Non-parametric Gravity*

James E. Anderson
Boston College and NBER

October 24, 2022

Abstract

Non-parametric gravity defined in this paper contains previous parametric forms in a wider class. Implications of spatial arbitrage in goods yield non-parametric sufficient statistics for arbitrage gains from trade and terms of trade. For world manufacturing trade 2000-2014, China's gains rose 2% yearly and terms of trade fell 8.3%. US gains fell 2% yearly and terms of trade rose 5.5%. A novel minimum distance estimator of the CES trade elasticity is illustratively applied to counterfactual industrial policy. A 1% rise in US 2014 world sales share raises gains by 0.4%.

*I am grateful to Danial Lashkari and Arthur Lewbel for helpful discussions and comments on a first draft. I also thank participants at the NBER 2021 Winter ITI meeting for helpful comments.

Bilateral trade shares deviate from world average trade shares in patterns suggestive of physical gravity, as first recognized by Tinbergen (1962). Economic gravity models subsequently deduced these properties from spatial arbitrage equilibrium under parametric structures that severely restrict income and substitution effects in demand, and impose similar strong restrictions on supply; e.g. Anderson (2011a) and Head and Mayer (2014). The restrictive assumptions raise doubts about results in the influential applied structural gravity literature. For example, the absence of net complementarity in standard parameterizations suggests that important third party effects may be suppressed. The class of models with gravity properties is greatly expanded in this paper. Non-parametric representation of spatial equilibrium yields consequential theoretical and empirical results.

Received structural gravity represents the implications of trade frictions for equilibrium bilateral trade patterns relative to as-if-frictionless trade in the form of relative resistances, ratios of bilateral resistance to the product of inward and outward multilateral resistance. (As-if-frictionless trade is defined as the case where all destinations spend an equal share on goods from each origin, equal to the world sales share of each origin.) The received class includes a variety of cases that reduce bilateral trade shares to a constant elasticity of relative resistance form. Non-parametric gravity as defined here is based on invertible demand systems that are common to multiple destinations. Invertible demand systems, the sufficient condition for non-parametric gravity, allow for both net complementarity and the widely observed zeros in bilateral trade at the sectoral level.[1] Spatial equilibrium distribution of equilibrium supplies is subject to market clearing constraints and budget constraints, as in the standard model. In the non-parametric gravity class, the effect of 'trade frictions' on bilateral trade reduces to bilateral relative resistances. The widely applied constant elasticity structure is a special case. As operationalized below, relative resistances are measured with non-parametric sufficient statistics, no elasticity parameters are required. The common features across the broad non-parametric class may allay some doubts about the robustness

[1] The 'connected substitutes' structure of Berry et al. (2013) is sufficient. Significantly, 'connected substitutes' allows for complementarity and also for zeros in bilateral trade.

of implications and inferences based on constant elasticity gravity models.

A pleasing by-product is revelation that economic gravity follows an inverse square law of economic distance, reconnecting economic gravity to its physical origin.[2] 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 equal to relative resistance (the ratio of domestic resistance to the product of inward and outward multilateral resistance). Relative resistance for domestic trade in turn is equal to the inverse square of a country's equilibrium economic distance to and from the world. Economic distance is defined to be equal to the geometric mean of inward and outward multilateral resistances normalized by the square root of internal distance.[3]

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. [4] The complex determination of equilibrium economic distance has previously hidden the simplicity of its inverse square role in characterizing spatial equilibrium.

Three applications of non-parametric gravity are derived based on an approximation to the general class. First is a nonparametric sufficient statistic for the arbitrage (exchange)

[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.

[3]The geometric mean is the natural average for a product. Internal distance itself is a spatial average of directionally asymmetric resistances between internal points. Economic distance is thus a natural intuitive summary of a country's relationship to the world as both buyer and seller.

[4]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.

gains from trade relative to as-if-frictionless trade. The parametric CES class[5] is used by Arkolakis et al. (2012) to infer gains from trade relative to autarky from the ratio of internal trade to global sales. Their insight is extended here to yield a non-parametric measure of the arbitrage gains from trade relative to as-if-frictionless trade based on observables only. This sufficient statistic is valid in a much wider class of demand structures.

Second, non-parametric gravity yields a sufficient statistic for terms of trade changes based on observables only. The terms of trade is equal to relative resistance for the domestic trade case. Non-parametric terms of trade is a useful measure because standard terms of trade calculations based on price comparisons are limited in scope and rife with measurement error. Systematic measurement error is especially problematic for export prices and their indexes. Many categories lack observable prices, while unit values are contaminated by aggregation bias. As a result, terms of trade measures are not much used except for countries with exports dominated by commodities. The theoretical importance of terms of trade in international economics suggests wide applicability of the non-parametric terms of trade statistic.

Illustrative applications are made to China and the US in manufacturing over the period 2000-2014 using the World Input-Output Database. China's overall terms of trade fall of 8.3% per year is associated with the near quadrupling of its world manufacturing share 2000-2014 accompanied by a small rise in its domestic share. The result was a 1.96% yearly rise in gains from trade relative to as-if-frictionless trade. US manufacturing terms of trade rose 5.5% yearly, associated with a near halving of world trade share while the domestic trade share fell slightly. The result was a 2.07% yearly fall in gains from trade relative to as-if-frictionless trade.

The terms of trade movements are driven by the respective changes in global market shares. A back-of-the-envelope partial equilibrium calculation suggests that a rise in US manufacturing share by 1% would induce a 0.6% fall in terms of trade. Counterfactually

[5]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.

pushing the model harder, consider a socially costless US industrial policy that induces the 1% rise in US share. The net result is a rise in the gains from trade of 0.4% because the gain in volume of exports is only partially offset by the increase in domestic sales due to the fall in the terms of trade. The counterfactual uses a CES trade elasticity that is estimated with novel data and and a novel method, explained next.

The third application of non-parametric gravity generates a CES trade elasticity as an example of the new method. Simonovska and Waugh (2014) use tariff and transport cost variation to identify a CES trade elasticity using the same data that would be used for counterfactuals. The proposed application uses the much larger bilateral variation in non-parametric measures of relative resistance to estimate the trade elasticity, a practical advantage in the use of available information. Methodologically there is an important difference. The non-parametric approach presumes that CES is false and seeks the least inaccurate elasticity to parameterize the projection model. The parametric econometrics perspective that presumes the CES is true implies an endogeneity bias in the elasticity estimate (explained below). The bias is a feature of the non-parametric procedure, not the bug it is for the parametric econometric approach.

The news for the large CES gravity literature is mixed. Trade elasticities associated with the US and China terms of trade movements and associated expenditure and sales shares are precisely estimated off time series variation, implying that the fit of non-parametric relative resistance to the CES equivalent is reasonably good. The best-fit trade elasticity is slightly larger than 1, implying a CES substitution elasticity equal to 2. The bad news is that application of the non-parametric model to *all* bilateral manufacturing trade flows results in negative relative resistances in a bit more than 20% of observations. The likely reason is very large sectoral composition variation across country pairs. This suggests a vastly larger study should apply the model to appropriately narrow sectors. Nested CES structure is indicated for trade elasticities to support full general equilibrium counterfactuals.

