Non-parametric Gravity*

James E. Anderson
Boston College and NBER

November 15, 2021

Abstract

Non-parametric gravity as defined in this paper characterizes operational implications of spatial arbitrage in goods. Intuition that structural gravity is more general than previous parametric forms is validated. Non-parametric sufficient statistics for gains from trade and terms of trade are derived. Terms of trade in manufacturing 2000-2014 reveal China's early improvement of 25% followed by an overall decline of 42% driven by its almost fourfold rise in world share. US manufacturing terms of trade declined 20%. Trade elasticities that best parameterize non-parametric gravity are less than half those commonly used in counterfactual exercises.

*I am grateful to Danial Lashkari and Arthur Lewbel for helpful discussions and comments.

Gravity models in economics and the inferences drawn from them rest on a class of parametric foundations that severely restrict income and substitution effects in demand. In contrast, economic intuition suggests that the logic of gravity holds in a much wider class. Non-parametric gravity as defined in this paper is a large subset of the class of spatial arbitrage equilibrium models with demand and supply structures that are consistent with efficient choice. Arbitrage equilibrium implies market clearing allocations from which no marginal reallocation of economic activity (goods or factors) between origin and destination pairs can yield a positive profit. Under invertibility restrictions on demand and supply structure,[1] the equilibrium distribution pattern of activity in the non-parametric gravity class is characterized by an inverse square law of economic distance. The inverse square law property pleasingly reconnects economic gravity to its physical origin.[2] The complex determination of equilibrium economic distance has previously hidden the simplicity of its inverse square role in characterizing spatial equilibrium.

The attractive force in economic gravity is the arbitrage gains from trade. In non-parametric gravity a country's gains from trade are locally one-to-one with its terms of trade. Its terms of trade are inversely proportional to the square of its equilibrium economic distance to and from the world, the ratio of the geometric mean of its internal distance (to and from domestic locations) to the geometric mean of inward and outward multilateral

[1]The 'connected substitutes' structure of Berry et al. (2013) is sufficient.

[2]Economic gravity was inspired by the metaphor of the physical two body problem of Newton. In physics the force of attraction between two objects centered at points A and B respectively is inversely proportional to the square of the distance between them. The reasoning is that the attraction of the mass at A toward the mass at B declines with the distance from A to B, while the attraction of B toward A declines with the distance from B to A. Physical distance being non-directional, the force of attraction declines with the square of distance between the two points. The inverse square law applies to many other physical phenomena such as radiation.

The economic gravity class of models reduces general interdependence to a set of pairwise equilibrium relationships of each location to a 'world' market. This effectively restores the physical logic. In contrast to physical distance, economic distance as eventually understood [Anderson (2011a)] is endogenous to the spatial equilibrium, is the geometric mean of directionally varying components, and reflects the spatial equilibrium interaction of economic activity flows between many origins and destinations rather than two. Economic gravity is focused on static equilibria, whereas physical gravity is focused on dynamics. The physical N body dynamics for $N > 2$ is described by a system of differential equations in which the inverse square property plays a role, but there is no reduction to a simple set of two body attractions. The dynamic system is generally not integrable. Stationary equilibrium requires very special restrictions.

resistances. In the special case of CES gravity that first provided an economic foundation for gravity, the gains from trade are globally inversely proportional to the square of economic distance. This analogy to physical phenomena is illustrated in Appendix 6.1.

Three sufficient statistics applications of non-parametric gravity are derived. First is a nonparametric sufficient statistic for the arbitrage (exchange) gains from trade that is valid in a wide subset of the general class. The Arkolakis et al. (2012) strategy to infer gains from trade (relative to autarky) from the ratio of internal trade to global sales uses the parametric CES class to yield a simple sufficient statistic based on observables combined with an estimate of the trade elasticity parameter.[3] Their strategy of focusing on the information in the ratio of internal sales to global sales is extended here to generate a non-parametric sufficient statistic for the arbitrage gains from trade in a much wider class of demand structures.

Non-parametric gravity also generates a sufficient statistic for terms of trade changes based on observables only. This is significant because standard terms of trade calculations based on price comparisons are limited in scope and rife with measurement error. (Many categories lack observable prices, while unit values are contaminated by aggregation bias.) From a gravity perspective, another deficiency of price comparison methods is the absence of any accounting for end user costs that surely vary across both users and product classes. The proposed measure remedies these deficiencies. A third practical application is a nonparametric method for estimating economic distance, given further restrictions. This is useful for applications where model parameters are unnecessary.

Illustrative applications of the proposed sufficient statistics are made to China and the US in manufacturing over the period 2000-2014 using the World Input-Output Database. China's near quadrupling of its world manufacturing share 2000-2014 resulted in an overall terms of trade decline of 42%. The initial fall in trade barriers induced an early terms of trade improvement of 25%. China's gains from manufacturing trade as a share of autarky manufacturing income fell from 17.7% in 2000 to 11.8% of the much larger base of 2014. US

[3]For the case of Ricardian supply with labor productivities drawn from a Fréchet distribution, the trade elasticity is the shape parameter of the productivity distribution.

manufacturing terms of trade declined 20%, while gains from trade as a share of autarky real income fell from 27.7% to 22.1%.

A fourth application of non-parametric gravity generates trade elasticity parameters for use in necessarily parametric counterfactual analysis (e.g. calculating the prospective effects of Brexit). The analyst faces two parameter questions, given a parametric specification. First, what is the best parameter value *for quantifying the objective of the analysis.* A typical example is the counterfactual change in the gains from trade to a shift in trade costs such as posed by Brexit. Second, what is an appropriate error band for the counterfactual. The non-parametric model implies time series changes in gains from trade and terms of trade. The parameter value that best fits changes in the non-parametric gains from trade to changes in the terms of trade over the available data provides answers to both questions.

The 'best-fit' estimate of the CES trade elasticity is based on the non-parametric gains from trade and terms of trade changes calculated for all years and countries in the WIOD manufacturing database. The result is a trade elasticity equal to 1.72, substantially below the usual estimates based on best fit to *all bilateral* trade flows. For a widely used example, Simonovska and Waugh (2014) estimate a trade elasticity approximately equal to 4. The downward adjustment implies that gains from trade effects of counterfactual policy changes are more than doubled.

Non-parametric gravity is related to a recent literature extending gravity via non-parametric approaches to more general parametric approximation models of demand and supply structures. The paper is closest in spirit to the Adão et al. (2017) non-parametric approach to reduced form spatial equilibrium exchange of embodied factors. Both papers assume the broad class of 'connected substitutes' demand systems of Berry et al. (2013). In Adão et al. (2017) the role of connected substitutes is to guarantee invertibility of the factor demand system. With multi-factor production models, derived factor demand systems do not generally satisfy invertibility, as the older literature on factor price equalization emphasized. Adão et al. (2017) therefore specialize to production with one inter-sectorally mobile composite factor endowment in each country. The narrower focus in this paper is on spatial equilibrium exchange in a model of sectoral goods markets. The goods outputs are given from static efficient equilibrium in supply. The sectoral focus is consistent with the political economy concerns that drive typical trade policy. The role of connected substitutes in demand in this paper is to justify application of the intermediate value theorem to characterize observed trade relative to 'as-if-frictionless' trade. This yields fully non-parametric gravity.

The paper also complements the Adão et al. (2020) non-parametric approach to modeling heterogeneity of firms productivities in Chaney-Melitz type gravity models. In contrast, this paper abstracts from selection and all other sources of endogenous supply shifts in order to cleanly develop the novel aspect of non-parametric specification of demand.[4] In principle, the method of this paper could be applied to endogenous supply to yield a non-parametric measure of the specialization gains from trade. The extension is only sketched here due to challenges explained in Section 4.

Section 1 develops the non-parametric gravity model that characterizes spatial arbitrage equilibrium in the distribution of given equilibrium supplies to many destinations with demands consistent with the weak axioms of revealed preference. Section 2 derives the non-parametric approach to terms of trade and gains from trade. Section 3 presents the applications to manufacturing trade 2000-2014. Section 3.3 applies the non-parametric gravity model to supply a measure of possible specification error in the necessarily parametric gravity models used in counterfactual exercises. Section 4 embeds the spatial arbitrage module of distribution of given supplies within a class of complete general equilibrium models with endogenous supply and endogenous trade costs. Section 5 concludes.

[4] Goods trade with selection of heterogeneous firms combines necessarily parametric selection structure with demand and supply structure. In Adão et al. (2020), the initial non-parametric probability distribution of productivities is approximated for quantitative evaluation with a flexible functional form.

1 Non-parametric Gravity

The non-parametric gravity model is approached from an initial intuitive graphical representation that suggests the generality of gravity representations of spatial equilibrium. It also provides perspective on the specialization of demand systems that permits non-parametric quantification. Section 1.1 sets out the graphical analysis of spatial equilibrium containing the essential ideas. Section 1.2 provides the formal analysis in class of invertible demand systems. Section 1.3 develops an operational nonparametric gravity approach to bilateral trade modeling. In Section 1.4, CES and Almost Ideal Demand System cases illustrate some parametric and semi-parametric uses.

Begin with the broad definition of the spatial arbitrage model.

Definition A:

(i) Equilibrium spatial arbitrage – at each destination the buyer's full price (including possible unobservable quality evaluation elements) deflated by trade frictions (including unobservable costs or resistance absorbed by the arbitrageur or the seller) is equal to a common net-of-frictions seller cost at each origin.[5]

(ii) Each origin ships an endowment of goods (a variety of a single product class or an aggregate bundle of product classes that differ in composition 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 sales of i to destination i) purchases.

(iv) Expenditures at each destination must be "rational", i.e. obey the weak axioms of revealed preference, and

(v) Trade frictions absorb a constant fraction of shipments (iceberg melting trade costs).

