conditions  $\sum_k L_{i,k} = \bar{L}_i$ .

Just as in the case of the Armington model in Section I, here function  $G_{i,k}$  goes either to infinity or a finite number depending on whether  $\alpha_k$  is smaller or larger than one. Intuitively, as the size of the industry becomes vanishingly small, the goods produced would be the ones for which the country's productivity is infinitely high, and this could overcome the Marshallian forces that push productivity to zero.<sup>40</sup> The difference is that in the Armington case,  $VMPL_{i,k}$  is not well defined when evaluated at  $L_{i,k} = 0$ , and so we defined it as  $w_i - \lim_{L_{i,k} \rightarrow 0} G_{i,k}$ ; whereas in the Eaton-Kortum case, as argued above,  $VMPL_{i,k}$  is trivially equal to zero when evaluated at  $L_{i,k} = 0$  as long as there is at least one country  $\ell$  with  $L_{\ell,k} > 0$ . Thus, in contrast to the Armington model, in the Eaton-Kortum model we need to exogenously assign which countries have positive labor allocations to arrive at the aggregate goods-market-clearing conditions (A8)–(A10).

It should be clear, however, that even though corner allocations can be equilibria in the Eaton-Kortum model when  $\alpha_k < 1$ , such equilibria are very peculiar in the following sense: given a wage  $w_i$ , the profit of hiring a worker in sector  $k$ ,  $-G_{i,k}$ , goes to infinity as employment in that sector becomes arbitrarily small; yet when evaluated directly at  $L_{i,k} = 0$ , the profit is  $-w_i$ . This is different from the typical case in which an equilibrium may be unstable—here, if  $\alpha_k < 1$ , then if  $L_{i,k}$  is very small, it is not just that the profits of hiring a worker in sector  $k$  in country  $i$  are positive; it is that they are very large.

## APPENDIX B. PROOF OF PROPOSITION 1

The case with  $\alpha_k = 0$  is trivial: given wages, labor allocations are explicitly obtained from the conditions  $L_{i,k} G_{i,k}(\mathbf{w}, \mathbf{L}_k) = 0$ . Below we focus on the case with  $\alpha_k > 0$ .

<sup>40</sup>The fact that the productivity of goods produced goes to infinity as the industry becomes vanishingly small is a standard trade-driven selection that arises in the multi-industry Eaton-Kortum model (see Costinot, Donaldson, and Komunjer 2012).

Consider a modified version of equations (12), (10), and (11) where we allow only countries from an exogenously given nonempty set  $\mathcal{N}_k \subseteq \{1, \dots, N\}$  to choose labor allocations. Formally, consider the following system of equations:

$$(B1) \quad \lambda_{ni,k}(\mathbf{w}, \mathbf{L}_k) = \frac{S_{i,k} L_{i,k}^{\alpha_k} (w_i \tau_{ni,k})^{-\varepsilon_k}}{\sum_{l \in \mathcal{N}_k} S_{l,k} L_{l,k}^{\alpha_k} (w_l \tau_{nl,k})^{-\varepsilon_k}}, \quad n = 1, \dots, N, i \in \mathcal{N}_k;$$

$$(B2) \quad G_{i,k}(\mathbf{w}, \mathbf{L}_k) = w_i - \frac{1}{L_{i,k}} \sum_{n=1}^N \lambda_{ni,k}(\mathbf{w}, \mathbf{L}_k) \beta_{n,k} w_n \bar{L}_n, \quad i \in \mathcal{N}_k;$$

$$(B3) \quad G_{i,k}(\mathbf{w}, \mathbf{L}_k) \geq 0 \perp L_{i,k} \geq 0, \quad i \in \mathcal{N}_k.$$

For brevity of notation suppress the subindex  $k$  and let  $a_{ni} \equiv S_i (w_i \tau_{ni})^{-\varepsilon} w_i^{-\alpha}$  and  $b_n \equiv \beta_n w_n \bar{L}_n$ . Combining equations (B1) and (B2), we then have

$$\frac{G_i(\mathbf{w}, \mathbf{L})}{w_i} = 1 - \frac{1}{w_i L_i} \sum_{n=1}^N \frac{a_{ni} (w_i L_i)^\alpha}{\sum_{l \in \mathcal{N}} a_{nl} (w_l L_l)^\alpha} b_n, \quad i \in \mathcal{N}_k.$$

