American Economic Review 2016, 106(10): 3159–3184
http://dx.doi.org/10.1257/aer.20141449

Trade, Domestic Frictions, and Scale Effects[†]

By Natalia Ramondo, Andrés Rodríguez-Clare, and Milagro Saborío-Rodríguez*

Because of scale effects, idea-based growth models imply that larger countries should be much richer than smaller ones. New trade models share the same counterfactual feature. In fact, new trade models exhibit other counterfactual implications associated with scale effects: import shares decrease and relative income levels increase too steeply with country size. We argue that these implications are largely a result of the standard assumption that countries are fully integrated domestically. We depart from this assumption by treating countries as collections of regions that face positive costs to trade among themselves. The resulting model is largely consistent with the data. (*JEL* F11, F14, F43, O47, R12)

Aggregate economies of scale are so central a feature of innovation-led growth theory that, in Jones' (2005, p. 1089) words, "rejecting one is largely equivalent to rejecting the other." Because of scale economies, idea-based growth models such as Jones (1995) and Kortum (1997) imply that larger countries should be richer than smaller ones.[1] There is some disagreement in the literature as to whether such scale effects are present in the data, but it is safe to say that they are very small compared to those implied by the theory.[2]

New trade models such as Krugman (1980), Eaton and Kortum (2001), and Melitz (2003) are also idea-based models, and carry the same counterfactual implication that real income per capita strongly increases with country size. One might expect

* Ramondo: School of Global Policy and Strategy, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, and NBER (e-mail: nramondo@ucsd.edu); Rodríguez-Clare: Department of Economics, University of California, Berkeley, 549 Evans Hall, Berkeley, CA 94720, and NBER (e-mail: andres@econ.berkeley.edu); Saborío-Rodríguez: Escuela de Economía, Universidad de Costa Rica, and CATIE, San Pedro de Montes de Oca, San Jose, Costa Rica (e-mail: msaborio@catie.ac.cr). We have benefited from comments from Jim Anderson, Treb Allen, Lorenzo Caliendo, Arnaud Costinot, Jonathan Eaton, Cecile Gaubert, Keith Head, Pete Klenow, Sam Kortum, David Lagakos, Thierry Mayer, Benjamin Moll, Peter Morrow, Steve Redding, and Mike Waugh, as well as seminar participants at various conferences and institutions, and three anonymous referees. We also thank David Schnholzer and Yui Yoshida for their excellent research assistance. All errors are our own. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

[†] Go to http://dx.doi.org/10.1257/aer.20141449 to visit the article page for additional materials and author disclosure statement(s).

[1] First-generation endogenous growth models such as Romer (1990) feature "strong" scale effects, whereby scale increases growth, whereas second-generation semi-endogenous growth models such as Jones (1995) and Kortum (1997) feature "weak" scale effects, whereby scale increases income levels rather than growth (see Jones 2005 for a detailed discussion). Models that do not display any scale effects, such as Lucas and Moll (2014), depart from the standard assumption that ideas are nonrival by assuming that knowledge can only be used in production when it is embodied in individuals with limited time endowments, and that individuals face search frictions in learning about better ideas.

[2] Rose (2006) finds no scale effects in the data.

scale effects in such models to be offset by the fact that small countries tend to gain more from trade than large ones. It turns out, however, that although small countries do gain more from trade, these gains are not large enough to neutralize the underlying scale effects. In fact, new trade models exhibit other counterfactual implications associated with scale effects; in particular, import shares decrease and relative income levels increase too steeply with country size.

One way in which the literature has reconciled innovation-led growth theories with the observation that income per capita does not increase with country size is by assuming that scale effects operate at the world rather than the country level (see Jones 2005 for a detailed discussion). Indeed, if foreign ideas can be used for domestic production, effective technology levels would vary less than proportionally with country size, weakening country-level scale effects.

In this paper we focus on an alternative channel which has been relatively unexplored, can be easily inferred from the data, and is arguably of first-order importance in reconciling the standard model with the data: the existence of domestic trade costs. Specifically, we depart from the usual assumption that countries are fully integrated domestically, as if they were a single dot in space, and reinterpret the Eaton and Kortum (2002) model of trade as applied to subnational economies, or "regions": as a result, each country is formed by a group of regions that face positive costs to trade among themselves. This assumption is consistent with the empirical evidence, which clearly shows that domestic trade costs are large.[3] We further allow technology levels to be proportional to the size of the economy, as in Krugman (1980); Jones (1995); Kortum (1997); Eaton and Kortum (2001); and Melitz (2003). As in the standard growth and trade models, this assumption leads to aggregate increasing returns to scale, but in our model these scale effects are (partially) offset by the presence of domestic trade costs. Intuitively, to the extent that large countries are composed of more regions, trade costs among regions reduce the advantage of country size and weaken scale effects; big counterfactual scale effects arise in the standard model precisely as a result of the crude treatment of geography.

We calibrate the model using data on population and geography for 287 metropolitan areas, international trade flows for 26 OECD countries, and intranational trade flows for the United States. We calibrate the key parameter determining the strength of economies of scale by appealing to the growth and trade literature. Trade costs between regions, both within and across countries, are estimated from distance and trade data between metropolitan areas. The calibration reveals that domestic trade costs cut in half the elasticity of productivity with respect to country size implied by the standard model, getting closer to the small elasticity we observe in the data. For a small country like Denmark, the model with no domestic trade costs implies that its productivity level would be 38 percent of the US level while in the data it is 94 percent. In contrast, our calibrated model implies a relative productivity level for Denmark of 80 percent.

[3] For the United States, Canada, and China, Tombe and Winter (2014) calculate a range between 100 and 140 percent; Agnosteva, Anderson, and Yotov (2014) calculate them in 109 percent for Canadian provinces; and Allen and Arkolakis (2014) estimate them in 55 percent for US metropolitan areas. Relatedly, Hillberry and Hummels (2008) find, for the United States, that manufacturing shipments between establishments in the same zip code are three times larger than between establishments in different zip codes.

Even with domestic trade costs, our calibrated model exhibits scale effects that are stronger than in the data. We argue that this difference could be interpreted as evidence of the importance of scale effects operating at the world level. Our paper shows that the presence of domestic trade costs makes international technology diffusion less important, but not irrelevant, in reconciling the theory with the data.

Our paper makes a contribution to an emerging literature exploring the interaction between international trade and domestic economic geography using quantitative models, such as Cosar and Fajgelbaum (2016); Allen and Arkolakis (2014); Fajgelbaum and Redding (2014); and Redding (2015). In particular, Redding (2015) also extends the Eaton and Kortum (2002) model by modeling each country as a collection of imperfectly integrated regions. The insight from his paper is that reallocation of labor across regions is consequential for the gains from trade in each location. In contrast, our focus is at the level of countries rather than regions: we focus on the extent to which including domestic trade costs improves the fit of the standard model with the country-level data, with special emphasis on country-level scale effects.

Our paper is also related to a literature that studies the relationship between country size, openness, and productivity. Anderson and van Wincoop (2004) and Anderson and Yotov (2010) show that in a standard gravity model, under some special conditions, home bias increases with country size, leading to lower import shares for larger countries. At the empirical level, Redding and Venables (2004) and Head and Mayer (2011) show that income increases with a measure of "market potential," which is increasing in country size, while Frankel and Romer (1999); Ades and Glaeser (1999); Alesina, Spolaore, and Wacziarg (2000); and Alcala and Ciccone (2004) document a positive effect of country size and trade openness on income levels. Other papers fail to find a positive effect of country size on productivity (Rose 2006). Our contribution to this literature is to show that, relative to the data, country-level scale effects are too strong in models without domestic trade costs, and that adding these costs allows the model to better matches the observed relationship between country size and productivity, import shares, relative income levels, and prices.[4]

Finally, Alvarez and Lucas (2007) and Waugh (2010) calibrate an Eaton and Kortum (2002) model to match observed trade flows and cross-country income levels. Both of these calibrations assume that there are no domestic trade costs, but allow technology or efficiency levels to vary across countries. In fact, strong scale effects are avoided in these two calibrated models by having technology levels decreasing rapidly with country size. Since it is hard to defend such systematic variation, we calibrate technology parameters to observed R&D intensities, which do not vary systematically with country size in our sample of OECD countries.

[4] When estimating market potential, Redding and Venables (2004) and Head and Mayer (2011) recognized the importance of domestic trade costs and estimated gravity equations that include the domestic trade pair and a measure of internal distance (e.g., a transformation of country area) to proxy for domestic trade costs. They did not explore, however, the role of domestic frictions on *cross-country* income levels and import shares.

I. The Model

Our starting point is the Ricardian trade model developed by Eaton and Kortum (2002)—henceforth, EK—with two key modifications: domestic trade costs and scale effects. We introduce these two modifications one by one in the next two subsections. The third subsection derives analytical results for the model with domestic trade costs and scale effects in the special case in which all regions within each country are fully symmetric.

A. *Domestic Trade Costs*

Domestic trade costs emerge naturally by applying the EK framework to subnational economies, or "regions." There are M subnational economies (regions) indexed by m and N countries indexed by n. Let Ω_n be the set of regions belonging to country n and M_n be the number of regions in that set. Labor is the only factor of production, available in quantity l_m in region m. There is no labor mobility within or across countries.[5]

There is a continuum of goods in the interval $[0, 1]$, and preferences are constant elasticity of substitution (CES) with elasticity of substitution σ. Technologies are linear with good-specific productivities in region m drawn from a Fréchet distribution with parameters $\theta > \sigma - 1$ and t_m. These draws are independent across goods and across countries. There are iceberg trade costs $d_{mk} \geq 1$ to export from k to m, with $d_{mm} = 1$ and $d_{mk} \leq d_{ms} d_{sk}$ for all m, s, k (triangular inequality). There is perfect competition.

