Appendix For "International Trade and Income Differences"
By Michael E. Waugh

A. Alternative Trade Models and Asymmetric Trade Costs

1. Anderson and van Wincoop (2003) Model

The model of Anderson and van Wincoop (2003) generates equation (4). To do so, assume that each country has constant returns technologies with competitive firms producing a good which is defined by its country of origin, i.e., the Armington assumption. These assumptions imply the unit cost (and price) to deliver a country j good to country i is $p_{ij} = \tau_{ij} T_j^{\frac{1}{1-\sigma}} c_j$. Here c_j is the cost of inputs to produce one unit of the country j good and $T_j^{\frac{1}{\sigma-1}}$ is total factor productivity in country j.

Preferences are equally simple. Each country has symmetric constant elasticity preferences over all the (country-specific) goods with common elasticity of substitution σ. The key result from this simple model are the expenditure shares

$$X_{ij} = \frac{T_j(\tau_{ij} c_j)^{1-\sigma}}{\sum_{\ell=1}^{N} T_j(\tau_{ij} c_j)^{1-\sigma}}. \tag{16}$$

The right-hand side is country i's imports from country j divided by country i's expenditure on all traded goods. The left-hand side relates the trade cost country i faces to import a good from country j and country j's unit cost of production relative to the sum of the prices paid for imported goods.[28]

Given preferences, each country faces the following price of tradable goods for each country i:

$$P_i \;=\; \Upsilon \left[\sum_{\ell=1}^{N} T_\ell \left(c_\ell \tau_{i\ell} \right)^{1-\sigma} \right]^{\frac{1}{\sigma-1}}. \tag{17}$$

Dividing equation (16) with the analogous equation for country j's expenditure on country j goods and noting the relationship between the denominator of equation (16) and the price index in equation (16) results in the following relationship:

$$\frac{X_{ij}}{X_{jj}} = \tau_{ij}^{1-\sigma} \times \left(\frac{P_j}{P_i} \right)^{1-\sigma}. \tag{18}$$

This is the same expression as in (4) relating the bilateral trade shares to trade costs and the relative aggregate price of tradables.

[28] Anderson and van Wincoop (2003) call the term P_i inward multilateral resistance because it is a summary measure of the difficulty for country i to import.

2. Melitz (2003)/ Chaney (2008) Model

As shown in Chaney (2008) and Helpman, Melitz and Rubinstein (2008), the Melitz (2003) framework can easily generate a "gravity-like" expression. As I show here, these frameworks generate a relationship similar to equation (4), which would yield the same conclusions as Section III, even though these frameworks have fixed costs and firms have market power. Below I generate the results from the model of Chaney (2008).[29]

The key components of this model are as follows: Each country has CES preferences over a measure of differentiated goods (determined in equilibrium) with a common elasticity of substitution σ. To produce a good for country i in country j, firms face a variable cost per quantity shipped $\tau_{ij}\frac{c_j}{z}$ and a fixed cost f_{ij} in units of the numeraire. The productivity levels z are specific to the firm and modeled as an idiosyncratic random variable drawn from a Pareto distribution with shape parameter γ,

$$G(z) = 1 - T_i z^{-\gamma}.$$

As in Helpman et al. (2008) and Chaney (2008), I assume that $\gamma > \sigma - 1$. Finally, firms are monopolistic competitors, and there is an unbounded measure of potential entrants.

This formulation generates the following expression for the share of country i expenditures on imports from country j:

$$X_{ij} = \frac{T_j\left(c_j\tau_{ij}\right)^{-\gamma}\left(\frac{f_{i,j}}{Y_i}\right)^{-\gamma\left(\frac{1}{\sigma-1}-\frac{1}{\gamma}\right)}}{\sum_{\ell=1}^{N} T_\ell\left(c_\ell\tau_{i\ell}\right)^{-\gamma}\left(\frac{f_{i,\ell}}{Y_i}\right)^{-\gamma\left(\frac{1}{\sigma-1}-\frac{1}{\gamma}\right)}},$$

where Y_i is total expenditures in country i. Similar to the approach above, note that the aggregate price of tradables in country i is

$$P_i = \kappa\left[\sum_{\ell=1}^{N} T_\ell\left(c_\ell\tau_{i\ell}\right)^{-\gamma}\left(\frac{f_{i,\ell}}{Y_i}\right)^{-\gamma\left(\frac{1}{\sigma-1}-\frac{1}{\gamma}\right)}\right]^{-\frac{1}{\gamma}}. \tag{19}$$