The model has implications for future applications in multiple areas. Spatial aggregation (of origin and destination locations at varying sizes) is a feature of all gravity applications. An approach to consistent aggregation is sketched below in Appendix Section 6.1. The model is developed for final demand systems for goods, but it also straightforwardly applies to demand systems for intermediate inputs. Appendix Section 6.2 verifies this claim. The model could also apply to supply of labor derived from expenditure systems including 'leisure'. Gravity models of migration have previously been rationalized by parametric discrete choice structures based on random preferences on the sellers' side of the labor market. Labor supply based non-parametric expenditure systems combined with non-parametric probability distributions for idiosyncratic elements of preferences[6] are plausible in principle and may be useful in thinking about migration and labor force participation.

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.

[6]The common use of extreme value distribution(s) is based on limiting distributions that are valid for large numbers of iid draws. The sparseness of many bilateral migration flows raises doubts about the validity of this practice.

The paper abstracts from selection and all other sources of endogenous supply shifts to allow a sharp focus on non-parametric specification of demand.[7] Efficient supply is assumed to be observed, meaning the supply vectors are associated with the equilibrium sellers price vectors. In principle, the method of this paper could be extended to apply to the explanation of endogenous supply, ideally yielding non-parametric measures of the gains from trade due to selection and specialization. The extension would build on the Adão et al. (2020) non-parametric approach to modeling heterogeneity of firms productivities in Chaney-Melitz type gravity models. Challenges to the extension are explained in Section 4.

Section 1 develops the non-parametric gravity model. Section 2 derives non-parametric measures of the gains from trade and terms of trade. Section 3 presents the applications to manufacturing trade 2000-2014. Section 3.1 applies the non-parametric gravity model to derive an appropriate method of estimating trade elasticity parameters for use in counterfactual exercises based on necessarily parametric gravity models . Section 4 discusses challenges to extending these methods to general equilibrium supply models. Section 5 concludes.

1 Non-parametric Gravity

The formal non-parametric gravity model is approached in steps that suggest the generality of gravity representations of spatial equilibrium. Gravity properties reduce to dependence on pairwise relative resistances. Section 1.1 presents an intuitive graphical analysis to illustrate how frictions determine the terms of trade and gains from trade. Section 1.2 provides a formal analysis such that the seemingly partial equilibrium graphical analysis actually illustrates general spatial equilibrium. The sufficient condition is that the demand system be invertible. Invertibility permits application of the intermediate value theorem to describe the relationship of every observed bilateral trade to its observable as-if-frictionless hypothetical value. This step reduces the representation of spatial equilibrium to an effectively pair-wise

[7] 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.

set of bilateral relationships. Section 1.3 applies further restrictions yielding an operational nonparametric gravity approach to quantifiable bilateral trade modeling. In Section 3.1, CES 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.[8]

(ii) Each origin ships an equilibrium supply (effectively 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 (potentially endogenous) fraction of shipments – equilibrium as-if iceberg melting trade costs.

The model is simplest to develop with each country distributing a single good. The multi-sector case where each origin i has an endowment vector leads to essentially the same non-parametric gravity model. The intuition is that on the buyer's side, multiple sectors and multiple origins of varieties of a single product class are much alike. Appendix Section 6.2 develops the argument.

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. For expositional convenience,

[8]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.

the sellers' unit cost is referred to as sellers' price, as in perfect competition. The formal model includes profit maximizing imperfect competition in long run equilibrium where the zero profit cutoff determines unit cost.[9]

Prices p_{ij} paid by buyers include trade costs and other frictions $\tau_{ij} \geq 1$. 'Frictions' include unobservable user costs and heterogeneity in preferences across destinations as well as endogenous trade services costs by profit-maximizing efficient trade services providers.[10] Allowing for endogeneity comes much closer to the reality of transportation costs association with congestion and mode choice variation.

In the arbitrage equilibrium $p_{ij}/\tau_{ij} = c_i, \ \forall i, j$. 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 $X_{ij} = p_{ij}x_{ij}$.

Analysis builds on a system of generic demand schedules to characterize the equilibrium allocation from origins i to destinations 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 $c_i\Pi_i$, the price paid by a hypothetical buyer in the 'world' market. Π_i is the sellers' incidence of trade frictions on world sales. 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

[9] Marginal revenue for imperfectly competitive firms exceeds the buyers price net of trade frictions by an endogenous markup. Non-parametric gravity includes the markup in trade frictions, as if it reflects rent charged by an intermediary. The focus of the model on implications of spatial arbitrage for the cross section pattern of trade also allows for endogenous markups to be split between buyer and seller side intermediaries. Much interesting economics is buried in the endogenous 'trade frictions'.

[10] 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.

the efficient spatial arbitrage $p_{ij} = c_i t_{ij},\ \forall j$. 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 $p_{ij}/P_j = c_i t_{ij}/P_j$ in arbitrage equilibrium. Thus the equilibrium shipment x^e_{ij} at the equilibrium point E is associated with relative price p_{ij}/P_j.

The aggregate demand $\sum_l x^D_{il}$ similarly has each demand function in the sum being a function of the location 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 equilibrium, it is as if each seller i sold to a world market with average buyer price $c_i\Pi_i$. In a particular equilibrium of observed trade flows, it is convenient to choose units such that world prices $c_i\Pi_i = 1,\ \forall i$.[11] Then $c_i = 1/\Pi_i$ and the equilibrium pattern of bilateral trade is determined by relative prices $\{t_{ij}/\Pi_i P_j\}$.[12] Then the price index P_j is interpreted as buyer j's incidence of the set of bilateral trade costs $t_{ij},\ \forall i$; an ideal index of the bilateral buyers incidences t_{ij}/Π_i.

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 $\{c_i\}$ along with $\{\Pi_i, P_i\}$ up to a normalization. The standard normalization of prices is $\sum_i c_i y_i / \sum_i y_i = 1$.

The denominator of $t_{ij}/\Pi_i P_j$ is a product. A natural mean of a product is its square root, the geometric mean, averaging asymmetric forces that are explicit in the denominator and

[11]For a time series of gravity equilibria, this is a base year units choice, with other years seller prices given as ratios to the base year prices.

[12]For a time series, the relative seller price variation is absorbed in the relative Π_i variation.

implicitly contained in the numerator.[13] 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. Define the economic distance between i and j as

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

The set of squares of economic distance $\{\sqrt{\Pi_i P_j/t_{ij}}\}$ determines the pattern of trade. In the special CES case, bilateral trade between i and j is determined inversely to the square of 'own distance' $\sqrt{t_{ij}/\Pi_i P_j}$ with elasticity parameter $\theta > 0$.

