Back to the Future: Gravity at Sixty

James E. Anderson
Boston College and NBER

1. Introduction

This essay bounces gravity off (what we would now call) its non-parametric origin forward into improved practice. [9] infers trade frictions from their effect on the observable pattern of trade with a model analogous to physical gravity. Trade flows from origin to destination are proportional to the product of economic masses at origin and destination. Taking out these scale effects, the residuals are inversely proportional to a function of economic distance, not necessarily its square. Then proxies for economic distance are fitted to the residuals. Subsequently [1] derived structural gravity from the spatial arbitrage equilibrium of exchange in a parametric CES demand system. [2] derives non-parametric gravity from spatial arbitrage equilibrium of exchange between economically rational agents alone. This structure generates trade frictions as residuals from the difference between welfare cost per unit in observable and as-if-frictionless equilibria. Comparability of the two equilibria is based on normalizing the world price vectors on the unit simplex, the standard practice in efficient general equilibrium theory based on the degree zero homogeneity in prices of excess demand systems.

Section 2 reviews Tinbergen's original development of gravity. Section 3 reviews the start of economic structural gravity in parametric form and its slow adoption in practice. Section 4 describes non-parametric structural gravity. The logic of [2] is illustrated here in diagrams with downward-sloping demand and the adding up properties of market clearing and budget constraints. En route, the essay's development of economic gravity prompts reflections on the history of economic thought and practice.

2. Origin Story

Tinbergen's quantitative model fit bilateral trade to his form of gravity.[1] For present purposes his model is given by

$$X_{ij} = M_i M_j / r_{ij} \tag{1}$$

where X_{ij} is the value of shipments from origin i to destination j at the buyers prices, M_i, M_j are the economic masses at i and j and r_{ij} is resistance to shipments from i to j due to economic distance. Think of r_{ij} as a residual to be explained by deeper determinants. The program resembles with productivity accounting as it developed to investigate the Solow residual.

Tinbergen proxied economic mass with GNP in origin and destination.[2] Economic distance was proxied by a list of variables associated with trade frictions. His strategy started from a base case with physical distance alone as a proxy for economic distance. The idea was to remove 'natural' variation as a first step, then analyze and account for residuals from the base case. The model was applied in log form with estimated coefficients to be thought of as due to an unknown reduced form. Tinbergen's focus was on quantifying the effects of frictions that could be amenable to policy, with dummy variables for membership in the predecessor to the EU in his case.[3]

In modern empirical practice terms, Tinbergen's economic mass proxies could have been replaced by fixed effects. In 1962, the sanitized term "fixed effects" lay in the future. They were called dummy variables, connoting that they could not inform us of their meaning. Dummy variable use was understood as a kludge, justified by controlling for irrelevant variation to achieve better fit of a structural model to data. The hundreds of dummy variables

[1] Interestingly, he does not use the term 'gravity' to describe his model, though the inspiration is unmistakable.

[2] In more recent times, GDP would probably be preferred, but in 1962 the difference between the two was small.

[3] Gravity appears in the Appendix of Tinbergen's book that is grandly entitled *Shaping the World Economy: Suggestions for an International Economic Policy.*

needed to control for economic mass at origins and destinations was literally unthinkable.

The applications that followed Tinbergen made use of various proxies for economic distance to quantify the effects of free trade agreements and other affinity measures following his 1962 example. But the results were difficult to understand in terms of economic theory. Where were the third party effects such as, for example, the trade diversion emphasized in customs union theory?

3. Parametric Structural Gravity

[1] offered a parametric model of arbitrage equilibrium of exchange that included modeling the missing third party effects. The Constant Elasticity of Substitution (CES) demand model focused attention on the trade elasticity parameter as conditioning the effect of frictions on bilateral trade. Economic masses were simply the value of sales from each origin at buyers prices and the total expenditures of buyers at each destination. Bilateral frictions were in relative form, deflated by the product of multilateral indexes of outward and inward frictions that included all missing third party effects. These were later called [[3]] outward and inward multilateral resistances in a structural application that estimated them. Relative resistance is the term used here for the ratio of bilateral resistance to the product of multilateral resistances and denoted R_{ij}. Relative resistance reduces trade in the CES model by $R_{ij}^{-\theta}$, $\theta > 0$ where θ is the trade elasticity parameter, the structural counterpart to Tinbergen's $1/r_{ij}$.