Bilateral trade flows between regions satisfy the standard expression in the EK model,

$$x_{mk} = \frac{t_k v_k^{-\theta} d_{mk}^{-\theta}}{\sum_s t_s v_s^{-\theta} d_{ms}^{-\theta}} x_m, \tag{1}$$

where v_k is the wage in region k and $x_m \equiv \sum_k x_{mk}$ is total expenditure in region m. In turn, price indices are

$$p_m = \gamma^{-1} \left(\sum_k t_k v_k^{-\theta} d_{mk}^{-\theta} \right)^{-1/\theta}, \tag{2}$$

where $\gamma \equiv \Gamma\left(\frac{1-\sigma}{\theta} + 1\right)^{1/(\sigma-1)} > 0$. Trade balance at the region level implies that $x_m = v_m l_m$, so that the labor market clearing condition in region m entails

$$v_m l_m = \sum_k \frac{t_m v_m^{-\theta} d_{km}^{-\theta}}{\sum_s t_s v_s^{-\theta} d_{ks}^{-\theta}} v_k l_k. \tag{3}$$

[5] In the working paper version we allow for perfect labor mobility within countries while assuming that workers are heterogeneous in their productivity across regions. The main aggregation result in Proposition 1 as well as the results discussed for the case of symmetric regions remain valid, while the quantitative results present negligible changes.

This constitutes a system that we can solve to determine equilibrium wages across all regions.

We introduce additional notation to keep track of country-level variables. Let $X_{ni} \equiv \sum_{k \in \Omega_i} \sum_{m \in \Omega_n} x_{mk}$ denote total trade flows from country i to country n, and let $X_n \equiv \sum_{m \in \Omega_n} x_m$, $L_n \equiv \sum_{m \in \Omega_n} l_m$ and $w_n \equiv X_n / L_n$ denote country n's total income, total labor, and average nominal income per worker, respectively. The average real income of workers in country n is $U_n \equiv \sum_{m \in \Omega_n} (l_m / L_n)\, v_m / p_m$. Finally, let $\lambda_{ni} \equiv X_{ni} / X_n$ denote country-level trade shares.

We now establish a basic aggregation result, namely that if there were no domestic trade costs, the model would generate the same country-level implications as the EK model. No domestic trade costs imply that $d_{mk} = 1$ for all $m, k \in \Omega_n$. By the triangular inequality, $d_{mk} = d_{m'k'}$ for all $m, m' \in \Omega_n$ and $k, k' \in \Omega_i$ (i.e., international trade costs are the same for all regions within a country), and $p_m = p_k$ for all $m, k \in \Omega_n$. Combined with (3), we get the following proposition (all proofs are in the online Appendix).

PROPOSITION 1: *If there are no domestic trade costs, country-level trade flows and price indices are, respectively,*

$$X_{ni} = \frac{T_i w_i^{-\theta} \tau_{ni}^{-\theta}}{\sum_j T_j w_j^{-\theta} \tau_{nj}^{-\theta}} X_n, \tag{4}$$

and

$$P_n = \gamma^{-1} \left(\sum_i T_i w_i^{-\theta} \tau_{ni}^{-\theta} \right)^{-1/\theta}, \tag{5}$$

where the country-level technology parameter is

$$T_i = \left(\sum_{m \in \Omega_i} (l_m / L_i)^{\theta/(1+\theta)} t_m^{1/(1+\theta)} \right)^{1+\theta}, \tag{6}$$

and country-level trade costs are

$$\tau_{ni} \equiv d_{mk} \text{ for } m \in \Omega_n \text{ and } k \in \Omega_i \text{ for } n \neq i, \tag{7}$$

with $\tau_{nn} = 1$. *Country-level welfare is given by*

$$U_n = \gamma \times T_n^{1/\theta} \times \lambda_{nn}^{-1/\theta}. \tag{8}$$

In the presence of domestic trade costs, a log-linear gravity equation for country-level trade flows as in (4) no longer holds, except for (at least) two special cases: when regions within each country are fully symmetric; and when trade costs are as in a hub-and-spoke trade system, so that all international trade is done through a single location (e.g., a port) in each country. The first case is studied in detail below, while the second case is studied in the working paper version of this paper (Ramondo, Rodríguez-Clare, and Saborío-Rodríguez 2014).

Domestic trade costs also matter for the gains from trade. Whereas in the case of frictionless trade or symmetric regions (below), the gains from trade, as defined in Arkolakis, Costinot, and Rodríguez-Clare (2012), are given by $1 - \lambda_{nn}^{1/\theta}$, this result no longer holds except for the special case with symmetric regions (see below and the discussion in Ramondo, Rodríguez-Clare, and Saborío-Rodríguez 2014).

B. *Scale Effects*

We have so far treated technology parameters t_m as exogenous. We now argue that technology levels should be allowed to depend positively on the size of the region. Building on Proposition 1, we then show that, in the case of no domestic trade costs, this dependency leads to aggregate economies of scale.

It is natural to expect larger regions to have better technologies. Suppose that we merged two identical regions with technology parameter t into a single region. It is easy to show that the new region would have a technology parameter $2t$.[6] Thus, if two regions m and k are identical except that region m is twice as large as region k, then $t_m = 2t_k$. Since labor is the only factor of production in our model, then the size of a region is given by its workforce, and so t_m will be proportional to l_m.[7] This relation between technology levels and population was derived formally by Eaton and Kortum (2001) in a model of endogenous innovation and Bertrand competition, and it also emerges from trade models with monopolistic competition, as we discuss below. This is captured formally by the following assumption.

ASSUMPTION 1 (Technology Scales with Population): $t_m = \phi_n l_m$ *for all* $m \in \Omega_n$.

We allow ϕ_n to vary with n to reflect differences in "innovation intensity" across countries. This parameter is calibrated to R&D employment shares in the quantitative analysis. The important part of this assumption, however, is that technology levels are proportional to population.

It is worth noting that equivalent formulations of our model plus Assumption 1 could be derived building on Krugman (1980) or Melitz (2003) rather than EK (derivations are in the online Appendix). With Krugman (1980), Assumption 1 follows immediately from the free entry condition combined with the standard assumption that the fixed cost of production is not systematically related to country size. With Melitz (2003), we would need to assume that the productivity distribution is Pareto, as in Chaney (2008). If the Pareto shape parameter is θ and either $\theta \approx \sigma - 1$ or the fixed cost of selling in market m is proportional to its population, l_m, then again Assumption 1 would hold because of free entry.

Assumption 1 leads to country-level scale effects: everything else equal, larger countries should exhibit higher real income levels. We can see this most directly in the case of no domestic trade costs. Proposition 1 combined with Assumption 1 and

[6] This result follows from the fact that if x and y are distributed Fréchet with parameters θ and t_x and t_y, respectively, then $\max\{x, y\}$ is distributed Fréchet with parameters θ and $t_x + t_y$.

[7] Formally, let a "technology" be a productivity ξ drawn from a Fréchet distribution with parameters θ and ϕ, and assume that the number of technologies per good is equal to the number of workers. It is then easy to show that the best technology for a good, $\max \xi$, is distributed Fréchet with parameters θ and ϕl_m .

the expression for T_n in (6) yields $T_n = \phi_n L_n$, which plugged into (8) reveals that the average real wage in country n is given by

$$U_n = \gamma \times (\phi_n L_n)^{1/\theta} \times \lambda_{nn}^{-1/\theta}. \tag{9}$$

Thus, conditional on trade shares and innovation intensity, average real income levels increase with country size with an elasticity of $1/\theta$. This is because a larger population is linked to a higher stock of nonrival ideas (i.e., technologies), and more ideas imply a superior technology frontier. The strength of this effect is linked to the Fréchet parameter θ: the lower is θ, the higher is the dispersion of productivity draws from this distribution, and the more an increase in the stock of ideas improves the technology frontier. These are the aggregate economies of scale that play a critical role in semi-endogenous growth models (Kortum 1997) and that underpin the gains from openness in EK-type models (Eaton and Kortum 2001; Arkolakis et al. 2008).

C. *A Special Case: Symmetric Regions*

We now study how domestic trade costs affect the strength of scale effects for the simple case in which regions within each country are fully symmetric, as captured by the following assumption:

ASSUMPTION 2 (Region-Level Symmetry): *$l_m = l_{m'}$ for all $m, m' \in \Omega_n$, and $d_{mk} = d_{m'k'}$ for all $m, m' \in \Omega_n$ and $k, k' \in \Omega_i$.*

We focus here on the case of symmetric regions because it is the only one for which we can provide analytical results for real wages given trade shares. This case also has the advantage that it leads to very similar results to the EK model for country-level trade flows, and yet shows clearly how domestic trade costs weaken scale effects. In Section II, we calibrate the model imposing Assumption 1, *but not* Assumption 2, and present quantitative results for the strength of scale effects for real wages. We also present results for the case with both Assumptions 1 and 2, which, thanks to Proposition 2, is easier to calibrate.

PROPOSITION 2: *Under Assumptions* 1 *and* 2, *country-level trade shares and price indices are as in* (4) *and* (5), *respectively*, *with country-level technology parameters given by*

$$T_i = \phi_i L_i, \tag{10}$$

international trade costs τ_{ni} as in (7), *and domestic trade costs given by*

$$\tau_{nn} \equiv \left(\frac{1}{M_n} + \left(1 - \frac{1}{M_n}\right)\delta_n^{-\theta}\right)^{-1/\theta}, \tag{11}$$

where $\delta_n \equiv d_{mk}$ for $m \neq k$ with $m, k \in \Omega_n$. In addition, country-level welfare is

$$U_n = \gamma \times \underbrace{\phi_n^{1/\theta}}_{\text{R\&D Intensity}} \times \underbrace{L_n^{1/\theta}}_{\text{Pure Scale Effect}} \times \underbrace{\tau_{nn}^{-1}}_{\text{Domestic Frictions}} \times \underbrace{\lambda_{nn}^{-1/\theta}}_{\text{Gains from Trade}}. \tag{12}$$

According to Proposition 2, domestic trade costs τ_{nn} are a weighted power mean with exponent $-\theta$ of the cost of intraregional trade, which we assume is 1, and the cost of trade between regions belonging to the same country, which is denoted δ_n, with weights given by $1/M_n$ and $1 - 1/M_n$. Note that countries with the same δ_n may have different τ_{nn} because of their different size; in particular, larger countries would have larger domestic trade costs.

The key result in Proposition 2 is in (12). This expression shows that there are four distinct forces that determine real wages across countries: innovation intensity, pure scale effects, domestic trade costs, and the gains from trade. In the presence of domestic trade costs, economies of scale depend on how τ_{nn} is affected by country size, L_n.[8] To derive sharper results, assume for the moment that country size scales with the number of regions, $L_n/L_i = M_n/M_i$, and that $\delta_n = \delta$, for all n so that all variation in τ_{nn} comes from variation in the number of regions M_n. In particular, τ_{nn} is increasing in M_n, so that domestic trade costs offset scale effects. More specifically, the strength of economies of scale under the presence of domestic trade costs, conditional on trade shares, is given by $\varepsilon \equiv \partial \ln U_n / \partial \ln L_n = (1/\theta)(\delta/\tau_{nn})^{-\theta}$: if $\delta = 1$, then $\tau_{nn} = 1$ and $\varepsilon = 1/\theta$; otherwise the term $(\delta/\tau_{nn})^{-\theta}$ is lower than 1 and offsets economies of scale, $\varepsilon < 1/\theta$.