The endowment of seller i is denoted y_i. Shipments from i to j are denoted x_{ij}. Prices received by sellers net of trade costs are denoted p_i. Prices p_{ij} paid by buyers include trade

[5]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.

costs and other frictions $\tau_{ij} \geq 1$. Frictions include unobservable user costs and heterogeneity in preferences across destinations. For simplicity, start by thinking of τ_{ij} as fixed, though subsequently it is endogenous.

In the arbitrage equilibrium $p_{ij}/\tau_{ij} = p_i, \ \forall i, j$. This condition is necessary and sufficient for zero arbitrage profits. Assumption (ii) takes supplies as given. The value of goods purchased at end user valuations (including any unobservable user costs) is $X_{ij} = p_{ij}x_{ij}$.

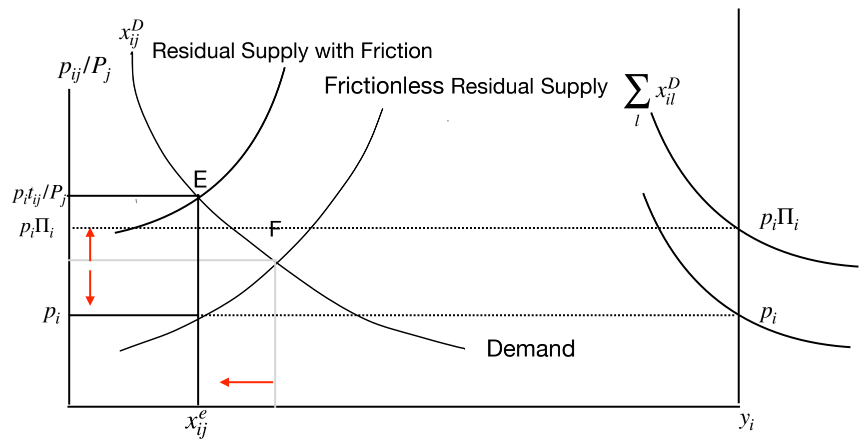
1.1 Graphics

The graphical analysis uses a goalpost diagram (familiar from the specific factors model). Start with a system of generic demand schedules to characterize 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 generic demand schedule for goods from i in j is labeled x_{ij}^D, downward sloping for standard reasons. 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, a hypothetical frictionless residual supply schedule is also drawn.

The worldwide aggregate demand for goods from i (defined under conditions specified below) is downward sloping and intersects the supply schedule y_i at price $p_i\Pi_i$, the price paid by a hypothetical buyer in the 'world' market. Π_i is the sellers' incidence of trade frictions on world sales. Intuitively, it is an index of the bilateral trade frictions faced by shipments from i to all destinations j including internal shipments to destination i. The index reflects the efficient spatial arbitrage $p_{ij} = p_i t_{ij}, \ \forall j$. The formal analysis in Section 1.2 derives the equilibrium incidences $\{\Pi_i\}$.

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 such that we may regard the left vertical axis in the figure below as measuring relative prices $p_{ij}/P_j = p_i t_{ij}/P_j$ in arbitrage equilibrium. Thus the equilibrium shipment x_{ij}^e at the equilibrium point E is associated with relative price p_{ij}/P_j. Comparing price indexes across destinations, assume initially that price indexes differ only because the trade frictions differ by destination, $t_{ij} \neq t_{il}$. Then the price index P_j varies across destinations j due to trade cost variation only. Then P_j is interpreted as buyer j's incidence of the set of bilateral trade costs t_{ij}, $\forall i$.

Equilibrium in the ij market at point E on the goalpost diagram below is associated with relative price price $p_{ij}/P_j = p_i t_{ij}/P_j$ and quantity x_{ij}^e. The quantity demanded is met by shipments $y_i - \sum_{l \neq j} x_{il}$ at price p_i projected from the right vertical axis. The buyers price in the hypothetical world market is $p_i \Pi_i$. The sellers' incidence Π_i on sales to the world deflates world buyers price $p_i \Pi_i$, yielding the net sellers price p_i. The projection of p_i from the right vertical axis to the quantity x_{ij}^e implies arbitrage equilibrium with Π_i being seller incidence. The remainder of bilateral trade cost $= t_{ij}/\Pi_i$ is buyer j's bilateral incidence on purchases from i. The goalpost diagram plots an as-if-frictionless residual supply schedule, with the essential implication that $x_{ij}^e < x_{ij}^F$ at the hypothetical as-if-frictionless shipment point. In contrast to the partial equilibrium analysis of incidence, the price associated with frictionless point F plays no role in the division of incidence. Because sales must add up to the total shipped, the implication is that $x_{ii}^e = y_i - \sum_{j \neq i} x_{ij}^e >> x_{ii}^F = y_i - \sum_{j \neq i} x_{ij}^F$. The effect of adjustment to frictions implies interdependence of the shifts in supply to markets as Π_i affects residual supply to all other markets. It also implies further shifts in demands as the price vector shifts demand schedules about. But the analysis suggests that all markets together determine the equilibrium incidences Π_i and the buyer incidences.



The aggregate demand $\sum_l x_{il}^D$ 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, it is convenient to choose units such that world prices $p_i\Pi_i = 1, \ \forall i$. Then $p_i = 1/\Pi_i$ and the equilibrium pattern of bilateral trade is determined by $\{t_{ij}/\Pi_i P_j\}$.

Trade varies negatively with $t_{ij}/\Pi_i P_j$. The denominator is a product. A natural mean of a product is its square root, the geometric mean. Define the economic distance between i and j as

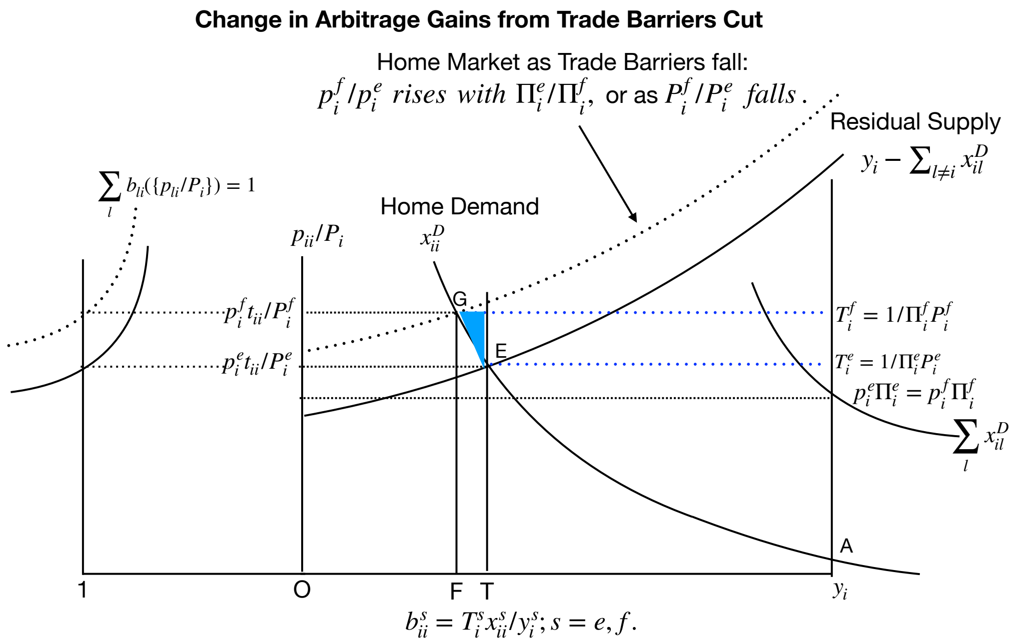
$$D_{ij} \equiv \sqrt{t_{ij}/\Pi_i P_j} \tag{1}$$

In the economic gravity context the equal exponents of the geometric mean reflect the equal forces of sales from i seeking higher net price and of purchases from j seeking lower price. Thus bilateral trade varies inversely to the square of economic distance. The set of inverse squares of economic distance $\{\sqrt{\Pi_i P_j / t_{ij}}\}$ determines the pattern of trade.

The full general equilibrium model that nests the gravity model of distribution requires links between expenditure and income in each location. The set of links (closures of the model) simultaneously with the gravity model determines equilibrium $\{p_i\}$ along with $\{\Pi_i, P_i\}$ up to a normalization. The standard normalization of prices is $\sum_i p_i y_i / \sum_i y_i = 1$.

Graphically, each of the N origin products has a market clearing condition represented by the intersection of aggregate demand with the right vertical axis supply schedule in the diagram. The equilibrium set of multilateral resistances and seller prices is efficient, with the no arbitrage profit efficiency condition represented by the intersection of bilateral demand and residual supply at $p_{ij} = p_i t_{ij}$.

The graphical model also illustrates the exchange gains from trade. Country i has supply of its product y_i, purchased by the rest of the world in amount $\sum_{j \neq i} x_{ij}$. These export sales in equilibrium equal $y_i - x_{ii}$, supply less domestic sales. The export supply schedule of country i is given by the upward sloping schedule $y_i - x_{ii}$, an increasing function of relative price $p_{ii}/P_i = p_i t_{ii}/P_i$. 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' $p_i = p_{ij}/t_{ij},\ \forall i, j$. The diagram shows initial export of Ty_i with domestic trade OT. The leftmost vertical axis depicts the adding up condition on purchases expressed as the sum of expenditure shares. This determines the buyers' incidence P_i.



Now consider a trade cost reduction due either to country i lowering its 'trade barriers' to imports or its partners lowering barriers to its exports.[6] The direct effect of reducing some t_{ij}s is that P_i must fall with i's import barriers falling (downward shift on the leftmost vertical axis) while Π_i must fall with a decline in barriers to its exports (downward shift on the rightmost vertical axis). The displaced domestic sales (x_{ii}) face external trade frictions so there are further endogenous changes in multilateral resistances. The net result in either case is a shift up and to the left in the export supply schedule along the home demand for i's goods. The analysis also reveals a qualitative symmetry between country i's export restrictions (affecting Π_i) and its import restrictions (affecting P_i). Lerner symmetry thus extends qualitatively to sectoral trade without trade balance.