Transforming variables with  $x_i \equiv w_i L_i$ , letting  $\mathbf{x}$  be the vector with indices from  $\mathcal{N}$ , and with a slight abuse of notation, we can write

$$G_i(\mathbf{x}) = 1 - \sum_{n=1}^N \frac{a_{ni} x_i^{\alpha-1}}{\sum_{l \in \mathcal{N}} a_{nl} x_l^\alpha} b_n, \quad i \in \mathcal{N}.^{41}$$

The system in (B3) can now be written as a nonlinear complementarity problem (NCP) in  $\mathbf{x}$ :

$$(B4) \quad G_i(\mathbf{x}) \geq 0 \perp x_i \geq 0, \quad i \in \mathcal{N}.$$

Note that if  $\mathbf{x}$  solves (B4), then  $\sum_{i \in \mathcal{N}} x_i G_i(\mathbf{x}) = 0$ ; hence,  $\sum_{i \in \mathcal{N}} x_i = \sum_{i=1}^N b_i$ . This implies that the solution to (B4) satisfies  $\mathbf{x} \in \Gamma \equiv \left\{ \mathbf{x} \in \mathbb{R}^{|\mathcal{N}|} \mid x_i \geq 0, i \in \mathcal{N}; \sum_{i \in \mathcal{N}} x_i = \sum_{i=1}^N b_i \right\}$  where  $|\mathcal{N}|$  is the number of elements in set  $\mathcal{N}$ .

It is easy to see that  $G_i$  is the derivative with respect to  $x_i$  of function  $F: \mathbb{R}_+^{|\mathcal{N}|} \setminus \{\mathbf{0}\} \rightarrow \mathbb{R}$  defined by

$$(B5) \quad F(\mathbf{x}) \equiv \alpha \sum_{n \in \mathcal{N}} x_n - \sum_{n=1}^N b_n \ln \left( \sum_{i \in \mathcal{N}} a_{ni} x_i^\alpha \right).$$

As we establish formally below, this makes it possible to solve the NCP in (B4) by way of solving  $\arg \min_{\mathbf{x} \in \Gamma} F(\mathbf{x})$  for  $\alpha \leq 1$  or  $\arg \max_{\mathbf{x} \in \Gamma} F(\mathbf{x})$  for  $\alpha > 1$ , where  $\Gamma$  is the compact set defined above.<sup>42</sup>

<sup>41</sup> Analogously to our treatment of the original functions  $G_{i,k}(\mathbf{w}, \mathbf{L}_k)$  and  $L_{i,k} G_i(\mathbf{w}, \mathbf{L}_k)$  at  $L_{i,k} = 0$  (see Section 1 of the online Appendix), we define values of  $G_i(\mathbf{x})$  and  $x_i G_i(\mathbf{x})$  at  $x_i = 0$  by their limits.

<sup>42</sup> We thank Anca Ciurte and Ioan Rasa for pointing us in this direction.

LEMMA 1: *If either (a)  $0 < \alpha < 1$  or (b)  $\alpha = 1$  and the matrix  $\{\tau_{ni,k}^{-\varepsilon_k}\}_{n,i}$  is nonsingular, then  $F(\cdot)$  defined in (B5) has a unique global minimum on  $\Gamma$ . If  $\alpha > 1$ , then  $F(\cdot)$  has a unique global maximum on  $\Gamma$ .*

PROOF:

Observe that for any  $n = 1, \dots, N$  function,  $\sum_{i \in N} a_{ni} x_i^\alpha$  is strictly concave for  $0 < \alpha < 1$  and is strictly convex for  $\alpha > 1$ .

If  $0 < \alpha < 1$ , then for any  $n = 1, \dots, N$  function,  $\sum_{i \in N} a_{ni} x_i^\alpha$  is a strictly concave function. Since the logarithm is a strictly concave function,  $F(\cdot)$  is strictly convex.

If  $\alpha > 1$ , then for any  $n = 1, \dots, N$  function,  $\sum_{i \in N} a_{ni} x_i^\alpha$  is a strictly convex function.