Proposition 2 lays out the implications of domestic trade costs on real wages *conditional on trade shares*. To derive analytical results for the unconditional effects of country size in the presence of domestic trade costs, we need to impose some additional restrictions. In particular, we assume that international and domestic trade costs are uniform and that countries are symmetric in terms of their innovation intensity and region size.

ASSUMPTION 3 (Country-Level Symmetry): *$l_m = l$ for all m, $\delta_n = \delta$ and $\phi_n = \phi$ for all n, and $\tau_{ni} = \tau$ for all $n \neq i$.*

Under this (admittedly very strong) assumption, which we maintain only for the next Proposition, we can characterize how country size matters also for import shares, nominal wages, and price levels.

PROPOSITION 3: *Assume Assumptions* 1, 2, *and* 3. *If $\tau > \delta$, then larger countries have lower import shares, higher wages, and lower price levels. If $\tau = \delta$, then larger countries have lower import shares, but wages and prices do not vary with country size.*

As expected, import shares decline with country size and large countries gain less from trade, but aggregate economies of scale are strong enough so that the overall effect is for real wages to increase with size. Proposition 3 also establishes that, if $\tau > \delta$, real wages increase with country size both because of higher wages and

[8] While the elasticity of U_n with respect to ϕ_n, L_n, and $1/\lambda_{nn}$ is $1/\theta$, the elasticity with respect to τ_{nn} is simply -1. The reason is that whereas ϕ_n, L_n, and $1/\lambda_{nn}$ affect U_n indirectly through T_n or trade, the effect of trade costs on utility is direct. This is most clearly appreciated in the limit as $M_n \to \infty$, in which case $\tau_{nn} = \delta_n$ and hence, a higher τ_{nn} leads to a proportional decrease in welfare.

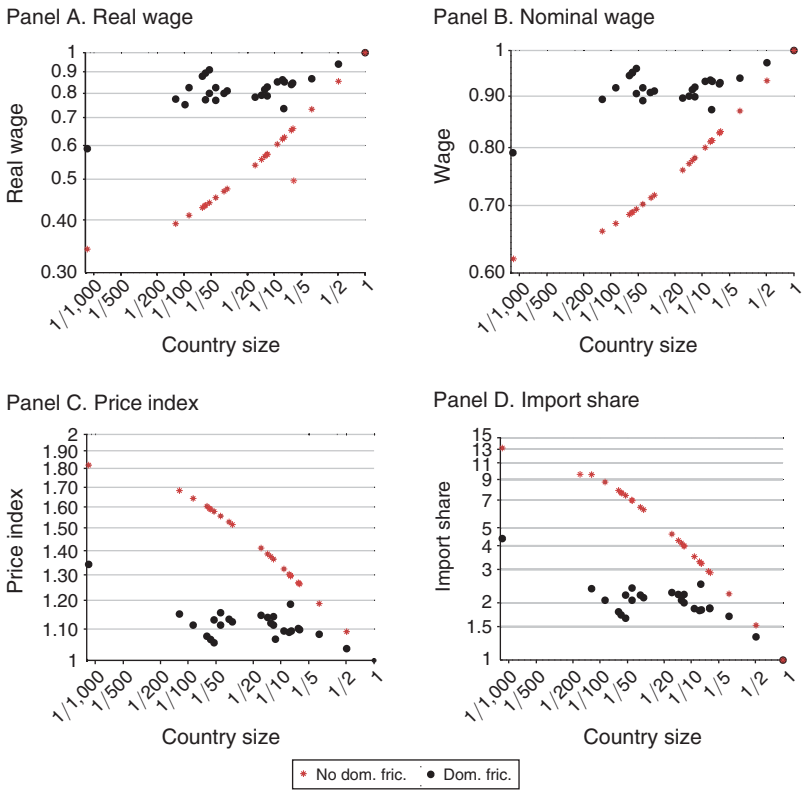

FIGURE 1. THE ROLE OF DOMESTIC FRICTIONS: NUMERICAL EXAMPLE (*Relative to US, in logs*)

Notes: No dom.fric. refers to the model without domestic frictions. Dom.fric. refers to the model with domestic frictions and symmetric regions. Both models have uniform international trade costs. Country size refers to equipped labor.

lower prices. More importantly, these scale effects disappear when $\tau = \delta$, suggesting that domestic trade costs weaken scale effects.

This result is illustrated more generally in Figure 1. For $\theta = 4$, we alternately fixed $\delta = 1$ and $\delta = 2.7$, and chose τ for each δ to match the same average import share (the one observed for our sample of 26 countries)—we want to make sure that both models match equally well this moment from the data. As we explain in more detail in the next section, we take the number of regions of a country to be the number of metropolitan areas with more than half a million habitants, and we take country size to be a measure of equipped labor. The figure shows the implied import shares, nominal wages, real wages, and price levels against country size for the model with no domestic trade costs (stars), and the model with those trade costs (circles). All four variables vary strongly with size in the model with no domestic trade costs, but this dependence is severely weakened when domestic trade costs are considered.

II. Quantitative Analysis

The goal of this section is to evaluate the role of domestic trade costs in reconciling the standard model with the data on real wages. Additionally, we show

the role of domestic trade costs for other variables such as nominal wages, price levels, and import shares, across countries of different size. We first only impose Assumption 1 (technology scales with population) and then we also explore the case with Assumption 2 (symmetric regions).

A. *Calibration Procedure*

We consider a set of 26 OECD countries for which all the variables needed are available. Additionally, we restrict the sample to this set of countries to ensure that the main differences across countries are dominated by size, geography, and R&D, rather than other variables outside the model. Importantly, as explained in detail below, among this set of countries, the definition in the data for "region" is fairly homogeneous.

We need to calibrate the parameter θ, the variables M_n and ϕ_n for all n, and l_m for all m, as well as the matrix of trade costs d_{mk}, for all m, k.

Parameter θ.—The value of θ is critical for our exercise. We turn to the growth literature to calibrate this parameter. Assuming that l_m grows at a constant rate $g_L > 0$ in all countries and invoking Assumption 1, the growth rate of t_m is equal to g_L, for all regions and countries. The long-run income growth rate is then $g = g_L/\theta$. With $g_L = 0.048$ (the growth rate of research employment), and $g = 0.01$ (the growth rate of TFP) among a group of rich OECD countries, both from Jones (2002), it follows that $\theta = 4.8$. Jones and Romer (2010) follow a similar procedure and conclude that the data support $g/g_L = 0.25$, which implies $\theta = 4$. We choose $\theta = 4$, which is also in the range of estimates from the trade literature.[9] Notice that even though our model is fully consistent with estimates of θ provided by the growth literature, it is not fully consistent with estimates provided by the trade literature since our general model does not deliver a log-linear gravity equation at the country level, except for the case of symmetric regions.

Number of Regions.—We assume that the number of regions for each country in the model, M_n, equals the number of metropolitan areas observed in the data, for each country. We use data on 287 metropolitan areas with a population of 500,000 habitants or more. For all countries, except Australia, New Zealand, Turkey, and Iceland, the data are from the OECD Metropolitan Areas Database; for these four remaining countries, we use data from the OECD Regional Database.[10] The data on population are for the year 2000. Column 8 in Table 1 presents the number of regions for each country. The number of metropolitan areas is strongly correlated with our measure of country size defined below (the correlation is 0.90).[11]

[9] Head and Mayer (2014) survey the estimates for the trade elasticity in the literature and conclude that, even though the variance is large, the mean estimate, for the subset of structural gravity estimates, is -3.78. Among the most recent estimates, Simonovska and Waugh (2014) place θ between 4 and 5.

[10] The OECD has developed a harmonized definition of urban areas, which includes metropolitan areas, to overcome limitations linked to administrative definitions (OECD 2012). An urban area is defined as a functional economic unit characterized by densely inhabited "urban cores" and "hinterlands" whose labor market is highly integrated with the cores.

[11] The population accounted for by our sample of metropolitan areas ranges from 25 and 27 percent for Norway and Finland, respectively, to 72 and 82 percent for Korea and Benelux, with 53 percent for the United States and

Table 1—Data Summary

	Domestic trade (1)	RGDP p.c. (2)	CGDP p.c. (3)	Price index (4) = (3)/(2)	R&D emp. (5)	Equipped labor (6)	Country size (7) = (5) × (6)	Number of regions (8)
Australia	0.83	0.97	0.78	0.80	0.68	7.92	0.05	8
Austria	0.44	1.12	1.09	0.98	0.49	2.92	0.01	3
Benelux	0.20	1.16	1.08	0.94	0.58	9.30	0.05	9
Canada	0.49	0.86	0.68	0.79	0.63	13.99	0.08	9
Switzerland	0.51	0.88	1.12	1.27	0.60	3.60	0.02	3
Denmark	0.42	0.94	1.15	1.22	0.63	2.25	0.01	1
Spain	0.71	1.14	0.83	0.73	0.38	10.76	0.04	8
Finland	0.68	0.84	0.92	1.09	1.23	2.06	0.02	1
France	0.68	1.07	1.08	1.01	0.59	20.08	0.10	15
Germany	0.70	0.92	0.96	1.04	0.60	33.73	0.18	24
Greece	0.63	0.90	0.63	0.70	0.30	2.90	0.01	2
Hungary	0.48	0.65	0.28	0.44	0.29	2.47	0.01	1
Ireland	0.28	1.32	1.25	0.95	0.51	1.04	0.005	1
Iceland	0.40	1.17	1.09	0.93	0.96	0.11	0.001	1
Italy	0.78	1.20	1.07	0.89	0.29	16.73	0.04	11
Japan	0.94	0.72	0.98	1.37	0.80	66.31	0.46	36
Korea	0.83	0.63	0.44	0.71	0.51	16.04	0.07	10
Mexico	0.63	0.78	0.40	0.51	0.06	16.60	0.01	33
Norway	0.57	1.11	1.07	0.97	0.81	2.21	0.02	1
New Zealand	0.64	0.74	0.61	0.82	0.48	1.48	0.01	1
Poland	0.68	0.50	0.24	0.48	0.32	10.07	0.03	8
Portugal	0.59	0.97	0.70	0.73	0.33	2.48	0.01	2
Sweden	0.59	0.81	0.97	1.20	0.83	3.90	0.03	3
Turkey	0.74	0.61	0.26	0.43	0.07	10.83	0.01	11
United Kingdom	0.64	1.00	0.98	0.98	0.53	20.83	0.10	15
United States	0.85	1.00	1.00	1.00	0.87	130.10	1.00	70

Notes: Domestic trade in manufacturing (column 1) is calculated as a share of absorption in manufacturing. RGDP p.c. (column 2) is PPP-adjusted real GDP divided by equipped labor (in column 6, in million of units). CGDP p.c. (column 3) is GDP in current US dollars divided by equipped labor. The price index (column 4) is GPP in current US dollars divided by real GDP. R&D employment (column 5) is calculated as a percentage of total employment. Column 8 shows the number of metropolitan areas with more than 500,000 habitants. Real and current GDP per capita, as well as the price index and country size, are relative to the United States. Variables are averages over the period 1996–2001.