The domestic share of sales shrinks from OT/Oy_i to OF/Oy_i. The new equilibrium relative price for buyers in the domestic market becomes $p_i^f t_{ii}/P_i^f$. Using the simplification $p_i^e \Pi_i^e = p_i^f \Pi_i^f$,[7] country i's terms of trade improve from $T_i^e = 1/\Pi_i^e P_i^e$ to $T_i^f = 1/\Pi_i^f P_i^f$.

[6]The simplest case is a trade friction cut with no revenue consequences, such as reasonably may apply to regulatory barriers.

[7]This is consistent with the partial equilibrium intuition of the diagram, and simplifies the more general

The net change in the gains from trade is non-parametrically negatively related to the domestic trade share

$$b_{ii} = \frac{p_i t_{ii}}{P_i} \frac{x_{ii}}{y_i}.$$

Graphically, $b^e_{ii} = T^e_i \times OT/Oy_i$ while $b^f_{ii} = T^f_i \times OF/Oy_i$. Assume for simplicity that the trade frictions removal accrues to sellers with no losses of revenue. Normalize by $Oy_i = 1$. The non-parametric measure of the gains from trade change is approximately $\Delta(p_i/P_i)(x^e_{ii} + x^f_{ii})/2$. This is represented by the area to the right of home demand schedule x^D_{ii} between the initial terms of trade T^e_i and the new higher terms of trade T^f_i, equal to the income rise $y_i(T^f_i - T^e_i)$ minus the loss of consumer surplus on domestic sales, the area to the left of home demand schedule x^D_{ii} between T^e_i and T^f_i. In application it is convenient to report the change as a percentage $\Delta\widetilde{T}_i\tilde{b}_{ii}$ where the tilde denotes evaluation at intermediate values for the domestic share b_{ii} and the percent change in T_i. Allowing for revenue/rent offsets to the 'free' box $(T^f_i - T^e_i) \times Ty_i$ may reduce the gain as far as the shaded triangle $\Delta(p_i/P_i)\Delta x_{ii}/2$. (The gains from trade relative to autarky is given by the area to the right of x^D_{ii} between T^e_i or T^f_i to point A, relative to the area of the rectangle formed by the product of Oy_i and relative price given by vertical line y_iA.)

The graphical treatment suggests a focus on the *ex post* change in gains from trade caused by movement of the terms of trade p_i/P_i and its association with how much trade is gained or lost. Any change shifts the equilibrium in each global market i, with changes in the terms of trade for all p_i/P_i and exchange gains from trade approximated non-parametrically by $\Delta(p_i/P_i)(x^f_{ii} + x^e_{ii})/2$. Conversion of the non-parametric gains measure to standard real income measures requires a fully parametric approach, subject to the burden of the accompanying assumptions. The parametric CES demand system resolves the issue of relating consumer surplus to real income of a representative agent.[8] The result is an exact measure of the relative change, shown by Arkolakis et al. (2012) to be a decreasing power function

conclusion that p_i is negatively related to Π_i.

[8] Treatment of heterogeneous agents and income distribution requires far more assumed structure, including an aggregate welfare function.

of b^e_{ii}/b^f_{ii} as discussed in Section 1.4. (The CES gains measure (see Section 3) falls inversely to the square of economic distance. Appendix 6.1 completes a graphical reconnection of the CES gravity model to its original physical inspiration.)

In principle, the same graphical analysis applies to exchange in derived factor demand systems where trade is in embodied factors, as in Adão et al. (2017). A special case is the one mobile factor (Ricardo or Ricardo-Viner) model of production. The diagrams above are reinterpreted by relabeling the quantity axis as embodied labor, the sellers' wage w_i as the sellers' price p_i while the buyers' price is replaced by $w_i t_{ij}/\Omega_j$ where Ω_j is a price index for the vector of embodied labor purchased by j. Arkolakis et al. (2012) show that the CES parametric expression for gains from trade (now including the specialization gains) applies to trade in embodied labor flows in the Eaton and Kortum (2002) Ricardian model of gravity. In this case the active labor productivities are generated as random draws from a Fréchet distribution. Adão et al. (2017) extend the interpretation of the single factor as a composite of multiple factors in their non-parametric setup.

The general non-parametric model yields non-parametric inference of changes in the terms of trade and the accompanying changes in gains from trade. A time series of such non-parametric terms of trade changes can usefully inform the parameterization that is required for *ex ante* counterfactual projection of the effects of changes in trade frictions.[9] These suggestions are applied based on the formal model that follows.

1.2 Formal Model

Spatial equilibrium in the non-parametric gravity model implies that the distribution of goods (the pattern of the b_{ij}s) is determined by the set of inverse squares of economic distances $\{D_{ij} \equiv \sqrt{\tau_{ij}/\Pi_i P_j}\}$ where τ_{ij} aggregates over h the underlying bilateral resistances $\{t_{ij}\beta^h_{ij}\}$ with β^h_{ij} indicating a household h specific quality shifter. D_{ii}, the case of internal

[9] Counterfactuals require parametric specification of the high dimensional mechanism that determines the endogenous movement of price indexes $\{P_i\}$ due to endogenous rises in the sellers prices $\{p_i\}$ and the simultaneous interaction of $\{P_i\}$ with $\{\Pi_i\}$.

trade, is a direct measure of the elusive "open-ness to trade" concept that is comparable across countries in the cross section and across non-nested modeling choices. A key implication is that region i's terms of trade are reduced relative to its as-if-frictionless value by D_{ii}.

Notice that *relative* frictions $\{\tau_{ij}/\sqrt{\tau_{ii}\tau_{jj}}\}$ are what determines the cross section pattern of trade:

$$\frac{\tau_{ij}}{\Pi_i P_j} = \frac{\tau_{ij}/\sqrt{\tau_{ii}\tau_{jj}}}{(\Pi_i/\sqrt{\tau_{ii}})(P_j/\sqrt{\tau_{jj}})}.$$

For the present purpose of characterizing spatial arbitrage, it is harmless to simplify notation with the relative trade friction form: henceforth $\tau_{ij} = t_{ij}\beta_{ij}/\sqrt{t_{ii}t_{jj}\beta_{ii}\beta_{jj}}$. The internal frictions are absorbed in the multilateral resistances.[10] With this simplification, $D_{ii} = \sqrt{\Pi_i P_i}$ understanding that the multilateral resistances are scaled by $1/\sqrt{\tau_{ii}}$. An important implication of this discussion is resolution of a spatial unit puzzle. Gravity applies to spatial arbitrage between units of any chosen size (countries, regions, commuting zones, ...). The natural asymmetries of directional distance are geometrically averaged in internal distances $\sqrt{\tau_{ii}}$ for the chosen unit size i, without consequence for characterizing spatial arbitrage between the units of the chosen sizes. Relatedly, small unit sizes are associated with smaller τ_{ii}, hence larger D_{ii}, contributing to a regularity observed in CES gravity model applications.

Assumption (iv) (buyer choices obey the weak axioms of revealed preference) implies that the expenditure of agent h in destination j is characterized by the value (expenditure or cost) function $e_j^h(\{p_{ij}\}, u_j^h)$, which is concave and homogeneous of degree one in the price vector and increasing in utility u_j^h. The effects of cross-section variation in agent h behavior on expenditure function $e_j^h(\cdot)$ are restricted here to define the general gravity class of equilibrium arbitrage models that satisfy

Definition G

1. *assumptions (i)-(v) hold along with*

[10] In applications to panel data where policy changes affect the ratio of internal to cross-border trade, the separate variation of internal and cross border frictions requires explicit treatment.

2. *assumption (vi): the demand system has the connected substitutes property of Berry et al. (2013).*

Let $\mathbf{p}^j$ denote the vector of p_{ij}s at destination j. Assumptions (vi)-(vii) imply that the common expenditure function $e(\mathbf{p}^j, u^{hj})$ is equal to $e(\mathbf{p}^{hj})u^{hj}$ for an appropriately chosen price vector $\mathbf{p}^{hj}$. Thus origin-destination-agent specific shifters $\beta_{ij}^h = p_{ij}^h / p_{ij},\ \forall i, j, h$ apply to each agent h in location j. The 'connected substitutes' assumption includes allowing tastes to differ by destination-agent varying quality shifters.The taste shifters incorporate the effect of price-dependent non-homotheticity *at the equilibrium.* Assumption (vi) and Shephard's Lemma imply a unique (up to a scalar) solution to β_{ij}^h from $e_i(\mathbf{p}^{hj}, u^{hj}) = e_i(\mathbf{p}^{hj})u^{hj}$. Thus a 'full price' in arbitrage equilibrium is $p_{ij}^h = p_i t_{ij} \beta_{ij}^h$, where β_{ij}^h is the inferred taste shifter that explains buyer choice when facing seller price p_i combined with trade cost t_{ij} in arbitrage equilibrium. All heterogeneity and non-homotheticity is hidden in the β_{ij}^h variables – origin-destination-agent fixed effects in the econometric sense of reduced form exogenous controls. Interpreting the properties of the equilibrium does not require unpacking the rich endogeneity concealed in the bilateral resistances. General gravity describes the spatial arbitrage equilibrium for given equilibrium values of the bilateral resistances $\tau_{ij}^h = t_{ij}\beta_{ij}^h$ and thus given real incomes u_j^h.

It is important in the gravity context to note that connected substitutes does not require that all goods be positively demanded at all destinations. [See Berry et al. (2013).] Zeros are prevalent in bilateral trade data. The occurrence of a zero in a destination is associated with a choke (reservation) price with an equilibrium value that influences all the positive demand shares. The vector of choke prices at any destination is effectively solved from the sub-system equating demand with the zero delivered supply and then carried into all the positively demanded goods.