The [1] spatial arbitrage model was effectively ignored for many years because it offered no applicability. As the most poignant example, I the author could not think of how to apply it. The structural gravity trade flow equation in this case is

$$X_{ij} = \frac{Y_i E_j}{Y} \left(R_{ij}\right)^{-\theta}, \tag{2}$$

where Y_i and E_j are origin sales and destination expenditures, $Y = \sum_i Y_i$ and all variables are valued at buyer prices. Relative resistance R_{ij} is replaced in 2) by the ratio of its components, bilateral resistance τ_{ij} and the product of outward multilateral resistance Π_i

and inward multilateral resistance P_j. System estimation of (2) combined with the equations for multilateral resistances have required the computing resources of the space program given the computer technology of the 70's .

Eventually the exchange model of [1] and related forms of parametric structural gravity [[7]] were further developed and applied. Computer technology improved enormously, and dummy variables became respectable fixed effects, exceptionally respectable in their tight link to economic gravity structure. Structural gravity uses origin and destination fixed effects to control for, respectively, the product of expenditure and the effective attractive effect of inward multilateral resistance P_j^θ and the product of origin sales and effective attractive effect of outward multilateral resistance Π_i^θ. The theoretical structure can then back out the effects of the multilateral resistances. Purely bilateral effects of resistance $\tau_{ij}^{-\theta}$ were proxied following Tinbergen's strategy – distance and membership of trade agreements along with other plausible proxies. The *size* of the relative resistance and its components as opposed to its effect on trade required the trade elasticity θ, identified (or not) by trade purchases response to variation of a measurable trade cost such as tariffs or transport cost.

Doubts linger about the general applicability of the model and its results because of its severe restrictions on functional form. Generally, expenditure shares depend on the *entire vector* of relative prices. What about third good effects on the relevant expenditure share, especially when these involve complementarity? How general are results that appear so dependent on a very restrictive parametric specification of demand?

The return to non-parametric methods in [2] calculates relative resistance sufficient statistics R_{ij} directly from the observable pattern of trade. Thus the elegant simplicity of relative resistance R_{ij} represents the effect of frictions on trade across a broad class of demand systems. From this perspective, CES gravity is a conveniently simple parametric case of a general principle.

Non-parametric relative resistances R_{ij} and their relationship to trade provide an opportunity to examine how accurate are CES applications in the received applied literature.

[2] reports a tightly fitted estimate of θ for aggregate manufacturing based on fitting the variation in observed expenditure shares to the CES predictions of those shares that should vary with $R_{ij}^{-\theta}$. The estimated θ is very close to 1, a pleasing coincidental nod to Tinbergen: the case $\theta = 1$ is equation (1), understanding the residual $r_{ij} = R_{ij}$.

The value $\theta = 1$ is low compared to the literature; e.g. [8]. One explanation is familiar – aggregation bias – aggregate manufacturing will naturally have lower trade elasticities than do disaggregated sectors. The other explanation is methodological – the inferred trade frictions R_{ij} have *much* more variation than do observable prices (in [8]) or tariffs and transport costs used in other studies. A lower elasticity is thus required to match the variation in relative resistances to variation in observable trade flows. The greater variation of relative resistances than of prices presumably applies regardless of the level of aggregation because important frictions are not reflected in observed prices, yet picked up by inferred relative resistances. The size of the trade elasticity matters because the gains from trade measure of [5] rises as θ falls.

4. Non-parametric Structural Gravity

Arbitrage exchange equilibrium imposes the zero arbitrage profits and adding-up conditions as in [1] and the succeeding structural gravity literature. [2] replaces the CES demand system with the class of invertible demand systems. Invertibility satisfies the requirement that widely different destination expenditure choices be comparable as responses to different relative resistances within a common structure. [6] prove that 'connected substitutes' is sufficient for invertibility. Importantly, 'connected substitutes' includes complementarity and admits zeros associated with choke prices. The latter is an essential inclusion for bilateral trade because of its many zeros. Previous parametric structural gravity models are contained in this wider class.

Operational non-parametric sufficient statistics for arbitrage (exchange) gains from trade changes and terms of trade changes are a payoff that follows in [2] under approximation restrictions. The approximation includes the general translog, allowing for possible complementarity with $N \times (N-1)/2$ substitution parameters. This essay presents the ideas in an intuitive diagrammatic analysis that is useful for teaching.

