Gravity, Productivity and the Pattern of Production and Trade*

James E. Anderson
Boston College and NBER

May 24, 2007

Abstract

Trade frictions have powerful implications for productivity. This paper extracts the implications and their intuition by embedding gravity in general equilibrium production models. Stronger implications are extracted for the specific factors model of production. A harvest of predictions about the pattern of production and trade is reaped from the specific factors model.

JEL Classification: F10, D24

*To be presented at Drexel University, Brandeis University and the NBER ITI Summer Institute.

Trade costs should intuitively have a big impact on productivity and the pattern of production and trade. Differential access to markets should alter the pattern of returns to countries' factors of production. Differential trade costs across products should tilt the pattern of production. The impacts should be big because trade costs are big. These implications have gone largely unexplored because the measurement of the productivity implications of trade costs is problematic. This paper develops a promising approach.

Productivity measurement in a world economy with trade frictions is problematic for three reasons. First, distribution costs affect prices on both the supply and demand sides of the market, but the incidence of the costs on the supply side alone is required for measuring their effect on productivity. Second, standard productivity measures in distribution sectors fail to capture the effects of globalization because quality changes produce savings that are hidden in the books of the ultimate buyers and sellers. Inference from gravity models of trade suggests that these effects are big. Third, outsourcing implies changes on the extensive margin that require different accounting than the intensive margin changes that standard methods measure.

All three problems are addressed in this paper by building on recent progress in understanding, interpreting and using the gravity model. The incidence problem is solved in this paper by extending the economic theory of gravity (Anderson, 1979; Anderson and van Wincoop, 2003, 2004). The concepts of outward and inward multilateral resistance solve the incidence problem in general equilibrium while at the same time being ideal aggregators of bilateral trade costs. Outward multilateral resistance is equivalent to a productivity penalty imposed on each sector in each economy, as if each producing sector traded with a single world market. Selection into trade on the extensive margin is incorporated in the gravity model by Helpman, Melitz and Rubinstein (2007), with implications for productivity drawn out here and related to outsourcing.

The specific gravity model — the specific factors model of production with gravity embedded — yields a harvest of tight and intuitive explanations of the equilibrium pattern of production and trade in terms of factor endowments, taste parameters and trade costs. This extends the predictive content of trade theory.

The previous literature contains only two cases in which the implications of production theory for equilibrium trade are developed. The Ricardian continuum model was recently extended to a world of costly trade by Eaton and Kortum (2002). Labor productivities are drawn from a Frechet probability distribution. In contrast, the specific factors model developed here features factor endowments as the basis of labor productivity. The Heckscher-Ohlin-Vanek factor content model also features the role of factor endowments, but only in a frictionless world.[1]

Section 1 sets the stage by assuming that the supply side incidence of trade frictions is equivalent to sectoral productivity penalties in the standard abstract model of production and trade. Multifactor productivity is the ratio of the usual Hicks neutral productivity parameter to the supply side incidence measure. Sectoral measures aggregate to a productivity measure for each country in the world economy for given equilibrium prices. But how are the sectoral incidence measures to be found? Intuitive clues are developed from a partial equilibrium analysis.

Section 2 provides the solution to incidence and aggregation in general equilibrium. Given the sectoral supply and expenditure shares determined by the model of Section 1, the bilateral allocation of trade determines the multilateral resistance indexes that determine incidence and deliver the sectoral allocation of production and expenditure. Full general equilibrium is achieved with mutual consistency of the two modules. Section 2 then draws some lessons for productivity measurement at given prices in a world of costly trade. Section 3 specializes the production model to the specific factors model in order to obtain definite analytic results that permit incidence analysis with endogenous prices. Section 4 applies the special case model to characterize the pattern of production and trade in world trade equilibrium. It also develops the consideration of productivity measurement in the special case trading world. Section 5 extends the discussion to treat intermediate products trade and selection into exporting. Section 6 concludes.

1 Trade Frictions and Productivity

Each country produces and distributes goods to its trading partners subject to trade frictions. Suppose that the effect of these frictions can be represented by an average friction Π_k^j for each product category k in each country j. With unit production cost $\widetilde{p}_k^j$ in country j, it is as if there was an average ('world') destination price for goods k delivered from j, $p_k^j = \widetilde{p}_k^j \Pi_k^j$. The incidence of

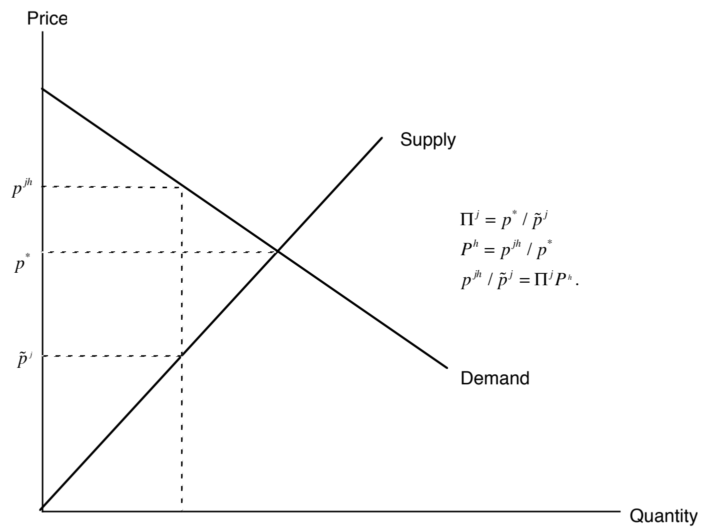
[1] Davis and Weinstein (2001) integrate trade costs inferred from gravity with the two factor Heckshcher-Ohlin continuum model, but only on the demand side. In future work I plan to draw out the implications of gravity in the HOV setting.

the trade frictions will be solved for in general equilibrium in the next section, but it is helpful to review the incidence problem in partial equilibrium as a preliminary.

1.1 Incidence

Partial equilibrium incidence analysis is reviewed in Figure 1.

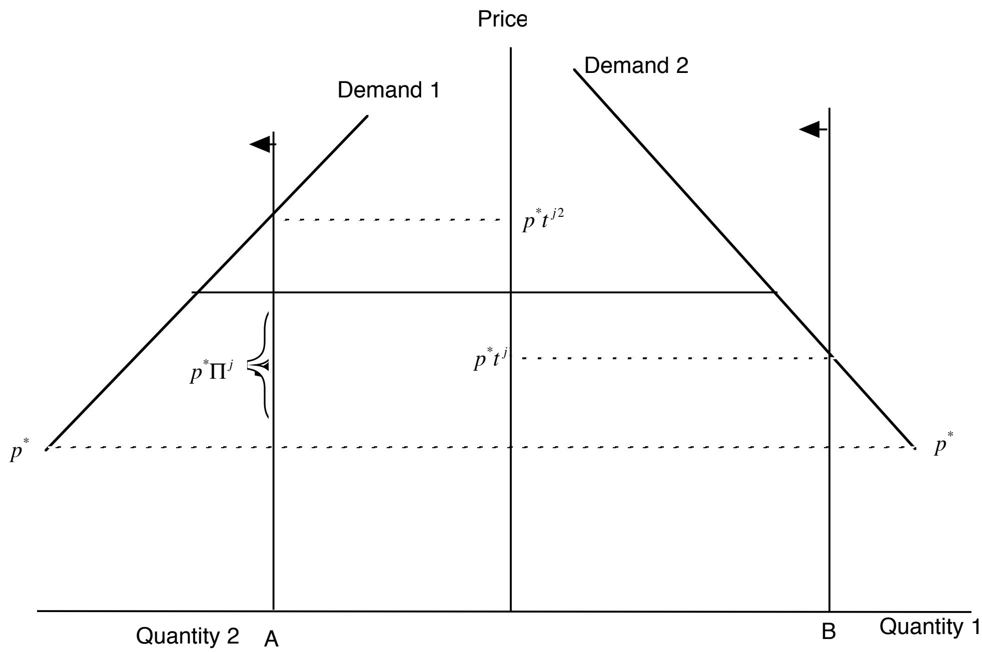
Figure 1. Incidence of Trade Cost



The price of the representative good from i at location j is p^{ij} while its unit cost is $\widetilde{p}^i$. The full trade cost friction marks up the unit production cost by $\Pi^i P^j$ to yield the buyer's price p^{ij}. The standard incidence analysis uses the hypothetical frictionless equilibrium with price p^* to split the cost into two components, of which Π^i falls on the supply side of the market and P^j falls on the demand side. The distribution cost component Π is conceptually identical to a productivity penalty that shifts the cost of production and distribution upward to where the hypothetical supply schedule intersects the horizontal line at p^* at the equilibrium quantity.