It is convenient in what follows to aggregate the household shifters β_{ij}^h by defining τ_{ij}

implicitly from

$$e(\{p_i\tau_{ij}\}) = e(\mathbf{p}^j) = \sum_h^{H_j} e(\{p_i\tau_{ij}^h\})u^{hj} / \sum_h u^{hj},$$

where H_j is the number of agents (households) in j. The explicit form is $\tau_{ij} = \sum_h X_{ij}^h \tau_{ij}^h / \sum_h X_{ij}^h$, from applying Shephard's Lemma with X_{ij}^h denoting expenditure in j by household h on goods from i. Thus $\mathbf{p}^j = \{p_i\tau_{ij}\}$. The aggregation is over many possible dimensions of heterogeneity. The well-known stability of estimated bilateral frictions in the empirical gravity literature suggests that the distribution of heterogeneous effects is reasonably stable.

The aggregate expenditure in the world under assumption (iv) combined with assumption (vi) implies that

$$E = \sum_j \sum_{h=1}^{H_j} e(\{p_i\tau_{ij}\})u^{hj} = \sum_j e(\{p_i\tau_{ij}\})u^j. \tag{2}$$

World equilibrium requires that expenditures add up to sales at end user valuation, hence $E - Y = 0$. It also requires world market clearance for each country's product. In a frictionless world equilibrium with the same demand structure and the same vector of supplies (endowments) $\{y_i\}$, the adding up conditions imply $E^* - Y^* = 0$. The price vectors for sellers differ. The standard normalization implies that relative prices are constrained such that $\sum_i (p_i - p_i^*)y_i = 0 = Y - Y^*$.

The normalization removes the common global gain in efficiency to isolate the relative effect on terms of trade and real incomes. The gravity model has no way to measure this absolute efficiency; it only reveals relative effects. The 'as if frictionless' pattern of trade is consistent with $\tau_{ij} = \tau_i\tau_j, \forall i, j$ for any values of $\tau_i, \tau_j \geq 1$.

Combine the adding up conditions for the actual and frictionless equilibria in

$$E - E^* - (Y - Y^*) = 0.$$

The properties of the expenditure function applied to this equilibrium condition yield relationships between the bilateral demands for each origin product in each destination relative to the same bilateral demand in frictionless equilibrium. This gives the general characterization of spatial arbitrage equilibrium.

Shephard's Lemma $[\partial e(\{p_i\tau_{ij}\})/\partial p_{ij} = x_{ij} = e_i(\cdot)u^j]$ gives demand for good i by the aggregate agent in j, and implies agent j's share of expenditure on good i, $b_{ij} = e_i(\{p_i\tau_{ij}\})p_{ij}/e(\cdot) = e_i(\{p_i\tau_{ij}/P_j\})p_i/P_j$. The last equation uses the true cost of living index $P_j = e(\{p_i\tau_{ij}\})$ and the homogeneity of degree zero of the demand system.

The world's expenditure share on good i, B_i, is in equilibrium a weighted average of the national agent shares b_{ij}, $B_i = \sum_j b_{ij}E_j/Y$. The fictitious world buyer faces 'price' vector $\mathbf{p}^*$ such that $e(\mathbf{p}^*)\sum_j\sum_{h=1}^{H_j} u^{hj} = E^* = Y^* = Y = E$. Under the assumption that the demand structure has the connected substitutes property, $\mathbf{p}^*$ is unique. Apply Shephard's Lemma to the aggregate expenditure E^* thus defined to solve for the common 'world market' price vector $\{p_i\Pi_i\} = \mathbf{p}^*$ that satisfies the market clearing conditions

$$\frac{Y_i}{Y} = B_i(\{p_i\Pi_i\}),\ \forall i. \tag{3}$$

Since $\sum_i Y_i/Y = 1$, system (3) solves for relative prices only. The adding up condition implies a normalization on prices such that the 'world price index' $e(\mathbf{p}^*) = 1$.

Use the previous choice of units such that the 'world price' vector in frictionless equilibrium is $p_i\Pi_i = 1,\ \forall i$. Then the buyer's price index for country j is $P_j = e(\{p_i\tau_{ij}\}) = e(\{\tau_{ij}/\Pi_i\})$. Thus P_j is the index of bilateral buyer's incidences τ_{ij}/Π_i, hence is interpreted as the buyers overall incidence of trade frictions with the world market. Π_i analogously is the sellers incidence of trade frictions with the world market. Take $e(\{\tau_{ij}/\Pi_i\}) = P_j$ into the arguments of b_{ij} using homogeneity of degree zero of e_i in the price vector, hence $b_{ij} = e_i(\{\tau_{ij}/\Pi_iP_j\})\tau_{ij}/\Pi_iP_j$. Thus the set of bilateral economic distances $D_{ij} = \sqrt{\tau_{ij}/\Pi_iP_j}$ determines the pattern of bilateral exchange. The next step reveals the inverse square law relationship by analyzing the difference between the equilibrium aggregate bilateral shares b_{ij} and the 'as-if-frictionless' shares B_i.

Return to the world adding up condition that combines the actual and frictionless equilibria $(E - E^*) - (Y - Y^*) = 0 = \sum_j e(\mathbf{p}^j)u^j - e(\mathbf{p}^*)\sum_j u^j - \sum_i (p_i - p_i^*)y_i$. Differentiate with respect to each p_i, apply Shephard's Lemma and multiply by p_i in the first two terms and sum over i and j while holding real incomes u^j constant. The result is:

$$\sum_{i,j} e_i(\mathbf{p}^j)p_{ij}u^j - \sum_i e_i(\mathbf{p}^*)p_i\Pi_i \sum_j u^j - (\sum_i Y_i - \sum_i Y_i^*) = 0. \tag{4}$$

Each component j of the double sum above simplifies using $b_{ij} = e_i(\mathbf{p}^j)p_{ij}/e(\cdot)$, $u^j = E_j/P_j$ and balanced trade, $E_j = Y_j$. Divide the jth component of (4) by u^j. The result is

$$\sum_i [b_{ij}P_j - B_i] - (P_j - Y_j^*/u^j), \ \forall j;$$

Note that $Y_j^* = u^{*j} = p_j^* y_j$ since $P_j^* = e(\mathbf{p}^*) = 1$, while $u^j = Y_j/P_j = p_j^* y_j/\Pi_j P_j$. Thus $u^{*j}/u^j = \Pi_j P_j$ and the percentage gain in real income for country j is given by

$$\sum_i [b_{ij}P_j - B_i] - (P_j - Y_j^*/u^j) = \Pi_j P_j - 1 = G_j, \ \forall j.$$

G_j is due to the terms of trade effect of the move to uniform frictions (observationally equivalent to frictionless trade). The move may create both winners and losers via the terms of trade gains and losses implied. But global average welfare ordinarily rises. Multiplying G_j by u^j and summing over j implies that

$$\sum_j \Pi_j P_j \frac{u^j}{\sum_j u^j} = \frac{\sum_j u^{*j}}{\sum_j u^j} > 1.$$

The inequality is established because (i) $\Pi_i > 1$ by $\tau_{ij} > 1, \ \forall i \neq j$. Second (ii), the average buyers price is preserved in the move to uniform frictions, $\sum_j P_j u^j / \sum_j u^j = e(\mathbf{p}^*) = 1$.[11] Third, $\sum_j \Pi_j P_j u^j / \sum_j u^j = \mathcal{E}(\Pi) \cdot 1 + Cov(\Pi, P)$ using the algebra of covariance.

[11]This follows from rearranging (4) and using the normalization $\sum_i (Y_i - Y_i^*) = 0$. This implies that $e(\mathbf{p}) = \sum_j e(\mathbf{p}^j)u^j / \sum_j u^j = \sum_j P_j u^j / \sum_j u^j$.

The covariance term is positive for plausible trade frictions structures. For example in the symmetric case $\Pi_j = P_j, \forall j$, the covariance term is a variance. For sufficient asymmetry the covariance could fall below zero but to reverse the inequality requires truly pathological asymmetry. Estimated multilateral resistances and bilateral frictions have no such properties.

Note that $\sqrt{\Pi_j P_j} = 1/D_{jj}$ (using the convention that the multilateral resistances Π_j, P_j are scaled by $1/\sqrt{\tau_{jj}}$). Then the global adding up property implies a normalization on the inverse squares of economic distance,

$$\sqrt{\sum_j D_{jj}^{-2} \frac{u^j}{\sum_j u^j}} = 1.$$

The "frictionless" trade pattern is equivalent to that associated with any uniform equivalent set of frictions denoted s such that $\tilde{\tau}_{ij}^s = D_{ii} d_i^s D_{jj} d_j^s,\ \forall i, j$, and $d_i^s > 0\ \forall i$. The normalization of economic distance resolves the indeterminacy.

1.3 Non-parametric Bilateral Trade Sufficient Statistics

From the point of view of potential non-parametric inference of gravity structure, economic distance consists of pair fixed effects τ_{ij} combined with origin and destination fixed effects $\Pi_i,\ P_j$. The first step toward sufficient statistics applies the intermediate value theorem to the expenditure function evaluated at each location's buyers prices differenced from the as-if-frictionless expenditure function. Non-parametric gravity is based on the elements of the resulting sum. Operational non-parametric sufficient statistics for gains from trade and terms of trade are approximations, exact under further restrictions. Section 3 reports results based on WIOD manufacturing data. More stringent restrictions lead to non-parametric inference of bilateral economic distance in Section 3.2.

The intermediate value theorem applied to the expenditure function for each country j [12]

[12] The theorem holds for price vectors in a connected set, a condition satisfied by the expenditure function under the connected substitutes restriction of Berry et al. (2013).

assures that for some intermediate price vector $\tilde{p}_{ij} = \lambda_j p_i \tau_{ij} + (1 - \lambda_j) p_i \Pi_i; \lambda_j \in [0, 1], \ \forall i, j$:

$$e(\{p_i \tau_{ij}\}) u^j - e(\mathbf{p}^*) u^j = \sum_i b_{ij}(\{\tilde{p}_{ij}\}) \frac{\tau_{ij}/\Pi_i - 1}{(\lambda_j \tau_{ij}/\Pi_i + (1 - \lambda_j))} u^j \tilde{P}_j, \ \forall j. \tag{5}$$

The price index $\tilde{P}_j = e(\{\tilde{p}_{ij}\})$, while world price $\Pi_i p_i$ is divided into both numerator and denominator to form the ratio on the right hand side of equation (5).