4.1. Basic Setup

Begin with a broad definition of the spatial arbitrage model.

Definition A:

(i) Equilibrium spatial arbitrage – at each destination the buyer's full price deflated by trade frictions is equal to a common net-of-frictions seller cost at each origin.[4]

(ii) Each origin ships an endowment of a variety of a single product class that differs by origin to many, potentially all, destinations.

(iii) Markets clear – the value of all shipments from origin i valued at destination full prices must equal the sum of bilateral (including domestic) purchases.

(iv) Expenditures by each destination minimize the cost of the bundle purchased, given the convex preferences (final goods) or technology (intermediate goods).

Assumption (iv) implies standard value functions, the expenditure function for final goods and the cost function for intermediate goods. The value functions apply at the sectoral level since the model focuses on the modular general spatial equilibrium of a sector, as with most of the applied gravity literature. Conditional on the observed equilibrium level of sub-utility (final goods) or output (intermediate goods), the value functions are concave and homogeneous of degree one in the destination prices. The gradient gives the vector of demand at each destination (Shephard's Lemma).

The focus on a single sector means that arbitrage gains from trade and terms of trade concepts are applied at the sectoral level. In contrast, the usual usage applies this conceptual terminology to the aggregate economy. The relationship of the sectoral to the aggregate in

[4] The focus is on bilateral trade over long intervals such as yearly, rather than bilateral price difference behavior over short intervals such daily. The assumption is that systematic deviations from arbitrage equilibrium are eliminated, remaining observed differences being independent of observed trade flows.

much applied work uses Cobb-Douglas aggregators to provide a simple weighted sum to relate the two. Connecting the sectoral value functions to an aggregate value function that applies to the common structure allowing comparability across countries requires weak separability of the upper level value function with respect to its partition into sectors.

Non-parametric characterization of sectoral spatial equilibrium is based on a common expenditure function to be associated with observable demand vectors at different prices follows from a restriction of differences across agents. A sufficient restriction is: *(v) the common demand system is invertible.* The exposition henceforth focuses on final goods for simplicity, but all operations are equivalent for intermediate goods.

Each country distributes a single origin-specific member of the sectoral class, a "good", henceforth without scare quotes. The endowment of seller i is denoted y_i. Shipments from i to j are denoted x_{ij}. Unit costs received by sellers net of trade 'costs' are denoted c_i.

The price p_{ij} paid by buyers at j on goods from i includes trade costs and other frictions represented by cost factors to sellers and distributors $\tau_{ij} \geq 1$. 'Frictions' may include unobservable user costs and heterogeneity in preferences across destinations as well as endogenous trade services costs by profit-maximizing efficient trade services providers.[5] In the arbitrage equilibrium

$$p_{ij}/\tau_{ij} = c_i, \ \forall i, j. \tag{3}$$

Condition (3) is necessary and sufficient for zero arbitrage profits.

A key additional insight about the discipline imposed by arbitrage equilibrium follows from considering *aggregate* demand by the world for the product of origin i. Aggregate demand is observable from production data evaluated at delivered prices by imposing the market clearing condition. Thus each origin's good is effectively sold to the hypothetical

[5]The usual simplification in gravity models is fixed iceberg costs, but endogenous equilibrium frictions are admissible. This is because arbitrage disciplines the equilibrium relationship between bilateral trade costs. Allowing for endogenous trade costs comes much closer to the reality of transportation costs association with congestion and mode choice variation. See [4] for a tractable example based on short run fixed bilateral capacities.

world market. These demands are interpreted as if generated in an 'as-if-frictionless' equilibrium such that each destination counter-factually faces the same as-if-frictionless world price. Buyers in the actual equilibrium on average pay $c_i\Pi_i$, where c_i is the seller cost (net payment to sellers) and Π_i is the equilibrium (endogenous) average friction factor on shipments from i to the 'world' market. Alternatively, in the counter-factual, each destination buyer faces the same vector of 'world' prices $\{c_i\Pi_i\}$. The world average vector of frictions is constant but bilateral heterogeneity across destinations is eliminated for each seller i.