If  $\alpha = 1$ , then we need to make sure that for any two vectors  $\mathbf{x} \neq \mathbf{y}$  we cannot have that  $\sum_i a_{ni} x_i = \sum_i a_{ni} y_i$  for all  $n$ . Otherwise, we would have  $F(\gamma \mathbf{x} + (1 - \gamma) \mathbf{y}) = \gamma F(\mathbf{x}) + (1 - \gamma) F(\mathbf{y})$  for any  $\gamma \in [0, 1]$ , and strict convexity would be violated. The assumption that the matrix  $\{\tau_{ni,k}^{-\varepsilon_k}\}_{n,i}$  is nonsingular guarantees that matrix  $A = (a_{ni})$  is also nonsingular. Hence, for any  $\mathbf{x} \neq \mathbf{y}$  we have  $\sum_i a_{ni} x_i \neq \sum_i a_{ni} y_i$  for at least one  $n$ . Hence, in the case of  $\alpha = 1$ ,  $F(\cdot)$  is also strictly convex under the regularity assumption on the matrix of trade costs. Note that without this assumption the function  $F(\cdot)$  is just convex but not necessarily strictly convex.

Observe that  $F(\cdot)$  is a continuous function on  $\Gamma$ , and  $\Gamma$  is a compact set. Hence,  $F(\cdot)$  has a global minimum and a global maximum on  $\Gamma$ . Below we show that function  $F(\cdot)$  is strictly convex on  $\Gamma$  for  $\alpha \leq 1$ . Then, given that  $\Gamma$  is a convex set,  $F(\cdot)$  has, at most, one global minimum on  $\Gamma$ .

LEMMA 2: *Let  $\mathbf{x}^*$  be the unique global minimum of  $F(\cdot)$  on  $\Gamma$ . If  $0 < \alpha < 1$ , then  $x_i^* > 0$  for all  $i = 1, \dots, N$ .*

PROOF:

Suppose without loss of generality that  $x_1^* = 0$ . Since  $\sum_n x_n^* = \sum_n b_n > 0$ , we can also suppose without loss of generality that  $x_2^* \neq 0$ . Consider the vector  $\mathbf{x}(\varepsilon) = (\varepsilon, x_2^* - \varepsilon, x_3^*, \dots, x_N^*)$ , where  $\varepsilon \in [0, x_2^*]$ . Clearly,  $\mathbf{x}(\varepsilon) \in \Gamma$ . Define

$$\tilde{F}(\varepsilon) \equiv F(\mathbf{x}(\varepsilon)) = \alpha \sum_n x_n^* - \sum_n b_n \ln \left[ a_{n1} \varepsilon^\alpha + a_{n2} (x_2^* - \varepsilon)^\alpha + \sum_{i=2}^N a_{ni} (x_i^*)^\alpha \right].$$

We now show that  $\tilde{F}(\varepsilon) < \tilde{F}(0)$  for small enough  $\varepsilon > 0$ . We have

$$\frac{\partial [a_{j1} \varepsilon^\alpha + a_{j2} (x_2^* - \varepsilon)^\alpha]}{\partial \varepsilon} = \alpha a_{j1} \varepsilon^{\alpha-1} - \alpha a_{j2} (x_2^* - \varepsilon)^{\alpha-1},$$

which, given  $\alpha \in (0, 1)$ , is positive for small enough  $\varepsilon$ . This implies that  $a_{j1} \varepsilon^\alpha + a_{j2} (x_2^* - \varepsilon)^\alpha + \sum_{i=2}^N a_{ji} [x_i^*]^\alpha > a_{j2} [x_2^*]^\alpha + \sum_{i=2}^N a_{ji} [x_i^*]^\alpha$  for small enough  $\varepsilon$ . Since  $\ln(\cdot)$  is a strictly increasing function, we then get that  $\tilde{F}(\varepsilon) < \tilde{F}(0)$  for small

enough  $\varepsilon$ , a contradiction. This implies that  $\mathbf{x}^*$  cannot be a global minimum of  $F(\cdot)$  on  $\Gamma$ . Hence, in the case of  $\alpha \in (0, 1)$ , we must have  $x_i^* > 0$  for all  $i$ .