The key contribution of this paper is to show in Sections 1.2 and 1.4 that bilateral trade is *locally* determined by the 'own distance' $\sqrt{t_{ij}/\Pi_i P_j}$ in a wide class of non-parametric gravity models.

1.1 Graphical Intuition

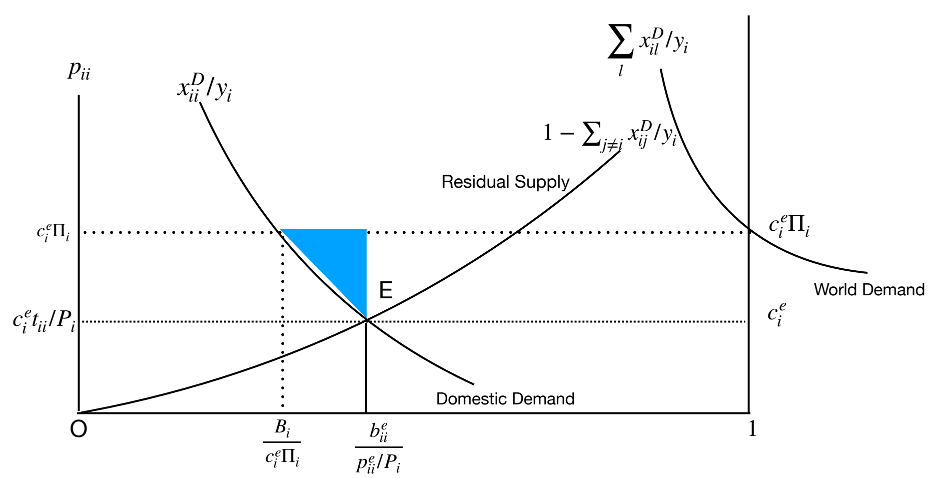
Intuition for the formal model uses a familiar supply and demand analysis in a goalpost diagram, as in exposition of the specific factors model. Each of the N origin products has a market clearing condition represented by the intersection of aggregate demand with the right vertical axis supply schedule in the goalpost diagram below. 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 $p_{ij} = c_i t_{ij}$.

The diagram illustrates the effect of frictions on the exchange gains from trade in the domestic trade case where $i = j$. Country i's supply of its product y_i is purchased by the rest of the world in amount $\sum_{j \neq i} x_{ij}$. The residual supply to the domestic market is $y_i - \sum_{j \neq i} x_{ij}$ and is an increasing function of relative price $p_{ii}/P_i = c_i t_{ii}/P_i$. The units on the horizontal

[13] The numerator, the direct bilateral friction, is typically controlled for with a combination of symmetric (physical distance) and asymmetric (border barriers) proxies. This point is expanded upon below in discussion of the interpretation of internal economic distance.

axis of the diagram are in sales shares x_{ii}/y_i for analytic convenience. For measurement purposes in section **??**, expenditure shares are used, b_{ii} in the observed equilibrium and B_i in the as-if-frictionless equilibrium where all buyers spend the same share of income on goods from seller i, equal to the observable share of world sales of all goods that are made by seller i. These expenditure shares are converted to quantity shares in the diagram. 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' $c_i = p_{ij}/t_{ij},\ \forall i, j$. For visual simplicity the projection of c_i^e to the left vertical axis is equal to p_{ii}^e/P_i^e.

Frictions and the Gains from Trade



The shaded triangle depicts the hypothetical exchange gain from trade when the existing trade frictions are converted to as-if-frictionless trade frictions. Alternatively, this is the loss of gains from trade due to the general equilibrium consequences of trade frictions as they affect country i. In as-if-frictionless equilibrium, all buyer locations would pay the same (relative) price for each good i, $c_i^e\Pi_i$. In the as-if-frictionless equilibrium, the price index $P_i^* = 1$ by a harmless normalization of its relative prices. Thus if, counterfactually, the relative buyers price rose from $c_i^e t_{ii}/P_i$ 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.

The graphical treatment suggests measurement via an approximation based on observables only. The shaded triangular area in the second diagram is the basis. Observed equilibrium domestic sales x_{ii}^e are compared to domestic sales x_{ii}^* in the as-if-frictionless equilibrium. The latter is also observable because in as-if-frictionless equilibrium, every country spends the same fraction of its income on i's good, equal to i's observable share of world sales from every origin.

For measurement purposes it is often necessary to switch from quantities to expenditures because natural quantity units are often not observable. Divide numerator and denominator of the relative domestic price by t_{ii}, so that multilateral resistance P_i is now deflated by t_{ii}. The shaded triangular area in the second diagram is approximated by $\Delta(c_i/P_i)\tilde{x}_{ii}$ where $\tilde{x}_{ii}$ is some intermediate average such as $(x_{ii}^f + x_{ii}^e)/2$. Switching to expenditure measures, the shaded trapezoidal area is approximated as

$$\frac{\Delta c_i/P_i}{\widetilde{c_i/P_i}} \tilde{P}_i \tilde{b}_{ii}. \tag{2}$$

Here, $\tilde{b}_{ii}$ is the domestic expenditure share at an approximated intermediate value between the observed b_{ii} and the 'world share' B_i equal to the observed sales share Y_i/Y. Also, $\tilde{P}_i$ is the price index associated with that intermediate value somewhere between the observed price index P_i and the hypothetical world price index $P^* = 1$.[14]

[14] The logic for the intermediate value of the price index is that a shift between equilibria must ordinarily affect the buyers price index, and multiplying the share by the price index is needed to convert to the level of nominal expenditure when the share is multiplied by expenditure.

The intermediate value theorem, if applicable, implies that expression (2) is equal to

$$P_i b_{ii} - B_i = \frac{\Delta c_i/P_i}{\widetilde{c_i/P_i}} \tilde{P}_i \tilde{b}_{ii} \tag{3}$$

for appropriately chosen intermediate values denoted by the tilde. The left hand side is observable directly. The formal approach below shows the conditions under which an appropriate intermediate value exists and the restrictions under which it is operational. See Sections 1.2 and 1.3 and the quantification of the shaded trapezoidal area in Section 3.

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 c_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 relative to frictionless 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.[15] These suggestions are applied based on the formal model that follows.

[15]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 $\{c_i\}$ and the simultaneous interaction of $\{P_i\}$ with $\{\Pi_i\}$.

1.2 Formal Model

Spatial equilibrium in the non-parametric gravity model as developed below 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 $\{t_{ij}\beta_{ij}^h\}$ with β_{ij}^h indicating a household h specific quality shifter. The taste shifters are isomorphic to transportation and other costs in their effect on the pattern of trade, hence the collective term resistance or friction is more accurate.

A prelimary simplification is justified by noting that *relative* frictions $\{\tau_{ij}/\sqrt{\tau_{ii}\tau_{jj}}\}$ are what determines the cross section pattern of trade:

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

The relative trade friction form is used henceforth: $\tau_{ij} = t_{ij}\beta_{ij}/\sqrt{t_{ii}t_{jj}\beta_{ii}\beta_{jj}}$. The internal frictions $t_{ii}\beta_{ii}$ are absorbed in the multilateral resistances.[16] With this simplification, $D_{ii} = \sqrt{\Pi_i P_i}$ understanding that the multilateral resistances are scaled by $1/\sqrt{\tau_{ii}}$.