$c_i\Pi_i$ is recognized as the arbitrageur's opportunity cost of shifting marginal quantities of goods from i into any destination j. Buyers at destination j must have willingness to pay that covers the additional cost of reaching j, τ_{ij}/Π_i. It is convenient to use relative prices to describe buyer costs $c_i\tau_{ij}/P_j$ in actual destinations, where P_j is the buyers price index in j. The ratio

$$R_{ij} = \frac{c_i\tau_{ij}}{c_i\Pi_i P_j} = \frac{\tau_{ij}}{\Pi_i P_j} \tag{4}$$

is the key relative resistance variable.

Trade data are in value form, so it is convenient to work with expenditure shares. Parametric structural gravity solved the comparability problem of inferring R_{ij} from differing shares with the CES parametric form and estimation of a CES trade elasticity. The share form comes from dividing equation (2) by E_j. Equilibrium trade expenditure falls as R_{ij} rises with elasticity θ, justified with the restriction that CES preferences apply to all destinations identical except for destination specific taste shifters that are absorbed in the trade frictions τ_{ij}.

Non-parametric gravity retains comparability across destinations with the invertibility restriction (v). Destination specific taste shifters are still part of τ_{ij} and in addition price dependent non-homotheticity allows for differences in per capita incomes that are absorbed in the taste shifters.[6] In equilibrium, the cross section comparison of observed to as-if-

[6]The income effects matter for shifts in equilibrium, but are given in equilibrium and hence absorbed in τ_{ij} for cross-section comparison of the effects of trade frictions. Similarly, elements of frictions such as

frictionless equilibrium holds utility in each location constant, so demand curves are compensated demand curves.

Higher relative resistance R_{ij} reduces trade volume x_{ij}, *ceteris paribus*, because demand curves slope downward.[7] How much volume decreases depends locally on a slope or an elasticity. Large changes are implied by the observed variation of bilateral trade shares from their observed as-if-frictionless equilibrium shares, so local approximation by parametric representation appears dubious. The key step forward to derived relative resistances uses invertibility to imply that the deviation of R_{ij} from 1, explains the deviation of observed bilateral shares from the as-if-frictionless shares (equal to the world expenditure shares on each country's outputs).

The hypothetical world buyer faces a vector of prices $c_i\Pi_i$ and has expenditure shares B_i. Market clearance implies that purchase of each origin's good must sum to its supply y_i, hence the common as-if-frictionless expenditure share $B_i = \sum_j c_i\Pi_i x_{ij}/\sum_i c_i\Pi_i x_{ij}, \ \forall i$ must equal the observed world sales share at buyer prices $c_i\Pi_i y_i/\sum_i c_i\Pi_i y_i = s_i, \ \forall i$. This observable world vector of sales shares is 'as-if-frictionless', buyers everywhere face the same price vector, with origin-specific trade frictions Π_i that are uniform across destinations.

The first step in comparing actual expenditure shares to as-if-frictionless shares is to solve a units problem with price levels. The standard way to do this in applied general equilibrium modeling is to normalize the price vector, in this case the world price vector. In the as-if-frictionless equilibrium, choose units such that $\sum_i c_i\Pi_i y_i = \sum_i y_i$. Thus the equilibrium world price vector is constrained to lie on the unit simplex: $\sum_i c_i\Pi_i y_i/\sum_i y_i = 1$. The normalization equation imposes the adding up condition that the value of world sales is equal to the value of world expenditures. The same normalization applies to the observed

transportation costs are given over the cross section but may vary with shifts in the general equilibrium.

[7] The deeper reason is that cost minimization implies concavity of the expenditure function.

equilibrium, which implies that the observed price indexes are normalized to satisfy

$$\sum_i P_i u^i = \sum_i y_i = \sum_i u^i.$$

Here, P_i is the normalized price index for country i in the observed equilibrium, $P_i u_i = E_i$ is country i's expenditure, equal to the expenditure function with $P_i = e(\mathbf{p}^i)$, the concave and homogeneous of degree one function of the effective delivered price vector $\mathbf{p}^i$ in country i.

Note several important implications of this accounting system for the effects of trade frictions. First, gravity cannot account for the *level* of trade frictions, just their relative variation. Second, the normalization emphasizes that the observed trade flows imply that for some countries $P_j > 1$ while for others $P_j < 1$ by necessity. The combination implies that while the level of trade frictions probably inflicts losses on buyers in all countries, the existing losses are relatively beneficial to buyers in some countries while being exceptionally costly to buyers in other countries.