Technology and Country Size.—We calibrate the parameter t_m imposing Assumption 1, $t_m = \phi_n l_m$.[12] We assume that the variable ϕ_n varies directly with the share of R&D employment observed in the data at the country level (since data on R&D by region are either very low quality or unavailable). We use data on R&D employment from the World Development Indicators averaged over the nineties. The size elasticity of R&D employment shares, for our sample of countries, is statistically indistinguishable from zero (−0.07, standard error 0.09), indicating that there is no systematic pattern between R&D intensity and size in our sample.[13]

68 percent for Japan. For some countries, these shares may seem low, but one has to keep in mind that these are the major urban areas (and their hinterlands) which concentrate most of the manufacturing activity and workers.

[12] Interestingly, we do find some evidence of scale effects operating at the level of metropolitan statistical areas (MSA), for the United States: the elasticity of real GDP per capita (from the Bureau of Economic Analysis) to population (from the census) across 349 MSAs, for 2007, is 0.12 (standard error 0.011). This evidence not only supports imposing Assumption 1 at the levels of metropolitan areas, but also the assumption that trade is frictionless within those same geographic units.

[13] Using the number of patents per unit of equipped labor registered by country n's residents, at home and abroad, from the World Intellectual Property Organization (average over the period 2000–2005), rather than R&D employment shares, as a proxy for ϕ_n, does not change our results below. Similarly to R&D employment shares, small countries do not have a systematically higher number of patents per capita.

We pair l_m in the model to equipped labor in the data to account for differences in physical and human capital per worker. This variable is available at the country level from Klenow and Rodríguez-Clare (2005). We treat equipped labor in region m as the product of country-level equipped labor (L_n) and the population share of region m in country n. The population share of region m is the population of a metropolitan area in our sample as a share of the total population in our sample of metropolitan areas belonging to country n.[14]

We refer to the term $\phi_n L_n$ as R&D-adjusted country size, and we adopt it as our measure of country size. The values for this variable are shown in Column 7 of Table 1.

Trade Costs.—We need to calibrate the whole matrix of trade costs between regions, d_{mk}, for $m \in \Omega_i$ and $k \in \Omega_n$, for all i, n. This amounts to a 287×287 matrix, i.e., $\left(\sum_n M_n\right)^2$. The obvious limitation is that data on trade flows between any two regions in our sample are not available, except for the United States and Canada. Hence, we proceed by imposing more structure on the trade costs and assume that

$$d_{mk} = \beta_0^{\iota_{mk}} \beta_1^{1-\iota_{mk}} dist_{mk}^{\beta_2 \iota_{mk} + \beta_3 (1-\iota_{mk})}, \tag{13}$$

with $d_{mm} = 1$. The variable $dist_{mk}$ denotes geographical distance between region m and k which is computed from longitude and latitude data for each metropolitan area in our sample. The variable ι_{mk} is a dummy variable that equals 1 if m and k belong to the same country, and 0 otherwise.

The distance elasticities within and across countries, β_2 and β_3, respectively, are disciplined invoking ordinary least squares (OLS) estimates of gravity equations. These estimates indicate that the distance elasticities for inter-regional and international trade flows are in the same range. Table 2 presents the results for different gravity specifications. For our sample of 26 countries, the distance elasticity ranges from -1.01 to -1.1. These estimates are within the range estimated in the literature, as surveyed by Head and Mayer (2014). Using data on trade flows among US regions, we get a distance elasticity that ranges from -1.02 to -1.06, as shown in columns 3–5 in Table 2. Allen and Arkolakis (2014) estimate a reduced-form distance elasticity between -0.95 and -1.35, using 122 CFS regions within the United States, when trade is restricted to road mode, for 2007. Tombe and Winter (2014) estimate a trade-elasticity of -1.25, for inter-provincial trade in Canada, for 2005.[15] Given this evidence, we impose $\beta_2 = \beta_3 = \beta_{dist}$, and target a trade-distance elasticity of -1.05.[16] For $\theta = 4$, the implied coefficient is $\beta_{dist} = 0.27$.

[14] Effectively, we are assigning the *total* country-level equipped labor to the regions *in our sample*, for each country, proportionally to their importance in terms of population.

[15] Using data for trade flows among regions of the European Union, as estimated by Thissen, Diodato, and van Oort (2013), and including a same-country dummy in the regression, we estimate a trade-distance elasticity of -1.14 (standard error 0.008).

[16] Since our general model does not deliver a log-linear gravity equation at the country level, we cannot take these trade-distance elasticities directly from the data and impose them in (13) using $\beta_{dist} = \beta_{dist}^{ols}/\theta$. Instead, given an initial guess for β_{dist}, and using $\theta = 4$, we compute the model's equilibrium, generate data on trade flows, and estimate the same gravity equation as in the data. We then iterate on β_{dist} until the trade-distance elasticity in the model is -1.05.

Table 2—Gravity Estimates

	International		US domestic			
	(1)	(2)	(3)	(4)	(5)	(6)
log distance	−1.084	−1.009	−1.038	−1.065	−1.023	−1.058
	(0.05)	(0.06)	(0.019)	(0.0165)	(0.0182)	(0.0173)
Common int. border		0.126				
		(0.13)				
Common language		0.38				
		(0.10)				
R^2	0.987	0.999	0.985	0.987	0.987	0.988
Observations	650	650	2,220	3,737	2,080	2,215

Notes: The dependent variable in columns 1 and 2 is (in logs) $\lambda_{ni} \equiv X_{ni}/X_n$ for $n \neq i$, for 26 OECD countries, while in columns 3 to 6 is (in logs) $x_{mk}/\sum_{r\in\Omega_n} x_{rm}$, for $m \neq k$, and $n = USA$. Regions in the United States refer to: in column 3, 55 metropolitan areas (present in both CFS and OECD data), for 2007; in column 4, 74 metropolitan and consolidated statical areas, from CFS, for 2007; and 51 states in columns 5 and 6, for 2002 and 2007, respectively. All regressions with importer and exporter fixed effects. Robust standard errors in parentheses.

The coefficients β_0 and β_1 are chosen to jointly match the share of total intraregional trade in total domestic trade for the United States ($\sum_{m\in\Omega_n} x_{mm}/X_{nn}$ in the model), and the average country-level bilateral trade shares in manufacturing observed in the data.

We use data from the Commodity Flow Survey (CFS) on manufacturing trade flows between subnational units within the United States. Intratrade shares range from 0.35 when we use 100 geographical areas, to 0.55 when we use the 55 metropolitan areas which are also included in the OECD dataset, for 2007.[17] We target a mid-value of 0.45, implying that 45 percent of domestic US trade flows are intraregional.

Data on country-pair level trade flows X_{ni} are from the OECD's STAN database, averaged over 1996–2001, while country-level absorption X_n is calculated (from the same source) as gross production minus total exports plus total imports from countries in our sample. In our sample, the average international (bilateral) trade share is 0.0156. The resulting coefficients are $\beta_0 = 2.33$ and $\beta_1 = 2.89$.[18]

Results.—The calibrated model captures well the patterns of trade observed in the data. In particular, the model captures 96 percent of the variation observed in the data on bilateral country-level trade shares; the correlation coefficient between country-level import shares in the data and the model is 0.80, and 0.65 if instead

[17] There is a discrepancy between the definition of metropolitan areas in the OECD and the CFS for the United States: of the 70 metropolitan areas recorded in the OECD dataset, only 55 can be matched with the ones in the CFS for which trade data are available. Our 100 geographical areas include 48 consolidated statistical areas (CSA), 18 metropolitan statistical areas (MSA), and 33 units represent the remaining portions of (some of) the states. For each of these 99 geographical units, we compute the total purchases from the United States and subtract trade with the 99 geographical units to get trade with the rest of the United States, which is considered the one-hundredth geographical unit.

[18] We are not including exogenous international trade imbalances in our calibration. Those imbalances may affect the terms of trade and the real wage, but since they are typically small, as a share of GDP, and not systematically correlated with country size, adding them to the standard model with no domestic trade costs does not change the results noticeably: implied scale effects remain basically the same. We strongly conjecture that the same would still be true in our model where each country is composed of multiple regions and domestic trade costs are positive. Doing this exercise formally, however, would be extremely difficult because we would need to assign country-level trade imbalances across the different regions in a country, and those data are simply not available.