Two useful implications follow. First 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. See Appendix Section 6.1 for details.

The second implication is that D_{ii}, region i's distance to and from the world market, is an inverse measure of the elusive "open-ness to trade" concept. Note that D_{ii} is comparable across countries in the cross section and over time, and is defined for the wide class of non-

[16] 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. See Agnosteva et al. (2019).

parametric gravity models. Also, i's terms of trade are reduced relative to the as-if-frictionless value of one in inverse proportion to the square of economic distance D_{ii}.

Now turn to the formal demonstration of the claim in the first paragraph. 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 $e^{hj}(\{p_{ij}\}, u^{hj})$, which is concave and homogeneous of degree one in the price vector and increasing in utility u^{hj}. The effects of cross-section variation in agent h behavior on expenditure function $e^{hj}(\cdot)$ are restricted here to define the general gravity class of equilibrium arbitrage models that satisfy

Definition G

1. *assumptions (i)-(v) hold along with*
2. *assumption (vi): the demand system is invertible.*

The purpose of assumption (vi) is to satisfy a crucial regularity condition that is implicit in the graphical analysis – all bilateral demand can be related to as-if-frictionless global demand by the same technique. A sufficient condition for assumption (vi) is that the demand system has the connected substitutes property of Berry et al. (2013). Assumption (vi) includes many plausible forms of non-homotheticity that are not price independent. This allows price elasticities to be functions of buyer income. Thus the assumptions in Definition G accommodate a great deal of heterogeneity of agents demand demand. For purposes of non-parametric implications of spatial arbitrage, a great many sources of heterogeneity have the same effect as classic iceberg trade costs.

To see this point, let $\mathbf{p}^j$ denote the vector of p_{ij}s at destination j. Assumption (vi) implies 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 $p_{ij}^h = c_i t_{ij} \beta_{ij}^h$, where β_{ij}^h is the inferred taste shifter that explains buyer choice when facing seller price c_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. The same reasoning applies to heterogeneity in the cost of serving customers, allowing for t_{ij}^h to vary endogenously over customers h for reasons related to trade volume as well as tastes. General gravity describes the spatial arbitrage equilibrium for given equilibrium values of the bilateral resistances $\tau_{ij}^h = t_{ij}^h \beta_{ij}^h$ and thus given real incomes u^{hj}.

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. (For nonparametric identification purposes, it is important that every good is demanded somewhere.)

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

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

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 $\mathbf{p}^j = \{c_i \tau_{ij}\}$. 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 Definition G implies that

$$E = \sum_j \sum_{h=1}^{H_j} e(\{c_i \tau_{ij}\}) u^{hj} = \sum_j e(\{c_i \tau_{ij}\}) u^j. \tag{4}$$

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 an as-if frictionless world equilibrium[17] 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 $\sum_i (c_i - c_i^*) y_i = 0 = Y - Y^*$.

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.[18]

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 as-if-frictionless equilibrium. This gives the general characterization of spatial arbitrage equilibrium.

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

[17] Replace the initial $\{\tau_{ij}\}$ frictions with hypothetical frictions $\{\tau_{ij}^*\} = \{\Pi_i P_j\}$ such that every country's expenditure share on each good i is equal to Y_i/Y.

[18] The 'as if frictionless' pattern of trade is consistent with $\tau_{ij}^* = \tau_i \Pi_i \tau_j P_j, \forall i, j$ for any values of $\tau_i, \tau_j \geq 1$.

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 (efficiency) 'price' vector $\mathbf{p}^*$[19] 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 is invertible, $\mathbf{p}^*$ is unique. Apply Shephard's Lemma to the aggregate expenditure E^* thus defined to solve for the common 'world market' price vector $\{c_i\Pi_i\} = \mathbf{p}^*$ that satisfies the market clearing conditions

$$\frac{Y_i}{Y} = B_i(\{c_i\Pi_i\}), \ \forall i. \tag{5}$$

Since $\sum_i Y_i/Y = 1$, system (5) 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 convenient choice of units such that the 'world price' vector in as-if-frictionless equilibrium is $c_i\Pi_i = 1, \ \forall i$.[20] Then the buyer's price index for country j is $P_j = e(\{c_i\tau_{ij}\}) = e(\{\tau_{ij}/\Pi_i\})$. 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 $b_{ij} = e_i(\{\tau_{ij}/\Pi_i P_j\})\tau_{ij}/\Pi_i P_j$. Thus the set of bilateral economic distances $D_{ij} = \sqrt{\tau_{ij}/\Pi_i P_j}$ determines the pattern of bilateral exchange.

Return to the world adding up condition that combines the actual and as-if-frictionless equilibria $(E - E^*) - (Y - Y^*) = 0 = \sum_j e(\mathbf{p}^j)u^j - e(\mathbf{p}^*)\sum_j u^j - \sum_i (c_i - c_i^*)y_i$. Differentiate with respect to each c_i, apply Shephard's Lemma and multiply by c_i in the first two terms

[19] Efficiency prices remove the effect of all the heterogeneous shifters that act on preferences via prices.

[20] In time series comparisons of trade frictions, the units choice is imposed on a base year set of world prices. The the other years sellers prices are relative to the base year prices, element by element.

and sum over i and j while holding real incomes u^j constant. The result is:

$$\sum_{i,j} e_i(\mathbf{p}^j)p_{ij}u^j - \sum_i e_i(\mathbf{p}^*)c_i\Pi_i \sum_j u^j - (\sum_i Y_i - \sum_i Y_i^*) = 0. \tag{6}$$

Each component j of the double sum above is operationalized using $b_{ij} = e_i(\mathbf{p}^j)p_{ij}/e(\cdot)$, $u^j = E_j/P_j$ and for simplicity[21] balanced trade, $E_j = Y_j$. $\sum_i(Y_i - Y_i^*) = 0$ by the standard normalization. Using balanced trade, equation (6) is rearranged as

$$\sum_{i,j} b_{ij}(\{p_{ij}/P_j\})Y_j/Y = \sum_i e_i(\mathbf{p}^*)c_i\Pi_i = \sum_i B_i.$$

Each buyer's component j suggests extracting meaning about frictions from the deviation of observed buyer shares b_{ij} from as-if-frictionless observed shares B_i.

1.3 Toward Operationality

The first step toward operationality applies the intermediate value theorem to the sum of differences between the expenditure function evaluated at each location's buyers prices and the expenditure function evaluated at as-if-frictionless prices on the hypothetical world market. Non-parametric gravity is based on the ijth elements of the resulting sum of differences. Operational non-parametric sufficient statistics for efficiency loss from trade and terms of trade are approximations, exact under restrictions explained below. Section 3 reports results based on WIOD manufacturing data.