The economic implications of (5) are revealed by eliminating u^j from both sides of the equation. $u^j \tilde{P}_j$ is equal to$\tilde{E}_j$, the expenditure required to support u^j facing price vector $\tilde{\mathbf{p}}^j$. Divide both sides of the equation by u_j and use $e(\mathbf{p}^*) = 1$ to give the percentage change in expenditure needed to support u^j relative to the frictionless equilibrium, $P_j - 1$. This is a measure of the loss due to frictions. Thus the percentage cost of trade frictions to country j implied by (5) is:

$$P_j - 1 = \sum_i b_{ij}(\{\tilde{p}_{ij}\}) \frac{(\tau_{ij}/\Pi_i - 1)}{[\lambda_j \tau_{ij}/\Pi_i + (1 - \lambda_j)]/\tilde{P}_j}, \ \forall j. \tag{6}$$

Shephard's Lemma also implies an equivalent expression in terms of observables:

$$P_j - 1 = \sum_i [b_{ij} P_j - B_i].$$

The equilibrium condition $B_i = Y_i/Y$ implies that B_i is observable.

Non-parametric gravity is derived using the elements of the sum on the right hand side of (6). The individual elements are equal to the observable expressions $b_{ij} P_j - Y_i/Y$, up to a non-parametric error term ϵ_{ij} that represents theoretically possible but unknowable deviations of the individual non-parametric gravity elements from their observable counterparts. ϵ is discussed further below. With this setup, non-parametric gravity is characterized by:

$$b_{ij}(\{p_{ij}\}) P_j - Y_i/Y = \tilde{b}_{ij} \tilde{P}_j \frac{(\tau_{ij}/\Pi_i - 1)}{[\lambda_j \tau_{ij}/\Pi_i + (1 - \lambda_j)]} + \epsilon_{ij}, \ \forall i, j. \tag{7}$$

The first term on the right hand side of (7) is the non-parametric predictor of the observable left hand side $b_{ij}P_j - Y_i/Y$. $\{\tau_{ij}\}$ is a set of bilateral fixed effects while $\{\Pi_i, P_j\}$ is a set of origin and destination fixed effects. $\tilde{b}_{ij} = b(\{\tilde{p}_{ij}\})$ where $\tilde{p}_{ij} = \lambda_j p_{ij} + 1 - \lambda_j$ using the as-if-frictionless price $p_j^* = 1 \, \forall j$. Under the 'connected substitutes' assumption for the expenditure function, $\tilde{b}_{ij}$ is a share intermediate between actual b_{ij} and frictionless $B_i = Y_i/Y$, capturing the general equilibrium effect of frictions 'on average' in shifting b_{ij} away from Y_i/Y. The ratio term on the right hand side of (7) isolates an 'own effect'. The numerator is interpreted as $(p_i\tau_{ij} - p_i\Pi_i)/p_i\Pi_i$, the percentage increase in the buyers price of good i, implicitly also relative to the as-if-frictionless price index $e(\mathbf{p}^*) = 1$. The denominator of the ratio is the intermediate value of the price of good i in destination j deflated by the world price, $p_i\tau_{ij}/p_i\Pi_i$ relative to the intermediate value of the price index $\tilde{P}_j$. (Move $\tilde{P}_j$ in (7) to the denominator of the denominator.) The ratio in this form is thus interpreted as an appropriate discrete form of the percentage change in p_{ij}/P_j, the relative price of good i in destination j implied by hypothetically moving from the observed situation to the as-if-frictionless equilibrium.

Non-parametric gravity expression in (7) uses all the non-parametric information implied by spatial arbitrage. Further restrictions yield a non-parametric class for which $\epsilon_{ij} = 0$. The non-parametric error term ϵ_{ij} represents possible deviations of non-parametric gravity from the observed data on the left hand side of (7).[13] Nevertheless, non-parametric gravity is correct 'on average' because it derives from an equilibrium model with adding up constraints. $\sum_i \epsilon_{ij} = 0$ by the budget constraint for each buyer and the application of the intermediate value theorem in (6). Similarly, the market clearing condition (3) implies that $\sum_j \epsilon_{ij} E_j/Y = 0$. Thus the first term on the right hand side of (7) is interpreted as correct 'on average', while ϵ_{ij} may be non-random. Parametric cases are sufficient to eliminate ϵ_{ij}, given that the parametric specification and the set of parameter values is 'true'.

(7) is qualitatively useful as a decomposition, but it is not operational because λ_j depends on the deep structure of equilibrium. The Törnqvist approximation $\bar{b}_{ij} \to \tilde{b}_{ij}$ sets $\lambda_j = 1/2$ to

[13] Note that $\tilde{b}_{ij}$ as constructed is not generally equal to the intermediate value of $b_{ij} - B_i$. Thus ϵ_{ij} reflects this potential difference.

achieve operationality. An approximation error η_{ij} is added to the non-parametric error ϵ_{ij} in this case. The translog demand system structure is a wide subset of non-parametric gravity models for which the Törnqvist approximation to (7) exactly reveals all the non-parametric information. Disregarding measurement and other random error sources, $\eta_{ij} = 0, \ \forall i, j$. If the translog is the 'true' model, then $\epsilon_{ij} = 0$ as well. Section 3 exploits this property to obtain non-parametric sufficient statistics for gains from trade and terms of trade. Section 3.3 uses the setup to propose a specification error measure for the parametric specifications needed to do counterfactuals.

1.4 Relation to Parametric Gravity

The standard parametric CES case implies that the equilibrium expenditure share is given by $b_{ij} = (p_i\tau_{ij}/P_j)^{-\theta}, \theta > 0$. The spatial equilibrium distribution is given by the closed form CES gravity expression

$$b_{ij} = \frac{Y_i}{Y}(\tau_{ij}/\Pi_i P_j)^{-\theta}. \tag{8}$$

The relationship of (8) to (7) is given by taking logs of (8) and rearranging:

$$\ln b_{ij} - \ln(Y_i/Y) = -\theta[\ln(\tau_{ij}/\Pi_i P_j) - \ln 1],$$

a parametric relationship in logs. The closed form (8) allows straightforward gravity estimation using bilateral friction proxies such as distance along with origin and destination fixed effects for each country. The qualitative difference of (8) from (7) is that cross effects play no role in determining equilibrium CES b_{ij} while cross effects matter in determining non-parametric $\tilde{b}_{ij}$, λ_j and $\tilde{P}_j$.[14]

For various parametric special cases that restrict the demand system in (7) but generalize

[14] Discrete changes in the CES case permit the comparative static hat algebra of Dekle et al. (2007), a property not generally available under Definition G. The gains from trade in the general case are monotonically decreasing in b_{ii} (because arbitrage equilibrium implies efficiency gains) but are not generally a power function as in the CES case. Moreover, the gains from trade are now a real-income-compensated concept unless preferences are homothetic.

from the CES case, b_{ij} has a closed form. In the translog case $b_{ij} - Y_i/Y$ is a translog form in the log $\tau_{ij}/\Pi_i P_j$ arguments. Non-parametric gravity (7) cannot generally determine the values of λ_j, but in the translog case the approximation $\lambda_j = 1/2$ is exact. Thus non-parametric gravity exactly contains the entire class of translog demand systems, obviating the need to infer $N \times (N-1)/2$ substitution parameters.

Assumption (vi) excludes any income effects in demand systems that are not associated with price. In contrast the PIGL class such as AIDS treats non-homotheticity with a price-independent income effect. Allowing for price independent effects of non-homotheticity in the AIDS case, a remainder term on the right hand side of (12) is equal to the change in internal shares due to the effect on b_{jj} of the change in real income from *all* sources between equilibrium 0 and equilibrium 1. With AIDS gravity estimates of the real income effects of income changes on b_{jj}, denoted ΔZ_{jj}, the change in gains from trade is given by $\Delta(b_{jj} - Y_j/Y) - \Delta Z_{jj}$. Anderson and Zhang (2019) have derived a closed form Almost Ideal Demand System (AIDS) gravity model, a log-quadratic approximation to the expenditure function $e(\cdot)$ used above. $b_{ij} - Y_i/Y$ is a linear closed form in $\{\ln(\tau_{ij}/\Pi_i P_j)\}$. (AIDS is a specialization of the PIGLOG class.) Zhang (2020) uses the AIDS gravity to model the role of taste heterogeneity β^h_{ij} proxied by immigrant proportions h of resident populations.

2 Gains from Trade and Terms of Trade

The well known gains from trade sufficient statistic approach of Arkolakis et al. (2012) is extended in non-parametric gravity to a much wider class of demand systems. Arkolakis et al. (2012) relate the gains from trade relative to autarky to the inverse of the internal share b_{jj} in a CES class of models that are observationally equivalent to (8). The gains from trade relative to autarky (where $b_{jj} = 1$) are given by

$$b_{jj}^{-1/\theta} = \left(\frac{Y_j}{Y}\right)^{-1/\theta} \frac{1}{\Pi_j P_j}. \tag{9}$$

The left hand side is the product of a $-1/\theta$ power transform of j's global sales share and the terms of trade $D^2_{jj} = 1/\Pi_j P_j$. A more widely useful application is to *ex post* evaluation of changes. It is convenient for purposes of comparison with non-parametric gravity methods to form log differences of (9):

$$\Delta lnG_j = -\frac{1}{\theta}\Delta lnb_{jj} = -\frac{1}{\theta}\Delta ln(Y_j/Y) + \Delta \ln T_j^{CES}. \tag{10}$$

The first term on the right controls for changes in global sales share of j while the second term is the change in log terms of trade $T_j^{CES} = 1/(\Pi_j P_j)^{CES}$.