Using this fact and dividing the analogous equation for country j's share of expenditures on country j

[29] Helpman et al. (2008) generate nearly the same expression except for the complications associated with the truncated Pareto distribution for productivity.

goods yields the following expression:

$$\frac{X_{ij}}{X_{jj}} = \hat{\tau}_{ij}^{-\gamma} \times \left(\frac{P_j}{P_i}\right)^{-\gamma}, \tag{20}$$

$$\text{where } \hat{\tau}_{ij} = \tau_{ij} \times \left(\frac{f_{ij}}{f_{jj}}\right)^{\left(\frac{1}{\sigma-1}-\frac{1}{\gamma}\right)} \times \left(\frac{Y_j}{Y_i}\right)^{\left(\frac{1}{\sigma-1}-\frac{1}{\gamma}\right)},$$

which is the similar to equation (4), but now what was one represented by variable trade costs is now a mixture of the variable trade costs, the relative fixed costs between the two markets, and market size.

Note that in this model, countries consume different varieties of goods. Hence, the price index in equation (19) is not comparable across countries. The data that I use in this paper are a common basket price index, which is the same across countries. Thus, an adjustment needs to be made because the objects in the model and the data do not correspond.

I will make an adjustment in the model by asking, if country n purchases the same basket of goods as the United States, than what would its price index be? I will use the basket in the United States as the reference basket, but using any country will generate the same results. The key is that we want to compute a *common* basket price index for all countries.

To implement this concept, I will assume that the fixed costs are multiplicatively decomposable, i.e., $f_{ij} = f_i \times f_j$ and $f_{ii} = f_i \times f_i$. With this assumption, one can show how country-specific price indices relate to country-common basket price indices:

$$P_n = (P_n^{us})^{\frac{\sigma-1}{\gamma}} \times \left(\frac{f_n Y_{us}}{f_{us} Y_n}\right)^{\left(\frac{1}{\sigma-1}-\frac{1}{\gamma}\right)}. \tag{21}$$

Then using equation (21) and substituting it into equation (21) yields the following expression:

$$\frac{X_{ij}}{X_{jj}} = \tau_{ij}^{-\gamma} \times \left(\frac{P_j^{us}}{P_i^{us}}\right)^{1-\sigma}, \tag{22}$$

which relates bilateral trade shares to relative aggregate prices of a common basket, variable trade costs, and bilateral trade shares. The only difference is that the power term on the variable trade costs is different from the power term on relative prices. Despite this difference, as long as $\sigma > 1$ and $\gamma > \sigma - 1$ (which the model requires), the same conclusions from Section III apply: to account for both bilateral trade shares and aggregate prices, trade costs must be systematically asymmetric with respect to development. Thus, my arguments about the structure of trade costs throughout this paper may be interpreted more generally.

B. Measuring Income per Worker in the Model as in the Data

Real GDP or income per worker is an important object of interest. This section describes my concept of real GDP in the model as measured in the data— specifically, real GDP in benchmark years of the Penn World Table (PWT). Feenstra, Heston, Timmer and Deng (2004) provide a very useful description of the mechanics behind the construction of real GDP in the PWT, and my presentation borrows from their analysis. First, the PWT collects GDP in current prices from each country's statistics agency. In the model, nominal GDP (in per worker terms) is

$$p_i^y c_i + ex_i - im_i, \tag{23}$$

where p_i^y is the price of the final good produced in country i, c_i is the quantity of the final good, and ex_i and im_i are nominal exports and imports, i.e., nominal final expenditures plus net trade balance at current prices. The PWT then uses the Geary-Khamis system to compute "reference prices" π_ℓ for each good ℓ and the purchasing power parities PPP_i for each country i used to deflate (23). Since there is only one final good, the Geary-Khamis system is the following set of equations:

$$\pi = \left(\sum_{j=1}^{N} \frac{p_j^y c_j}{PPP_j^e} \right) \left(\sum_{j}^{N} c_j \right)^{-1} \text{ and } PPP_j = \left(\frac{p_j^y c_j}{\pi c_j} \right).$$