The intermediate value theorem applied to the expenditure function for each country j [22] assures that for some intermediate relative price vector with components $\tilde{p}_{ij} = \lambda_j c_i \tau_{ij}/P_j +$

[21]Relaxing the balanced trade restriction while preserving a static framework, let $E_i = \phi_i Y_i,\ \forall i$ subject to $\sum_i \phi_i Y_i/Y = 1, \phi_i > 0 \forall i$.

[22]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).

$(1-\lambda_j)c_i\Pi_i;\ \lambda_j \in [0,1],\ \forall i,j$:

$$e(\{c_i\tau_{ij}\})u^j - e(\mathbf{p}^*)u^j = \sum_i b_{ij}(\{\tilde{p}_{ij}\})\frac{\tau_{ij}/P_j - \Pi_i}{\lambda_j\tau_{ij}/P_j + (1-\lambda_j)\Pi_i}u^j\tilde{P}_j,\ \forall j. \tag{7}$$

Equation (7) normalizes the hypothetical 'world market' price index =1. The intermediate price index $\tilde{P}_j = e(\{\tilde{p}_{ij}\})$.

Equation (7) is usefully rewritten in terms of relative resistances rather than relative prices. On the right hand side of (7), divide numerator and denominator by Π_i to yield $(R_{ij}-1)/[\lambda_j R_{ij} + (1-\lambda_j)]$. Here $\tilde{R}_{ij} = \lambda_j R_{ij} + (1-\lambda_j)$ is the intermediate value of relative resistance.

The economic implications of (7) 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 (relative to the as-if-frictionless equilibrium) that is due to frictions. Thus the percentage cost of trade frictions to country j implied by (7) is $P_j - 1$ equal to:

$$\sum_i [b_{ij}P_j - B_i] = \sum_i b_{ij}(\{\tilde{p}_{ij}\})\tilde{P}_j\frac{R_{ij}-1}{[\lambda_j R_{ij} + (1-\lambda_j)}],\ \forall j. \tag{8}$$

The logic of non-parametric approaches to inference from data suggests using observables only to draw implications. Non-parametric gravity as defined here extends this logic to "observable in principle", meaning observable from experiments on the economy. The equilibrium condition $B_i = Y_i/Y$ implies that B_i is observable. On the left hand side, the shares b_{ij} are observed at initial points $b_{ij}(\{p_{ij}\})$ and B_i, while on the right hand side the share is potentially observable at the intermediate point $b_{ij}(\{\tilde{p}_{ij}\})$. Given invertibility (the 'connected substitutes' assumption is sufficient), $\tilde{b}_{ij}$ 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 price index is observed at the initial points P_j

and 1 on the left while potentially observable at the intermediate point $\tilde{P}_j$ on the right. An approximation assumption leads to observability in practice for $\{\tilde{b}_{ij},\ \tilde{P}_j\}$.

The individual elements of the sum on the left hand side of (8), the observable expressions $b_{ij}P_j - Y_i/Y$, are equal to the elements of the sum on the right hand side 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.

This setup suggests characterizing non-parametric gravity with equations for the individual elements as:

Proposition 1

$$b_{ij}(\{p_{ij}\})P_j - Y_i/Y = \tilde{b}_{ij}\tilde{P}_j \frac{R_{ij} - 1}{[\lambda_j R_{ij} + (1 - \lambda_j)]} + \epsilon_{ij},\ \forall i, j. \tag{9}$$

The adding up condition of the expansion implies that $\sum_i \epsilon_{ij} = 0$. Similarly, the market clearing condition (5) implies that $\sum_j \epsilon_{ij} E_j/Y = 0$. Thus the first term on the right hand side of (9) is interpreted as correct 'on average'. ϵ_{ij} may be non-random but this is unknowable. ϵ_{ij} and other error sources are discussed below.

The economic interpretation of (9) is straightforward. For cases $i \neq j$, the left hand side of (9) is typically negative and the deterministic term on the right hand side measures the loss due to relative buyer price being pushed above its as-if-frictionless value. For the case $i = j$, domestic trade, the left hand side of (9) measures the loss due to equilibrium relative resistance pushing relative domestic price below its hypothetical as-if-frictionless value. Rearrange the left hand side for the domestic trade case $i = j$ as $s_j[P_j b_{jj}/s_j - 1]$. $P_j - 1 > 0$ gives the unavoidable cost of frictions per unit of relative utility even if $b_{jj}/s_j = 1$. Elsewhere (as typically is the case for sectoral trade and aggregate trade due to trade imbalances), $b_{jj}/s_j > 1$ raises the proportion of sales diverted into domestic trade, hence it raises the loss measure.

The relative resistance difference term on the right hand side of (9)

$$\frac{R_{ij}-1}{\lambda_j R_{ij}+(1-\lambda_j)}$$

reduces the complex general equilibrium effects of resistance on bilateral trade to an 'own effect' of relative resistance equal to the percentage change in the buyer j's price of good i over its as-if-frictionless value. The ratio

$$(R_{ij}-1)/(\lambda_j R_{ij}+1-\lambda_j)$$

is an appropriate discrete form of the percentage change in R_{ij} implied by hypothetically moving to the observed situation from the as-if-frictionless equilibrium.

Equation (9) has useful implications. A theoretical implication developed in Appendix Section 6.1 applies (9) to non-parametric spatial aggregation. This clarifies the relationship between gravity applications with different spatial units. The main implications developed below are empirical. Section 1.4 solves (9) for R_{ij} to generate 'data' on relative resistances using the Toïnqvist approximation to the demand system, $\lambda_j = 1/2$. The applications reported below focus on $i = j$ case and the rates of change in the loss measure $P_j b_{jj} - s_j$ and the terms of trade R_{jj}. The rate of change focus eliminates the normalization of economic distances.

1.4 Operational Measures

(9) is qualitatively useful as a decomposition, but it is not operational because λ_j depends on the deep structure of equilibrium. The Törnqvist approximation $\lambda_j = 1/2$ achieves operationality. An approximation error η_{ij} is added to the non-parametric error ϵ_{ij} in this case. (The spatial aggregation analysis based on (22) suggests aggregation error is included in ϵ_{IJ}.) The translog demand system structure is a wide subset of non-parametric gravity models for which the Törnqvist approximation to (9) exactly reveals all the non-parametric

information.[23] Then $\epsilon_{ij} = 0$ while $\tilde{b}_{ij} = \bar{b}_{ij} = (b_{ij} + Y_i/Y)/2$ and $\tilde{P}_j = \exp(\ln P_j/2 + \ln 1/2) = \sqrt{P_j}$.[24] 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.