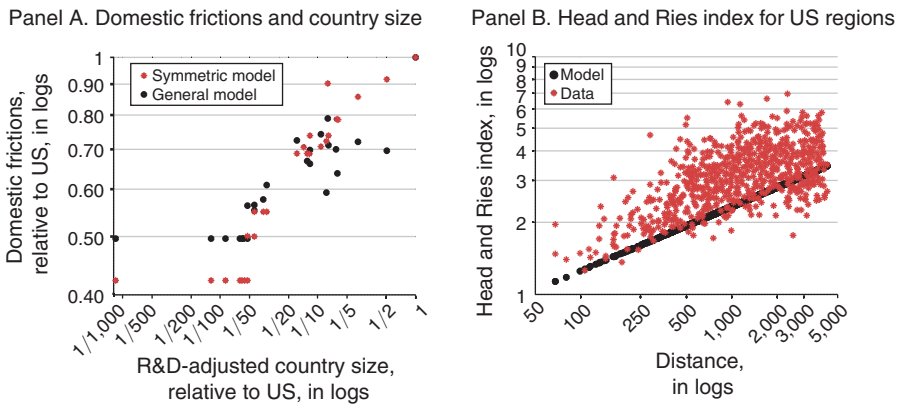

FIGURE 2. DOMESTIC TRADE COSTS: CALIBRATION

Notes: Panel A: domestic frictions in symmetric model refer to τ_{nn} calculated using (11); in general model, they refer to τ_{nn} calculated applying the aggregation in (14) to the calibrated domestic costs d_{mk}. R&D-adjusted country size refers to $\phi_n L_n$, where ϕ_n is the share of R&D employment and L_n is equipped labor. Panel B: The Head and Ries index refers to: in the data, $d_{mk}^{hr} \equiv \left((x_{mk}\, x_{km})/(x_{kk}\, x_{mm})\right)^{-1/(2\theta)}$; in the model, the expression in (13). Both variables are calculated for 55 metropolitan areas in the United States.

export shares are considered; and the share of total intraregional trade in total domestic trade for Canada—the other country for which we have these trade data—is 0.77 in the model and 0.79 in the data (computed using trade between 13 Canadian provinces, for 2007, from British Columbia Statistics).[19]

What are the implications of our calibrated model for domestic trade costs? For expositional purposes only, we aggregate our calibrated domestic trade costs at the country level using a procedure in Agnosteva, Anderson, and Yotov (2014),

$$\tau_{nn} = \sum_{m \in \Omega_n} (l_m/L_n) \left(\sum_{k \neq m,\, k \in \Omega_n} (l_k/L_n) d_{mk}^{-\theta} \right)^{-1/\theta}. \tag{14}$$

Figure 2 shows this index against our measure of country size, relative to the United States. Larger countries have larger domestic trade costs among its regions: the correlation between τ_{nn} and our measure of country size $(\phi_n L_n)$ is 0.70, and 0.86 if we consider the number of regions in each country, M_n. For instance, while a small country like Denmark has $\tau_{DNK,DNK}$, almost one-half of the US level, a large country like Japan has $\tau_{JPN,JPN}$ of around 70 percent of the US level. The six smallest countries in our sample have domestic costs that are one-half of the US level, while the corresponding number for the six largest countries is 75 percent. The fact that domestic trade costs are increasing in country size already suggests that these costs will undermine the strength of aggregate scale effects; by how much is what we analyze next.

Finally, our estimates of domestic trade costs can be compared with estimates coming from the index developed by Head and Ries (2001) applied to domestic

[19] If we use the data on regional trade flows for the European Union, estimated by Thissen, Diodato, and van Oort (2013), to compute the intraregional share for the 15 EU countries in our sample, our calibrated model captures 90 percent of the variation observed in that data.

trade for the United States. Using (1), and assuming symmetric trade costs between any two regions, $(d_{mk}^{hr})^{-2\theta} \equiv (x_{mk}x_{km})/(x_{kk}x_{mm})$. Panel B of Figure 2 shows trade costs across regions of the United States against distance, using our calibrated trade costs and the Head-Ries index for the subset of 55 US metropolitan areas for which trade flows are available from the CFS, for 2007, and $\theta = 4$. While the distance elasticity for the model's domestic costs is 0.27, as calibrated above, the one for the domestic costs calculated using the Head and Ries index is 0.28 (including origin and destination fixed effects). Additionally, d_{mk}^{hr} has a higher mean than d_{mk} (3.29 versus 2.55), suggesting that our estimates of domestic trade costs, at least for the United States, are on the conservative side.

B. *The Role of Domestic Trade Costs*

We use the calibrated model to explore how the presence of domestic trade costs affects the strength of scale effects on real wages. We ask two related questions: does the model with scale effects, international trade, and domestic trade costs match better the cross-country data on real wages vis-à-vis the model with no domestic trade costs? By how much do domestic trade costs contribute to match the data on cross-country income? We also evaluate both models on their prediction with respect to import shares, prices, and nominal wages, across countries of different size.

In the data, the real wage is computed as real GDP (purchasing power parity adjusted, or PPP-adjusted) from the Penn World Tables (7.1), divided by our measure of equipped labor so that it is simply TFP. In turn, import shares are calculated using international trade data from STAN; nominal wages are GDP at current prices, from the World Development Indicators, divided by our measure of equipped labor; and price indices are simply nominal wages divided by real wages. All variables are averages over the period 1996–2001 and are summarized in Table 1.

Figure 3 shows a decomposition of the real wage (relative to the US) implied by: the full model (dots); the model with only scale effects (crosses) which implies imposing $\beta_0 = 1$, $\beta_1 = \infty$, and $\beta_2 = 0$; the model with both scale effects and international trade but no domestic trade costs (squares) which implies imposing $\beta_0 = 1$ and $\beta_2 = 0$; and the data (stars). Real wages are plotted against our measure of R&D-adjusted country size. Table A1 in the online Appendix shows the numbers behind the figure.

The model with only scale effects severely underestimates the real wage for the smallest countries (crosses versus stars): the real wage for a small country like Denmark is predicted to be only 33 percent of the level in the United States, reflecting very strong scale effects at work. In contrast, the observed relative real wage of Denmark is 94 percent. For the six smallest countries in our sample, the model with only scale effects implies a relative real wage of 30 percent, whereas in the data these countries have an average real wage almost equal to the one in the United States. This is a reflection of a very high income-size elasticity in the model ($1/\theta = 0.25$) in comparison with the one in the data which is not statistically different from zero (-0.006, standard error 0.03).

How much does adding international trade and domestic trade costs to the model help in matching the observed real wages for countries of different sizes? As the squares indicate in Figure 3, adding trade openness, but no domestic trade costs,

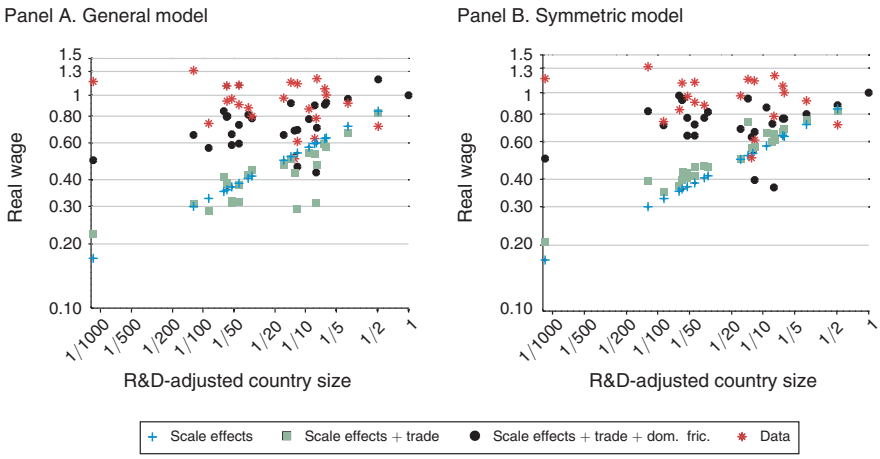

FIGURE 3. SCALE EFFECTS, TRADE OPENNESS, AND DOMESTIC TRADE COSTS (*Relative to US in logs*)

Note: R&D-adjusted country size refers to $\phi_n L_n$, where ϕ_n is the share of R&D employment and L_n is equipped labor.

does not help much in bringing the model closer to the data. For example, the relative real wage for Denmark is 38 percent, only a small improvement over the model with only scale effects. A similarly small improvement is obtained for the six smallest countries (from 30 to 33 percent), while the income-size elasticity is 0.22 (standard error 0.007).

It is important to clarify that, as expected, small countries do gain more from trade than large countries (see column 6 in Table A1 in the online Appendix). It is just that these gains are not large enough to have a substantial effect in closing the gap between the model with only scale effects and the data. Continuing with our previous example, Denmark has much larger gains from trade than the United States (22 versus 2.2 percent), but this almost ten-fold higher gains only increase its implied relative real wage from 33 to 38 percent.[20]

Comparing the dots and stars in Figure 3 reveals that the main contribution in getting the full model to better match the data comes from adding domestic trade costs. The mechanism comes from the fact that domestic trade costs are higher for larger countries, and hence they undermine scale effects. Again, for Denmark, our full model implies a relative real wage of 80 percent, much closer to the 94 percent observed in the data than the 33 percent implied by the model with only scale effects. A similar pattern is found for the six smallest countries in the sample: the full model is able to close by more than one-half the gap in the real wage between the data and the model with only scale effects, while openness to trade only closes such gap by around four percent. More generally, inspecting the income-size elasticity implied by the full model reveals that, even though still significantly positive (0.13, standard

[20] The gains from trade are calculated, in the general model with domestic frictions, as the average change in real wages, across regions in country n, from autarky to openness to trade, weighted by the population share of each region. See Ramondo, Rodríguez-Clare, and Saborío-Rodríguez (2014) for a theoretical derivation of these gains under the general case in which labor is mobile across regions.

error 0.02), it is one-half of the elasticity implied by the model with only scale effects (0.25).

Figure 4 compares *calibrated* versions of the models with and without domestic trade costs regarding real and nominal wages, import shares, and price indices, across countries.[21] Panel A makes clear that in the calibrated model with no domestic trade costs, real wages rise too rapidly across countries of different size: the size elasticity of the real wage is 0.20 (standard error 0.01), much higher than the zero elasticity observed in the data and double the one delivered by the model with domestic trade costs. Panels B and C show that the behavior of the real wage in the model with no domestic trade costs is the result of nominal wages that rise—and prices that fall—too rapidly with size. The model with no domestic trade costs implies a size elasticity for the nominal wage and price index that are too high (in absolute value) relative to the ones in the data. Both elasticities are halved as we allow for domestic trade costs. Table A2 in the online Appendix reports the size elasticities and averages for each variable.

Even though both calibrated models match well the average import share in the data, the model with no domestic trade costs implies import shares that decrease too rapidly with country size (the size elasticity almost doubles the one in the data), while adding domestic trade costs decreases the trade-size elasticity by one-third, as suggested by our theoretical example in Figure 1.