For non-parametric gravity, the corresponding change in gains expression is derived from (7) with $i = j$ using the Toïnqvist approximation. Rises in $[b_{jj}P_j - Y_j/Y]$ measure a loss relative to as-if-frictionless trade,

$$g_j = [P_j b_{jj} - Y_j/Y] = \bar{b}_{jj}\frac{1 - 1/\Pi_j}{(1/\Pi_j P_j)^{1/2}}. \tag{11}$$

Equation (11) specializes (7) with $i = j$ to the translog case, implying $\bar{b}_{jj} = [b_{jj} + Y_j/Y]/2$, $\tilde{P}_j = (P_j^1)^{1/2}$ and $\widetilde{1/\Pi_j} = (1/\Pi_j)^{1/2}(1)^{1/2}$, based on translog expenditure function properties. The term $(1 - 1/\Pi_j)/(1/\sqrt{\Pi_j P_j})$ is the discrete percentage change in terms of trade implied in the move to as-if-frictionless trade. Using observable price index data P_j along with the other observables, equation (11) can be solved for Π_j. Then *ex post* changes can be evaluated with declines in loss

$$-\Delta \ln g_j = -\Delta\bar{b}_{jj} + \Delta \ln T_j. \tag{12}$$

Non-parametric gains expression (12) is fully non-parametric, in contrast to CES expression (10) that requires an estimate of the trade elasticity θ.

(12) is interpreted as a proportionate change in real income. (Its derivation via (5)-(6) uses the property that the level of real income cancels from both sides of the equations.) $g_j u_j$ scales by the base level of real income u_j to give a compensating variation measure.

The proportionate change $\Delta \ln g_j$ is invariant to the change in real income from u_j^0 to u_j^1.

Note that (12), like (9), incorporates changes in Y_j/Y. Thus it reflects changes in specialization due to terms of trade changes along with any other supply side forces at work.[15] In contrast, for the case of comparison to frictionless trade (11) the endowments are constant and the loss is interpreted as a loss of the exchange gain.

The change in the log terms of trade implied by (12) is

$$\Delta \ln T_j = -\Delta \ln(P_j b_{jj} - Y_j/Y) - \Delta \ln \bar{b}_{jj}. \tag{13}$$

In comparison, the change in the log terms of trade implied by the CES restriction is

$$\Delta T_j^{CES} = -\Delta \ln(\Pi_j P_j)^{CES} = -\frac{1}{\theta}[\Delta \ln b_{jj} - \Delta \ln(Y_j/Y)]. \tag{14}$$

Terms of trade inferred from gravity are useful because standard measures of the terms of trade are deficient. Price comparison is rife with measurement error and incomplete coverage for well known reasons. Less obviously but perhaps more importantly prices do not contain unobserved user costs, costs that vary across users and product types. Measure (13) uses structural gravity and usually high quality observations on value of production and trade while also making use of observed P_j that is subject to the same problems with to price comparison.

The general case for *ex post* evaluation avoids having to know "the" trade elasticity or any other parameter by appeal to the Törnqvist approximation. Quantifying the changes in real income implied by changes in $P_j b_{jj} - Y_j/Y$ is relatively easy in the general setting here. Begin with the measurement of the loss of gains relative to as-if-frictionless equilibrium. The right hand side of (7) for the case $i = j$ requires an estimate of $\tilde{b}_{jj}$. The Törnqvist approximation is $\bar{b}_{jj} = (b_{jj} + Y_j/Y)/2 \approx \tilde{b}_{jj}$. For the translog case, the approximation is

[15] A subtlety here is that situation 1 is not the as-if-frictionless equilibrium, so the world price vector $\tilde{\mathbf{p}} \neq \mathbf{1}$. The effect of changes in $\tilde{\mathbf{p}}$ is captured by $\Delta(b_{jj} - Y_j/Y)$.

exact. Thus $\epsilon_{ij} = 0,\ \forall i, j$

3 Practical Applications

General gravity (7) has important practical applications to nonparametric measures of gains from trade, terms of trade and economic distance. A fourth application is to inference of parameters for counterfactual exercises that necessarily use parametric gravity.

The applications are based on the WIOD data for manufacturing 2000-2014 and use the Törnqvist approximation. The data do not permit treatment of final demand and intermediate input demand separately, so the cost function $e(\mathbf{p}^j)$ is assumed to be identical for both uses.

Price indexes from the WIOD are consistently associated with the production and expenditure flows. The buyers side price indexes of the theory suggest using the intermediate input price indexes of the WIOD. The adding up condition on bilateral shares to world market shares, implies that the normalization of the price indexes is $\sum_j E_j P_j / \sum_j E_j = 1$.[16] Thus the observed price indexes $\hat{P}_j$ are deflated to form the normalized $P_j = \hat{P}_j / \sum_j E_j \hat{P}_j$.

Both the approximation error η_{ij} and the unknowable non-parametric error ϵ_{ij} may be substantial and non-random. Inability to treat final and intermediate demand systems separately introduces further specification error. All methods are subject to measurement error, but in contrast to CES gravity the non-parametric method additionally relies on buyer price indexes subject to error.

3.1 Gains from Trade and Terms of Trade: US and China

Non-parametric sufficient statistics for changes in gains from trade and terms of trade are calculated below for the US and China. Demand is interpreted as being the derived demand

[16] The adding up condtion is $\sum_j P_j u^j / \sum_j u^j = 1$, and $u^j = E_j / P_j$. The WIOD data do not report a P_j for the rest-of-world category, which is generated here by assuming that the missing price is equal to the expenditure-weighted average of the reported prices.

for intermediate goods. Thus u^j is interpreted as the real activity in destination j for the set of intermediate goods being produced, and $e(\cdot)$ is interpreted as the cost function for the intermediate goods. The good produced by each country is identified with a sector. *Sectoral* trade is a natural focus for gravity analysis, and is readily justified in the preceding setup by assuming that intersectoral budget shares are constant – the technology (or preferences) nests general sectoral structure within an upper level Cobb-Douglas structure. For this case, all the preceding analysis applies sector by sector.[17]

The ubiquity of unbalanced trade requires a modification of the gains from trade measure to consistently account for it. A procedure that is consistent with the focus of typical counterfactual analysis is to assume that the ratio of expenditure to income ϕ_j, $\forall j$ remains constant at its observed base value as the static equilibrium is perturbed.[18] For countries that borrow ($\phi_j > 1$), a rise in seller's price p_j yields a rise in spending power $\phi_j p_j$, with real purchasing power, the terms of trade, changing to $\phi_j p_j / P_j$. Relative changes in terms of trade are invariant to the value of ϕ_j with constant ϕ_j.

The model notation in (7) implicitly extends to multiple goods distributed from each origin. The i notation now refers to an origin-sector product, explicitly designated with subscript ik. The effects of economic distance on the terms of trade now imply aggregation across sectors as well as partners, for both sellers and buyers. Apply the Törnqvist approximation to the WIOD data for total manufacturing for the US. The translog case implies exact aggregation across sectors k.

The application implies that US manufacturing experienced a 5.6% fall in gains from trade relative to autarky in manufacturing from 2000 to 2014 (from 27.7% to 22.1%). This was due to a 20% deterioration in US manufacturing terms of trade.

A second application of gains from trade and terms of trade change is to China. The high growth rate of China's world manufacturing share (from 8% to 32%) in the same period drove

[17] The constant shares assumption avoids having to deal with connections between sectors in the comparative statics of (7).

[18] The adding up constraint for the world implies a consistency constraint on the set of ϕ_js: $E = \sum_j E_j = Y \Rightarrow \sum_j \phi_j Y_j = Y$.

a manufacturing terms of trade deterioration of 42%, a partial offset to the doubtless high social return on China's reallocation of factors from agriculture to manufacturing. China's gains from manufacturing trade relative to autarky fell from 17.7% in 2000 to 11.8% in 2014. Intriguingly, if autarky is regarded as the threat point in bilateral bargaining, the comparison suggests that China has less to lose than the US.

3.2 Nonparametric Estimation of Economic Distance

(7) under the Törnqvist approximation also suggests a nonparametric approach to estimating economic distance. This is advantageous for purposes that do not require demand parameters.

The translog case implies that (7) rearranged to isolate an error term $\mu_{ij} = \eta_{ij} + \epsilon_{ij}$ on the right hand side is given by

$$1 + 2\frac{b_{ij}P_j - Y_i/Y}{b_{ij} + Y_i/Y} - \frac{\tau_{ij} - 1/\Pi_i}{(\Pi_i P_j)^{1/2}} = \mu_{ij}, \ \forall i, j. \tag{15}$$

Non-parametric 'fit' to $\sum_{i,j} \mu_{ij}^2$ is over-determined with $N^2 - N$ τ_{ij}s plus $2N - 1$ multilateral resistances. Combining the cross-section data on b_{ij} and Y_i/Y with time series variation, non-parametric minimum distance estimation can potentially identify time varying multilateral resistances Π_i, P_j and time-invariant τ_{ij}s. It is important to allow for some relevant time variation in the latter, disciplined by information about trade agreements and similar events. See Anderson and Yotov (2016) for an example of non-parametric estimation using fixed effects only, interpreted in a CES structural setting. Time invariant τ_{ij}s are less plausible as price-dependent non-homothetic effects and time-varying composition effects of heterogeneous populations are important. Parametric methods are unavoidable then.

Concern about price independent income effects may be ameliorated using Almost Ideal gravity to yield a closely related nonparametric measure of economic distance. This requires parametrically estimated income effects on shares as a prior step. The left hand side of (15) is then adjusted to take out the income effects, after which the right hand side delivers the economic distances $\forall i, j$. If the income effect parameters are not externally provided, their estimation requires fully parametric methods applied to the data.

Since roughly half of world trade is in intermediate goods, the demand systems for intermediates are derived from cost minimization, hence inherit the properties of the final demand systems used above. Details are in Section 4.