Applying the Toïnqvist approximation to (9) implies $\tilde{P}_j = \sqrt{P_j}$ and $\tilde{b}_{ij} = (b_{ij} + Y_i/Y)/2 \equiv \bar{b}_{ij}$. The result (suppressing the error term ϵ_{ij} is

Proposition 2

$$P_j b_{ij} - Y_i/Y = 2\bar{b}_{ij}\sqrt{P_j}\frac{R_{ij} - 1}{(R_{ij} + 1)}; \ \forall i, j. \tag{10}$$

Equation (10) solves for relative resistances

$$R_{ij} = \frac{2\bar{b}_{ij}\sqrt{P_j} + (P_j b_{ij} - Y_i/Y)}{2\bar{b}_{ij}\sqrt{P_j} - (P_j b_{ij} - Y_i/Y)}; \ \forall i, j. \tag{11}$$

All else equal, equation (11) implies that relative resistance R_{ij} is normally increasing in Y_i/Y for $j \neq i$ and decreasing in Y_i/Y for $j = i$:

$$\frac{\partial R_{ij}}{\partial (Y_i/Y)} = -\frac{1 + R_{ij}}{2\sqrt{P_j}\bar{b}_{ij} - (P_j b_{ij} - Y_i/Y)}.) \tag{12}$$

This intuitive sharp result suggests that in the cross section, larger countries (sellers) have worse terms of trade and face higher relative resistance to their exports.[25] Also, over time faster growing sellers experience worsening terms of trade and rising relative resistance to their exports. General equilibrium effects of course blur this intuition, but it helps explain the results reported below on rates of change in terms of trade and sales shares for China and the US.

Note that when $b_{ij} = 0$, the R_{ij} measure is defined, but its meaning is the lower bound

[23] Note that in the present context of non-parametric gravity, the effects of non-homotheticity are absorbed in price 'parameters', implying a much wider class than the standard general translog specification. Note also that this is the general translog with $N \times (N-1)/2$ substitution parameters and origin-destination-shifters absorbing taste differences and endogenous idiosyncratic trade costs as well as non-homotheticity.

[24] The 'observed' P_j is usually approximated with the Stone price index in the literature.

[25] This is because the denominator of (12) is negative for cross-border trade and positive for domestic trade.

of the range of relative resistances sufficient to choke off trade – the reservation relative resistance. The reservation R^C_{ij}s (C for choke resistance) influence the observed shares and observed P_js and thus indirectly influence the inferred active R_{ij}s, but they play no role in solving for the active relative resistances. In evaluating the effects of *ex post* changes between time t and time $t+1$ that involve extensive margin changes, the difference $R^C_{ij,t} - R_{ij,t+1}$ becomes relevant. Section 1.4 develops operational gains from arbitrage measures that in principle can include extensive margin changes.

Section 3 exploits (10) to obtain non-parametric sufficient statistics for loss of gains from trade relative to as-if-frictionless trade and terms of trade using (11). Section 3.1 considers use of the full set of $\ln R_{ij}$s from (11) to estimate the least inaccurate CES trade elasticity.

2 Gains from Trade and Terms of Trade

For non-parametric gravity, the loss relative to as-if-frictionless trade expression is derived from (9) with $i = j$. Operationalizing with the To̎rnqvist approximation implies equation (11) for $i = j$. The loss from frictions compared to as-if-frictionless equilibrium using the translog closed form solution is given by

$$L_j = [P_j b_{jj} - Y_j/Y] = \bar{b}_{jj}\sqrt{P_j}\frac{R_{jj}-1}{(R_{jj}+1)/2}. \tag{13}$$

On the right hand side of (13), the term $\bar{b}_{jj}\sqrt{P_j} = \bar{b}_{jj}\tilde{P}_j$. In the diagram in Section 1.1, L_j is equal to the shaded loss-to-frictions area as a proportion of E_j.

Ex post changes in loss can be non-parametrically evaluated with the percentage change in loss relative to as-if-frictionless trade:

$$\Delta \ln L_j = \Delta \ln[P_j b_{jj} - Y_i/Y]. \tag{14}$$

The second equation in (13) implies that improvements in the terms of trade $T_j = R_{jj} =$

$1/\Pi_j P_j$ will raise the gains from trade of country j. $\Delta \ln[P_j b_{jj} - Y_j/Y]$ in equation (14) is fully non-parametric, in contrast to log changes in CES expression (16) that require an estimate of the trade elasticity θ.

Note that (14) 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. Note also that the basic logic implies that the formulae incorporate the effects of changes in both the intensive and extensive margins of trade. In the cross-section comparison of terms of trade changes, relative increases in Y_i improve terms of trade, as do decreases in b_{ii}. In contrast, for the case of comparison to as-if-frictionless trade (13) the endowments are constant.

The formal treatment here and the application below both suppress treatment of multiple goods. This is a harmless simplification for looking at a single equilibrium, following the argument at the end of Section 6.1. For a time series comparison where the sectoral composition of trade is shifting, the aggregated approach conceals the effect of the composition shifts.

Log terms of trade inferred from non-parametric gravity for sector k is measured by the log of equation (11) at the sectoral level for the case $i = j$. The multi-sector terms of trade is measured by

$$\ln \bar{R}_{jj} = \sum_k \frac{\bar{b}^k_{jj}}{\sum_k \bar{b}^k_{jj}} \ln R^k_{jj}, \tag{15}$$

where the linear aggregation is justified by the the linearity of equations (9) and (13).

Operational log terms of trade measure equal to the log of (11) for $i = j$ and its multi-sector extension (15) may be the most widely useful result in the paper. Potential applications range far beyond the gravity literature. Standard measures of the terms of trade have well known deficiencies. Price comparison is widely based on unit values and associated measurement error while incomplete coverage for exports is especially salient for the exports of diversified economies.[26] Less obviously but perhaps more importantly prices do not con-

[26]For this reason, terms of trade are not much used except for countries with exports dominated by commodities.

tain unobserved user costs, costs that vary across users and product types. Non-parametric gravity measure (15) uses usually high quality observations on value of production and trade combined with observed P_j data that is subject to the standard problems of price comparison indexes.

2.1 Relationship to CES Gravity

The well known gains from trade relative to autarky sufficient statistic approach of Arkolakis et al. (2012) is based on the insight that under strong conditions two key observables, the observed domestic share b_{jj} and the hypothetical autarky share equal to 1 are sufficient statistics to quantify gains from trade. The conditions are that *changes come from foreign sources only*, the 'true' model is CES and the trade elasticity itself is known. Thus

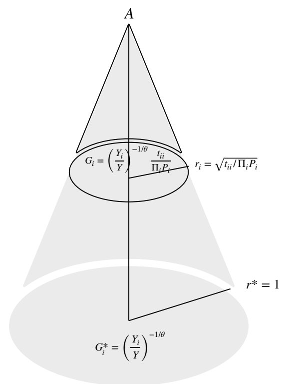
$$G_j = \left(\frac{Y_j}{Y}\right)^{-1/\theta} R_{jj}, \tag{16}$$

where $R_{jj} = 1/(\Pi_j P_j) = T_j$, the terms of trade of country j.

Their CES gains measure falls globally in inverse proportion to the square of economic distance. The measure is also useful for comparison of *ex post* changes when foreign sources of change are dominant. Along with observables the measure requires only one parameter value.