C. *Symmetry*

The model with symmetric regions (Assumption 2) allows for a simpler calibration. This is because, as shown in Proposition 2, the model delivers a log-linear gravity relationship for trade flows at the country level, a simple expression for internal trade costs in each country, and a simple formula for country-level gains from trade as in Arkolakis et al. (2012). We directly use (12) to compute the model-implied real wage, (8) to compute the real wage implied by the standard model without domestic trade costs, and $U_n = \gamma(\phi_n L_n)^{1/\theta}$ for the model with only scale effects. We keep the value of θ at 4, which can now be *directly* obtained from estimates of the trade elasticity in the trade literature. We use the same values for M_n, $L_{n,}$, and ϕ_n as above and set $T_n = \phi_n L_n$ directly from (10).

The only task left is to calibrate δ_n, which can then be combined with our calibrated values for θ and M_n to compute τ_{nn} using (11). We use data on domestic trade flows for the United States and the model's equilibrium equation $\tau_{nn}^{\theta} = M_n \sum_{m \in \Omega_n} x_{mm}/X_{nn}$, which can be obtained from manipulating (1) under Assumption 2. As above, we set $\sum_{m \in \Omega_n} x_{mm}/X_{nn} = 0.45$ for the United States, which together with $M_{US} = 70$ and $\theta = 4$ implies $\tau_{US,US} = 2.37$ and hence, using (11), $\delta_{US} = 2.75$.

Lacking data on domestic trade flows for other countries in our sample, we impose $\delta_n = \delta_{US}$ for all n. Note that we are still allowing for differences in τ_{nn} across countries that come from differences in the number of regions in each country, M_n; this is precisely what will offset the economies of scale in the symmetric model with domestic trade costs.

[21] The model without domestic trade costs is calibrated using the procedure described above, but assuming $d_{mk} = 1$ for all $m, k \in \Omega_n$. The R^2 between bilateral trade shares in the data and the calibrated model is 0.94.

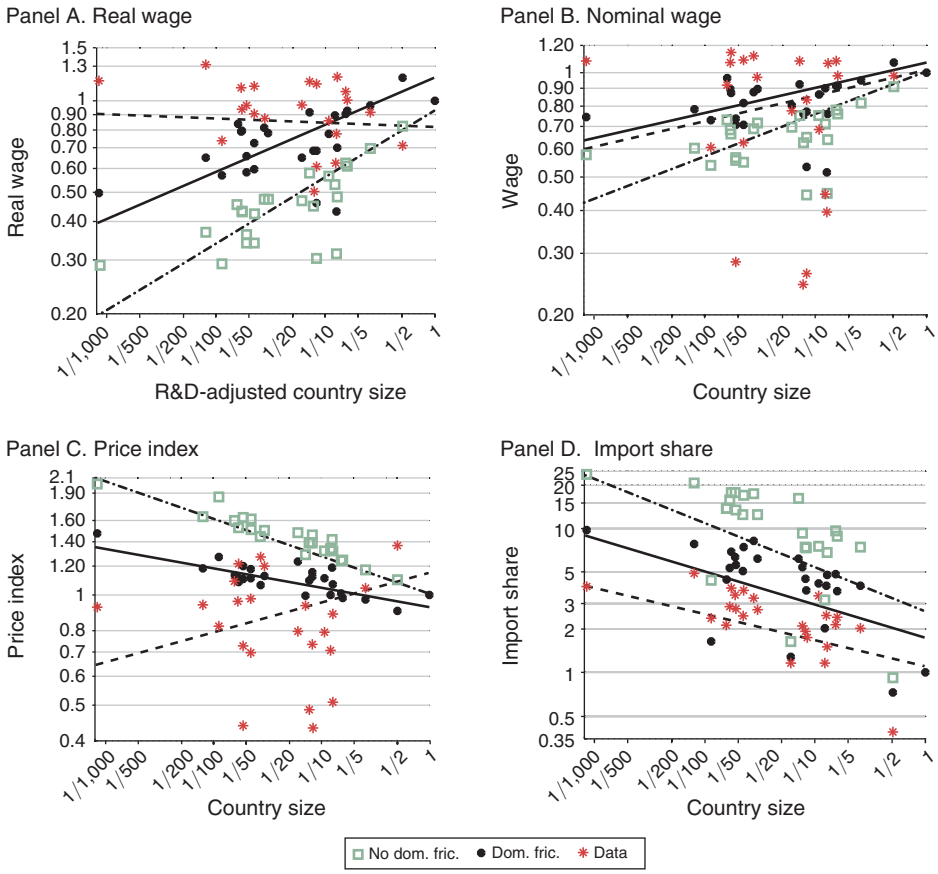

FIGURE 4. CALIBRATED MODEL AND DATA (*Relative to US in logs*)

Notes: No dom.fric. refers to the model without domestic trade costs; dom. fric. refers to the model with domestic trade costs. In the data: the real wage is computed as real GDP (PPP-adjusted) divided by equipped labor, L_n; the nominal wage is calculated as GDP at current prices divided by equipped labor, L_n; the price index is calculated as the nominal wage divided by the real wage; and import shares refer to total imports, as share of absorption, in the manufacturing sector. R&D-adjusted country size refers to $\phi_n L_n$, where ϕ_n is the share of R&D employment and L_n is equipped labor. The dashed line is fitted through the data; the solid line is fitted through the model with domestic frictions; and the dot-dashed line is fitted through the model without domestic frictions.

Panel A of Figure 2 shows the calibrated τ_{nn}'s, relative to the United States, against country size (stars). Estimates are remarkably similar to the ones stemming from our general model (circles). On average, domestic trade costs calibrated under Assumption 2 are not very different from the ones calibrated using the general model (0.64 versus 0.63), but their size elasticity is higher (0.14 versus 0.09).

Inspecting panel B of Figure 3 reveals that the model with symmetric regions delivers a similar decomposition to the one obtained using the general model.[22] The size-elasticity implied by the symmetric model is 0.08 (standard error 0.02), a bit

[22] There is a slight difference in the nature of the exercises behind panels A and B of Figure 3 regarding how the case with trade but no domestic trade costs is computed. In panel A, trade shares result from taking the calibrated model and shutting down domestic trade costs, so that the resulting trade shares do not necessarily match the ones in the data. In contrast, in panel B, trade shares for the case with no domestic trade costs are exactly the ones in the data because the real wage is calculated directly, using (12). One may prefer to refer to the exercise in panel A as a

lower than for the general model (0.13), and one-third of the size elasticity implied by the model with only scale effects (0.25).[23]

III. Discussion

In this section, we discuss two issues: how our calibration procedure and results are related to other calibrations of trade models in the literature, namely Eaton and Kortum (2002); Alvarez and Lucas (2007); and Waugh (2010), and how the introduction of nontradable goods, multinational production, and diffusion may affect our results regarding scale effects.

A. *Domestic Trade Costs versus International Trade Cost Asymmetries*

The effects of domestic trade costs could be replicated in a model without these costs but in which international trade costs and innovation intensity were chosen appropriately. As we show next, the key is whether one allows for asymmetries in international trade costs, and whether one deviates from Assumption 1 by allowing for a systematic pattern between innovation intensity and country size.

We compare the implications of three models that differ in terms of the assumptions on trade costs: symmetric international trade costs with domestic trade costs (RRS), as in this paper; asymmetric international trade costs with asymmetries arising from importer-specific terms (EK), as in EK; and asymmetric international trade costs with asymmetries arising from exporter-specific terms (W), as in Waugh (2010). To proceed, let $\alpha_{ni} = \alpha_{in}$ for all $i \neq n$ be the symmetric component of trade costs and consider the following alternative assumptions for trade costs:

ASSUMPTION 4A (Symmetric Trade Costs with Domestic Frictions): $\tau_{ni}^{RRS} = \alpha_{ni}$ *for all* $i \neq n$, *and* τ_{nn}^{RRS} *as in* (11).

ASSUMPTION 4B (Trade Costs with Asymmetries from Importer Effects): $\tau_{ni}^{EK} = F_n^{EK}\alpha_{ni}$ *for all* $i \neq n$ *and* $\tau_{nn}^{EK} = 1$ *for all* n.

ASSUMPTION 4C (Trade Costs with Asymmetries from Exporter Effects): $\tau_{ni}^{W} = F_i^{W}\alpha_{ni}$ *for all* $i \neq n$ *and* $\tau_{nn}^{W} = 1$ *for all* n.

We use the same parameter θ and the same country sizes, L_n for all three models, but we allow for differences in technology levels and trade costs. The RRS model has technology levels T_n^{RRS} satisfying Assumption 1, domestic trade costs as implied by Assumption 2, and international trade costs satisfying Assumption 4A.

decomposition of forces in the calibrated model, and to the exercise in panel B as a comparison of models calibrated to the same data.

[23] Even under Assumption 2, exploring the implications of domestic trade costs on nominal wages, prices, and trade shares requires calibrating the full matrix of international trade costs. This is done in a way that mimics our calibration for the general case (see the online Appendix). The results are very similar to the ones for the general model: see Figure A1 and Table A2.

The EK model has the same technology levels as the RRS model, $T_n^{EK} = T_n^{RRS}$, and trade costs satisfying Assumption 4B with $F_n^{EK} = 1/\tau_{nn}^{RRS}$. The W model has technology levels $T_n^W = T_n^{RRS}\left(\tau_{nn}^{RRS}\right)^{-\theta}$ and trade costs satisfying Assumption 4C with $F_n^W = 1/\tau_{nn}^{RRS}$.

Proposition 4 follows directly from the expression for trade flows in (4) and price levels in (5).

PROPOSITION 4: *The RRS, EK, and W models generate the same equilibrium wages and trade flows. The equilibrium price levels are the same for the RRS and W models, but they differ in the EK model.*

One can further show that implied prices in EK are systematically high in small countries when compared with implied prices in the RRS and W models, since $P_n^{EK} = P_n^{RRS}/\tau_{nn}^{RRS}$ and τ_{nn}^{RRS} increases with size.[24] This point is analogous to the one made by Waugh (2010), but applied here to large versus small as opposed to rich versus poor countries.

While Proposition 4 says that models W and RRS have the same implications for trade flows, wages, and prices, the mechanisms are different. To see this, note that $T_n^W = T_n^{RRS}\left(\tau_{nn}^{RRS}\right)^{-\theta}$ combined with the fact that τ_{nn}^{RRS} is increasing in country size implies that T_n^W increases less steeply with L_n than T_n^{RRS}. In other words, larger countries are not richer in model W because their T_n/L_n falls with L_n, whereas in model RRS this happens because τ_{nn}^{RRS} increases with L_n.