**LEMMA 3:** *If  $0 < \alpha \leq 1$ , then  $\mathbf{x}^*$  is a global minimum of  $F(\cdot)$  on  $\Gamma$  if and only if  $\mathbf{x}^*$  is a solution to (B4).*

**PROOF:**

The proof of this lemma is almost trivial, because the conditions in (B4) are just the first-order conditions for the minimization of  $F(\cdot)$  on  $\Gamma$ . The only complication is that to invoke the first-order conditions, we need to have differentiability of  $F(\cdot)$  on  $\Gamma$ , which is understood as differentiability of  $F(\cdot)$  on some open set containing  $\Gamma$ . In the case of  $\alpha < 1$ , any such open set necessarily includes points  $\mathbf{x}$  with  $x_i \leq 0$  at which  $F(\cdot)$  is not differentiable. The proof below formally deals with this complication.

Let us start with the simpler case of  $\alpha = 1$ . We can take the set  $D \equiv \{\mathbf{x} \in \mathbb{R}^N \mid \sum_i a_{ni}x_i > 0, \forall n\}$  as the domain of  $F(\cdot)$  and use  $\Gamma$  as the constraint set. Clearly,  $\Gamma \subset D \cap \{\mathbf{x} \in \mathbb{R}^N \mid \sum_i x_i = \sum_i b_i\}$  and  $F(\cdot)$  is differentiable on  $D$  for  $\alpha = 1$ . Consider the minimization problem

$$(B6) \quad \min_{\mathbf{x} \in D} F(\mathbf{x}) \quad \text{s.t.} \quad \sum_i x_i = \sum_i b_i \quad \text{and} \quad \mathbf{x} \geq 0.$$

Its first-order conditions, after some manipulations, can be written as

$$x_i \geq 0, \quad \frac{\partial F(\mathbf{x})}{\partial x_i} \geq 0, \quad x_i \frac{\partial F(\mathbf{x})}{\partial x_i} = 0, \quad i = 1, \dots, N, \quad \sum_i x_i = \sum_i b_i.$$

These conditions are also sufficient because  $F(\cdot)$  is convex and the set that satisfies the constraints of the minimization problem is a convex set. Then, since  $G_i(\mathbf{x}) = \frac{\partial F(\mathbf{x})}{\partial x_i}$ , we get that any solution of the minimization problem (B6) is also a solution of the NCP in (B4), and vice versa.

Let us now turn to the case with  $0 < \alpha < 1$ . Let  $\mathbf{x}^*$  be a minimum of  $F(\cdot)$  on  $\Gamma$ . By Lemma 2,  $x_i^* > 0$  for all  $i$ . Let  $\delta > 0$  be some number such that  $\delta \leq x_i^*$  for all  $i = 1, \dots, N$ . Define the domain of  $F$  by  $\tilde{D} \equiv \{\mathbf{x} \in \mathbb{R}^N \mid x_i > \delta/2, i = 1, \dots, N\}$ . Consider the minimization problem

$$(B7) \quad \min_{\mathbf{x} \in \tilde{D}} F(\mathbf{x}) \quad \text{s.t.} \quad \sum_i x_i = \sum_i b_i.$$

Since  $\tilde{D} \cap \{\mathbf{x} \in \mathbb{R}^N \mid \sum_i x_i = \sum_i b_i\} \subset \Gamma$ , if  $\mathbf{x}^*$  minimizes  $F$  on  $\Gamma$ , it also solves the minimization problem (B7). Since  $x_i^* > \delta/2$  for all  $i$ , the first-order conditions for (B7) are given by

$$(B8) \quad \frac{\partial F(\mathbf{x}^*)}{\partial x_i} = 0, \quad i = 1, \dots, N, \quad \text{and} \quad \sum_i x_i^* = \sum_i b_i.$$