Figure 2 portrays the aggregation of supply side incidence in partial equilibrium for the case of two markets. The reference equilibrium price p^* is preserved by maintaining the total quantity shipped while replacing the nonuniform trade costs with the uniform trade cost Π^j. An analogous diagram can be drawn to illustrate the aggregation of demand side incidence.

Figure 2. Supply-Preserving Aggregation



Total shipments AB are preserved by moving the goalposts left such that a uniform markup is applied to each shipment.

1.2 Productivity

Accounting for multilateral resistance as a productivity component uses familiar principles. The key building block is the gross domestic product (GDP) function. It is written as $g(\widetilde{p}^j, v^j)$ where v^j is the vector of factor endowments. g is convex and homogeneous of degree one in prices, by its maximum value properties.

The supply vector to *final demand* is given by $g_{\widetilde{p}}$, by Hotelling's lemma, using the convention that subscripts with variable labels denote partial differentation with respect to the variable. The production share of k in country j is given by

$$s_k^j = g^j_{\widetilde{p}^j_k} \widetilde{p}^j_k / g_j.$$

Now consider productivity accounting with trade frictions. Take p^j as a given vector of 'world' prices. Then $\widetilde{p}^j_k = p^j_k / \Pi^j_k, \forall k$. The aggregate productivity penalty due to trade frictions is given by

$$\bar{\Pi}^j \equiv g^j(p^j, v^j)/g^j(\widetilde{p}^j, v^j).$$

The logic uses the distance function. The reference GDP is that for a frictionless economy, $\Pi_k = 1, \forall k$. The uniform productivity penalty that is equivalent to the vector of productivity penalties satisfies

$$g(p/\bar{\Pi}, v) = g(\widetilde{p}, v).$$

GDP is homogeneous of degree one in the prices, hence $\bar{\Pi}$ has the explicit solution given. $\bar{\Pi}$ is equal to the cost of delivered goods relative to the cost of production. The GDP price deflator (i.e., the price index for delivered goods) is given by

$$g(p, \cdot) = \bar{\Pi} g(\widetilde{p}, \cdot).$$

$\bar{\Pi}$ is homogeneous of degree one in $\{\Pi_k\}$. In rates of change the aggregate penalty to productivity imposed by trade frictions is given by $\sum_k s_k \hat{\Pi}_k$. Various approximations to the rate of change of $\bar{\Pi}$ can be used, such as a Laspeyres index. Section 4 offers an exact index based on a special case of production technology, the specific factors/Cobb-Douglas model.

The analysis here extends to encompass the effects of Hicks neutral technological differences across goods and countries. Let a_k denote the productivity parameter (relative to some benchmark) in sector k. The $\widetilde{p}$'s are then reinterpreted as 'efficiency' unit costs, $\widetilde{p}_k = p_k a_k/\Pi_k$. Supply is given by $g_{\widetilde{p}_k} = g_{p_k} a_k/\Pi_k$. Total factor productivity is measured by

$$T = \frac{g(\widetilde{p}, v)}{g(p, v)}.$$

T is decomposable into $\bar{a}/\bar{\Pi}$ based on $\bar{a} = g(\widetilde{p}, v)/g(\{p_k/\Pi_k\}, v)$ and the analogous operation for $\bar{\Pi}$.

How are the mysterious Π's to be found? The Π's must be derived as indexes of the underlying trade frictions, with ideal index properties that satisfy the present purposes, which include resolving the incidence of trade costs on the supply and demand sides of the markets.

The next section will show that the multilateral resistance variables $\{\Pi_k^j, P_k^h\}$ properly decompose the supply and demand side incidence of the aggregated trade frictions in general equilibrium. In contrast, productivity analysis based on a trade-weighted index number of the full bilateral trade costs would overstate their impact unless the incidence fell entirely on the supply side, illustrated by the case where demand is infinitely elastic at price p^* and $P^j = 1$ in Figure 1.

The explanation of differing levels of productivity over time or space involves incorporating differences in equilibrium prices. Figure 1 again illustrates. Imagine shifts in supply or demand, or in the full trade cost. The new solutions for Π and P yield differences to be explained in the reduced form model by the shifts in the exogenous variables, incorporating the shifts in the equilibrium price. Dealing with incidence properly in a reduced form global general equilibrium setting requires building a specific production structure, dealt with in Section 4.

2 Determination of Multilateral Resistance

When trade is costly, the incidence of costs on productivity must be determined in a general equilibrium model. To develop the insight, specializing assumptions are useful. The analysis relies on the assumption of *trade separability* — the composition of expenditure or production within a product group is independent of prices outside the product group. On the supply side, separability is imposed by the assumption the goods from j in class k shipped to each destination are perfect substitutes in supply. On the demand side, separability is imposed by assuming that expenditure on goods class k

forms a separable group containing shipments from all origins. This setup enables two stage budgeting analysis. Subsection 2.1 sets out the upper level allocation of expenditure and production. Subsection 2.2 derives multilateral resistance from the lower level allocation of goods across trading partners.

A specific factors Cobb-Douglas general equilibrium production case developed in Section 3 and applied in Section 4 allows a thorough characterization of how factor endowments and demand parameters determine equilibrium prices and productivity.

2.1 General Equilibrium Allocation

Suppose that expenditure can be represented by identical homothetic preferences for all countries and assume the absence of trade taxes or foreign owned factors or international transfers. There are iceberg trade frictions $t_k^{jh} \geq 1$ that lead to prices at destination $p_k^{jh} = \widetilde{p}_k^j t_k^{jh}$ for goods in class k delivered from j to market h. Suppose that preferences are separable with respect to the partition between classes. Then exact price aggregators P_k^h are defined, eventually specified as CES aggregators of the prices of goods from all origins to destination h in class k. Let the domestic price vector at location h be given by $q^h = \{P_k^h\}, \forall k$.

Expenditure is given by $e(q^h)u^h$. Then by Shephard's Lemma, the expenditure on k from j in location h is given by $e_{P_k^h} \partial P_k^h / \partial p_k^{jh}$. The share of expenditure on k from all origins at destination h is given by

$$\theta_k^h = e_{P_k^h} \frac{P_k^h}{e(q^h)}.$$

The international budget constraint implies that expenditure is equal to GDP for each country, hence $u^h = g^h / e(q^h)$.

Market clearance in the world economy requires that for each good k from source j the value of shipments (at delivered prices) equal the expenditure on the good over all destinations. Let the value of shipments from origin h in product class k be given by $Y_k^h = s_k^h g^h$. Let the expenditure in destination h on product class k be given by $E_k^h = \theta_k^h g^h$. Then market clearance requires:

$$Y_k^j = \sum_h E_k^h \frac{\partial P_k^h p_k^{jh}}{\partial p_k^{jh} P_k^h}. \tag{1}$$

Replacing p_k^{jh} with $\widetilde{p}_k^j t_k^{jh}$, (1) solves for $\{\widetilde{p}_k^j\}$.

2.2 Multilateral Resistance

The gravity model determines inward and outward multilateral resistances $\{P_k^j, \Pi_k^j\}$ as an implicit function of the underlying iceberg trade frictions $\{t_k^{jh}\}$ and the upper level allocations $\{E_k^h, E_k^j\}$.

The gravity module yields the multilateral resistances as the solution to the system of (sub-)budget constraints and market clearance equations. The CES sub-expenditure function case yields:

$$(P_k^h)^{1-\sigma_k} = \sum_j \left\{\frac{t_k^{hj}}{\Pi_k^j}\right\}^{1-\sigma_k} \frac{Y_k^j}{\sum_j Y_k^j} \tag{2}$$

and

$$(\Pi_k^j)^{1-\sigma_k} = \sum_h \left\{\frac{t_k^{hj}}{P_k^h}\right\}^{1-\sigma_k} \frac{E_k^h}{\sum_h E_k^h}. \tag{3}$$

Here, σ_k is the elasticity of substitution parameter for goods class k and θ_k^j is the expenditure share for goods of class k from every origin in location j. Inward multilateral resistance P_k^j is also the CES price aggregator for class k at location j that enters the upper level expenditure function as the relevant price argument to determine the expenditure shares $\{\theta_k^j\}$. (3) is an alternative representation of the market clearance condition and (2) is an alternative representation of the budget constraint.