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



The non-parametric loss measure L_i is conceptually distinct from the Arkolakis et al. (2012) measure G_i because the counterfactual alternatives differ (as-if-frictionless trade in the first case and autarky in the second case). More importantly they differ because the non-parametric loss measure allows for changes in domestic frictions and endowments as well as foreign ones. Nevertheless, the loss measure builds on the Arkolakis et al. (2012) insight that the domestic share b_{ii} is a key observable variable that is negatively related to the gains from trade. The two measures are similar in depending on the ratio of observed domestic spending share to a hypothetical alternative. Formally, the loss measure in the CES case is

$$P_i b_{ii} - s_i = P_i \frac{c_i t_{ii}}{P_i}^{-\theta} - (c_i \Pi_i)^{-\theta} = (c_i \Pi_i)^{-\theta} \left[P_i \left(\frac{t_{ii}}{\Pi_i P_i} \right)^{-\theta} - 1 \right]. \tag{17}$$

Relative to the preceding diagram, convert the loss measure to a gains measure with the power transform $-1/\theta$ applied to equation (17). The maximum gain $s_i^{-1/\theta}$ remains the area of the circle at the base of the cone. The area is reduced by increasing economic distance at the rate

$$\left[P_i\left(\frac{t_{ii}}{\Pi_i P_i}\right)^{-\theta}-1\right]^{-1/\theta},$$

rising economic distance no longer reduces gains globally in proportion to the inverse square law.

3 Practical Applications

General gravity (9) in its operational form (10) suggests a number of applications. Four are illustrated below using the WIOD data for manufacturing 2000-2014. Non-parametric measures of gains from trade, terms of trade and economic distance are reported . The fourth application is to infer CES trade elasticity parameters as would be needed for counterfactual exercises. Treatment of final demand and intermediate input demand separately is suspect for familiar reasons, 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 $\sum_j E_j P_j / \sum_j E_j = 1$.[27] 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 Törnqvist 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.

Non-parametric sufficient statistics for percentage changes in gains from trade relative to as-if-frictionless trade and terms of trade are reported below for manufacturing trade of the

[27] 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.

US and China. (The aggregation of sectors to all of manufacturing conceals the effects of compositional change on relative resistances, but the lens of the model still provides a sharp interpretation.) The discrete percentage change in gains is $2(g_j^1 - g_j^0)/(g_j^1 + g_j^0)$ for any years 0 and 1 where equation (13) is applied to calculate g_j in any year. Terms of trade discrete percentage change $2(R_{jj}^1 - R_{jj}^0)/(R_{jj}^1 + R_{jj}^0)$ is is calculated from equation (11) for the case $i = j$.

Demand is interpreted as being the derived demand for intermediate goods. Thus u^j is reinterpreted as the real expenditure in destination j for the set of intermediate goods being purchased, and $e(\cdot)$ is interpreted as the cost function for the intermediate goods. The good produced by each country is identified with the manufacturing sector. *Sectoral* trade is a natural focus for gravity analysis.

The ubiquity of unbalanced trade requires a modification of the gains from trade measure to consistently account for it. A simple procedure is to assume that the ratio of expenditure to income $\phi_j,\ \forall j$ remains constant at its observed base value as the static equilibrium is perturbed.[28] Relative changes in sectoral terms of trade are invariant to the value of ϕ_j with constant ϕ_j.

The application implies that US manufacturing from 2000 to 2014 experienced a 2.07% annual average fall in gains from trade relative to as-if-frictionless trade. This was accompanied by a 5.5% annual average rise in US manufacturing terms of trade. Both are associated with the near halving of the US share of world manufacturing trade while the US domestic share fell only slightly. [See equations (14) and (12) and the discussion following the latter.] China's gains from trade relative to as-if-frictionless trade rose an annual average 1.96%, accompanied by an annual average 8.3% fall in terms of trade. Both are associated with a near quadrupling of China's share of world manufacturing trade while its domestic share rose slightly. The experience of both countries is consistent with global effects of changes in competitiveness, an improvement in China's case and a deterioration in the US case. In per-

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

spective, note that the gains from trade change is a gross benefit or cost. Net benefit or cost adjusts the gross benefit by the cost or benefit associated with the change in competitiveness.

3.1 Parameter Inference and Model Selection

Counterfactual measurement of projected changes in gains from trade and related effects is the object of a large recent literature. Projection requires parametric gravity. The purpose of the projection presumably should condition the specification of the model. The non-parametric perspective suggests that ignorance of the true model should condition parametric implementation. A new approach is illustrated in this paper in an application motivated by the above reported changes in non-parametric gains from trade and terms of trade.

Parameterization is subject to two important sources of error: (i) error in the specification (i.e., model selection error); and (ii) error in the parameter estimate, given the parametric specification. Ignorance of the true model [problem (i)] suggests that fidelity to the information in relative resistance statistics (11) should guide parametric inference. The purpose of the counterfactual exercise should also condition model choice, since [problem (iii)] context-specific unimportance may be assigned to some errors.

A new approach to parameter estimation [problem (ii)] for the CES case is developed and applied below. Issues with model selection [problem (i)] that arise even in the CES case are discussed. The application is to a counterfactual back-of-the-envelope evaluation of US industrial policy that counterfactually raises US world market share of sales. The application and discussion reveals puzzles and challenges for future research relating to model selection [problem (i)] and error toleration [problem (iii)].

Start with problem (ii), parameter estimate error in the CES case. The standard econometric approach assumes the CES specification is true. The goal is an unbiased trade elasticity θ that best fits the bilateral trade data in the stochastic version of equation (18)[29]. Bias avoidance requires procedures designed to satisfy independence of regressors from the

[29] Simonovska and Waugh (2014) recommend use of trade elasticities inferred from fitting the assumed CES gravity model to the bilateral trade data to be used in the counterfactual, a practice followed here.

error term. The non-parametric approach to CES parameterization suggests inference of the trade elasticity based on fitting the log of non-parametric changes in relative resistance R_{ij} given by the log of equation (11) to the right hand side of CES equation (19). The non-parametric approach assumes the CES specification is false and obtains the least inaccurate trade elasticity for counter-factuals based on the non-parametric sufficient statistics on relative resistance $\ln R_{ij}$.

The key difference in assumptions, whether the CES specification is true or false, is shown below to imply different methods. A practically important difference in application is that the econometric best fit trade elasticity is identified off variation in tariffs or other directly observed trade costs while the non-parametric approach fits the trade elasticity to the (much larger and potentially more informative) variation in non-parametric relative resistance statistics $\ln R_{ij}$.