Hence,  $\mathbf{x}^*$  solves NCP in (B4). Conversely, if  $\mathbf{x}^*$  solves NCP, then  $x_i^* > 0$  for all  $i$  because the condition  $G_i(\mathbf{x}^*) \geq 0$  cannot be satisfied for  $x_i^* = 0$  if  $0 < \alpha < 1$ . Hence,  $\mathbf{x}^*$  satisfies conditions (B8), which are the first-order conditions for an interior solution of (B7) with an appropriately chosen  $\delta > 0$ . Since these first-order conditions are also sufficient,  $\mathbf{x}^*$  solves (B7). Now, suppose by contradiction that the minimum of  $F(\cdot)$  on  $\Gamma$  is some  $\mathbf{x}^{**} \neq \mathbf{x}^*$ . Then, by Lemma 2,  $x_i^{**} > 0$  for all  $i$ . Therefore we can extend the open set on which  $F(\cdot)$  is differentiable to include both  $\mathbf{x}^*$  and  $\mathbf{x}^{**}$ . Then both  $\mathbf{x}^*$  and  $\mathbf{x}^{**}$  satisfy the first-order conditions (B8), which gives a contradiction given that  $F(\cdot)$  is strictly convex and the constraint set is convex. Hence,  $\mathbf{x}^*$  is the minimum of  $F(\cdot)$  on  $\Gamma$ .

Combining the results from Lemmas 1–3, we get a proof of the part of Proposition 1 concerning the case of  $0 < \alpha \leq 1$ .

## APPENDIX C. SCALE ECONOMIES, WELFARE, AND TRADE FLOWS

### C1. Proof of Proposition 3

The argument just above Proposition 3 has already established that, setting  $w_i = 1$  by choice of numeraire, if  $L_{i,k} > 0$ , then  $P_{i,k} < P_{i,k}^A$ , and so there are strictly positive gains from trade in industry  $k$  for country  $i$ . Let us now consider the case with  $L_{i,k} = 0$ . In this case we know that in equilibrium we must have  $G_{i,k} \geq 0$ , and using  $w_i = 1$  again, this can be rewritten as  $P_{i,k} \leq \mu_{i,k} \left( \frac{1 - \Delta_{i,k}}{S_{i,k} \beta_{i,k} \bar{L}_i} \right)^{1/\varepsilon_k}$ , where  $\Delta_{i,k} \equiv \sum_{n \neq i} S_{i,k} \tau_{ni,k}^{-\varepsilon} \mu_{n,k}^{-\varepsilon_k} P_{n,k}^{\varepsilon_k} \beta_{n,k} w_n \bar{L}_n$ . But in autarky we have  $P_{i,k}^A = \mu_{i,k} (S_{i,k} \beta_{i,k} \bar{L}_i)^{-1/\varepsilon_k}$ ; hence,  $P_{i,k}/P_{i,k}^A \leq (1 - \Delta_{i,k})^{1/\varepsilon_k}$ . With trade we have  $0 < \Delta_{i,k} < 1$ ; hence,  $P_{i,k} < P_{i,k}^A$ . ■

### C2. Data on Tariffs and Estimates of Elasticities of Scale

*Data on Tariffs.*—We use data on ad valorem tariffs for the year 2008 from the United Nations Statistical Division, Trade Analysis and Information System (UNCTAD-TRAINS, n.d.). For some countries 2008 tariff data are missing. In these cases we use tariff data for the nearest available year. Following Caliendo and Parro (2015), we focus on effectively applied tariffs and use simple tariff line average as the measure of tariff for each importer-exporter-sector combination.

The original UNCTAD-TRAINS data for 2008 do not cover individual European Union (EU) member countries, but they include the EU as a reporter. We infer tariffs for the EU member countries by using the fact that the EU members enjoy free circulation of goods within the EU and exert a common external tariff on all goods entering the EU market.

The UNCTAD-TRAINS data cover 31 two-digit (in ISIC rev.3) sectors, 181 reporters, and 245 partners. We aggregate these data into 16 sectors and 34 countries that we use in the current paper: we first aggregate tariffs across reporters by using reporters' total imports as weights, then we aggregate tariffs across partners by