For a unique solution, some normalization is necessary. Setting $\pi = 1$, it is straightforward to show that PPP_i for country i equals p_i^y, or the price of the final good. This results in the flowing definition of real GDP per worker:

$$c_i + \frac{ex_i - im_i}{p_i^y}.$$

With balanced trade, the last term equals zero. Notice that the net trade balance is deflated by the price index for domestic absorbtion. This is actual contrary to the advice of the United Nations 1993 System of National Accounts. They advise that the net trade balance be deflated with both export and import price indices. Hence, the PWT measure of real GDP is more closely related to gross domestic income or command-basis GDP as defined in the United Nations 1993 System of National Accounts. Imposing market clearing and balanced trade, real GDP per worker in my model—as measured in the PWT—is,

$$y_i = \frac{w_i}{p_i^y} + \frac{r_i k_i}{p_i^y},$$

in which income from wages and capital are deflated by each country's final goods price.

C. Derivation of a Useful Representation of Income per Worker

Suppressing some notation and rearranging (2), using (1), and the fact that the rental rate on capital is pinned down by the expression $r_i = \frac{\alpha}{1-\alpha} w_i k_i^{-1}$ provides an expression for each country's home trade share:

$$X_{ii} = \frac{\left[k_i^{-\alpha\beta} w_i^{\beta} p_i^{q(1-\beta)}\right]^{\frac{-1}{\theta}} \lambda_i}{p_i^{q\left(\frac{-1}{\theta}\right)} \Psi}. \tag{24}$$

Further rearrangement of (24) provides the expression

$$\left(\frac{w_i}{p_i^q}\right) = \Psi \left(\frac{\lambda_i}{X_{ii}}\right)^{\frac{\theta}{\beta}} k_i^{\alpha}, \tag{25}$$

in which wages, deflated by the intermediate goods price, are a function of each country's home trade share and its capital-labor ratio.

Given the definition of real income per worker defined above and using a representative firm's first-order conditions determining the rental rate as a function of the wage, I express income per worker as a function of only the wage and the final goods price:

$$y_i = \frac{1}{1-\alpha} \frac{w_i}{p_i^y}. \tag{26}$$

Since my interest is only in relative income differences, constant terms are abstracted from. Combining the expression for the price of final goods and (26), real income per worker is expressed as

$$y_i = \left(\frac{w_i}{p_i^q}\right)^{1-\gamma} k_i^{\alpha\gamma}. \tag{27}$$

Combining equations (25) and (27), real income per worker is now

$$y_i = X_{ii}^{\frac{-\theta(1-\gamma)}{\beta}} \lambda_i^{\frac{\theta(1-\gamma)}{\beta}} k_i^{\alpha}. \tag{28}$$

Here real income per worker is only a function of each country's home trade share X_{ii}, its technology parameter λ_i, and its capital-labor ratio.

D. Zeros

An implication of the Eaton and Kortum (2002) framework is that, in aggregate, every country should purchase some nonzero amount of goods from all other countries. In fact, the bilateral trade matrix has many recorded zeros. For the sample considered there are 5,929 possible trading combinations; 1,610 (27 percent) show no trade at all. This presents both an estimation issue and a computational issue.

Regarding the estimation, I will omit any zero observed trade flows from the estimation of equation (8). This has been a standard approach in the empirical trade literature. Santos-Silva and Tenreyro (2006) propose a Poisson pseudo-maximum-likelihood (PPML) estimator to alleviate any bias from the log-linear specification in equation (8) due to the presence of heteroskedasticity and the omission of zero observed trade flows. I employed their technique of estimating the gravity equation in levels and including zero observed trade flows, and found that the quantitative results and counterfactual exercises do not differ dramatically relative to using ordinary least squares. This does not contradict their findings. Consistent with their results, I find that OLS exaggerates the distance elasticity, suggesting that the bias they emphasize is present. For example using PPML, the percentage effect on cost is 129, 140, 141, 177, 223, and 263 percent for each distance category. Compared to Table 2, shorter distances are more costly and longer distances are less costly relative to OLS. This is consistent with the lower distance elasticity Santos-Silva and Tenreyro (2006) find when using PPML relative to OLS.

