

The Incidence of Gravity *

James E. Anderson
Boston College and NBER

January 10, 2009

Abstract

The high trade costs inferred from gravity are rarely used in the wide class of trade models. Two related problems explain this omission of a key explanatory variable. First, national seller and buyer responses to trade costs depend on their incidence rather than on the full cost. Second, the high dimensionality of bilateral trade costs requires aggregation for most practical uses in interpretation or standard trade modeling. This paper provides an intuitive description of a resolution to the aggregation and incidence problems. For each product, it is as if each province or country sells to a world market containing all buyers and buys from from that market containing all sellers, the incidence of aggregated bilateral trade costs being divided between sellers and buyers according to their location. Measures of incidence described here give intuitive insight into the consequences of geography, illustrated with results from Anderson and Yotov (2008). The integration of the incidence measures with standard general equilibrium structure opens the way to richer applied general equilibrium models and better empirical work on the origins of comparative advantage.

JEL Classification: F10.

Contact information: James E. Anderson, Department of Economics, Boston College, Chestnut Hill, MA 02467, USA.

*I am grateful to Adrian Wood for helpful comments.

The gravity model is one of the great success stories of economics. The success of the model is its great explanatory power: the equations fit well statistically and across many different data sets give quite similar answers: inferred bilateral trade costs are big, varying with distance and border crossings.

Despite this success, the inferred trade costs have had little impact on the broader concerns of economics until very recently. The costs have been hard to integrate with other models used to understand trade.¹ There are two difficulties. First, national buyer and seller responses to bilateral trade costs depend on their *incidence* instead of the full cost. Second, the high dimensionality of bilateral trade costs requires aggregation, both for elementary comprehension of magnitude and for use in the wide class of trade models that focus on resource and expenditure allocation as sectoral aggregates.

This paper discusses a solution to both problems. Measures of aggregate incidence described here provide intuitive guides to the consequences of geography, illustrated with results drawn from a study of Canada's changing economic geography (Anderson and Yotov, 2008). The paper goes on to show how the aggregated incidence measures can be used in a standard class of applied general equilibrium trade models. This opens the way to richer applied work, both in simulation and in econometric inference.

The solution to incidence and aggregation problems exploits the properties of the structural gravity model. Gravity was initially developed by analogy with the physical gravity model, using only its '2-body' representation. Anderson (1979) derived an economic theory of gravity from demand structure, at the same time providing the economic analogy to a solution for the physical N-body problem. See also Anderson and van Wincoop (2003, 2004), who coin the term 'multilateral resistance' for the key solution concept that captures the N-body properties of the trade system.

Inward and outward multilateral resistance are, respectively, the demand and supply side aggregate incidence of trade costs. For each product, it is as if each country shipped its output to a single 'world' market and shipped home its purchases from the single world market. Each country's multilateral resistances depend on *all* bilateral trade costs in the world system, not just the bilateral cost between country i and country j and not just i 's costs with

¹Eaton and Kortum (2002) develop a Ricardian many country many good trade model with bilateral trade costs. The methods used in this paper apply to their model as well, but the promise of the methods lies in integrating gravity with the much wider class of general equilibrium trade models in the common toolkit.

all its partners and j 's costs with all its partners. Multilateral resistance thus embeds the effect of trade costs between third and fourth parties, meeting an objection to the earlier gravity model raised by Bikker in Chapter 2.² For example, via its impact on multilateral resistances around the world, the implementation of the NAFTA agreement should theoretically have an effect on the trade of its members with EU countries through multilateral resistance, and moreover it also has an effect on the trade of the EU countries with Japan and China.

The integration of aggregated incidence measures with standard general equilibrium models builds on the structure sketched in Anderson and van Wincoop (2004). Allocation between sectors is separated from allocation within sectors by the simplifying assumption of 'trade separability'. Production for sale in all destinations depends only on the 'average' sellers incidence of trade costs while purchases from all origins depend only on the 'average' buyers' incidence. Within sectors the global bilateral distribution of shipments is conditioned on each region's aggregate expenditure and production allocations to the sector. This paper develops the implications of the structure further.

1 Trade Frictions and Incidence

Some impediments to trade are due to the sellers' side of the market (like export taxes or export infrastructure user costs), some are due to the buyers' side of the market (like import taxes or import infrastructure user costs) while still others are difficult to identify with either side of the market (such as transport costs, information costs, or costs due to institutional insecurity). Whatever their origin, however, all trade impediments will have economic incidence that is shared between buyer and seller. Economic incidence differs from naive incidence based on who initially pays the bill: the seller typically may pay for transport cost and insurance, but the sellers' price to the

²Bikker claims, misleadingly, that previous gravity models do not account for substitution between trade flows. This is wrong, because the class of theoretical models following Anderson (1979) are based on CES demand structure. But the empirical literature until very recently did not act on Anderson's original point that ordinary gravity models were biased estimators because they omitted the influence of what now are called multilateral resistances. Bikker additionally proposes an Extended Gravity Model with a demand structure less restricted than the CES, a procedure that has both advantages and disadvantages.

buyer includes a portion of these costs with the portion being economically determined.