TABLE C1—AVERAGE TARIFFS BY IMPORTERS

Country	Average tariff, %	Country	Average tariff, %	Country	Average tariff, %	Country	Average tariff, %
AUS	2.71	ESP	3.12	ITA	3.16	SVK	3.05
AUT	3.06	FIN	3.06	JPN	6.68	SVN	3.05
BEL	3.07	FRA	3.15	KOR	12.64	SWE	3.06
BRA	9.66	GBR	3.12	MEX	5.35	TUR	6.18
CAN	3.42	GRC	3.06	NLD	3.08	TWN	5.97
CHN	8.92	HUN	3.06	POL	3.08	USA	2.81
CZE	3.06	IDN	7.16	PRT	3.06	ROW	6.50
DEU	3.21	IND	13.83	ROM	3.06		
DNK	3.06	IRL	3.06	RUS	8.20		

Notes: Original data on tariffs are from UNCTAD-TRAINS for the year 2008. Average tariffs are calculated across exporters and sectors with exporters weighted by their total exports and sectors weighted by total sector imports.

TABLE C2—LIST OF SECTORS WITH TRADE AND SCALE ELASTICITIES, AVERAGE TARIFFS BY SECTORS

Sector	$\varepsilon_k$	$\psi_k$	Average tariff, %
Agriculture, hunting, forestry, and fishing			12.05
Mining and quarrying			0.94
(M) Food, beverages, and tobacco	4.40	0.16	16.47
(M) Textiles and textile products; leather, leather products, and footwear	7.70	0.12	10.27
(M) Wood and products of wood and cork	8.70	0.11	4.15
(M) Pulp, paper, printing, and publishing	7.80	0.11	3.10
Coke, refined petroleum, and nuclear fuel			2.87
(M) Chemicals and chemical products	3.40	0.21	3.76
(M) Rubber and plastics	2.90	0.26	5.59
(M) Other nonmetallic mineral	6.80	0.14	6.02
(M) Basic metals and fabricated metal	7.15	0.13	3.55
(M) Machinery, not elsewhere classified	6.20	0.13	3.38
(M) Electrical and optical equipment	9.75	0.09	5.82
(M) Transport equipment	5.55	0.16	6.31
(M) Manufacturing, not elsewhere classified; recycling			5.82
Electricity, gas, and water supply			0.60
Construction			
Sale and repair of motor vehicles and motorcycles; retail sale of fuel; wholesale trade, except of motor vehicles and motorcycles			
Retail trade and repair, except of motor vehicles and motorcycles			
Hotels and restaurants			
Inland transport			
Water transport			
Air transport			
Other supporting transport activities			
Post and telecommunications			
Financial intermediation			
Real estate activities			
Renting of machinery and equipment and other business activities			
Education			
Health and social work			
Other community, social and personal services; private households with employed persons; public administration and defense; compulsory social security			

Notes: Manufacturing sectors are marked with (M). Values for  $\varepsilon_k$  and  $\psi_k$  for manufacturing sectors are based on estimates from BCDR (see the description in the text of how these estimates are used to get the values for  $\varepsilon_k$  and  $\psi_k$ ). Original data on tariffs are from UNCTAD-TRAINS for the year 2008. Average tariffs are calculated across importers and exporters with importers weighted by their total imports and exporters weighted by their total exports. Empty entries in the columns for  $\varepsilon_k$  and  $\psi_k$  mean that either BCDR do not have estimates for the corresponding sectors or their estimates are not used in the current paper. Empty values in the column for average tariffs mean that UNCTAD-TRAINS does not have data for the corresponding sectors.

using partners' total exports as weights, and, finally, we aggregate tariffs across sectors by using sectors' total imports as weights. When calculating the weights we use imports and exports only in the sectors for which UNCTAD-TRAINS has data on tariffs.

Tables C1 and C2 present average tariffs for importers and sectors. Averaging is done only for the sectors that have data on tariffs in UNCAD-TRAINS. In the exercises with tariffs in Section VB, we assume zero tariff for all sectors that have empty values in the "Average Tariff, %" column of Table C2.

*Estimates of Trade and Scale Elasticities.*—We use industry-level estimates of trade and scale elasticities obtained by BCDR. These values are provided in Table C2. Several comments are in order. First, in their paper the list of industries is more disaggregated than the one in the current paper. In particular, three industries from our list—"basic metals and fabricated metal," "electrical and optical equipment," and "transport equipment"—are broken into subindustries in BCDR, each with its own estimated elasticities. To assign values of elasticities to industries from our list, we compute simple averages over corresponding industries in BCDR. Second, we only use BCDR's estimates for the manufacturing industries.<sup>43</sup> Finally, BCDR do not have an estimate for one industry that we classify as manufacturing: "Manufacturing, N.E.C.; recycling." In our counterfactual exercises we assign the same trade and scale elasticities to this industry as to the nonmanufacturing industries.