The expenditure shares θ_k^j are determined in the general equilibrium of the preceding sub-section along with the supply shares $\{s_k^j\}$ and the GDP's $\{g^j\}$. The full general equilibrium is determined by the solution values for $\{\widetilde{p}_k^j, P_k^j, \Pi_k^j\}$ in the system (1)-(3) as functions of endowments, trade costs and the parameters of technology and preferences.

Given estimates of gravity models that yield inferred t's and with the accompanying data on the shares, inference on the multilateral resistances is readily operational. Computation of the multilateral resistances from (2)-(3) given the shares is quite fast for modern computers. That is because the system is essentially quadratic in the power transforms of the multilateral resistances.

Anderson and van Wincoop (2004) show that the multilateral resistance indexes are ideal indexes of trade frictions in the following sense. Replace all the bilateral trade frictions with the hypothetical frictions $\widetilde{t}_k^{jh} = \Pi_k^j P_k^h$. The market clearance and budget constraint equations continue to hold at the same prices, even though individual bilateral trade volumes change.

There are two important implications. First, the incidence of bilateral trade costs is decomposed *on average* by $\widetilde{t}$. Thus the gravity model extends the incidence analysis of Figure 1 to general equilibrium. The demands and supplies of the upper level allocation remain constant, as they do in Figure 1, while the 'equilibrium world prices' analogous to p^* in Figure 1 are those prices that support the aggregate quantities when the average incidence is imposed according to the solution of (2)-(3). No single bilateral cost t_k^{jh} is accurately measured or decomposed in this way, but on average the outward incidence is given by the Π's and the inward incidence is given by the P's.

The second implication is that, for each good k in each country j, bilateral distribution costs aggregate to an ideal average outward multilateral resistance Π_k^j, with market clearance at the same producer price and volume. It is as if a single shipment was made at the average cost. This justifies the treatment of multilateral resistance as a productivity penalty on the activity of production and delivery.

There are important regularities in the cross section pattern of multilateral resistance. For each good:

Proposition 1 *If trade costs are uniform border barriers, the multilateral resistances (inward and outward) of large supply share economies are lower than those for small supply share economies. Conversely, the multilateral resistances (inward and outward) of large expenditure share economies are higher than those for small expenditure share economies. For given expenditure shares, multilateral resistances are increasing in net import shares.*[2]

The proof of the proposition is in the Appendix. The intuition is this. On the supply side, the larger the supply share, the lower the markup Π must be on the cost $\widetilde{p}$ in order to meet the hypothetical uniform frictionless price p. On the demand side, the larger the expenditure share, the higher the markup P must be to bring the buyers' price down to the hypothetical uniform frictionless price p. General equilibrium links the inward and outward multilateral resistances together so that the changing shares on the supply or demand sides act to move both multilateral resistances in the same direction. While the uniform border barriers assumption is very special, the intuition explaining Proposition 1 suggests that its characterization of the relationship of incidence and market shares tends to hold more generally.

[2]The proposition extends that of Anderson and van Wincoop (2003), which deals with a the introduction of a small uniform border barrier in a one good balanced trade economy for which $P^j = \Pi^j$.

As an illustration of the economic significance of multilateral resistance and its implications for productivity, and to illustrate how far Proposition 1 may inform more general structures, see Table 1, a reproduction of Table 8 in Anderson and van Wincoop (2004). The numbers are interpreted as Π's. Bilateral resistance in the estimated model is due to distance and the presence (or absence) of the international border. Multilateral resistance is calculated here in a setting where each region is endowed with a unique good, trade costs are symmetric and trade is balanced. Thus inward and outward multilateral resistance are equal. Internal trade costs are normalized by assuming that Maryland's internal trade is frictionless. The Table shows that Canadian provinces face higher multilateral resistance because they are on average further from their markets and because a much higher proportion of their markets require crossing the international border between the US and Canada. The table also shows that larger regions (states or provinces) tend to have lower multilateral resistance, as per the suggestion drawn from Proposition 1. But this effect is blurred by locational advantages — even small Northeastern economies such as Massachusetts have relatively low Π because they are close to the big markets of the Northeast. Aggregate national multilateral resistance can be calculated using the fixed regional endowment assumption for each country. This implies that $\bar{\Pi}$ is given by a regional-shipments- weighted average of the regional Π's. $\bar{\Pi}$ for Canada is equal to 1.55 while for the US it is 1.38.

Another important implication of Proposition 1 is that productivity analysis that neglects the structure of trade frictions will tend to confound economies of scale with the effects of trade frictions. Large economies tend to have lower multilateral resistance and thus higher productivity even in the absence, as here, of conventional scale economies. Conversely, neglecting scale economies will tend to overstate the effect of trade frictions.

The outward multilateral resistance variables $\{\Pi_k^j\}$ are endogenous with respect to $\{Y_k^j, E_k^j\}$, with consequent implications for the analysis of changes in productivity in the economy. The implications can be drawn out by considering simple benchmark cases.

One important benchmark is invariance. Invariance occurs the case of uniform factor endowment growth everywhere in the world. Multilateral resistances are constant because all shares are constant. A second benchmark case gauges the significance of trade frictions for productivity. Imagine a pure globalization shock in which trade costs fall uniformly by 4 percent — the world gets literally smaller by 4 percent. Since (2)-(3) is homogeneous of degree 1/2 in the t's, all Π's fall by 2 percent and hence productivity rises by 2 percent everywhere in the world.[3] All relative prices remain constant, hence all shares are constant. Welfare rises everywhere by 4 percent because all P's fall by 2 percent while GDP rises by 2 percent due to the 2 percent fall in the Π's.

A third benchmark is given by Anderson and van Wincoop (2003), who show that the introduction of a small uniform border barrier in a frictionless world will raise the multilateral resistance of small countries by more than that of large countries. Proposition 1 suggests that reductions in trade frictions from any level tend to reduce the multilateral resistance of small supply share participants relative to large supply share participants. Moreover, it suggests that increases in supply shares tend to lower the multilateral resistance of the market share gainers. Simulation is required to examine this suggestion.

In general, asymmetric declines in trade frictions and growth in factor endowments have asymmetric effects on multilateral resistance and productivity. The complexity of the model defeats all but the most simple analytics. The very simple cases above do provide some insight to guide future simulations.

The future application of multilateral resistance to productivity studies will have to confront a puzzling empirical regularity in gravity studies. Previous investigators have pointed to results showing that trade costs inferred from gravity apparently do not fall over time. This puzzle is probably due to failure to apply the gravity model appropriately. Anderson and van Wincoop (2004) point to a number of problematic aspect of the literature and propose solutions.

For present purposes, two problems stand out. First, aggregation bias (most of the literature infers trade cost estimates from aggregate bilateral trade flows) is likely to be masking effects that disaggregation can uncover. The limited amount of disaggregated estimation available reveals widely varying trade costs over goods classes. Second, until very recently, empirical researchers had no very good way to treat fixed trade costs, a considerable impediment to understanding a trading world in which movement on the extensive margin (from zero to positive trade in various bilateral pairings)

[3]Another important implication of homogeneity is that gravity models alone can only provide information about *relative* trade costs. It is convenient to normalize by some *presumptively* small bilateral cost t_k^{jj} where region j is selected for its plausibly low internal distributive frictions.

appears to be important. Helpman, Melitz and Rubinstein (2007) provide a model of selection and show that it makes considerable difference to trade cost estimates in an aggregate trade flows setting. The implications for productivity are drawn out in Section 5.

3 The Specific Factors Model

The causal links between multilateral resistance and the allocation of production and expenditure are clarified by considering the special case of the specific factors model. The implications for the equilibrium pattern of production and trade are drawn out.

Labor is intersectorally mobile, while there are sector specific factors in fixed supply. The latter can be regarded as possibly mobile in the long run, an interpretation used below for reference. The sectorally fixed supply can be motivated by adjustment costs of various sorts. One form useful for future developments will be fixed costs of entry, providing a link to recent theories focused on firm heterogeneity and its implications for productivity and trade.