The gravity model is used first to predict a benchmark of what bilateral trade flows would look like in a frictionless world. Then the deviations of actual from benchmark flows are econometrically related to a set of proxies for trade costs. The results are used to infer unobservable trade costs associated with such frictions as distance and trade policy barriers, discriminatory regulatory barriers and other variables related to national borders — different languages or legal systems, differential information about opportunities, extortionist actions at border bottlenecks, all have been found to impede trade very significantly. Chapters 3 and 5 pursue some of the many important issues of specification of the proxies themselves and the way in which they enter the trade cost relationship.

Each producing nation in each product class provides a variety³ to the world market. The total value of outward and inward shipments is given. Each consuming nation buys available varieties from the world market.

The gravity model makes the enormously useful simplifying assumption that tastes (and for intermediate products, technology) are the same everywhere in the world. In a truly frictionless world, this would mean that the shares of expenditure falling on the varieties of products from every origin would be the same in every destination. That is, destination h would spend the same share of its total expenditure on goods class k as would destination $i \neq h$, and within goods class k would spend the same share on goods from origin j as would destination $i \neq h$.

Bilateral trade costs shift the pattern of bilateral shipments from the frictionless benchmark. Below, the expenditure shares are assumed to derive from Constant Elasticity of Substitution (CES) preferences. Different expenditure shares are explained by different bilateral trade costs in a tightly specified fashion.

It is analytically convenient to develop the gravity model logic by focussing on shipments from ‘the factory gate’ to a world market that includes local distribution. This differs from many standard treatments of costly trade, so it is important to keep in mind that ‘shipments to the world market’ in this paper include shipments that end up being sold at home and

³For purposes not germane here, free entry monopolistic competition can endogenize the number of firms that together provide each nation’s varieties. See Bergstrand (1985) and his subsequent work.

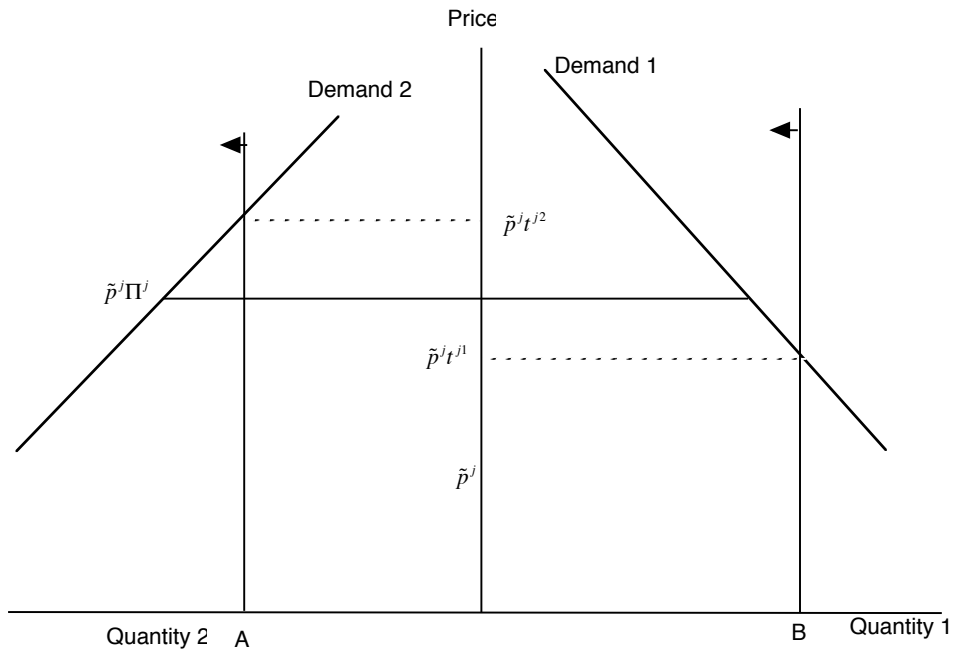
‘purchases from the world market’ include products produced at home. Local distribution costs are ordinarily less than distribution costs to the rest of the world, but are in fact substantial (as discussed in Anderson and van Wincoop, 2004).

Arbitrage ties together prices in different locations. For shipment from j to h in goods class k , the zero-profit arbitrage condition implies that $p_k^{jh} = \tilde{p}_k^j t_k^{jh}$ where p_k^{jh} denotes the buyers’ price in h for goods in class k purchased from source j , \tilde{p}_k^j denotes the cost of good k at j ’s ‘factory gate’ and $t_k^{jh} > 1$ is the trade cost markup factor.