We now compare models W and RRS quantitatively by using the expression in (12), $U_n = T_n^{1/\theta} \times \lambda_{nn}^{-1/\theta} \times \tau_{nn}^{-1}$, together with data for domestic trade shares and real wages to compute the values of T_n. We set $\tau_{nn} = 1$ to compute $\phi_n^W \equiv T_n^W/L_n$, and $\tau_{nn} > 1$ as calibrated in Section C, to compute $\phi_n^{RRS} \equiv T_n^{RRS}/L_n$, where L_n is equipped labor from the data. We use the results to plot ϕ_n^W (squares) and ϕ_n^{RRS} (dots), both relative to innovation intensity in the data (ϕ_n^{data})—proxied by R&D employment shares—against our R&D-adjusted measure of country size ($\phi_n^{data}L_n$) in Figure 5.

The figure shows that the dashed line is significantly steeper than the solid line; the size elasticity for ϕ_n^W/ϕ_n^{data} is -0.92 (standard error 0.10), while the one for $\phi_n^{RRS}/\phi_n^{data}$ is -0.34 (standard error 0.13). In other words, to be consistent with the data, a model without domestic trade costs requires research intensities to decrease rapidly with country size, while this relationship is mitigated when domestic trade costs are included. As the figure clearly shows, research intensities in the data present a much flatter relationship across countries of different size than the ones implied by the model with no domestic trade costs.[25] Of course, the finding that $\phi_n^{RRS}/\phi_n^{data}$ also decreases with country size is just the flip side of the finding in Section C that domestic trade costs do not fully reconcile the model with the data when we impose Assumption 1.

[24] It is not possible to achieve a full equivalence between RRS and EK by deviating from $T_n^{EK} = T_n^{RRS}$, since the only way in which (4) holds for the two models is by imposing $T_n^{EK} = T_n^{RRS}$ and $F_n^{EK} = 1/\tau_{nn}^{RRS}$.

[25] Results are similar if research intensities are proxied by patents per capita, rather than by R&D employment shares.

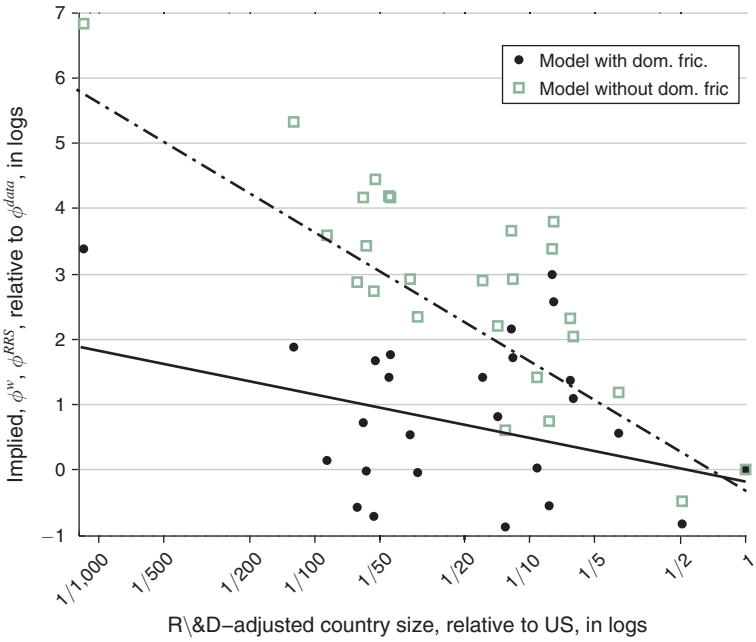

FIGURE 5. RESEARCH INTENSITIES AND SIZE: MODEL AND DATA

Notes: $\phi^{s}_{n} \equiv T^{s}_{n}/L_n$, with $s = RRS$ for the model with domestic trade costs ($\tau_{nn} > 1$), and $s = W$ for the model without domestic trade costs ($\tau_{nn} = 1$). Elasticities are -0.34 (solid line), and -0.92 (dashed line), significant at 5 percent and 1 percent, respectively. The variable ϕ^{data}_{n} is the share of R&D employment and L_n is equipped labor. R&D-adjusted country size refers to $\phi^{data}_{n} L_n$.

We finish this subsection by briefly relating our results to those in Alvarez and Lucas (2007), whose calibrated model without domestic trade costs matches fairly well the relationship between size and import shares across countries. Like us, they impose Assumption 1 such that there is no variation in innovation intensity with country size, but they calibrate country size to match nominal GDP in the data. This is equivalent to allowing for differences in efficiency per unit of equipped labor across countries. For our sample of countries, the resulting efficiency parameters are decreasing in country size (i.e., equipped labor) with an elasticity of 0.10 (significant at 1 percent). This implies that small countries have a significantly higher efficiency per unit of equipped labor than large countries. Again, this is hard to defend given that small countries are not systematically better in terms of variables related to higher efficiency, such as schooling levels, corruption in government, rule of law, and bureaucratic quality.[26]

[26] The size elasticities, calculated using OLS (and robust standard error), are nonsignificant for all the variables, except for bureaucratic quality (0.06, standard error 0.02). The data are the average years of schooling from Barro and Lee (2000), and indices ranging from zero (worst) to six (best) on corruption in government, rule of law, and bureaucratic quality, from Beck et al. (2001).

B. *Nontradable Goods, Multinational Production, and Diffusion*

The benchmark model introduced in Section I that we calibrate in Section II ignores intermediate goods, nontradable goods, multinational production, and international technology diffusion. How do these considerations affect our conclusions regarding scale effects?

Consider an extension of our model with final and intermediate goods in which (a CES aggregate of) intermediate goods are used in the production of each intermediate and final good. Allowing for scale effects as in our baseline model (Assumption 1), assuming that regions are symmetric, as in Assumption 2, assuming that both types of goods are tradable, and using superscript g for intermediates and superscript f for final goods, the real wage can be written as

$$(15)\quad U_n = \tilde{\gamma} \times \left(\phi_n L_n\right)^{(1+\eta)/\theta} \times \left(\tau_{nn}^g\right)^{-\eta} \times \left(\tau_{nn}^f\right)^{-1} \times \left(\lambda_{nn}^g\right)^{-\eta/\theta} \times \left(\lambda_{nn}^f\right)^{-1/\theta},$$

where

$$(16)\qquad \tau_{nn}^s = \left(\frac{1}{M_n} + \frac{M_n - 1}{M_n}\left(\delta_n^s\right)^{-\theta}\right)^{-1/\theta}, \text{ for } s = g, f,$$

and η is a parameter that captures the importance of intermediate goods in the production of both intermediate and final goods and is smaller than 1.[27]

If there were no domestic trade costs (i.e., $\tau_{nn}^g = \tau_{nn}^f = 1$) and final goods were nontradable across countries (i.e., $\lambda_{nn}^f = 1$), as in Alvarez and Lucas (2007), equation (15) would collapse to

$$U_n = \tilde{\gamma} \times \left(\phi_n L_n\right)^{(1+\eta)/\theta} \times \left(\lambda_{nn}^g\right)^{-\eta/\theta}.$$

Calibrating θ such that we target a growth rate of one-fourth, as in the baseline model (i.e., $(1+\eta)/\theta = 1/4$), this expression implies that scale effects are as strong as before but the gains from trade are lower ($\eta/\theta < 1/\theta$), implying an even bigger puzzle for the standard model.

Suppose instead that final goods are also nontradable among regions within a country. This is a version of the idea that domestic trade costs weaken scale effects. With $\delta_n^f \to \infty$, $\tau_{nn}^f = (1/M_n)^{-1/\theta}$. Assuming that $l_m = \bar{l}$, for all m, so that all variation in country size comes from the number of regions, we have

$$U_n = \tilde{\gamma} \times \left(\phi_n \bar{l}\right)^{1/\theta} \times \left(\phi_n L_n\right)^{\eta/\theta} \times \left(\lambda_{nn}^g\right)^{-\eta/\theta}.$$

Not surprisingly, assuming that final goods are nontradable across regions within countries implies weaker scale effects than in the baseline model ($\eta/\theta < 1/\theta$).

One problem with the assumption that final goods are nontradable is that we are missing the possibility of multinational production (MP), which is another source of openness, and relevant for final goods even among regions within a country. The

[27] Formally, as in Alvarez and Lucas (2007), $\eta = (1-\alpha)/\beta$, where α and β are the (Cobb-Douglas) shares of labor in final and intermediate good production.

simplest way to proceed is to allow for MP but not trade in final goods and for trade but not MP in intermediate goods.[28] Following the analysis in Ramondo and Rodríguez-Clare (2013), and assuming for simplicity that domestic costs and shares for trade and MP are the same, the real wage would be

$$U_n = \tilde{\gamma}(\phi_n L_n)^{(1+\eta)/\theta} \times \tau_{nn}^{-(1+\eta)} \times \lambda_{nn}^{-(1+\eta)/\theta}.$$

With $(1+\eta)/\theta = 1/4$, as before, the strength of the scale effects would be the same as in the baseline model; adding MP leaves the key role of domestic trade costs in reconciling data and model unaltered.[29] In the general case, the key question is whether the gains from MP are able to bring the standard model closer to the data; as we show in the online Appendix, the answer is negative because large countries, like the United States, do not have systematically lower gains from MP than smaller countries, like Denmark.

Next, consider the role of another source of gains from openness: the diffusion of ideas across countries. We introduce diffusion in the simplest way by ignoring trade and assuming that an idea originated in country i can be used in country n but subject to an iceberg-type friction given by $h_{ni} \geq 1$, with $h_{ii} = 1$. The real wage would now be

$$U_n = \gamma^D \times (\phi_n L_n)^{1/\theta} \times \xi_{nn}^{-1/\theta},$$

where $\xi_{nn} \equiv \phi_n L_n / \left(\sum_i \phi_i L_i h_{ni}^{-\theta}\right)$ is the share of domestic ideas in the total stock of ideas diffused to country n.[30] If there were no frictions to the diffusion of ideas across countries (i.e., $h_{ni} = 1$ for all n, i), the real wage would be $U_n = \gamma^D \left(\sum_i \phi_i L_i\right)^{1/\theta}$, and scale effects would operate at the world level with no consequence for the income of individual countries. This result suggests that including international technology diffusion would further bring our baseline model closer to the data regarding the strength of scale effects.[31]

Unfortunately, the big challenge is that diffusion is not directly observable in the data. Except for the small part that happens through licensing, technology diffusion does not leave a paper trail that can be used to *directly* measure the value of production done in a country by domestic firms using foreign technologies.[32] Our approach

[28] See the online Appendix for a detailed discussion of multinational production with domestic trade costs.