## REFERENCES

- Abdel-Rahman, Hesham, and Masahisa Fujita.** 1990. "Product Variety, Marshallian Externalities, and City Sizes." *Journal of Regional Science* 30 (2): 165–83.
- Allen, Treb, and Costas Arkolakis.** 2014. "Trade and the Topography of the Spatial Economy." *Quarterly Journal of Economics* 129 (3): 1085–140.
- Allen, Treb, Costas Arkolakis, and Yuta Takahashi.** 2020. "Universal Gravity." *Journal of Political Economy* 128 (2): 393–433.
- Alvarez, Fernando, and Robert E. Lucas.** 2007. "General Equilibrium Analysis of the Eaton-Kortum Model of International Trade." *Journal of Monetary Economics* 54 (6): 1726–68.
- Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare.** 2012. "New Trade Models, Same Old Gains?" *American Economic Review* 102 (1): 94–130.
- Arkolakis, Costas, Svetlana Demidova, Peter J. Klenow, and Andrés Rodríguez-Clare.** 2008. "Endogenous Variety and the Gains from Trade." *American Economic Review* 98 (2): 444–50.
- Armington, Paul S.** 1969. "A Theory of Demand for Products Distinguished by Place of Production." *IMF Staff Papers* 16 (2): 159–78.
- Baqae, David Rezza, and Emmanuel Farhi.** 2020. "Productivity and Misallocation in General Equilibrium." *Quarterly Journal of Economics* 135 (1): 105–63.
- Bartelme, Dominick G., Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare.** 2021. "The Textbook Case for Industrial Policy: Theory Meets Data." NBER Working Paper 26193.
- Behrens, Kristian, Andrea R. Lamorgese, Gianmarco I.P. Ottaviano, and Takatoshi Tabuchi.** 2009. "Beyond the Home Market Effect: Market Size and Specialization in a Multi-Country World." *Journal of International Economics* 79 (2): 259–65.
- Caliendo, Lorenzo, Robert C. Feenstra, John Romalis, and Alan M. Taylor.** 2015. "Tariff Reductions, Entry, and Welfare: Theory and Evidence for the Last Two Decades." NBER Working Paper 21768.

<sup>43</sup>For example, BCDR have estimates for "Coke, refined petroleum, and nuclear fuel," which we classify as nonmanufacturing, and as a result we exclude BCDR's estimates for this case.