The average (across all destinations h) incidence of trade frictions on the supply side — the average impact on the sellers’ price of the set of t_k^{jh} ’s that contains all destinations h — is represented by an index Π_k^j for each product category k in each country j . The index is derived from the simple notion that the complex actual shipment pattern is equivalent in its impact on sellers to a hypothetical world economy that behaves as if there was a ‘world’ destination price for goods k delivered from j , $p_k^j = \tilde{p}_k^j \Pi_k^j$.

Figure 1 illustrates the hypothetical equilibrium for the case of two markets, suppressing the goods class index k for clarity. Market 1 to the right may be thought of as the home market with market 2 to the left being the export market. Distribution costs are lower in the home market than in the export market. The equilibrium factory gate price \tilde{p}^j is preserved by maintaining the total quantity shipped while replacing the nonuniform trade costs with the uniform trade cost Π^j .

Figure 1. Quantity-Preserving Aggregation



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

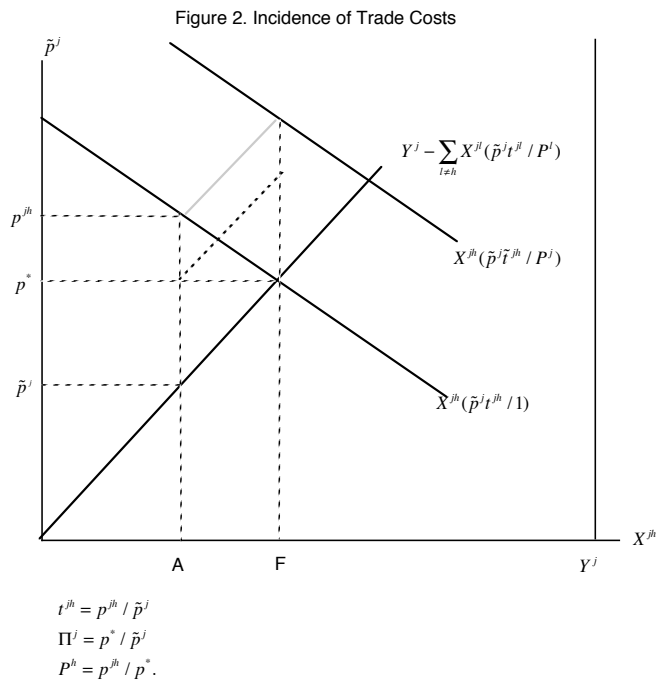
The incidence of bilateral trade costs on the buyers' side of the market is given by t_k^{jh}/Π_k^j , taking away the sellers' incidence. The average buyers' incidence of all bilateral costs to h from the various origins j is given by the buyers' price index P_k^h . The balance of the effects of all bilateral trade costs on the trade flow in goods class k from origin j in destination h , X_k^{jh} , is given by $t_k^{jh}/\Pi_k^j P_k^h$. This relative trade cost incidence form is intuitive, but strictly valid only for the special Constant Elasticity of Substitution (CES) form of demand structure, detailed in the next section.

The usage of 'incidence' here to describe the indexes Π_k^j and P_k^h is derived from the familiar partial equilibrium incidence analysis of the first course in economics. Π_k^j and P_k^h will be given an exact formal description below that reveals how they can be calculated in practice and how they indeed capture incidence in conditional general equilibrium (i.e., preserve the same factory gate prices, conditional on observed aggregate shipments and expenditures).

Figure 2 illustrates the division of a single trade cost into its incidence on the buyers' and sellers' prices. The demand and supply schedules of the first course in economics are converted here into value functions, the revenue from sales for sellers and the expenditure for buyers. This conversion aids connection later on with the gravity model. For simplicity the goods class subscript k is again dropped. The actual trade flow is given by the value at F, which is conveniently chosen to be equal to the frictionless value of trade. F represents an average effect on bilateral trade (including local trade) from the system of trade costs. This is because the value of shipments Y^j is fixed and the bilateral costs simply shift the pattern around in such a way that the effects of bilateral trade costs average out.

A hypothetical partial equilibrium bilateral trade flow is given by point A projected down from the 'demand' schedule labeled $X^{jh}(\tilde{p}^j t^{jh}/1)$, the expenditure associated with the buyers' price p^{jh} when the index $P^h = 1$, the frictionless value of the price index. The 'demand' schedule expresses the dependence of demand on the sellers' price \tilde{p}^j , holding constant the price index P , which in reality is possible only when h spends a very small share on goods from j . It is downward sloping because the elasticity of demand is assumed to be greater than one (an assumption that is empirically sound based on the extensive gravity literature.) The 'supply' schedule is a residual from supply to all other destinations for j 's good: $Y^j - \sum_{l \neq h} X^{jl}(\tilde{p}^j t^{jl}/P^l)$. It is upward sloping in \tilde{p}^j because all the bilateral 'demand' schedules are downward sloping. The 'supply' schedule expresses dependence on sellers' price

while holding constant a combination of all the price indexes, a constancy which is possible only in a very special case.