Supply understood as deliveries to final demand in product class k is given by

$$X_k = f^k(L_k, K_k)a_k/\Pi_k, \forall k \tag{4}$$

where the country index superscript j is omitted for notational ease. K_k is the specific factor endowment, possibly a bundle of such factors. f^k is a concave (usually homogeneous of degree one) production function. Labor L_k is mobile across sectors, with efficient allocation implied by the value of marginal product conditions

$$w = f^k_L p_k a_k/\Pi_k, \forall k \tag{5}$$

Labor market clearance implies

$$\sum_k L_k = L. \tag{6}$$

Gross domestic product $\sum_k \widetilde{p}_k X_k$ is given by the maximum value GDP function $g(\widetilde{p}, L, \{K_k\})$. As a reminder, $\widetilde{p} = \{p_k a_k/\Pi_k\}$, the vector of efficiency unit costs of production while p is the vector of 'world' prices. Hotelling's Lemma implies that the supply in general equilibrium is given by $X_k =$

$g_{\tilde{p}_k} a_k/\Pi_k$ while the equilibrium wage is given by g_L. This is the specific gravity model of production.[4]

A very useful closed form solution for g arises when f^k has the Cobb-Douglas form with identical share parameters across the sectors. While this is an extreme simplification, it is consistent with the stability of aggregate labor shares across periods of time when the composition of GDP has altered tremendously. Let $K = \sum_k K_k$ and let α be the parametric share parameter for labor.

For the special Cobb-Douglas case,

$$g = L^\alpha K^{1-\alpha} G \tag{7}$$

where G is given by

$$G = [\sum_k \lambda_k (p_k a_k/\Pi_k)^{1/(1-\alpha)}]^{1-\alpha}, \tag{8}$$

and $\lambda_k = K_k/K$, the proportionate allocation of specific capital to sector k.[5] GDP is the product of real activity in production and distribution $R = L^\alpha K^{1-\alpha}$ and the real activity deflator G. The effect of prices and productivity on the real activity deflator G decompose neatly: G is equal to the GDP price deflator times the aggregate technology factor divided by the trade cost productivity penalty:

$$G = [\sum_k \lambda_k (p_k a_k/\Pi_k)^{1/(1-\alpha)}]^{1-\alpha} = [\sum_k \lambda_k p_k^{1/(1-\alpha)}]^{1-\alpha} \bar{a}/\bar{\Pi}. \tag{9}$$

Notice that (9) implies a solution for the aggregate productivity penalty $\bar{\Pi}$ in terms of the initial Π's, the technology factors $\{a_k\}$ and the 'world' prices $\{p_k\}$. In what follows it is convenient to suppress explicit accounting for the technology factors, so henceforth the a's are set equal to one (or equivalently, the Π's can be understood as $\Pi_k/a_k, \forall k$).

The Cobb-Douglas specific factors model thus yields a GDP function with the constant elasticity of transformation (CET) form. The elasticity of

[4] In physical science, specific gravity refers to the density of an object normalized by the density of water. In economics, stretching the reference, specific gravity refers to the opportunity cost of a good as its marginal labor requirement normalized by the labor requirement of a constant labor requirement frictionless good, in a setting where the latter is equivalent to opportunity cost.

[5] Solve the labor market clearance condition for the equilibrium wage, then use the Cobb-Douglas property $wL/\alpha = g$.

transformation is equal to $\alpha/(1-\alpha)$, the ratio of labor's share to capital's share, while the distribution parameters $\{\lambda_k\}$ are capital allocation shares.[6] The supply share for any good k is given by:

$$s_k = p_k X_k/g = \frac{\lambda_k \widetilde{p}_k^{1/(1-\alpha)}}{\sum_k \lambda_k \widetilde{p}_k^{1/(1-\alpha)}} \tag{10}$$

where $\widetilde{p}_k = p_k/\Pi_k$.

Because the Cobb-Douglas special case will be used extensively to obtain clean results, it is important to consider how representative are its properties. For present purposes, these properties do seem to be representative. Let g_k denote the partial derivative of g with respect to $\widetilde{p}_k$. The specific factors model in general implies $g_{kj} < 0; k \neq j$, $g_{kk} > 0$ and $sum_j g_{kj}\widetilde{p}_j = 0$. The supply share function (10) represents a wider class that cannot deviate very much from the properties of (10). The wider class of share functions must be homogeneous of degree zero in the prices and retain the derivative properties of g, hence:

$$\frac{\partial s_k}{\partial \widetilde{p}_j}\frac{\widetilde{p}_j}{s_k} = \delta_{kj} - s_j + \frac{g_{kj}\widetilde{p}_j}{g_k}.$$

In the Cobb-Douglas case, $g_{kj}\widetilde{p}_j/g_k = -s_j\alpha/(1-\alpha); k \neq j$ and $g_{kk}\widetilde{p}_k/g_k = (1-s_k)\alpha/(1-\alpha)$. In the general case the same sign properties obtain, as do the same adding up properties.

The characterization of prices and productivity presented in this section is useful in summarizing the implications of trade costs for productivity in partial equilibrium. But since prices and multilateral resistances are endogenous in the world general equilibrium, it is useful to build a general equilibrium model of the relation of trade costs and productivity.

4 World Trade Equilibrium

Within each goods class, a CES sub-expenditure function allocates expenditure across trading partners. For goods class k, the expenditure share on shipments from j delivered to h is given by

$$\left\{\frac{\beta_{kj}\widetilde{p}_k^j t_k^{jh}}{P_k^h}\right\}^{1-\sigma_k}.$$

[6]The CET form is commonly used in applied general equilibrium modeling. The microfoundations provided here for the CET structure may prove useful in this context.

Here, the unit cost of production $\widetilde{p}_k^j$ is augmented by iceberg trade cost factor t_k^{jh} to yield the destination h price of k from j.

Market clearance with balanced trade implies

$$s_k^j g_j - \sum_h \theta_k^h \Big\{ \frac{\beta_{kj} \widetilde{p}_k^j t_k^{jh}}{P_k^h} \Big\}^{1-\sigma_k} g^h = 0, \tag{11}$$

$\forall k, j$. This system of equations determines the set of unit costs, $\widetilde{p}_k^j$, one for each k and j. The system is homogeneous of degree zero in the unit costs (understanding that the inward multilateral resistance indexes are homogeneous of degree one in the unit costs, being CES price indexes), hence relative unit costs only are determined.

For concreteness, the upper level expenditure allocation that determines the θ's may be specified as based on CES preferences for final goods.[7] For each country j, its expenditure share on goods of class k is given by

$$\theta_k^j = \gamma_k \Big\{ \frac{P_k^j}{P^j} \Big\}^{1-\epsilon} \tag{12}$$

where P_k^j is the inward multilateral resistance for country j in goods class k and P^j is the CES index of inward multilateral resistances for country j. The distribution parameters γ_k and the substitution parameter ϵ are common across countries.

4.1 Production and Trade Patterns

The model implies strong restrictions on the cross section pattern of trade. Using the market clearance equations and the definition of the supply shares, the unit costs are determined as an increasing power function of a shift 'parameter' in demand relative to supply:

$$\widetilde{p}_k^j = \Big\{ \frac{D_k^j}{\omega^j \lambda_k^j (\Pi_k^j)^{\sigma_k - 1}} \Big\}^{\frac{1-\alpha}{\alpha + \sigma_k(1-\alpha)}} (G^j)^{1/(\alpha + \sigma_k(1-\alpha))} \tag{13}$$

where

$$D_k^j = \beta_{kj}^{1-\sigma_k} \Theta_k,$$

[7] Any other separable form yields qualitatively similar results.

$$\Theta_k = \sum_h \theta_k^h \omega^h,$$

$$\omega^h = g^h / \sum_h g^h.$$

(13) is derived using (3) in (11). Here the demand shifter D_k^j is the product of a k specific component Θ_k reflecting tastes in the global economy for good k and a (j,k) specific 'quality' parameter $\beta_{kj}^{1-\sigma_k}$ reflecting tastes within goods class k for varieties from origin j. In the Cobb- Douglas case of (12), where $\epsilon \to 1$, Θ_k is a parameter, hence so is D_k^j.

The incidence of trade costs in varieties from j in class k, Π_k^j, drives the unit cost (suppliers' price) lower for the empirically relevant case $\sigma_k > 1$. Unit costs are increasing in the demand side drivers D_k^j, aggregate demand Θ_k and quality $\beta_{kj}^{1-\sigma_k}$. On the supply side, bigger country size ω^j and bigger sectoral allocations of specific factors λ_k^j both reduce unit costs. Finally, G^j represents the influence of the overall cost push on unit costs coming from all sectors in location j.