Helpman et al. (2008) particularly focus on zero trade flows, building on the model of Melitz (2003) with fixed costs and firm heterogeneity. When firm-level productivity is drawn from a truncated Pareto distribution, they can deliver zero trade flows between country pairs. Their results suggest that any bias arising from the omission of zero trade flows is quantitatively small.

Regarding the computation, when computing equilibrium prices and counterfactuals, I will set trade costs for the instances in which X_{ij} is zero to an arbitrarily large value to approximate what appears to be a trade cost of infinity.

E. Technology Heterogeneity: Evidence

A concern is that the distributional assumptions generate implausible differences in productivity at the micro or good level. This is important because the degree of dispersion in productivity affects the response of aggregate TFP and income differences as trade costs change. In this section, I argue that the dispersion in productivity implied by my model is reasonable and conservative based on the available empirical evidence.

To make this argument, I first performed the following exercise. I assumed that there were 100,000 goods and generated a productivity term $z^{-\theta}$ for each good from the calibrated distribution for the United States, United Kingdom, and Uganda. I then asked two questions: (i) how much does $z_i^{-\theta}$ vary within a country? and (ii) For a given good, how much does $z_{us}^{-\theta}$ differ relative to $z_{uga}^{-\theta}$ or $z_{uk}^{-\theta}$? After recording various measures of dispersion, I repeated this process 500 times. Table 4 reports the means of these measures across the simulations.

Table 4: Dispersion in Productivity

Model: Variation Within Countries, Across Goods			
	U.S.	U.K.	Uganda
Estimated $\lambda_{us}^{\theta}/\lambda_{i}^{\theta}$	—	1.10	6.30
Simulated mean relative to U.S.	—	1.12	6.30
$99-\ \ 1$ ratio of $z_i^{-\theta}$	3.04	3.04	3.04

Model: Variation Within Goods, Across Countries		
	U.S. / U.K.	U.S. / Uganda
99th percentile of $(z_{us}/z_j)^{-\theta}$	2.58	14.5
1st percentile of $(z_{us}/z_j)^{-\theta}$	0.48	2.73

1. Variation Within Countries

The top panel of Table 4 reports the degree of variation in productivity within a country. For all countries, the ratio of productivity in the top 99th percentile over the bottom 1th percentile is about a factor of 3. This value is only a function of θ, with increases in θ increasing the difference between percentiles.

Relative to the available evidence, Table 4 shows my model implies a degree of dispersion in productivity that is conservative. For example, Hsieh and Klenow (2007a) employ plant-level data from China, India, and the United States and construct estimates of TFP at the plant level. My point of comparison is what they call "TFPQ", which is most closely related to how I would measure TFP given my model. They report a dramatic amount of dispersion in TFP. For example, the ratio of the 90th percentile to the 10th percentile in China in 1998 is 15.18—10 times the amount of dispersion in my model. In India in 1998, this same ratio is 31.18—20 times the amount of dispersion in my model. Measuring TFP at this level is difficult and comes with many caveats; however, this evidence suggests that amount the of dispersion in productivity implied by my model is very conservative.

2. Variation Across Countries

The bottom panel of Table 4 reports the variation across countries in productivity to produce a particular good. For example, compare the productivity to produce blue tennis shoes in the United States, United Kingdom, and Uganda in the model. The top row presents the 99th percentile of the distribution of relative productivities between countries to produce the same good. For the 99th percentile, the United States is 2.58 times more productive than the United Kingdom and 14.5 times more productive than Uganda to produce the same good. The bottom row present the 1st percentile. Here, the United Kingdom is 2 times more productive than the United States and Uganda is 2.7 times *less* productive than the U.S. to produce the same good.