The elementary analysis of incidence decomposes t^{jh} into the sellers' incidence Π^j and the buyers' incidence t^{jh}/Π^j based on the frictionless equilibrium price p^* generated by the intersection of demand and supply with no frictions. The vertical line segment between the demand and supply schedules

at trade level A divides into the sellers' incidence $p^*/\tilde{p}^j = \Pi^j$ and the buyers' incidence p^{jh}/p^* . Incidence depends on the relationship between demand and supply elasticities. For large supply elasticities, as might be expected in the context of bilateral trade where supply is diverted from many other large markets, the sellers incidence falls toward its limit of 1 and all incidence is borne by buyers.

The gravity model incorporates the effect of all bilateral trade frictions on Π^j , as will be detailed below; and on the demand for goods from j through their effect on average prices of substitute goods from all sources acting through the price index P^h . Replacing the actual system of trade costs with the hypothetical system $\tilde{t}^{jh} = \Pi^j P^h$ results in the frictionless quantity demanded on the vertical line erected from point F on Figure 2.

For the very special case chosen to simplify the presentation of principles, the effect of the switch to hypothetical trade costs on the supply schedule to market h is nil. More importantly, the special case assumes that the actual trade cost is equal to the hypothetical trade cost: $t^{jh} = \tilde{t}^{jh}$. The dashed lines parallel to the supply schedule project the standard textbook incidence decomposition northeast to the line segment between the two demand schedules on the vertical line from frictionless sales point F. Thus the sellers incidence is Π^j and the buyers' incidence is P^h .

It is tempting to think that incidence is determined by the same ratio of demand to supply elasticities as in the textbook case, and indeed for the case drawn intuition is aided by exactly this analogy drawn from projecting backward and forward between the line segment above A and the line segment above F.⁴ Unfortunately, the general equilibrium relationship between markets is far too complex for the relative elasticity intuition of the diagram to be illuminating.

Thinking about the sales pattern of supplier j in terms of Figure 2, but for cases where the bilateral flow is not equal to the frictionless flow, some actual trade flows will be above average (to the right of the frictionless level on Figure 1). This is certainly true for the local shipments from j to itself, the well-known phenomenon of home bias in sales patterns.

Most trade flows, likely all but local ones, will be to the left of the frictionless level in Figure 2. In an average sense, across markets the trade flow shifts

⁴Using a linear approximation, the standard algebra of incidence implies that $P = t(\beta + \delta)/(\beta t + \delta)$ and $\Pi = (\beta t + \delta)/(\beta + \delta)$, where $-\beta$ is the demand slope and δ is the supply slope.

from the frictionless level cancel out because the volume of goods shipped from j must add up to the given amount Y^j . (This is true in conditional general equilibrium: the effect of actually changing trade costs would result in reallocations of resources such that the amounts produced would change in full general equilibrium.)

The sellers' incidence Π is conceptually identical to a productivity penalty in distribution. j 's factors of production must be paid less in proportion to Π in order to get their goods to market. For intermediate goods demand, P is a productivity penalty reflecting the distribution cost incidence that falls on users of intermediate goods. This link to productivity is exploited by Anderson and Yotov (2008) to convert their incidence results into TFP measures.

2 Determination of Incidence

The incidence of trade costs within each sector is determined in a conditional general equilibrium that distributes bilateral shipments across origin-destination pairs for given bilateral trade costs and given total shipments and total expenditures. See the next section for a defense of this separation and the validity of conditioning on total shipments and expenditures.

Impose CES preferences on the (sub-)expenditure functions for each goods class k , where σ_k is the elasticity of substitution parameter for goods class k and $(\beta_k^j)^{1-\sigma_k}$ is a quality parameter for goods from j in class k . (For intermediate products demand impose the analogous CES structure for the cost functions.) Then the bilateral trade flow from j to h in goods class k , valued at destination prices, is given by

$$X_k^{jh} = \left(\frac{\beta_k^j \tilde{p}_k^j t_k^{jh}}{P_k^h} \right)^{1-\sigma_k} E_k^h \quad (1)$$

where \tilde{p}_k^j is the cost of production of good k in the variety produced by j , $t_k^{jh} > 1$ is the trade cost markup parameter in class k from j to h , E_k^h is the expenditure on goods class k in destination h , and P_k^h is the true cost of living index for goods class k in location h . P_k^h is defined by

$$P_k^h \equiv \sum_j [(\beta_k^j \tilde{p}_k^j t_k^{jh})^{1-\sigma_k}]^{1/(1-\sigma_k)}.$$

Let the value of shipments at *delivered* prices from origin h in product class k be denoted by Y_k^h .