The GDP shares are may be expressed as reduced form equations in the international equilibrium using (13). Let $\eta_k = \alpha + \sigma_k(1-\alpha)$. Then:

$$s_k^j = (\lambda_k^j)^{1-1/\eta_k} (\Pi_k^j)^{(1-\sigma_k)/\eta_k} (D_k^j)^{1/\eta_k} (\omega^j)^{-1/\eta_k} G_j^{1-\sigma_k}. \qquad (14)$$

The adding up condition on the shares implies that G^j can be solved for in terms of the shares, the Π's and the D's: $1 = \sum_k s_k^j$. The solution exists and is unique since $\eta_k = \alpha + \sigma_k(1-\alpha) \geq 1$ and hence the elasticity $(1-\alpha)/\eta_k$ in (13) is less than one for $\sigma_k \geq 1$, the empirically relevant case. The solution for the G's yields the solution for the GDP's and thus for the world GDP shares $\{\omega^j\}$.

A special case clarifies. A closed form solution for G^j in terms of the ω's and the parameters arises in the case where $\sigma_k = \sigma, \forall k$:

$$G^j = \{(\omega^j)^{-1/\eta} \sum_k (\lambda_k^j)^{1-1/\eta} (\Pi_k^j)^{(1-\sigma)/\eta} (D_k^j)^{1/\eta}\}^{1/(\sigma-1)}. \qquad (15)$$

Define $\Lambda^j \equiv \sum_k (\lambda_k^j)^{1-1/\eta} (\Pi_k^j)^{(1-\sigma)/\eta} (D_k^j)^{1/\eta}\}^{1/(\sigma-1)}$. Then $g^j = R^j \Lambda^j \omega^{j-1/\eta}$, where $R^j = (L^j)^\alpha (K^j)^{1-\alpha}, \forall j$. The GDP shares simplify to

$$\omega^j = \left\{ \frac{\omega^j}{[\sum_j R^j \Lambda^j \omega^{j-1/\eta}]^{-\eta}} \right\}^{-1/\eta}$$

This equation system has a unique solution for the GDP shares. The denominator is interpreted as a CES deflator for real activity in the world R^W :$\sum_j g^j/[\sum_j R^j\Lambda^j\omega^{j-1/\eta}]^{-\eta} = R^W$.

In the case of $\sigma_k = \sigma$ the reduced form GDP share equations simplify to

$$s_k^j = \frac{(\lambda_k^j)^{1-1/\eta}(\Pi_k^j)^{(1-\sigma)/\eta}(D_k^j)^{1/\eta}}{\sum_k(\lambda_k^j)^{1-1/\eta}(\Pi_k^j)^{(1-\sigma)/\eta}(D_k^j)^{1/\eta}}. \tag{16}$$

As compared to (14), (16) eliminates the effect of country size on the equilibrium pattern of production. This simplification is not likely to distort the implications much, since the effect of country size tends to cancel out even in the more general case. Based on (16):

Proposition 2 *In the special case of equal elasticities of substitution in expenditure (with $\sigma > 1$) and uniform Cobb- Douglas production functions, the equilibrium production share is*

1. *increasing in the capital allocation share λ_k^j;*
2. *increasing in the demand 'parameter' D_k^j, the product of global market size Θ_k and national quality $\beta_{kj}^{1-\sigma_k}$;*
3. *increasing in the dispersion of D_k^j/λ_k^j and*
4. *decreasing in the incidence of trade costs Π_k^j.*

The third item follows because the deflator in (16) is concave in D/λ for $\eta\geq 1$. The economic intuition for this effect is essentially that the dispersion of D/λ measures the extent to which the sectoral allocation of the specific factor is inefficiently matched to demand. The analysis is drawn out at the end of Section 4.1.

If Proposition 1 applies, as is plausible despite non-uniform trade costs as argued above, then (16) implies that capital infusion reaps an externality via the resulting decline in the incidence of trade costs on supply. Larger capital allocations λ_k^j gain market share through their direct effect in (16) and through their knock-on effect in lowering Π_k^j.

The reduced form unit cost equations simplify when $\sigma_k = \sigma$ to

$$\widetilde{p}_k^j = (\omega^j)^{-1/(\sigma-1)} \frac{(D_k^j/\lambda_k^j\Pi_k^j)^{\sigma-1})^{(1-\alpha)/\eta}}{\left\{\sum_k(D_k^j)^{1/\eta}(\Pi_k^j)^{(1-\sigma)/\eta}(\lambda_k^j)^{1-1/\eta}\right\}^{1/(\sigma-1)}} \tag{17}$$

Compared to (13), the special case (17) implies that larger countries have uniformly lower unit production costs.

The implications of (17) for equilibrium 'competitiveness' are very intuitive and sharp:

Proposition 3 *In the special case model, all else equal:*

1. *larger specific endowments lower costs;*
2. *larger world demand for a good raises its cost;*
3. *higher quality costs more;*
4. *higher incidence of trade costs lowers unit costs;*
5. *bigger countries have lower costs.*
6. *higher dispersion of D_k^j/λ_k^j raises unit costs.*

The third point, higher quality costs more, is less obvious than it might seem. The CES model of preferences implies that some of each variety will be demanded, so it is not true that lower quality must have a lower price to be purchased by anyone.[8] Proposition 3 (3) states that in general equilibrium, higher quality goods have higher unit costs, all else equal. The sixth point reflects the concavity of the deflator in (17) in D/λ, with economic intuition that the mismatch of the sectoral allocation of capital with the pattern of demand lowers GDP, hence the relative size effect in the world economy improves the terms of trade.

The model implies very strong restrictions on the equilibrium pattern of trade. The ratio of gross exports to GDP in the special case of equal elasticities of substitution is given by

$$s_k^j - \theta_k^j \left\{ \frac{\beta_{kj} t_k^{jj} \widetilde{p}_k^j}{P_k^j} \right\}^{1-\sigma}. \tag{18}$$

Using (17) and the definition of θ_k^j, (12), and imposing $\epsilon = \sigma$ this reduces to

$$s_k^j - (s_k^j)^{(\eta-1)/\eta} \frac{\Delta_k^{jj}}{(D_k^j (\Pi_k^j)^{1-\sigma})^{1-1/\eta^2}} \tag{19}$$

[8]The interpretation of $\beta_{kj}^{1-\sigma_k}$ as a quality parameter is natural from examining the sub-utility function that lies behind the CES expenditure function: starting from equal consumption of each variety, the consumer's willingness to pay is higher the larger is $\beta_{kj}^1 - \sigma_k$.

where $\Delta_k^{jj} \equiv \gamma_k(\beta_{kj} t_k^{jj}/P^j)^{1-\sigma}$ and s_k^j is given by (16). The implications are that:

Proposition 4 *in the special case model the ratio of gross exports to GDP is*

1. *increasing in s_k^j, which moves according to Proposition 2;*
2. *increasing in D_k^j,*
3. *decreasing in the incidence of trade costs Π_k^j, and*
4. *decreasing in Δ_k^{jj}.*

Each item in the proposition is intuitive.

The levels of trade follow from scaling up the GDP shares by national GDP's. The implications of specific gravity model for the cross country pattern of aggregate production and wages are very strong.

For any pair of countries j and h, the ratio of their GDP's is given by

$$\frac{g^j}{g^h} = \frac{R^j}{R^h}\frac{G^j}{G^h},$$

where $R^j = (L^j)^\alpha (K^j)^{1-\alpha}, \forall j$. The influence of prices and productivity on the relative GDP's comes through the relative G's. Using the decomposition form (9),

$$\frac{G^j}{G^h} = \frac{[\sum_k \lambda_k^j (p_k^j)^{1/(1-\alpha)}]^{1-\alpha}}{[\sum_k \lambda_k^h (p_k^h)^{1/(1-\alpha)}]^{1-\alpha}} \frac{1/\bar{\Pi}^j}{1/\bar{\Pi}^h} \tag{20}$$

The first ratio on the right hand side is the terms of trade. The second ratio is the relative productivity of the two economies due to trade costs.