4.2. Gains from Trade and Relative Resistances

For each country j, given the setup in Section 4.1, the term $P_j - 1$ measures the relative loss per unit of real income to buyers due to frictions. From the social point of view, payments by domestic buyers go to domestic sellers. The fraction of expenditures on the domestic product is b_{jj} in destination j. $P_j - 1$ is the average incidence of (normalized) frictions borne by buyers in j. Country j is both buyer and seller, and faces higher than average trade frictions on imports. Its gains as seller domestically are offset by the loss to buyers. The loss per unit of utility of country j due to *cross-border* trade frictions is given by rearranging $P_j - 1 = \sum_i (P_j b_{ij} - s_i)$ to yield

$$L_j = P_j b_{jj} - s_j = P_j - 1 - \sum_{i \neq j} (P_j b_{ij} - s_i).$$

The loss due to frictions is simplified in the first equation above to a measure based on domestic sales. L_j is related to the familiar gains from trade measure of [5] in its comparison of the domestic trade share to an observable alternative. It differs in being a compensating variation measure rather than a real income measure as well as in other significant ways described in [2].

Relative resistance is based on obtaining R_{ij} from each $P_j b_{ij} - s_i$ using approximation techniques developed and discussed in [2]. Casual empiricism reveals that domestic expenditure shares b_{jj} are generally much larger than in as-if-frictionless equilibrium shares $B_j = s_j$, while import expenditure shares are much less than in as-if-frictionless equilibrium. Corresponding to this, intuition suggests $R_{jj} < 1$ and $R_{ij} > 1$; $\forall i \neq j$. The diagrammatic exposition below illustrates how the spatial arbitrage mechanism works to give intuition, and to give perspective on the quantification method for R_{ij} developed in [2].

4.3. Diagrammatic Analysis

The non-parametric model is illustrated by a supply and demand analysis in the goalpost diagram in Figure 1.[8] A system of generic demand schedules is restricted by assumptions (i)-(v) that guarantee comparability across countries. Figure 1 characterizes the equilibrium allocation from origin i to a particular destination j. Region i's residual supply to j is given by $x_{ij}^{RS} = y_i - \sum_{l \neq j} x_{il}$. The compensated demand schedule for goods from i in j is labeled x_{ij}^{D}, downward sloping due to the assumption of cost minimization by rational buyers. The residual supply schedule with frictions slopes upward because it is the difference between the endowment y_i and the sum of downward sloping demands being filled in all destinations other than j. For reference, the as-if-frictionless residual supply schedule is also drawn.

Worldwide aggregate demand for goods from i intersects the supply schedule y_i at price $c_i \Pi_i$, the price paid by a hypothetical buyer in the 'world' market. Π_i is the sellers' incidence of trade frictions on world sales, an index of the bilateral sellers' trade friction incidences

[8] The goalpost diagram is familiar to many economists from the standard exposition of the allocation of labor in the context of the specific factors model.

τ_{ij}/P_j including internal shipments τ_{ii}/P_i. The index reflects efficient spatial arbitrage $p_{ij} = c_i t_{ij}$, $\forall j$ for markets that are served.

Demand systems of the general class considered here are characterized by homogeneity of degree zero in prices $\{p_{ij}\}$. This implies that each destination has an ideal price index P_j and the allocation of expenditure is equivalently characterized by relative prices. Thus the left vertical axis in Figure 1 measures relative prices p_{ij}/P_j. The equilibrium shipment quantity x_{ij}^e is at the arbitrage equilibrium point E where demand intersects the residual supply function. x_{ij}^e is associated with relative price $p_{ij}/P_j = c_i\tau_{ij}/P_j$. $c_i\Pi_i$ is projected from the right vertical axis where world demand intersects the supply of good i, y_i. The units on the horizontal axis are in shares of shipments, so the right vertical axis is erected at $y_i/y_i = 1$. The upward sloping residual supply schedule imposes the requirement that willingness to pay $p_{ij}/P_j = c_i\tau_{ij}/P_j$ must cover the opportunity cost of diverting delivery of the marginal unit from the rest of the world $c_i\Pi_i$. The equilibrium relative resistance that must be met is their ratio $R_{ij} = \tau_{ij}/\Pi_i P_j$.

Equilibrium in the ij market at point E on Figure 1 below is associated with relative price price $p_{ij}/P_j = c_i\tau_{ij}/P_j$ and quantity x_{ij}^e. All quantities on the horizontal axis are normalized by the total supply y_i, hence the points are quantity shares. The quantity demanded is met by residual supply $y_i - \sum_{l\neq j} x_{il}^D$ at seller unit cost c_i.