Market clearance requires:

$$Y_k^j = \sum_h \left(\frac{\beta_k^j \tilde{p}_k^j t_k^{jh}}{P_k^h} \right)^{1-\sigma_k} E_k^h. \quad (2)$$

Now solve (2) for the quality adjusted efficiency unit costs $\{\beta_k^j \tilde{p}_k^j\}$:

$$(\beta_k^j \tilde{p}_k^j)^{1-\sigma_k} = \frac{Y_k^j}{\sum_h (t_k^{jh}/P_k^h)^{1-\sigma_k} E_k^h}. \quad (3)$$

Based on the denominator in (3), define

$$(\Pi_k^j)^{1-\sigma_k} \equiv \sum_h \left(\frac{t_k^{jh}}{P_k^h} \right)^{1-\sigma_k} \frac{E_k^h}{\sum_h E_k^h},$$

where $\sum_h E_k^h$ replaces $\sum_j Y_k^j$.

The outward multilateral resistance term Π_k^j gives the supply side incidence of bilateral trade costs to origin j . It is as if j ships to a single world market at cost factor Π_k^j . To see this crucial property, divide numerator and denominator of the right hand side of (3) by total shipments of k and use the definition of Π , yielding:

$$(\beta_k^j \tilde{p}_k^j \Pi_k^j)^{1-\sigma_k} = Y_k^j / \sum_j Y_k^j. \quad (4)$$

The right hand side is the world's expenditure share for class k goods from country j . The left hand side is a 'global behavioral expenditure share', understanding that the CES price index is equal to one due to the normalization implied by summing (4):

$$\sum_j (\beta_k^j \tilde{p}_k^j \Pi_k^j)^{1-\sigma_k} = 1. \quad (5)$$

The global share is generated by the common CES preferences over varieties in the face of globally uniform quality adjusted efficiency unit costs $\beta_k^j \tilde{p}_k^j \Pi_k^j$.

Outward and inward multilateral resistances can readily be computed once the empirical gravity model has been estimated and the implied trade

costs t_k^{jh} have been constructed from its results. Substitute for quality adjusted efficiency unit costs from (3) in the definition of the true cost of living index, using the definition of the Π 's:

$$(P_k^h)^{1-\sigma_k} = \sum_j \left\{ \frac{t_k^{hj}}{\Pi_k^j} \right\}^{1-\sigma_k} \frac{Y_k^j}{\sum_j Y_k^j}. \quad (6)$$

Collect this with the definition of the Π 's:

$$(\Pi_k^j)^{1-\sigma_k} = \sum_h \left\{ \frac{t_k^{hj}}{P_k^h} \right\}^{1-\sigma_k} \frac{E_k^h}{\sum_h E_k^h}. \quad (7)$$

These two sets of equations jointly determine the inward multilateral resistances, the P 's and the outward multilateral resistances, the Π 's, given the expenditure and supply shares and the bilateral trade costs, subject to a normalization. A normalization of the Π 's is needed to determine the P 's and Π 's because (6)-(7) determine them only up to a scalar.⁵

Notice that the Π 's and the P 's generally differ, even if bilateral trade costs are symmetric: $t_k^{hj} = t_k^{jh}$.

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 $\tilde{t}_k^{jh} = \Pi_k^j P_k^h$. The budget constraint (6) and market clearance (7) equations continue to hold at the same prices, even though individual bilateral trade volumes change.

Thus for each good k in each country j , from the point of view of the factory gate, it is as if a single shipment was made to the 'world market' at the average cost. On the demand side, similarly, inward multilateral resistance consistently aggregates the demand side incidence of inward trade and production frictions. From the point of view of the 'household door' it is as if a single shipment was made from the 'world market' at the average markup. This discussion is the formal counterpart to the intuitive claim that Figure 2 represents an 'average' good in a partial equilibrium representation of the incidence decomposition.

The CES specification of within-class expenditure shares, after substitu-

⁵If $\{P_k^0, \Pi_k^0\}$ is a solution to (6)-(7), then so is $\{\lambda P_k^0, \Pi_k^0/\lambda\}$ for any positive scalar λ ; where P_k denotes the vector of P 's and the superscript 0 denotes a particular value of this vector, and similarly for Π_k .

tion from (3), implies the gravity equation

$$X_k^{jh} = \left\{ \frac{t_k^{jh}}{\Pi_k^j P_k^h} \right\}^{1-\sigma_k} \frac{Y_k^j E_k^h}{\sum_j Y_k^j}. \quad (8)$$

For a ‘representative’ trade flow with $t_k^{ij} = \Pi_k^j P_k^h$, the gravity equation implies that the flow is equal to the frictionless flow, conditional on $Y_k^j E_k^h$. The normalization (5) in combination with a frictionless equilibrium normalization $\sum_j (\beta_k^j \tilde{P}_k^{j*})^{1-\sigma_k} = 1$ completes the extension of the partial equilibrium theory of incidence to conditional general equilibrium.⁶

3 Incidence in Practice

Computing the multilateral resistances is readily operational, given estimates of gravity models that yield the inferred t ’s and given the global shares, $\{E_k^h / \sum_h E_k^h, Y_k^j / \sum_j Y_k^j\}$. The multilateral resistance indexes permit calculation of Constructed Home Bias $(t_k^{hh} / \Pi_k^h P_k^h)^{1-\sigma_k}$ indexes. These are equal to the ratio of predicted trade to frictionless trade in the home market.