Relative prices and relative productivities are both determined in general equilibrium by the full structure of the model. In the Cobb-Douglas special case, the reduced form (15) applied to the ratio of G's implies:

$$\frac{G^j}{G^h} = \left\{\frac{R^j}{R^h}\right\}^{-1/(1+\eta(\sigma-1))} \left\{\frac{\sum_k \lambda_k^j (D_k^j/\lambda_k^j (\Pi_k^j)^{\sigma-1})^{1/\eta}}{\sum_k \lambda_k^h (D_k^h/\lambda_k^h (\Pi_k^h)^{\sigma-1})^{1/\eta}}\right\}^{1/(\sigma-1)(1+\eta(\sigma-1))}. \tag{21}$$

The first term on the right in (21) is a relative size effect, implying in light of (20) that bigger countries in real terms have lower wages and GDP price deflators, but not so much as to lower relative nominal GDP. The second term on the right in (21) is a composition effect reflecting the match of sector specific factor allocations to the pattern of demand in the global economy.

The model has very strong implications for relative wages. The relative wage (using $w = g_L$) is given by

$$\frac{w^j}{w^h} = \left\{\frac{K^j/L^j}{K^h/L^h}\right\}^{1-\alpha}\frac{G^j}{G^h}$$

Using (21), relative wages are obviously increasing in the relative capital/labor ratio and decreasing in the relative inefficiency of the match of sectoral allocations to demand. One key influence on the latter is the average outward multilateral resistance: the relative wage is lower the higher is the relative outward multilateral resistance.

Intuition about the composition effect is provided by considering the hypothetical efficient allocation of the K's. In partial equilibrium, for given prices, the identical Cobb-Douglas production structure assumed here in the special case will usually imply that countries completely specialize in one good, a corner solution to the maximization problem. In general equilibrium, however, due to the love of variety structure of preferences, prices will adjust to permit diversification. With efficient allocation of the K's, it is readily shown that $D_k^j(\Pi_k^j)^{1-\sigma}/\lambda_k^j = 1, \forall k, j$.[9] Then $\sum_k \lambda_k^j(D_k^j\Pi_k^{1-\sigma}/\lambda_k^j)^{1/\eta}$ measures how far short of efficient allocation of its specific factors economy j is. Relative wages and GDP's are increasing in this inefficiency due to the relative size effect operating on terms of trade in the world economy.

4.2 Productivity, Prices and Trade Frictions in General Equilibrium

Trade frictions and productivity are linked through their effects on world prices. A coherent account of the effect of trade frictions on productivity at given equilibrium prices is given by $\bar{\Pi} = g(p,\cdot)/g(\widetilde{p},\cdot)$. In making cross section comparisons of productivity levels, it is also useful, and for some

[9] The λ's must be chosen to equate the value of marginal product of capital in each sector. Using the GDP function this implies: $g_{\lambda_k}/K = g_K, \forall k$. For the special Cobb-Douglas case this implies that $s_k = \lambda_k$. Then in general equilibrium it must be true that $D_k\Pi_k^{1-\sigma}/\lambda_k = \bar{c}$, a constant. Moreover, by (10) and (16), $\bar{c} = 1$.

This also implies that $\widetilde{p}_k = 1, \forall k$. This property is no surprise in light of the Ricardian property of the specific factors model when the capital/labor ratios are all equal in the long run.

purposes more useful, to incorporate the effect of the trade frictions on the prices.

The point is most sharp in the special case model when the specific factors are efficiently allocated. In general equilibrium for this special case, the GDP's are just equal to the real activity indexes $L^{\alpha}K^{1-\alpha}$ because the activity deflators G are all equal to 1. There is *no* effect of globalization on productivity; all gains are passed on to consumers in the form of consumer price index declines (as iceberg costs fall). In effect, supply is infinitely elastic for this case (the supply schedule in Figure 1 is horizontal) so all the incidence of distribution friction falls on consumers. All the Π's are equal to 1. The size of the gains in real income depend on the pattern of trade cost declines as they affect each consuming location.

Most reasonable models in contrast imply shared incidence of trade frictions due to finite elasticity of supply in general equilibrium. For example, the special case in demand where $\sigma_k = \sigma$ along with Cobb-Douglas production, sector specific capital and identical labor shares yields the real activity deflator in equilibrium as (15):

$$G^j = \{\omega^{j-1/\eta} \sum_k (\lambda_k^j)^{1-1/\eta} (\Pi_k^j)^{(1-\sigma)/\eta} (D_k^j)^{1/\eta}\}^{1/(\sigma-1)}.$$

The effect of the incidence of trade frictions on G can be calculated and decomposed by sector with this formula (for the given equilibrium GDP shares and expenditure patterns E_k). The overall productivity effect is captured with the uniform Π that is equivalent to the actual set of $\{\Pi_k^j\}$'s:

$$\bar{\Pi}^j = \left\{ \frac{\sum_k (\lambda_k^j)^{1-1/\eta} \beta_{kj}^{(1-\sigma)/\eta} E_k^{1/\eta}}{\sum_k (\Pi_k^j)^{(1-\sigma)/\eta} (\lambda_k^j)^{1-1/\eta} \beta_{kj}^{(1-\sigma)/\eta} E_k^{1/\eta}} \right\}^{\eta/(\sigma-1)}.$$

This formula or its predecessor provide an explanation for cross section differences in productivity in world trading equilibrium that incorporates the effect of multilateral resistance on world prices.

5 Intermediate Inputs and Selection

Intermediate products trade comprises a large and growing share of world trade. Vertical disintegration is apparent — integrated production broken apart, with components produced in one location and assembled in another.

Presumably the gains from vertical disintegration have a powerful impact on productivity. A simple extension of the specific factors model of production readily encompasses intermediate products trade.

The boundary between components still produced within one location and those produced elsewhere and traded is central to the phenomena to be explained. In the multi-country context, essentially the same boundary phenomenon arises because only a small portion of the potential bilateral trade links have any positive trade flows. Countries able to fill more of the links in intermediate products trade presumably reap a productivity benefit.

The data show that larger markets are served by more suppliers. The natural explanation is that fixed costs impose a barrier that selects only those markets large enough to be profitable to serve. Helpman, Melitz and Rubinstein (2007) treat selection in a gravity model of final goods trade. Their method is extended here to cover selection in intermediate products trade, and the gravity model is embedded in the specific factors model of production. The treatment of selection in this section naturally applies to the final goods trade of preceding sections.

Firms are assumed to be competitive in factor markets (the common labor market and the sector specific factor markets, one for each sector) and monopolistic competitors in product markets. The marginal firms earn zero profits in production and distribution to each destination's market. Due to the CES specification, all firms mark up their unit costs by a common factor, inducing a distribution of prices that mirrors the distribution of productivities of extant firms.

The analysis is static for simplicity, so the fixed costs of entering production and trade are treated as sunk. Different countries will have different total factor productivities, due to their differing entry decisions in production, while the fixed costs of trade further alter total factor productivities through their effects on multilateral resistance.

The simplification to a static model shuts down an important general equilibrium factor market linkage between entry decisions and current production costs. See Melitz (2003) for a proper dynamic treatment in a one factor production model. The pragmatic reason for treating the static case here is to avoid the complexity of simultaneously treating entry and the reallocation of the specific factors.

5.1 Specific Factors Production with Intermediates

The production function for each industry k is comprised of the production functions of those firms that earn non- negative profits. At a prior stage, firms choose to enter production and then receive a Hicks-neutral productivity draw from a probability distribution. Those firms unlucky enough to receive draws too low to allow breaking even exit from production. The average productivity in industry k, $1/\bar{a}_k$, is determined by the cutoff productivity of the marginal firm in combination with the parameters of the productivity draw distribution. Average productivity is for present purposes taken as given. Since average productivity is Hicks-neutral, it enters the specific factors general equilibrium model multiplicatively with multilateral resistance.

The average productivity is associated with an average price, a constant markup over the the average unit cost of extant firms. See Melitz (2003) for details. This setup allows treatment of the heterogeneous firm model as a representative firm model easily linked to the general equilibrium production theory of preceding section. Profits are earned by inframarginal firms, and form part of the rents earned by the sector specific factors.