The quantity shares on the horizontal axis are backed out from trade data as it usually exists: equal to the product of country j's world expenditure share E_j/Y and its observable expenditure shares b_{ij} divided by their relative price p_{ij}/P_j. Due to the normalization of national price vectors, the inward multilateral resistance P_j is scaled by E_j/Y, thus absorbing it. Then

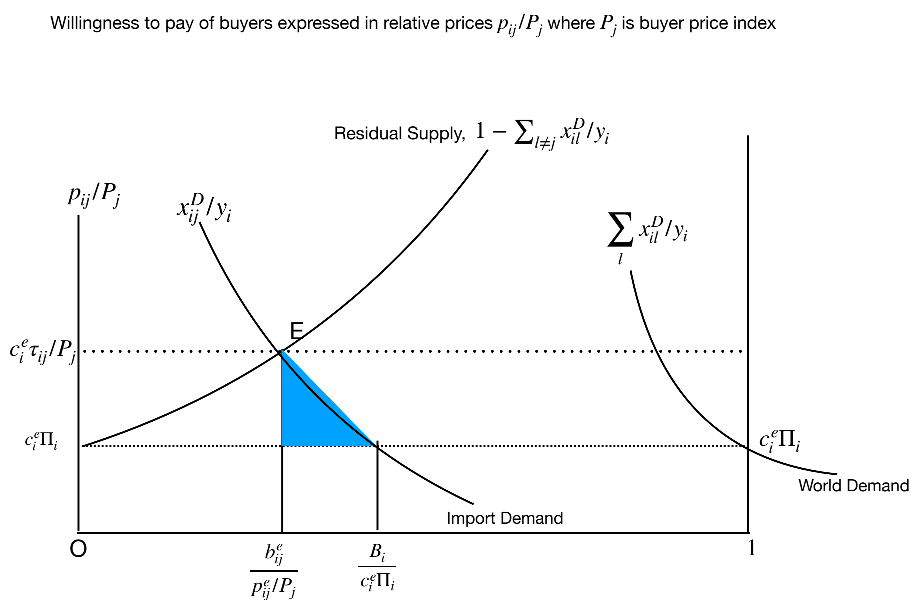
$$x_{ij}^e = \frac{b_{ij}^e}{p_{ij}^e/P_j}.$$

Similarly, $B_i/c_i\Pi_i$ is every destination buyer's quantity share in the as-if-frictionless equilibrium where all buyers spend the same share of income B_i on goods from seller i, $c_i\Pi$ is the price of good i, and the as-if-frictionless price index is equal to 1 from the normalization constraint $\sum_i B_i(\{c_i\Pi_i\}) = 1$. The hypothetical B_i is equal to the observable share of world sales of all goods that are made by seller i.

The generality of this setup deserves emphasis. The buyers price in the hypothetical world market is $c_i\Pi_i$, equal to a real world counterpart, the opportunity cost of sending a unit of good i to any particular destination. In an as-if-frictionless world equilibrium,[9] all buyers would face the same vector of prices $c_i\Pi_i, \ \forall i$ and hence have identical expenditure shares on each good i, B_i. The effects of heterogeneous tastes and price dependent non-homotheticity interacting with income differences is absorbed in the trade frictions τ_{ij}.

The setup allows inference of the counterfactual as-if-frictionless quantity share of imports x^*_{ij}/y_i compared to the observed share with $R_{ij} > 1$.[10]

Figure 1: Bilateral Gains from Trade



The shaded triangle area measures the gain in consumer surplus enjoyed from the lower price of imports from i in market j due to the counterfactual shift to as-if-frictionless equilib-

[9] As-if-frictionless equilibrium follows when trade frictions are counterfactually set to $t^*_{ij} = \tau_i\tau_j, \ \forall i, j$. For example, replace τ_{ij} with $\Pi_i P_j$.

[10] The validity of this setup implicitly depends on the invertibility of the demand system. The very large shifts in shares between the observed and the as-if-frictionless equilibria is uniquely determined, given invertibility.

rium. The measure is technically valid as a compensating variation when the demand curve is real-income-compensated. The shaded area is a non-parametric construct that only approximates the 'true' integral value of the compensating variation. From the nonparametric perspective, the 'true' measure requires unattainable knowledge of the compensated demand function in form and in parameterization.