Multilateral resistance is equivalent to a Total Factor Productivity (TFP) penalty. The Π ’s push below the world price the ‘factory gate’ price \tilde{p} that sellers receive, which determines what they can pay their factors of production. Similarly, the P ’s raise the price that buyers must pay for final or intermediate goods. The effect of changes in multilateral resistance on real GDP can be captured by a linear approximation to the TFP change:

$$-\sum_k \frac{Y_k^j}{\sum_k Y_k^j} \hat{\Pi}_k^j - \sum_k \frac{E_k^j}{\sum_k E_k^j} \hat{P}_k^j.$$

In practice it will often be useful to avoid having to solve for the equilibrium quality adjusted efficiency unit costs needed for normalization (5). A units choice can always be imposed — for example, $\beta_k^i P_k^{*i} = 1, \forall k$ or $P_k^h = 1, \forall k$ for some convenient reference country i or h . (The former convention implies $(\Pi_k^i)^{1-\sigma_k} = Y_k^i / Y_k, \forall k$.) In any case, *relative multilateral resistances are what matters for allocation in conditional general equilibrium.*

⁶The demands and supplies of the upper level allocation remain constant, as in the partial equilibrium Figure 2. The aggregation of bilateral t ’s into ‘ Π ’s at constant \tilde{p} is analogous to the aggregation shown in Figure 1.

The normalization choice can be freely made for convenience in calculation and interpretation.

Anderson and Yotov (2008) construct multilateral resistances for Canadian provinces 1992-2003 using these procedures. They first estimate gravity coefficients, then construct the implied t 's and then use (6)-(7) with a normalization to calculate the provincial multilateral resistances for each year and province for 18 goods classes. They find that outward (sellers' incidence) multilateral resistance is around 5 times bigger than inward (buyers' incidence) multilateral resistance. The sellers' incidence is negatively related to sellers' market share, and it falls significantly over time. Theoretical reasons are offered for these results.

The fall in sellers' incidence over time suggests a powerful and previously unrecognized force of globalization: specialization in production is driving a fall in the sellers' incidence of trade costs despite constant gravity coefficients and hence t 's. There is a big fall in Constructed Home Bias. The fall in sellers' incidence results in rises in real GDP that are around 1% per year for star performing provinces. These are big numbers.

4 General Equilibrium

Multilateral resistances permit a useful integration of gravity with general equilibrium production and expenditure structures. Consider an iterative process that moves back and forth between the lower level allocation of shipments within sectors and the upper level allocation of resources across sectors. In full general equilibrium computations that simulate equilibria away from the initial conditional equilibrium analyzed in preceding sections, the upper level general equilibrium model at initial Π 's and P 's yields the new global shares $\{E_k^h / \sum_h E_k^h, Y_k^j / \sum_j Y_k^j\}$ and the normalized quality adjusted efficiency unit costs $\{\beta_k^j \tilde{p}_k^j\}$. The new shares and efficiency unit costs are then inputs into the computation of the new multilateral resistances, the Π 's and P 's from (6)-(7) subject to a normalization. The new multilateral resistances are then plugged into the resource allocation module to solve for the new shares and efficiency unit prices. The iterative process continues until convergence.

The determination of supply to the market from each origin and of expenditure at each destination is specified here in general equilibrium using the standard toolkit. It is analytically very convenient to exploit the trade

separability property assumed throughout to first describe allocation in each country for given world prices. Subsequently, global market clearance is used to determine the world prices.

4.1 Allocation for Given World Prices

Total supply from each origin and total expenditure at each destination are determined in an upper stage of general equilibrium. In each country, on the supply side each product class draws resources from the common pool and on the demand side each expenditure class draws from the common income. These total supply and total expenditure variables are taken as given in this stage of the model, in order to focus on the key determinants of the distribution of supply and expenditures across origin-destination pairs.

A tremendous simplification is achieved with the specializing assumption of *trade separability* — the composition of expenditure or production within a product group is independent of prices outside the product group. Separability permits consistent aggregation and a simple solution to the incidence problem.

On the supply side, separability is imposed by the assumption that 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. Goods are differentiated by place of origin, an assumption that has a deeper rationale in monopolistic competition. This setup enables two stage budgeting. A further specialization to CES structure for the separable groups yields operational multilateral resistance indexes that capture the inward and outward incidence of trade costs.

See Anderson (2008) for a full description of the upper level allocation of expenditure and production and characterization of the global equilibrium pattern of production and trade in the case of the specific factors model of production.

In each country, on the supply side each product class draws resources from the common pool and on the demand side each expenditure class draws from the common income. The simplifying assumption of *trade separability* allows treatment of these allocation decisions at the sectoral level, abstracting from bilateral patterns of shipment. Separability means that the composition of expenditure or production within a product group is independent of prices outside the product group.