Intermediate products enter for simplicity as just a single intermediate product, potentially produced as a variety at each location.[10] The CES aggregate of the varieties is an input into production of all final goods and the intermediate good at each location. To ease notation, suppress country indexes. The production function for product k is given by

$$X_k = f^k(L_k, K_k, M_k), k = 1, ..., m;$$

where M_k is the quantity of the CES aggregate intermediate input used in sector k and sector m is the intermediate goods production sector. Specializing to the Cobb-Douglas case,

$$f^k = L_k^{\alpha} K_k^{1-\alpha-\nu} M_k^{\nu} / \bar{a}_k \Pi_k$$

Let P_m denote the price of the intermediate input used by the home country, a CES aggregate of the intermediate products purchased from all trading origins while p_m denotes the price of the intermediate input produced at home. Cost minimization combines with the labor market clearance condition to yield the GDP function $g(\widetilde{p}, P_m, L, K, \{\lambda_k\})$ with a closed form in

[10]The methods used here readily scale up to any number of intermediates.

the special case given by

$$L^{\alpha/(1-\nu)}K^{1-\alpha/(1-\nu)}[(\sum_{k=1}^{m}\lambda_k\widetilde{p}_k^{1/(1-\alpha-\nu)})^{1-\alpha-\nu}P_m^{-\nu}]^{1/(1-\nu)}c. \tag{22}$$

Here, c is a constant term combining the parameters, while $\widetilde{p}_k = p_k/\bar{a}_k\Pi_k$, the 'efficiency unit cost' in sector k.

5.2 Selection to Trade

Now allow for a fixed cost of exporting to each market, a cost that must be borne by each firm active in county i in goods class k. Helpman, Melitz and Rubinstein (2007) derive the gravity model with selection and show that there are two consequences that contrast with the preceding model: first, some bilateral trade flows may be zero due to no firm being able to pay the cost of entry, and second, where bilateral trade is positive it reflects both substitution on the intensive margin as in the preceding model and substitution on the extensive margin due to selection into exporting by marginal firms. The exposition below reviews their model, and reformulates it to highlight the role of multilateral resistance in both intensive and extensive margins.

The firm is assumed to be a monopolistic competitor facing a continuum of other firms selling to a market characterized by Dixit-Stiglitz love of variety preferences (and the analogous setup for intermediate products) represented by a CES expenditure function in each goods class. The mass N_i^k of firms that enter into production is given. Firms that enter receive a productivity draw from a Pareto distribution. In any market worth serving, the profit maximizing firms price to market with a constant markup over their costs $\mu_k = \sigma_k/(\sigma_k - 1)$.

The cost of a firm to serve its own market (assuming that $t_{ii}^k = 1$ for simplicity) is given by $\widetilde{p}_k^i$ times a_k^i, the inverse of the firm's productivity draw. Denote aggregate expenditure on product class k at destination j by E_k^j and use the CES expenditure system to allocate expenditure across origins. Sales by i to country $j \neq i$ are profitable only if $a_k^i \leq a_k^{ij}$ where a_k^{ij} is defined by the zero profit condition:

$$(1 - 1/\mu_k)(\frac{t_{ij}\widetilde{p}_k^i a_k^{ij}}{\mu_k P_k^j})^{1-\sigma_k} E_k^j = f_k^{ij}$$

Here, f_k^{ij} denotes the fixed cost.

It eases notational clutter in what follows to temporarily suppress the separate accounting for each goods class k, and to move the location indexes to the subscript position. Define the selection variable $V_{ij}(a_{ij})$ where

$$V_{ij} = \int_{a_L}^{a_{ij}} a^{1-\sigma_k} dF(a)$$

for $a_{ij} \geq a_L$ while

$$V_{ij} = 0$$

otherwise. Here, F is the cumulative density function. Denote the expenditure in location j on the generic good shipped from all origins as E_j while the value of shipments to all destinations from location i is denoted Y_i. Then the bilateral import value of shipments is given by

$$X_{ij} = (\frac{\widetilde{p}_i t_{ij}}{P_j/\mu})^{1-\sigma} E_j N_i V_{ij}.$$

Now derive the gravity model. The total value of shipments is

$$Y_i = \sum_j X_{ij} = \widetilde{p}_i^{1-\sigma} N_i \sum_j (\frac{t_{ij}}{P_j/\mu})^{1-\sigma} V_{ij} E_j.$$

First, solve market clearance for $\widetilde{p}_i^{1-\sigma} N_i$. Next, define

$$\Pi_i^{1-\sigma} = \sum_j (\frac{t_{ij}}{P_j/\mu})^{1-\sigma} V_{ij} E_j / Y \qquad (23)$$

where $Y = \sum_i Y_i = \sum_j E_j$. Substitution yields the gravity model:

$$X_{ij} = (\frac{t_{ij}}{P_j \Pi_i/\mu})^{1-\sigma} V_{ij} Y_i E_j / Y,$$

where

$$P_j^{1-\sigma} = \sum_i (\frac{t_{ij}}{\Pi_i/\mu})^{1-\sigma} V_{ij} Y_i / Y. \qquad (24)$$

A further simplification is helpful. Notice that $\{P, \Pi\}$ are homogeneous of degree 1/2 in μ, hence X_{ij} is homogeneous of degree zero in μ. Thus μ can be eliminated by a normalization.

The selection equation can be restated to highlight the role of multilateral resistance. Using normalized P, Π, selection is controlled by:

$$(1 - 1/\mu)(\frac{a_{ij}t_{ij}}{P_j \Pi_i})^{1-\sigma} E_j y_i / Y = f_{ij}. \tag{25}$$

Here, $y_i = Y_i / N_i Y$, the country i average sized firm's shipment share as a fraction of world shipments.

There are three implications. First, notice that the gravity model with selection combines the effects of trade costs on the intensive margin with their effects on the extensive margin acting through V_{ij}. Higher fixed costs reduce volume while larger markets draw more entrants. Second, despite the normalization, μ plays a role in selection. Incorporating variation across goods class, higher markup (lower elasticity) goods classes will have more firms selected into exporting, all else equal. Third, most importantly, the multilateral resistance variables incorporate both the productivity penalty imposed by the incidence of trade costs and the productivity gain garnered by the incidence of selection into trade.

The formal model is completed by specifying a distribution function for G. With the Pareto distribution used by Helpman, Melitz and Rubinstein, let the Pareto parameter be κ. Then

$$V_{ij} = \frac{\kappa a_L^{\kappa - \sigma + 1}}{(\kappa - \sigma + 1)(a_H - a_L)} W_{ij}$$

$$W_{ij} = max[(a_{ij}/a_L)^{\kappa - \sigma + 1} - 1, 0].$$

Helpman, Melitz and Rubinstein estimate selection with a Probit regression, then use these estimates to control for selection in the second stage gravity model regression with positive trade flows. Identification is achieved with an exclusion restriction that readers may find unconvincing (common religion affects fixed costs but not variable costs). The proposed research aims at more convincing exclusion restrictions by thinking of fixed costs as sunk (hence for example exchange rate variability and expropriation risk will affect selection but not variable cost) and by exploiting commodity class characteristics.

5.3 Selection, Productivity and Trade Patterns

The solution for the multilateral resistance terms in (23)-(24) determines the productivity and comparative advantage implications of the incidence of trade costs. The fixed export costs are treated as sunk costs for simplicity. Then the special case Cobb-Douglas model yields the GDP function (22). Higher incidence of trade costs in intermediate inputs penalizes GDP more heavily.

Due to the separability of the GDP function, the reduced form production shares are independent of the incidence of trade costs on intermediate inputs P_m. This separability implies that all the production and trade pattern results of Section 4 apply.

The influence of selection on production and trade patterns is isolated in the outward multilateral resistance terms, the incidence of trade costs on productivity, while the effect of selection on aggregate productivity also enters through the inward multilateral resistance for intermediates. The last formal task of the paper is to extract the implications of fixed trade costs for productivity, looking for an extension of Proposition 1.

Proposition 1′ *With selection in the uniform border barriers case, Proposition 1 continues to hold, and in addition, multilateral resistance is lower the more firms that are selected into trade, all else equal.*

The Appendix provides the proposition in the case of a uniform trade cost between countries.

Selection is endogenous, with (25) permitting a characterization conditional on the expenditure and production shares. More firms will be selected from i to trade with j the larger is the market in j, the larger is i's share of world shipments, and the fewer (hence larger) are i's firms. Thus selection reinforces the productivity implications of trade costs: big market share shippers bear lower incidence of trade costs. In contrast, selection tends to offset the higher incidence induced by larger expenditure shares.

6 Conclusion

This paper provides a platform for consistent aggregation of the fine structure of trade costs into productivity measures that are suitable for the analysis of productivity differences across goods, countries and time. The implications of the productivity differences at a point in time for the pattern of production and trade are explored in detail for the special case of the specific factors model.