3.3 Parameter Inference from Non-parametric Gravity

Counterfactual measurement of projected changes in gains from trade and related effects is the object of a large recent literature. Projection requires parametric gravity. Results are subject to two important sources of parameterization error: (i) error in the parameter estimate, given the parametric specification; and (ii) error in the specification. Sensitivity analysis with respect to the estimated parameter error is often crude in the literature. Simonovska and Waugh (2014) recommend use of trade elasticities inferred from the parametric gravity model estimated on the data to be used in the counterfactual. Some of the literature follows this practice. In contrast, the literature typically does not treat specification error.

Non-parametric gravity suggests a method to deal with both sources of error. As to (i), parameter estimate error, the standard gravity estimation method selects the parameter that best fits *bilateral* trade data. In the usual counterfactual context where the change in the gains from trade is the objective, the appropriate focus is instead on the fit of the parametric model to the *domestic* trade share. This suggests inference based on fitting the non-parametric changes in gains from trade with the best fit parametric representation. As to (ii), specification error, the treatment below for optimizing the fit to the gains from trade extends naturally to evaluating the fit of different parametric specifications.

The widely used CES gains from trade change measure of Arkolakis et al. (2012) is a convenient example. The CES model requires a value of the trade elasticity. To be consistent with the non-parametric gains from trade change estimate, the value of θ used for the CES estimate in (14) must satisfy equality with $\Delta \ln T$ in equation (13):

$$H_j \equiv (-1/\theta)[\ln(b^1_{jj}/s^1_j) - \ln(b^0_{jj}/s^0_j)] - \Delta \ln T_j = 0. \tag{16}$$

The CES measure of change in the China's manufacturing gains from trade is consistent with the translog measure of change in the gains from 2000 to 2014 for the value of $\theta = 2.46$ that exactly satisfies equation (16). $\theta = 2.46$ is in the range of estimated θ in the literature based on the CES specification, though it is below most estimates.

A single value of θ cannot normally satisfy $H_j = 0$ in US or Chinese manufacturing changes for each year from 2001 to 2014. A natural quantitative approach is to find the value of θ that fits best on average for the sample years. Moreover, for counterfactuals the analysis is usually focused on the changes in gains from trade for multiple countries. This suggests that the common value of θ should be a best fit for experienced gains from trade effects across the group of countries. A key issue is the uneven process of globalization over 2000-2014. A non-parametric approach to control for this is a set of time fixed effects that adjust H_j for common globalization shocks.[19]

Denote $H_{j,t}$ as the difference in year t of the sample, where H_j is the difference defined in equation (16) above. Solve for the minimum distance parameter estimate $\hat{\theta}$ that satisfies

$$\min_{\theta} \sum_{j,t} H^2_{j,t}. \tag{17}$$

Apply the minimum distance technique (17) for the WIOD countries from 2001-2014, amending $H_{j,t}$ to include time fixed effects. The solution yields the fitted value of $\theta = 1.72$, significantly below the value of 2.46 reported above for the solution to (16) for China 2001-14. The within goodness-of-fit R^2 is 0.31 (while the between R^2 is 0.015).[20] The overall goodness-

[19]This can be extended to allow for a limited amount of country specific time fixed effects to allow for asymmetries. See Anderson and Yotov (2020) for an example of non-parametric estimation of this type that is consistent with a structural model in which bilateral 'marketing capital' adjusts over time.

[20]OLS regressions solves (17). The case with no time fixed effects yields $\theta = 2$ and $R^2 = 0.14$.

of-fit of the parametric model $\hat{\theta}$ represents how closely the parametric gains measure comes to the non-parametric measure across the variation in the sample. The within $R^2 = 0.31$ suggests that richer parametric models may do better. Relative to the extensive literature on counterfactual evaluation of trade policy effects on gains from trade, the lower trade elasticities implied by minimum distance consistency with non-parametric gravity suggest doubling the gains from trade effects.

Parametric models more general than CES imply a parameter vector $\boldsymbol{\Theta}$ in the analog to (16)-(17). In this case the goodness of fit of the counterfactual gains measure based on using $\hat{\boldsymbol{\Theta}}$ still applies. In practice, the goodness-of-fit of various specifications of the Θ vector can be evaluated using this technique, with a standard statistical penalty for increasing the number of parameters. Imposing a distribution assumption for the random error, the log likelihood function for each specification can be deployed to generate an Aikake Information Criterion (AIC) score to rank the alternatives.[21]

The suggested comparison treats all parametric specifications equally except for the statistical penalty for increase in the number of parameters. Gravity models are attractive for counterfactuals in being sparsely parameterized, so there may be reason to favor sparser specifications that is not reflected in the AIC score. Even so, the goodness-of-fit of various specifications relative to the non-parametric model remains a measure of credibility of the counterfactual exercise.

The minimum distance technique permits statistical inference when the residuals equal to $H_{j,t}$ evaluated at $\hat{\theta}$ are random. For example, standard 95% confidence intervals for the time fixed effects estimate yield $-1/\hat{\theta} \in [-0.465, -0.695]$. Even if standard statistical inference is not applicable,[22] the minimum distance interpretation suggests an informative confidence band for purposes of evaluating counterfactual projections. Appendix 6.2 discusses non-randomness of error terms.

[21] A believable error variance for estimated $\hat{\boldsymbol{\Theta}}$ requires the un-knowable covariance structure of the specification error.

[22] Non-randomness may be due to the approximation error or to specification error relative to the unknowable 'true' specification as well as any systematic error in the trade flows and the price indexes.

4 Endogenous Supply and Specialization

The specialization gains from trade can in principle be non-parametrically measured with the technique of Section 1.3: apply the intermediate value theorem to a supply side value function, the GDP function or profit function. Succeeding steps allow non-parametric calculation of the specialization gains from trade due to endogenous supply of outputs and sourcing of intermediate inputs in a setting that includes endogenous trade costs. The method for potential non-parametric quantification of changes in specialization gains in the composite factor case is sketched below.

Two obstacles discourage the extension, one theoretical and the other practical. The theoretical obstacle is that the connected substitutes property of Berry et al. (2013)) must be taken to apply to the maximum value GDP or profit function. For GDP functions this requires dubious restrictions on the technology and/or the endowment differences of countries. Adão et al. (2017) assume a single composite primary factor of production in the GDP function applied to generate their factor demand system. Given a composite factor, connected substitutes is no more restrictive on the supply side than on the demand side. In the context of the extension, the single composite factor effectively treats GDP as if based on a joint product technology. All non-jointness effects are buried in implicit endogenous productivity shifters that act like the non-homotheticity shifters of Section 1.2. Thus the main drivers of specialization in factor proportions and heterogeneous firms models are implicit in the non-parametric approach.

The practical obstacle to a parallel treatment is the well-known dubious quality of data on imported intermediate inputs. Input-output table builders allocate imported intermediate goods to sectors in the same proportions as the observed allocation of domestic counterparts. This is known to be seriously erroneous in the few cases where it is possible to check the practice against observed direct data.

The GDP function under the single composite factor assumption is the product of the GDP deflator function and the aggregator function of the primary factor endowment vector.

The GDP deflator function is convex and homogeneous of degree one in prices, the vector of seller prices $\mathbf{p}_y^j$ and input buyer prices $\mathbf{p}_m^j$. The factor aggregator function is concave and homogeneous of degree one in the endowment vector $\mathbf{v}^j$. World GDP is the sum of country GDPs.

Applying the method of Section 1.2, the actual world GDP can be related to the as-if-frictionless world GDP with common price vectors using the intermediate value theorem applied to the GDP deflator function. The specialization gains from trade can be non-parametrically calculated from the domestic sales share, in parallel to the exchange gains calculation based on the domestic expenditure share in Section 1.2.

Restriction of cross-country differences in technology to output- and factor-augmenting technology differences mimics the treatment of taste differences above. Endogenous technology shifters admit selection among heterogeneous firms, endogenous markups and some forms of returns to scale. The endogenous sellers prices that generate the endogenous supply vectors are generated in the spatial arbitrage equilibrium as in Section 1.

The extension of gravity to general GDP functions permits a very general representation of trade frictions. This is important because distribution surely involves complicated interaction with pure production. The current understanding of gravity in practice is mainly limited to iceberg trade costs as in assumption (v), with only very limited extension.[23] Appendix section 6.3 has more details.

Panel data changes in non-parametric specialization gains from trade and terms of trade measures analogous to (17) could in principle be applied to parameterize GDP functions for use in counterfactuals. This project faces challenges in selecting the technology and selection forces to be parameterized. For example, the constant elasticity of transformation GDP function equivalent to (17) is implausibly restrictive, since it is associated with specific

[23] In the general case, endogenous trade frictions soak up a potentially enormous amount of economic action. Head and Mayer (2014) call gravity trade frictions 'dark' in appropriating the gravity metaphor of cosmology. Special tractable cases may shed some light and reduce the unexplained magnitude of the frictions. See Arkolakis (2010) and Anderson and Yotov (2020) for gravity model examples of endogenously increasing trade costs.

factors and mobile labor allocation based on identical Cobb-Douglas production functions (Anderson (2011b)).

5 Conclusion

Economic gravity describes the static equilibrium of bilateral trade between N^2 pairs of regions where N is any positive integer. The attractive force driving to equilibrium is rational profit maximizing arbitrage drawn by the gains from trade between locationally separated supplies and demands by rationally motivated cost minimizing buyers. Adding up conditions on sales and expenditures constrain the possible bilateral trades. For a wide class of demand systems, the equilibrium depends on a set of the inverse squares of bilateral equilibrium economic distances. Mild restrictions are provided under which non-parametric sufficient statistics for gains from trade and terms of trade may be calculated. Under further regularity conditions, non-parametric estimation of economic distances is possible.

Essentially, arbitrage aggregates the equilibrium distribution of goods such that a set of two body relationships can characterize the equilibrium. The two body relationships take the form of Newton's two body law – the inverse of the square of bilateral distance. The two body property of equilibrium appears most clearly in characterizing the equilibrium terms of trade of any region. Its terms of trade are driven below its as-if-frictionless terms of trade in proportion to the inverse of the squared economic distance between that region and the world market.