The loss triangle area is quantified as a positive number by $-(P_j b_{ij} - B_i)$. Divide the height by $c_i\Pi_i$ and multiply the base of the triangle by $c_i\Pi_i$, preserving the same area with height $R_{ij} - 1$ and base $b^e_{ij}/R_{ij} - B_i$. Then the formula for the area of a triangle suggests an approximation formula for relative resistance as the positive root of the quadratic expression below evaluated at $\epsilon_{ij} = 0$:

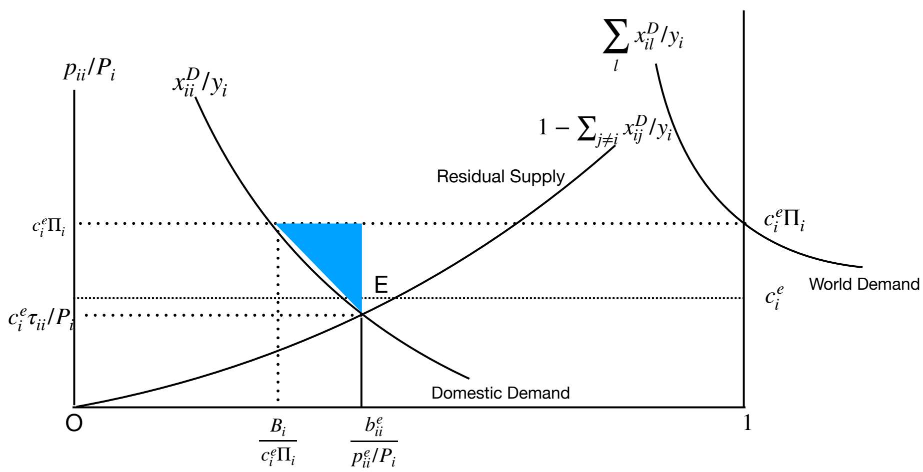
$$-(P_j b_{ij} - B_i) = \frac{1}{2}(R_{ij} - 1)(b^e_{ij}/R_{ij} - B_i) + \epsilon_{ij}.$$

The approximation error ϵ_{ij} is the area between the demand curve and the hypotenuse of the triangle. Because the difference between $B_i = s_i$ and b_{ij} is large, ϵ_{ij} is likely to be large. [2] develops approximation that improves on the triangle method without imposing a functional form on the demand function.

A similar diagram depicts each bilateral market, $\forall i \neq j$. The aggregate demand $\sum_l x^D_{il}$ similarly has each demand function in the sum being a function of the l specific vector of relative prices $\{p_{il}/P_l\}$. The arbitrage equilibrium conditional on given total expenditure or real income in each destination is reached by finding the equilibrium set of $\{\Pi_i, P_j\}$ that is consistent with zero arbitrage profit. Essentially, the aggregate seller incidence Π_i affects buyer incidence P_l in all destinations l, so the bilateral demands and aggregate demands shift about until equilibrium is found. In a particular equilibrium of observed trade flows, any bilateral shipment of good i competes with a 'world market' willingness to pay $c_i\Pi_i$. Payment rise to cover the cost of serving destination j is $c_i\tau_{ij}/P_j$ relative to opportunity cost $c_i\Pi_i$ on the world market, equal to relative resistance $R_{ij} \equiv \tau_{ij}/\Pi_i P_j$. Demand falls relative to its as-if-frictionless value (on the 'world market'). The equilibrium pattern of bilateral trade is determined by relative resistances $R_{ij} \equiv \{\tau_{ij}/\Pi_i P_j\}$.

The next diagram illustrates the effect of frictions on the exchange gains from trade in the domestic trade case where $i = j$. The residual supply to the domestic market i is $y_i - \sum_{j \neq i} x_{ij}$ and is an increasing function of relative price $p_{ii}/P_i = c_i \tau_{ii}/P_i$. The units on the horizontal axis of the diagram are in sales shares x_{ii}/y_i. As before, the general equilibrium of distribution determines the sellers' incidence Π_i from the adding up condition for sales on the rightmost vertical axis, hence the common 'no arbitrage profit' $c_i = p_{ij}/\tau_{ij},\ \forall i, j$. For visual simplicity the projection of c_i^e to the left vertical axis is equal to p_{ii}^e/P_i^e.