The paper points to future empirical work. First, it will be valuable to estimate multilateral resistance indexes for an appropriately disaggregated set of goods for a set countries and years. Second, the paper points to use of the multilateral resistance indexes as an explanatory variable for the pattern of production and trade.

The paper also points to future theoretical refinement. The extreme simplicity of the model buys strong results, while hinting that the results hold in less restrictive cases. How robust is t he model?

7 References

Anderson, James E. (1979), "A Theoretical Foundation for the Gravity Equation," *American Economic Review*, 69, 106-16.

Anderson, James E. and Eric van Wincoop (2003), "Gravity with Gravitas: A Solution to the Border Puzzle", *American Economic Review*, 93, 170-92.

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

Davis, Donald and David Weinstein (2001), "An Account of Global Factor Trade", *American Economic Review*, 91, 1423-53.

Eaton, Jonathan and Samuel Kortum (2002), "Technology, Geography and Trade", *Econometrica*, 70(5), 1741-1779.

Helpman, Elhanan, Marc J. Melitz and Yonah Rubinstein (2007), "Trading Partners and Trading Volumes", NBER Working Paper No. 12927.

Melitz, Marc J. (2003), "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity", *Econometrica*, 71(6), 1695-1725.

8 Appendix: Multilateral Resistance and Size

8.1 Proof of Proposition 1

Proposition 1 states that with uniform border barriers, multilateral resistance is decreasing in the supply share of the countries, increasing in the expenditure share of countries and increasing in the net imports of countries.

The assumption of the Proposition imposes a *uniform* international trade cost t on all trades across borders, will internal trade (from j to j) assumed to be frictionless. A slightly different notational convention is used here than in the text. The analysis takes a representative good, so the subscript k is suppressed. It then eases notation slightly to move the location indexes from the superscript to the subscript position. Let s_j denote country j's share of world shipments of the generic good, while b_i denotes the expenditure share of country i of the generic good.

The system of equations that determine P_i, Π_i for all i, j is given by the simple case version of (2)-(3), which reduces under the changes in notation and the implications of the uniform international trade cost to:

$$P_j^{1-\sigma} = t^{1-\sigma}\bar{h} + (1 - t^{1-\sigma})s_j/\Pi_j^{1-\sigma} \tag{26}$$

$$\Pi_i^{1-\sigma} = t^{1-\sigma}\bar{h}' + (1 - t^{1-\sigma})b_i/P_i^{1-\sigma}. \tag{27}$$

Here, $\bar{h} = \sum_i s_i \Pi_i^{\sigma-1}$ and $\bar{h}' = \sum_j P_j^{\sigma-1} b_j$. Global market clearance requires $\sum_j b_j = \sum_j s_j$, which implies that $\bar{h} = \bar{h}'$. $\bar{h}$ can be solved for eventually but the essential step is to provide the closed form solution for multilateral resistance given $\bar{h}$.

Multiply both sides of (26) by $\Pi_i^{1-\sigma}$ and multiply both sides of (27) by $P_i^{1-\sigma}$. Use the resulting equality to solve

$$\Pi_i^{1-\sigma} = P_i^{1-\sigma} + \frac{(1 - t^{1-\sigma})(b_i - s_i)}{\bar{h}t^{1-\sigma}}.$$

Then substitute into (26) and extract the positive root[11] of the resulting quadratic equation in the transform $P_i^{1-\sigma}$:

$$2P_i^{1-\sigma} = \gamma + [\gamma^2 + 4(1 - t^{1-\sigma})b_i]^{1/2} \tag{28}$$

where

$$\gamma = \bar{h}t^{1-\sigma} - \frac{(1 - t^{1-\sigma})(b_i - s_i)}{\bar{h}t^{1-\sigma}}.$$

[11]The positive root of the quadratic is necessary for P to be positive.

At this solution

$$2\Pi_i^{1-\sigma} = \bar{h}t^{1-\sigma} + [\gamma^2 + 4(1-t^{1-\sigma})b_i]^{1/2}.$$

The solution for $\bar{h}$ is implicit in the next expression, obtained from using the definition of $\bar{h}$ and the preceding solution for Π_i,

$$\bar{h} = \sum_i s_i[\bar{h}t^{1-\sigma} + (\gamma^2 + 4(1-t^{1-\sigma})b_i)^{1/2}]^{-1},$$

where γ is given as a function of $\bar{h}$ above.

Multilateral resistance (inward and outward) is unambiguously decreasing in supply share s_i at equilibrium and unambiguously increasing in expenditure share b_i in equilibrium. It is unambiguously increasing in the net import share $b_i - s_i$ for given expenditure shares.

8.2 Proof of Proposition 1′

Assume that trade costs take the form of uniform border barriers. With the selection mechanism of Section 5, the multilateral resistance indexes must satisfy

$$\Pi_i^{1-\sigma} = \bar{h}'_i t^{1-\sigma} + (1-t^{1-\sigma})V_{ii}b_i/P_i^{1-\sigma}$$
$$P_j^{1-\sigma} = \bar{h}_j t^{1-\sigma} + (1-t^{1-\sigma})V_{jj}s_j/\Pi_j^{1-\sigma},$$
$$\forall i, j.$$

Here,

$$\bar{h}_j = \sum_i V_{ij}s_i\Pi_i^{-1+\sigma},$$
$$\bar{h}'_i = \sum_j V_{ij}b_jP_j^{-1+\sigma}.$$

The market clearance condition implies that $\bar{h}_j = \bar{h}'_j$, as in the preceding case of multilateral resistance without selection. The main difference from Proposition 1 lies in the fact that $\bar{h}_i$ now varies by country. Thus all the results of Proposition 1 continue to hold as *ceteris paribus* propositions.

The implications of selection are run by the variation of the aggregate indexes $\bar{h}_j$ over countries. *The larger is the set of links with positive trade, and the larger the number of firms that participate in trade in each link, the larger is* $\bar{h}_i$.

Multilateral resistance has the same form of solution as in the no-selection case, but now the quadratic solution 'parameter' γ is a function not just of market size s_i or b_i but also of $\bar{h}_i$. Thus

$$2P_i^{1-\sigma} = \gamma_i + [\gamma_i^2 + 4(1 - t^{1-\sigma})b_i]^{1/2}$$

$$2\Pi_i^{1-\sigma} = \bar{h}_i t^{1-\sigma} + [\gamma_i^2 + 4(1 - t^{1-\sigma})b_i]^{1/2}$$

where

$$\gamma_i = \bar{h}_i t^{1-\sigma} - \frac{(1 - t^{1-\sigma})(b_i - s_i)}{\bar{h}_i t^{1-\sigma}}.$$

Multilateral resistance is decreasing in $\bar{h}_i$ directly. It is also ordinarily decreasing via the influence of $\bar{h}_i$ on the 'parameter' γ_i:

$$\frac{\partial \gamma_i}{\partial \bar{h}_i} \bar{h}_i = 2\bar{h}_i t^{1-\sigma} - \gamma_i,$$

which is positive for $b_i - s_i > 0$ and likely to be positive for all but large exporters $s_i - b_i >> 0$. This implies that *the more firms that are selected into trade, the lower the multilateral resistance.*

Table 1: Multilateral Resistance for US and Canadian Regions

Provinces	
Alberta	1.65
British Columbia	1.55
Manitoba	1.65
New Brunswick	1.62
Newfoundland	1.74
Nova Scotia	1.63
Ontario	1.51
Prince Edward Island	1.67
Quebec	1.53
Saskatchewan	1.66
States	
Alabama	1.39
Arizona	1.49
California	1.31
Florida	1.38
Georgia	1.37
Idaho	1.49
Illinois	1.35
Indiana	1.36
Kentucky	1.37
Louisiana	1.41
Maine	1.43
Maryland	1.31
Massachusetts	1.33
Michigan	1.37
Minnesota	1.42
Missouri	1.38
Montana	1.50
New Hampshire	1.39
New Jersey	1.33
New York	1.30
North Carolina	1.37
North Dakota	1.47
Ohio	1.35
Pennsylvania	1.33
Tennessee	1.38
Texas	1.41
Vermont	1.40
Virginia	1.35
Washington	1.43
Wisconsin	1.39
Rest USA	1.50

Notes : This table reports multilateral resistance indices for US states and Canadian provinces, based on data and trade cost estimates from Anderson and van Wincoop (2003). The assumed elasticity of substitution is 8.