On the supply side, separability is implied by the assumption that 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. Goods are differentiated by place of origin, an assumption that has a deeper rationale in monopolistic competition. This setup enables two stage budgeting. As the previous section shows, further specialization to CES structure for the separable groups yields yields operational multilateral resistance indexes that capture the inward and outward incidence of trade costs.

Given the separable setup the sellers' incidence Π is conceptually identical to a productivity penalty in distribution. Region j 's factors of production are paid less because of distribution frictions: metaphorically a portion $\Pi - 1$ of their pay melts away. Demand for intermediate goods faces a similar productivity penalty P reflecting the trade cost incidence on users of intermediate goods. This link to productivity is exploited by Anderson and Yotov (2008) to convert their incidence results into TFP measures. In final demand the sectoral incidence P acts like a uniform tax on consumption.

The implications are drawn out here in a generic general equilibrium model with final goods only. See Anderson (2008) for a full description of the upper level allocation of expenditure and production and characterization of the global equilibrium pattern of production and trade in the case of the specific factors model of production, including treatment of intermediate goods.

Each country produces and distributes goods to its trading partners. Production is more costly than with best practice by a Hicks neutral multiplier $a_k^j \geq 1$ for product k in country j . In other words, at the factor prices relevant for product k in country j , $a_k^j - 1$ more factors are used than needed with the best practice. Distribution to destination h requires additional factors to be used, in the proportion $t_k^{jh} - 1$ to their use in production (iceberg-melting distribution costs). Otherwise, products sold to all destinations are identical: perfect substitutes in production.⁷

The cost of product k from origin j at destination h is given by $p_k^{jh} = a_k^j t_k^{jh} \tilde{p}_k^j$, where \tilde{p}_k^j is the unit cost of production using best practice tech-

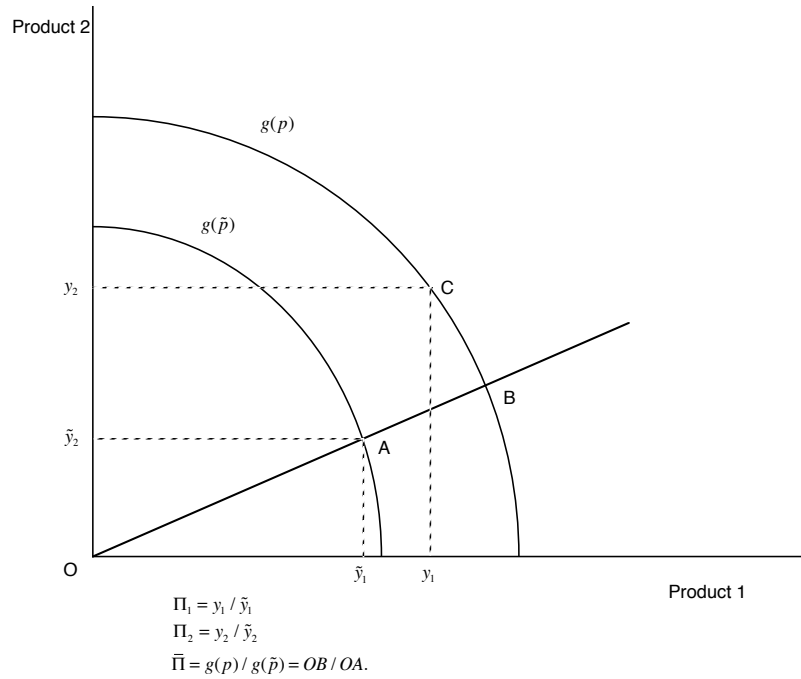
⁷The setup can easily be generalized to allow for a separable joint output structure in which products for each destination are imperfect substitutes via a Constant Elasticity of Transformation structure.

niques, called the ‘efficiency unit cost’. Since the a ’s and T ’s enter the model multiplicatively, they combine in a productivity penalty measure $T_k^{jh} = a_k^j t_k^{jh}$ that represents both trade frictions and frictions in the assimilation of technology. The preceding section derives the aggregate supply side incidence of trade costs Π_k^j for each product category k in each country j . Incorporating the technology penalty a into Π to conserve notation, let Π ’s now represent both distribution and technology penalties.

The key building block describing the supply side of the economy is the gross domestic product (GDP) function. It is written as $g(\tilde{p}^j, v^j)$ where v^j is the vector of factor endowments. g is convex and homogeneous of degree one in prices, by its maximum value properties. Take p^j as a given vector of ‘world’ prices. Then $\tilde{p}_k^j = p_k^j / \Pi_k^j, \forall k$. Using Hotelling’s lemma, $g_k^j \tilde{p}_k^j \Pi_k^j = Y_k^j$, where g_k^j denotes $\partial g^j / \partial \tilde{p}_k^j$.