Again, this degree of dispersion in TFP across countries is conservative relative to available empirical evidence on the variation in productivity within industries across developed countries. Baily and Gersbach (1995) show that value added per worker within the same manufacturing industries varies by as much as a factor of 3 between the United States, Japan, and Germany in the early 1990s.[30] And they argue that differences in the use of physical capital played little role in explaining these differences. Relative to the bottom panel of Table 4, this suggests that between rich countries I am slightly understating the dispersion in productivity differences.

Because even the best producer in the poor country is still nearly 3 times *less* productive than the United States, I am again understating the dispersion in productivity differences relative to the evidence. There is less micro-evidence regarding poor countries, but a theme that emerges is that some firms in poor countries are as productive as firms in rich countries. Banerjee and Duflo (2005) discus evidence from India on TFP at the industry level from the McKinsey Global Institute (2001). Banerjee and Duflo (2005) highlight the fact that the best firms in several of the manufacturing industries studied are basically using the global best practice technologies. Lewis (2004) argues that this same pattern prevails in other developing countries such as Brazil, Russia, and Korea. Though the focus of his paper is not in manufacturing, Lagakos (2008) provides evidence for the retail sector that the most productive firms in developing countries are nearly as productive as the firms in developed countries.

F. Alternative Distributional Assumptions

In this section, I consider alternative distributional assumptions and their quantitative implications.

Without making the distributional assumptions, the model loses analytical expressions mapping the data to parameters of the model, i.e., trade costs and technology parameters. To solve the model under alternative distributional assumptions, I employed the following approach. First, I assumed there was a large number (100,000) of potentially tradable goods. For each country, good-level efficiencies were drawn from the country-specific distribution and assigned to the production technology for each good. Then, for each importing country, the low-cost supplier across countries is found for each good and the aggregate bilateral pattern of trade is computed. In the examples below, I assume there are only 10 countries and adjusted the data under the assumption that these are the only countries in the economy. I did not consider the full 77-country example because of the computational burden associated with estimating/calibrating the model.

I considered two alternative distributional assumptions: (i) a log-normal distribution with country-specific center parameter μ_i and common parameter σ and (ii) Pareto distribution with country-specific center

[30] See Parente and Prescott (2002) for a nice discussion of this evidence.

term κ_i and common shape parameter ν. Because the common parameters σ and ν play roles similar to that of θ, I calibrated them for these examples such that all models have the same coefficient of variation. I calibrated the country-specific parameters (μ_i or κ_i) and ex_i for each country and common parameters relating to the effects of distance and shared borders to best fit the data. For comparison purposes, I used the same approach in the benchmark case with a Fréchet distribution rather than exploiting its analytical convenience.

Table 1: Alternative Distributional Assumptions

Model Fit			
Distribution	M.S.E.	var(log(y))	y_{90}/y_{10}
Data	—	1.98	25.11
Fréchet (benchmark)	1.99	1.43	17.04
Log-normal	2.12	1.47	17.96
Pareto	2.25	1.53	19.25
Gain from Trade, $\tau = 1$			
	Mean Δ y (%)	var(log(y))	y_{90}/y_{10}
Fréchet (benchmark)	68	0.82	8.10
Log-normal	43	1.12	11.58
Pareto	69	0.86	8.82

The top panel of Table 1 presents some summary measures of the fit. The first column is the mean squared error between the data and model in logs. The second and third columns report the implied difference in income per worker across the two scenarios. In terms of the fit of the data and measures of dispersion in income per worker, all three models perform similarly.[31]

The bottom panel of Table 1 presents the average increase in income per worker after removing all trade costs, the variance in log income per worker, and the 90/10 ratio in income per worker. Between the Fréchet and Pareto case there is little difference in the change across all three measures. Both generate reductions in cross-country income differences of approximately 44 percent. The log-normal distribution results in less reduction in cross-country income differences—approximately a 24 percent decline in the variance of log income per worker. As the log-normal case illustrates, the distributional assumptions obviously play a role. More research is needed on the implications of alternative distributions and evidence supporting these assumptions. But in all these cases, poor countries gained the most from reductions in trade costs, thus reducing cross-country income differences.

[31] For reference, had I used the analytical convenience when calibrating the economy with the Fréchet distribution, the mean square error would have been 1.31, the variance in log income per worker 1.62 and the 90/10 ratio 25.76.