Counterfactual calculations require parametric representations of gravity. For such exercises, the non-parametric gains from trade calculations in Section 3 of this paper may provide a useful specification bias test on the values of parameters used. For example, as in the CES case developed there, use several years of data to calculate changes in the gains, then solve for the implied trade elasticity parameter that is consistent with those changes in gains. The specification bias test can supplement the usual sensitivity checks with respect to values of the parameters within a given parametric representation such as CES.

References

Adão, Rodrigo, Arnaud Costinot, and Dave Donaldson, "Nonparametric Counterfactual Predictions in Neoclassical Models of International Trade," *American Economic Review*, 2017, *107* (3), 633–689.

_ **, Costas Arkolakis, and Sharat Ganapati**, "Aggregate Implications of Firm Heterogeneity: A Nonparamatric Analysis of Monopolistic Competition Trade Models," 2020. Yale University.

Anderson, James E., "The Gravity Model," *Annual Review of Economics*, 2011, *3*, 133–160.

_ , "The Specific Factors Continuum Model, with Implications for Globalization and Income Risk," *Journal of International Economics*, 2011, *85* (2), 174–185.

_ **and Penglong Zhang**, "Latent Exports: Almost Ideal Gravity and Zeros," 2019. Boston College.

Anderson, James E. and Yoto V. Yotov, "Terms of Trade and Global Efficiency Effects of Free Trade Agreements, 1990-2002," *Journal of International Economics*, 2016, *99* (C), 279–298.

Anderson, James E. and Yoto V. Yotov, "Short Run Gravity," *Journal of International Economics*, 2020.

Arkolakis, Costas, "Market Penetration Costs and the New Consumers Margin in International Trade," *Journal of Political Economy*, 2010, *118* (6), 1151–1199.

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

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

Dekle, Robert, Jonathan Eaton, and Samuel Kortum, "Unbalanced Trade," *American Economic Review: Papers and Proceedings*, 2007, *97*, 351–355.

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

Head, Keith and Thierry Mayer, "Gravity Equations: Workhorse, Toolkit, and Cookbook," Chapter 3 in the Handbook of International Economics Vol. 4, eds. Gita Gopinath, Elhanan Helpman, and Kenneth S. Rogoff, Elsevier Ltd., Oxford, 2014.

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

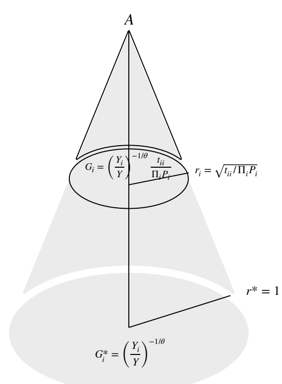
Zhang, Penglong, "Home-biased gravity: The role of migrant tastes in international trade," *World Development*, 2020, *129* (May).

6 Appendix

6.1 Lost Gains from Frictions

The CES measure of gains from trade in Section 3 falls inversely to the square of economic distance. The figure below translates this property into a physical analogy with phenomena like radiation (as well as Newton's Law). Economic distance is measured as height on the axis of the cone. Gains are measured as the areas of circles associated with cross sections at various heights. The radius of the circles falls proportionally with height above the base of the cone.

Gains and Economic Distance



6.2 Non-random Error Consideration

Non-randomness in the error term is an important obstacle for statistical inference. The error structure in $H_{j,t}$ has its source in (7). In the illustrative case of the trade elasticity for manufacturing reported above, composition effects are likely to be a big source of non-randomness. Disaggregation is the appropriate treatment. The analytical considerations below apply to remaining sources of non-randomness.

Temporarily disregard measurement and other sources of error in the variables on the left hand side. The right hand side of (7) for $i = j$ gives ϵ_{jj} as the error of the non-parametric model defined here from the unknowable true model. By construction $\sum_j \epsilon_{ij} = 0$. Given this, random ϵ_{jj} appears defensible by the principle of insufficient reason. The Törnqvist approximation $\bar{b}_{jj}$ for the unknown $\tilde{b}_{jj}$ in (11) introduces an approximation error to the operational version of (7). It may not be random if the true model is not translog. The change in log terms of trade (13) is also subject to the possibly non-random approximation error as (7). Then the key variable $H_{j,t}$ defined in (16) is subject to those same source of error.

Now consider non-random error in the observable variables on the left hand side of (7). Non-randomness is most problematic for the price index P_j, though non-random error in the other observable variables may also be plausible. Based on current understanding of trade frictions inferred from gravity models, the theoretical buyer multilateral resistance P_j is partly due to unobservable user costs. This means that observable buyer prices and the associated observable price indexes are downward biased. Unsystematic downward bias in P_j will net out of (16). Otherwise, non-randomness is due to both systematic approximation error and non-random error in the observables.

Given a commitment to CES gravity, a potential procedure to deal with error in observed P_js for the purpose of inferring the trade elasticity from (7) uses the CES gravity estimator to provide instruments. First deflate the destination fixed effects estimated from standard CES gravity by observed expenditure E_j to yield an estimate of $\widehat{P_j^{\theta}}$. Next, replace P_j in (16) with $\left(\widehat{P_j^{\theta}}\right)^{1/\theta}$. The resulting minimum distance estimator for this variant of (17) fully exploits the assumed CES structure as part of identification of the trade elasticity that best fits the variation of gains from trade changes.

From the standpoint of non-parametric gravity, the preceding trade elasticity inference procedure throws away potential information in observed P_js. A full investigation of the relationship of observed P_j to inferred inward multilateral resistance is far beyond the scope of this paper.[24] A simple compromise procedure allows a role for the information in P_js and is easier to use.[25] An alternative compromise instrument is constructed from a bivariate regression of the expenditure-deflated destination fixed effects on observed P_j. In this case $\left(\widehat{P_j^\theta}\right)^{1/\theta}$ is replaced by $\left(\widehat{P_j^\theta}\right)^{1/\tilde{\theta}}$ where $1/\tilde{\theta}$ is the regression slope. The bivariate regression controls for the potential downward bias in observed P_j with a constant term and its residuals may be informative about remaining systematic error structure.

In some plausible cases where non-randomness in the observable data is suspected, a pseudo data approach may be feasible for creating standard errors, etc. Suppose that the time series variation in the data is due to variation in the supply data, with the τ_{ij} being time invariant. (For periods where there are no regional trade agreement implementations, this is plausible provided that price dependent non-homotheticity is rule out.)[26] In the cross section, supply and associated expenditure is exogenous, but on the time dimension of the panel data structure there may be serial correlation. The plausibility of the pseudo data approach depends on getting the random draw mechanism right.

The pseudo data approach to deal with this source of non-randomness in the sample generates the pseudo data as random draws $\tilde{t}$ from the actual data:

$$H_{j,\tilde{t}} = \left[\frac{b_{jj}^{\tilde{t}}/(Y_j^{\tilde{t}}/Y^{\tilde{t}})}{b_{jj}^0/(Y_j^0/Y^0)}\right]^{-1/\theta} - T_j^{\tilde{t}}/T_j^0.$$

Using the pseudo data generated by many such draws, calculate the minimum distance estimator defined in (17) where t now runs across all the pseudo data samples.

[24] Inferred multilateral resistances from CES gravity investigations show that outward multilateral resistance is much larger and has much more variation over time and space than does inward multilateral resistance.

[25] The high nonlinearity of the first procedure may be computationally problematic.

[26] A simple type of time dependence of τ_{ij} is absorbed by the time variation of $b_{jj,t}$. See Anderson and Yotov (2020) for an example of a cross-border-time fixed effect that absorbs globalization effects in a reduced form gravity model.

6.3 GDP Function Approach

The convex technology is formalized with vectors $\mathbf{x}^i = \{x_{ij}^k\}$ of sector-origin-destination final outputs, $\mathbf{m}^i = \{m_{ji}^k\}$ sector-origin-destination intermediate (produced) inputs, and origin-primary-factors $\mathbf{v}^i = \{v_l^i\}$. Restrict locational differences in technology in parallel to Definition G.

Definition T

Technologies differ across locations only by augmentation shifters.

Technology differences in general gravity that are origin-sector specific are 'frictions' that are absorbed in sellers' incidences (outward multilateral resistances). There is no need for separate accounting here. Definition T as it applies to primary factors implies that v_l^i is measured in efficiency units.

Let y_{i0}^k denote production of sector k output in origin i "at the factory gate", while $y_{ij}^k, \ \forall i, j, k > 0$ denotes delivery of sector k output in origin i to destination j. Output in origin i requires produced inputs $m_{ji}^k, \ \forall j, i, k$ and primary factors $v_l^i, \ \forall l, i$; all measured in efficiency units, under Definition T. The technology comprises feasible vectors $\mathbf{y}^i, \mathbf{m}^i, \mathbf{v}^i \in \mathcal{T}(\mathbf{y}^i, \mathbf{m}^i, \mathbf{v}^i)$ where $\mathcal{T}$ is a convex set. All productivity differences across origins and destinations are absorbed in 'distribution frictions' by sector-origin-destination and by primary factor augmentation shifters embedded in the v_l^i variables.

Efficient production results in a GDP function $R^i(\mathbf{p}_y^i, \mathbf{p}_m^i, \mathbf{v}^i)$ that is convex and homogeneous of degree one in the price vector $(\mathbf{p}_y^i, \mathbf{p}_m^i)$, and concave and homogeneous of degree one in $\mathbf{v}^i$. The trade frictions are due to the technology, with their equilibrium values revealed by p_{ij}^k / p_{i0}^k for both final and intermediate products (with some abuse of sector notation allowing k to refer to either final or intermediate production of sector k). The joint product restriction implies that the GDP function becomes $r^i(\mathbf{p}_y^i, \mathbf{p}_m^i) f(\mathbf{v}^i)$ where the composite factor aggregator function $f(\cdot)$ is concave and homogeneous of degree one.

World GDP is $R^W = \sum_i R^i = \sum_i r(\mathbf{p}_y^i, \mathbf{p}_m^i) f(\mathbf{v}^i)$.