- Caliendo, Lorenzo, and Fernando Parro.** 2015. "Estimates of the Trade and Welfare Effects of NAFTA." *Review of Economic Studies* 82 (1): 1–44.
- Chipman, John S.** 1965. "A Survey of the Theory of International Trade: Part 2, The Neo-Classical Theory." *Econometrica* 33 (4): 685–760.
- Costinot, Arnaud, Dave Donaldson, and Ivana Komunjer.** 2012. "What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas." *Review of Economic Studies* 79 (2): 581–608.
- Costinot, Arnaud, and Andrés Rodríguez-Clare.** 2014. "Trade Theory with Numbers: Quantifying the Consequences of Globalization." In *Handbook of International Economics, vol. 4*, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, 197–261. Amsterdam: Elsevier.
- Dekle, Robert, Jonathan Eaton, and Samuel Kortum.** 2008. "Global Rebalancing with Gravity: Measuring the Burden of Adjustment." *IMF Staff Papers* 55 (3): 511–40.
- Dornbusch, Rudiger, Stanley Fischer, and Paul A. Samuelson.** 1977. "Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods." *American Economic Review* 67 (5): 823–39.
- Eaton, Jonathan, and Samuel Kortum.** 2001. "Technology, Trade, and Growth: A Unified Framework." *European Economic Review* 45 (4–6): 742–55.
- Eaton, Jonathan, and Samuel Kortum.** 2002. "Technology, Geography, and Trade." *Econometrica* 70 (5): 1741–79.
- Ethier, Wilfred J.** 1982. "Decreasing Costs in International Trade and Frank Graham's Argument for Protection." *Econometrica* 50 (5): 1243–68.
- Feenstra, Robert C., Philip Luck, Maurice Obstfeld, and Katheryn N. Russ.** 2018. "In Search of the Armington Elasticity." *Review of Economics and Statistics* 100 (1): 135–50.
- Galle, Simon, Andrés Rodríguez-Clare, and Moises Yi.** 2023. "Slicing the Pie: Quantifying the Aggregate and Distributional Effects of Trade." *Review of Economic Studies* 90 (1): 331–75.
- Graham, Frank D.** 1923. "Some Aspects of Protection Further Considered." *Quarterly Journal of Economics* 37 (2): 199–227.
- Grossman, Gene M., and Esteban Rossi-Hansberg.** 2010. "External Economies and International Trade Redux." *Quarterly Journal of Economics* 125 (2): 829–58.
- Hanson, Gordon H., and Chong Xiang.** 2004. "The Home-Market Effect and Bilateral Trade Patterns." *American Economic Review* 94 (4): 1108–29.
- Helpman, Elhanan, and Paul Krugman.** 1985. *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy*. Cambridge, Massachusetts: MIT Press.
- Krueger, Alan B., and Lawrence H. Summers.** 1986. "Efficiency Wages and the Wage Structure." NBER Working Paper 1952.
- Krugman, Paul.** 1980. "Scale Economies, Product Differentiation, and the Pattern of Trade." *American Economic Review* 70 (5): 950–9.
- Krugman, Paul.** 1995. "Increasing Returns, Imperfect Competition and the Positive Theory of International Trade." In *Handbook of International Economics, vol. 3*, edited by Gene M. Grossman and Kenneth Rogoff, 1243–77. Amsterdam: Elsevier.
- Kucheryavyy, Konstantin, Gary Lyn, and Andrés Rodríguez-Clare.** 2021. "Spatial Equilibria: The Case of Two Regions." NBER Working Paper 29592.
- Kucheryavyy, Konstantin, Gary Lyn, and Andrés Rodríguez-Clare.** 2023. "Replication Data for: Grounded by Gravity: A Well-Behaved Trade Model with Industry-Level Economies of Scale." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. <https://doi.org/10.3886/E174021V1>.
- Lashkaripour, Ahmad, and Volodymyr Lugovskyy.** 2021. "Profits, Scale Economies, and the Gains from Trade and Industrial Policy." Unpublished.
- Lyn, Gary, and Andrés Rodríguez-Clare.** 2013a. "External Economies and International Trade Redux: Comment." *Quarterly Journal of Economics* 128 (4): 1895–905.
- Lyn, Gary, and Andrés Rodríguez-Clare.** 2013b. "Marshallian Externalities, Comparative Advantage and International Trade." Unpublished.
- Marshall, Alfred.** 1890. *Principles of Economics, vol. 1*. London: Macmillan.
- Melitz, Marc J.** 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71 (6): 1695–725.
- Redding, Stephen J.** 2016. "Goods Trade, Factor Mobility and Welfare." *Journal of International Economics* 101: 148–67.
- Somale, Mariano.** 2021. "Comparative Advantage in Innovation and Production." *American Economic Journal: Macroeconomics* 13 (3): 357–96.

- Timmer, Marcel P., Erik Dietzenbacher, Bart Los, Robert Stehrer, and Gaaitzen J. de Vries.** 2015. "An Illustrated User Guide to the World Input-Output Database: The Case of Global Automotive Production." *Review of International Economics* 23 (3): 575–605.
- UNCTAD-TRAINS.** "United Nations Statistical Division, Trade Analysis and Information System." World Integrated Train Solution. <https://wits.worldbank.org> (accessed September 1, 2017).
- Venables, Anthony J.** 1987. "Trade and Trade Policy with Differentiated Products: A Chamberlinian-Ricardian Model." *Economic Journal* 97 (387): 700–17.
- Yatsynovich, Yury.** 2017. "Technological Spillovers and Dynamics of Comparative Advantage." Unpublished.