[29] Inspecting (16), τ_{nn} depends on θ and is calibrated to match domestic trade flows in the data. Hence, as long as $(1+\eta)/\theta$ does not change, $\tau_{nn}^{-(1+\eta)}$ does not change either.

[30] One can allow for both trade and diffusion by assuming that each idea is a vector of productivities, one for each country, drawn independently from a Fréchet distribution, as in Ramondo and Rodríguez-Clare (2010). In that case the expression above is augmented with the gains from trade, $U_n = \gamma^D \times (\phi_n L_n)^{1/\theta} \times \xi_{nn}^{-1/\theta} \times \lambda_{nn}^{-1/\theta}$.

[31] In Ramondo and Rodríguez-Clare (2010), we present a model with trade, MP, and diffusion of ideas, but no domestic frictions. The analysis including those frictions reveals that the role for diffusion is smaller than the one found in our previous paper.

[32] According to the data published by the Bureau of Economic Analysis, royalties and licenses paid to US parents and foreign affiliates by unaffiliated parties for the use of intangibles represented only one percent of total affiliates sales, in 1999. Some indirect evidence points to the importance of international diffusion for growth. Eaton and Kortum (1996, 1999) develop a quantitative model that allows them to use international patent data to indirectly infer diffusion flows. They estimate that most of the productivity growth in OECD countries, except for the United States, is due to foreign research: between 84 percent and 89 percent in Germany, France, and the United Kingdom, and around 65 percent for Japan. Keller (2004) also finds that, for nine countries that are smaller than the United Kingdom, the contribution of domestic sources to productivity growth is about ten percent.

in this paper is to go in an another, potentially more promising, direction: how much lower are the implied economies of scale at the country level once we add observable sources of the gains from openness and domestic trade costs? We can attribute any remaining difference in TFP between the data and our model to nonobservable diffusion. Hence, our framework can be thought as an indirect way to discipline this unobservable flow.

IV. Conclusion

Models in which growth is driven by innovation naturally lead to aggregate economies of scale. This has the counterfactual implication that larger countries should be much richer than smaller ones. These scale effects are also present in new trade models. We argue that these implications are largely a result of the standard assumption that countries are fully integrated domestically, as if they were a single dot in space. We depart from this assumption by treating countries as collections of regions that face positive costs to trade among themselves. We calibrate the model and evaluate the role of domestic trade costs in reconciling the theory with the data. By weakening scale effects, domestic trade costs are key to explain the discrepancy between the standard model and the data on TFP levels across countries. These domestic trades also make the model better match observed import shares, relative income levels, and price levels across countries.

REFERENCES

Ades, Alberto F., and Edward L. Glaeser. 1999. "Evidence on Growth, Increasing Returns, and the Extent of the Market." *Quarterly Journal of Economics* 114 (3): 1025–45.

Agnosteva, Delina E., James E. Anderson, and Yoto V. Yotov. 2014. "Intra-national Trade Costs: Measurement and Aggregation." NBER Working Paper 19872.

Alcala, Francisco, and Antonio Ciccone. 2004. "Trade and Productivity." *Quarterly Journal of Economics* 119 (2): 613–46.

Alesina, Alberto, Enrico Spolaore, and Romain Wacziarg. 2000. "Economic Integration and Political Disintegration." *American Economic Review* 90 (5): 1276–96.

Allen, Treb, and Costas Arkolakis. 2014. "Trade and the Topography of the Spatial Economy." *Quarterly Journal of Economics* 129 (3): 1085–140.

Alvarez, Fernando, and Robert E. Lucas, Jr. 2007. "General Equilibrium Analysis of the Eaton-Kortum Model of International Trade." *Journal of Monetary Economics* 54 (6): 1726–68.

Anderson, James E., and Eric van Wincoop. 2004. "Trade Costs." *Journal of Economic Literature* 42 (3): 691–751.

Anderson, James E., and Yoto V. Yotov. 2010. "The Changing Incidence of Geography." *American Economic Review* 100 (5): 2157–86.

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.

Aten, Bettina, Alan Heston, and Robert Summers. 2009. "Penn World Table Version 7.1" Center for International Comparisons of Production, Income, and Prices at the University of Pennsylvania. https://knoema.com/PWT2012/penn-world-table-7-1.

Barro, Robert J., and Jong-Wha Lee. 2000. "International Data on Educational Attainment: Updates and Implications." Center for International Development at Harvard University Working Paper 42.

Beck, Thorsten, George Clarke, Alberto Groff, Philip Keefer, and Patrick Walsh. 2001. "New Tools in Comparative Political Economy: The Database of Political Institutions." *World Bank Economic Review* 15 (1): 165–76.

Chaney, Thomas. 2008. "Distorted Gravity: The Intensive and Extensive Margins of International Trade." *American Economic Review* 98 (4): 1707–21.

Cosar, A. Kerem, and Pablo D. Fajgelbaum. 2016. "Internal Geography, International Trade, and Regional Specialization." *American Economic Journal: Microeconomics* 8 (1): 24–56.

Eaton, Jonathan, and Samuel Kortum. 1996. "Trade in Ideas: Patenting and Productivity in the OECD." *Journal of International Economics* 40 (3-4): 251–78.

Eaton, Jonathan, and Samuel Kortum. 1999. "International Technology Diffusion: Theory and Measurement." *International Economic Review* 40 (3): 537–70.

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.

Fajgelbaum, Pablo, and Stephen J. Redding. 2014. "External Integration, Structural Transformation and Economic Development: Evidence from Argentina 1870–1914." NBER Working Paper 20217.

Frankel, Jeffrey A., and David Romer. 1999. "Does Trade Cause Growth?" *American Economic Review* 89 (3): 379–99.

Head, Keith, and Thierry Mayer. 2011. "Gravity, Market Potential and Economic Development." *Journal of Economic Geography* 11 (2): 281–94.

Head, Keith, and Thierry Mayer. 2014. "Gravity Equations: Workhorse, Toolkit, and Cookbook." In *Handbook of International Economics,* Vol. 4, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, 131–95. Amsterdam: Elsevier.

Head, Keith, and John Ries. 2001. "Increasing Returns versus National Product Differentiation as an Explanation for the Pattern of US–Canada Trade." *American Economic Review* 91 (4): 858–76.

Hillberry, Russell, and David Hummels. 2008. "Trade Responses to Geographic Frictions: A Decomposition Using Micro-Data." *European Economic Review* 52 (3): 527–50.

Jones, Charles I. 1995. "R&D-Based Models of Economic Growth." *Journal of Political Economy* 103 (4): 759–84.

Jones, Charles I. 2002. "Sources of US Economic Growth in a World of Ideas." *American Economic Review* 92 (1): 220–39.

Jones, Charles I. 2005. "Growth and Ideas." In *Handbook of Economic Growth*, Vol. 1, edited by Philippe Aghion and Steven Durlauf, 1064–111. Amsterdam: Elsevier.

Jones, Charles I., and Paul M. Romer. 2010. "The New Kaldor Facts: Ideas, Institutions, Population, and Human Capital." *American Economic Journal: Macroeconomics* 2 (1): 224–45.

Keller, Wolfgang. 2004. "International Technology Diffusion." *Journal of Economic Literature* 42 (3): 752–82.

Klenow, Peter J., and Andrés Rodríguez-Clare. 2005. "Externalities and Growth." In *Handbook of Economic Growth*, Vol. 1A, edited by Philippe Aghion and Steven N. Durlauf, 817–61. Amsterdam: Elsevier.

Kortum, Samuel S. 1997. "Research, Patenting, and Technological Change." *Econometrica* 65 (6): 1389–419.

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

Lucas, Robert E., Jr., and Benjamin Moll. 2014. "Knowledge Growth and the Allocation of Time." *Journal of Political Economy* 122 (1): 1–51.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71 (6): 1695–725.

Organisation for Economic Co-operation and Development. 2012. "Redefining 'Urban': A New Way to Measure Metropolitan Areas." Paris: Organisation for Economic Co-operation and Development.

Ramondo, Natalia, and Andrés Rodríguez-Clare. 2010. "Growth, Size, and Openness: A Quantitative Approach." *American Economic Review* 100 (2): 62–67.

Ramondo, Natalia, and Andrés Rodríguez-Clare. 2013. "Trade, Multinational Production, and the Gains from Openness." *Journal of Political Economy* 121 (2): 273–322.

Ramondo, Natalia, Andrés Rodríguez-Clare, and Milagro Saborío-Rodríguez. 2014. "Scale Effects and Productivity across Countries: The Role of Openness and Domestic Frictions." NBER Working Paper 18532.

Ramondo, Natalia, Andrés Rodríguez-Clare, and Milagro Saborío-Rodríguez. 2016. "Trade, Domestic Frictions, and Scale Effects: Dataset." *American Economic Review.* http://dx.doi.org/10.1257/aer.20141449.

Redding, Stephen J. 2015. "Goods Trade, Factor Mobility, and Welfare." National Bureau of Economic Research Working Paper 18008.

Redding, Stephen, and Anthony J. Venables. 2004. "Economic Geography and International Inequality." *Journal of International Economics* 62 (1): 53–82.

Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98 (5): S71–S102.

Rose, Andrew K. 2006. "Size Really Doesn't Matter: In Search of a National Scale Effect." *Journal of the Japanese and International Economies* 20 (4): 482–507.

Simonovska, Ina, and Michael E. Waugh. 2014. "The Elasticity of Trade: Estimates and Evidence." Journal of International Economics 92 (1): 34–50.

Thissen, Mark, Dario Diodato, and Frank G. van Oort. 2013. *Integrated Regional Europe: European Regional Trade Flows in 2000*. The Hague: PBL Netherlands Environmental Assessment Agency.

Tombe, Trevor, and Jennifer Winter. 2014. "Fiscal Integration with Internal Trade: Quantifying the Effects of Equalizing Transfers." http://econ.ucalgary.ca/sites/econ.ucalgary.ca.manageprofile/files/unitis/publications/1-4681118/tombewinter2016.pdf.

Waugh, Michael E. 2010. "International Trade and Income Differences." *American Economic Review* 100 (5): 2093–124.