Figure 3 illustrates. The actual bundle delivered to the market is $(\tilde{y}_1, \tilde{y}_2)$, point A. The value of delivered goods, GDP, is $g(p, \Pi) = p_1 \tilde{y}_1 + p_2 \tilde{y}_2$. This is also equal to the GDP available if the most efficient technology were to be used facing the prices $(\tilde{p}_1, \tilde{p}_2)$, hence $g(p, \Pi) = g(\tilde{p}, \iota) = p_1 y_1 / \Pi_1 + p_2 y_2 / \Pi_2$. Point C represents (y_1, y_2) , the hypothetical most efficient production and delivery possible based on the resources allocated to achieve the actual deliveries of point A.

Figure 3. Incidence and TFP



Two limiting cases clarify the conceptual basis of productivity used here. The case where trade costs are absent gives the standard TFP measurement. The production possibilities frontier through points C and B represents the most efficient technology where $a_i = 1 = t_{ij}; \forall i, j$. With $a_i > 1, t_{ij} = 1$, the production possibilities frontier through point A represents the actual technology frontier. For this case, $\Pi_i = a_i$. The case where productivity

frictions are absent in production is the case of pure iceberg trade costs, $a_i = 1, t_{ij} = T_{ij} \geq 1; \forall i, j$; and Π_i is the incidence of trade costs on sector i . Point A gives the bundle actually delivered while point C represents the bundle as it leaves the factory gate. The shrinkage is due to the iceberg melting trade costs. Digging into the iceberg metaphor to relate the concepts to national income accounting, gross output in each sector y_1, y_2 (represented by point C) effectively uses some of its own output as an intermediate good in order to achieve deliveries to final demand \tilde{y}_1, \tilde{y}_2 (represented by point A). The discussion shows that the metaphor of iceberg melting trade costs can be extended to productivity frictions that ‘melt’ the resources applied to produce before the shipments begin their journey to market.

The illustration in Figure emphasizes two separate aspects of the productivity frictions. Along ray OAB, the ratio OB/OA represents the average (across industries) productivity penalty, equivalent to the usual aggregate TFP notion. But point B differs from point C, the bundle that would be produced facing the same world prices that result in actual production A, but with frictionless production and distribution. The difference between B and C is due to $\Pi_1 > \Pi_2$, causing substitution in production away from the relatively penalized good 1.

4.2 Equilibrium World Prices

Now turn to the determination of global equilibrium prices. Let the value of shipments at *delivered* prices from origin h in product class k be denoted by Y_k^h . At efficiency production prices, the supply is valued at $g_k^j \tilde{p}_k^j$. Y_k^j is margined up from $g_k^j \tilde{p}_k^j$ to reflect the ‘average’ cost of delivery. Thus

$$Y_k^j = g_k^j \tilde{p}_k^j \Pi_k^j.$$

The link of the conditional general equilibrium in (6)-(7) to full general equilibrium uses the unified world market metaphor and $Y_k^j = g_k^j \tilde{p}_k^j \Pi_k^j$. The market clearance conditions in (4) can be rewritten as:

$$(\beta_k^j \tilde{p}_k^j \Pi_k^j)^{1-\sigma_k} = \tilde{p}_k^j \Pi_k^j g_k^j(\cdot) / \sum_j \tilde{p}_k^j \Pi_k^j g_k^j(\cdot). \quad (9)$$

For given Π 's and β 's, (9) solves for the \tilde{p} 's, the origin prices. The full general equilibrium obtains when the world production shares that arise with equilibrium \tilde{p} 's in (9) are consistent with the world production shares used

to solve for the Π 's in (6)-(7) subject to (5) while the Π 's that arise from (6)-(7) for given Y 's are consistent with the Π 's used in (9).

5 Conclusion

This paper describes a framework for decomposing trade costs into their incidence on buyers and sellers, aggregated as if all shipments are made to or from a world market. The results of its implementation for Canada's provinces demonstrate that most incidence falls on sellers. Over time, sellers' incidence is falling due to a previously un-noticed force of globalization — specialization of production. On-going work by Anderson and Yotov extends the empirical work to a world of more than 100 countries and 28 manufacturing goods.

6 References

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

Anderson, James E. (2008), "Gravity, Productivity and the Pattern of Production and Trade", Boston College.

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.

Anderson, James E. and Yoto V. Yotov (2008), "The Changing Incidence of Geography", NBER Working Paper No. 14423.

Bergstrand, Jeffrey H. (1985), "The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence", *Review of Economics and Statistics* 67 (1985), 471-481.

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

Feenstra, Robert C. (2004), *Advanced International Trade*, Princeton: Princeton University Press.

Helpman, Elhanan, Marc J. Melitz and Yonah Rubinstein (2008), "Trading Partners and Trading Volumes", *Quarterly Journal of Economics*, 123: 441-487.

Krugman, Paul R and Maurice Obstfeld (2006), *International Economics: Theory and Policy*, 7th Edition, Addison-Wesley.