Figure 2: Frictions and the Gains from Trade



The shaded triangle approximates the hypothetical loss of exchange gains from trade due to the existing trade frictions deviation from as-if-frictionless trade frictions, incorporating the general equilibrium consequences of *all* trade frictions as they affect country i, directly and indirectly. In as-if-frictionless equilibrium, all buyer locations pay the same relative price for each good i, $c_i^e \Pi_i$. Thus if, counterfactually, the relative buyers price rose from $c_i^e \tau_{ii}/P_i$ to $c_i^e \Pi_i/P_i^*$, the reduced domestic demand releases supply to the external market with willingness-to-pay $c_i\Pi_i$, greater than domestic willingness to pay on all the infra-marginal units. The net effect is an increase in real income represented by the area of the shaded triangle. As in Figure 1, the loss area is a compensating variation, holding real income constant at the observed equilibrium value.

5. Future Directions

Relative resistance statistics represent a new starting point for Tinbergen's strategy of explaining determinants of the residual resistances by regressing them on proxies. The first step in investigating relative resistances is using origin-time and destination-time fixed effects in the regression to identify the outward and inward multilateral resistances. Then various proxies can be used to pull further information from the residuals.

Data limitations present important challenges. Non-parametric gravity requires matched trade and production data because it depends on shares (in expenditure and production alike). It also requires buyer price indexes (for both intermediate and final goods) and distribution costs that convert seller sales to buyer valuation. While these requirements apply to all of the gravity literature, non-parametric gravity introduces dependence on the buyer price indexes.

The WIOD data is a good example of the issues, exploited by [2] using the manufacturing aggregates. Measurement error in the data is a key concern at any level of aggregation, as is aggregation bias. The price index data is a particular concern, both for compromises reached in its original construction and in the normalization applied in order to compare general arbitrage equilibria. The non-parametric statistics for $R_{ij,t}$ have the effects of approximation error as well as measurement error. Future work may make progress on appropriate treatments of at least measurement error and perhaps their combination.

The relative resistance accounting program resembles the productivity accounting program in investigating the determinants of the Solow residual. Also, outward and inward multilateral resistances are seller and buyer incidences of trade frictions, with close parallels to the productivity variables of that program. A key difference is that relative resistances are generated from discrete changes (observed shares difference from as-if-frictionless shares), so volume effects matter a lot. In contrast, the Solow residual is usually generated from local changes using observed shares. Nevertheless, there may be useful parallels to draw on in developing the relative resistance program.

The practical usefulness and simplicity of CES structural gravity argue for its continued use in projection modeling applications where it remains reasonably accurate. Non-parametric gravity nests the CES within it and can provide a check on its (in)accuracy. The relative resistance sufficient statistics can be fitted to CES specifications and enable model selection within nested CES specifications along with best-fit parameter values.

[2] uses generated panel 'data' on relative resistances $R_{ij,t}$ to calculate the minimum distance estimate of a common CES trade elasticity θ. The simple CES structure comes surprisingly close to the data and the estimated elasticity is tightly fitted at slightly greater than 1, much smaller than most of the literature. In contrast to standard econometric practice, the minimum distance method views the objective to be finding the least information losing representation of the gravity model under the false CES specification. The obvious two way causality between generated $R_{ij,t}$ and the determinants of trade flow shares is a feature, not a bug. Is there a way to improve on this practice when values of trade elasticity paramter(s) are needed for projection purposes?

References

[1] **Anderson, James E.**, "A Theoretical Foundation for the Gravity Equation," *American Economic Review*, 1979, *69* (1), 106–116.

[2] ___ , "Gains fron Scale in Costly Trade: A Non-parametric Gravity Accounting," 2023.

[3] ___ **and Eric van Wincoop**, "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 2003, *93* (1), 170–192.

[4] ___ **and Yoto V. Yotov**, "Short Run Gravity," *Journal of International Economics*, 2020.

[5] **Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare**, "New Trade Models, Same Old Gains?," *American Economic Review*, 2012, *102* (1), 94–130.

[6] **Berry, Steven, Amit Gandhi, and Philip Haile**, "Connected Substitutes and Invertibility of Demand," *Econometrica*, 2013, *81* (5), 2087–2111.

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

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

[9] **Tinbergen, Jan**, *Shaping the World Economy: Suggestions for an International Economic Policy*, New York: The Twentieth Century Fund, 1962.