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

$$b_{ij} = \frac{Y_i}{Y}(\tau_{ij}/\Pi_i P_j)^{-\theta} = \frac{Y_i}{Y}(R_{ij})^{-\theta}. \tag{18}$$

The relationship of (18) to (11) is given by first rearranging (18) to isolate R_{ij} on the left hand side:

$$R_{ij}^{CES} = \left(\frac{b_{ij}}{Y_i/Y}\right)^{-1/\theta}$$

and then taking logs. The result is

$$\ln R_{ij}^{CES} = -(1/\theta)[\ln b_{ij} - \ln s_i]. \tag{19}$$

Here $s_j = Y_j/Y$ is used for notational ease. The right hand side of (19) uses the CES functional form to explain the level of R_{ij} by movement of b_{ij} away from as-if-frictionless $s_i = Y_i/Y$.

In the non-parametric case, $\ln R_{ij}$ is given by the log of (11). The movement of $\ln R_{ij}$ is explained by movement of b_{ij} away from as-if-frictionless Y_i/Y while using information about P_j and its non-parametric movement away from as-if-frictionless equilibrium. Turning to the non-parametric approach to parameter inference, the least inaccurate fit CES trade elasticity (inverse) to non-parametric relative resistance minimizes the sum of squared residuals η_{ij}^2 from the cross-section 'regression' equation:

$$\ln R_{ij} = (-1/\theta)[\ln(b_{ij} - \ln s_i] + \ln \eta_{ij}. \tag{20}$$

Here $\ln \eta_{ij}$ represents the effect of specification error as well as measurement error.

From the econometric perspective, regression (20) yields a biased estimate of the trade elasticity because the error term $\ln \eta_{ij}$ cannot be orthogonal to the regressor $\ln(b_{ij}/s_i)$. Endogeneity bias is guaranteed because b_{ij} and s_i both determine R_{ij} given by (**??**) and appear on the right hand side of (20). In contrast, the non-parametric approach assumes that the CES is false but the least inaccurate CES elasticity to fit the observed variation in $\ln R_{ij}$ is desired. Endogeneity bias from the econometric perspective is a feature, not a bug when viewed from the non-parametric perspective.

Similarly, the econometric perspective suggests that regression (20) typically yields trade elasticities subject to selection bias because typically some bilateral trade shares $b_{ij} = 0$.[30] The non-parametric perspective suggests that selection bias is again a feature, not a bug. The reasoning here brings in the issue of model selection. Note first that (10) gives a reservation value R_{ij} when $b_{ij} = 0$. The reservation values are uninformative for fitting the variation of positive trade flows in any specification, CES or not. From the non-parametric perspective, zeros should be dropped from the CES 'regression' (20).

The non-parametric approach generally comes at the cost of inability to make probability statements about the results. The minimum distance technique only permits statistical inference when the residuals equal to $\ln \eta_{ij,t}$ evaluated at $\hat{\theta}$ are random. Even with standard

[30] In the WIOD aggregate manufacturing data, there are no zeros.

statistical inference not applicable,[31] the minimum distance method provides an informative percentage of explained variation as context for evaluating counterfactual projections. Looking toward standard inference, measurement error affects the variables on both sides of equation (20). Given knowledge of the measurement error structure, it might be possible to improve on both the efficiency and measurement error bias of the minimum distance estimator.

Equation (20) is usefully extended to a panel setting, adding the time subscript t. The minimum distance CES elasticity estimated from panel data solves

$$\min_{\theta} \sum_{i,j,t} \ln \eta_{ij,t}^2. \tag{21}$$

An application example based on the terms of trade results for the US and China reported above yields a tightly estimated θ equal to 1.01 with standard deviation 0.01 in the US subsample, and 1.06 with standard deviation 0.01 in the China sub-sample. The adjusted R^2 is .93 in both cases. The elasticity θ is estimated off time variation, suggesting it should be interpreted as a short run elasticity. The very small time variation of calibrated θs suggests it may be close to a long run elasticity, meaning it is somewhat low (elasticity of substitution equal to 2) relative to other methods.

The minimum distance trade elasticity estimate is applied to a counterfactual analysis of industrial policy by the US. One policy would act against deindustrialization by exogenously increasing its manufacturing share at the cost of a terms of trade deterioration. Another policy would reduce or eliminate its annual fall in gains from trade a the cost of deteriorating terms of trade. The objective is to quantify how costly are such policies? A back-of-the-envelope approach is to quantify a local elasticity of R_{jj} with respect to Y_j/Y, where j is

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

the US for a given year. The elasticity is based on equation (12). Thus

$$\frac{d\ln R_{jj}}{d\ln s_j} = -\frac{1+R_{jj}}{R_{jj}} \frac{s_j}{2\sqrt{P_j b_{jj}} - (P_j b_{jj} - s_j)}.$$

(Properly doing the job requires a full general equilibrium approach that is far beyond the aim of this paper.)

The US 2014 terms of trade elasticity with respect to the share is equal to -0.6. Using the US 2014 terms of trade elasticity implies that a 10% rise in US manufacturing share (from 12.5%, a 1.25% addition to 13.75%) induces a 6% fall in the US terms of trade. Another application focuses on the US share change relative to other forces. The US manufacturing share in world sales declines over the period 2000-2014 at a 4.8% annual exponential rate (from 0.234 to 0.125). The 'own effect' of this fall on the rise in US terms of trade is 2.8%, about half of the 5.5% rise in the estimated results. Much of remaining rise can be attributed to the spectacular rise in China's share, along with all the other exogenous changes and their general equilibrium consequences. Applying (12) to calculate China's 2014 terms of trade elasticity with respect to its share reveals elasticity equal to -0.67, so a 10% rise in its share (from 31.9% to 35%) induces a 6.7% fall in its terms of trade. This improves the terms of trade of an average of its trade partners.

Quantification of the effect of changes in seller shares on the gains from trade measure uses the CES gravity structure. The CES gravity equation implies $b_{jj} = R_{jj}^{-\theta}$. The CES loss measure is $L_j^{CES} = P_j R_{jj}^{-\theta} - s_j$. Log differentiate L_j with respect to the share s_j using the terms of trade elasticity:

$$\frac{d\ln L_j^{CES}}{d\ln s_j} = -1 - \theta P_j R_{jj}^{-\theta} \frac{d\ln R_{jj}}{d\ln s_j}.$$

Round to $\theta = 1$ and use $d\ln R_{jj}/d\ln s_j = -0.6$ for the US 2014 manufacturing sector. Plugging in those values on the right hand side yields a value of -0.4 on the left hand side. Thus a 1% rise in US manufacturing share induces a 0.4% rise in the gains from trade (fall in the efficiency loss). Interpreted as the shift in the right hand side of equation (13), more is gained on increased volume of exports as $\bar{b}_{jj}$ falls than is lost by the fall in terms of trade. The terms of trade deterioration and its relative income effect is a social cost of the industrial policy. The assumed efficiency of supply choice implies that the crude exercise abstracts from any added social cost of the rise in supply. It likewise makes no assessment of possible social benefits (redistributive or inefficiency-reducing).

Counterfactual industrial policy exercises depend on the trade elasticity θ since the terms of trade effect is directly proportional to θ. At a deeper level the quantification depends on the specification of the parametric demand model. Sticking to the CES model, the tight fit of the initial trade elasticity example is a bit surprising in light of the very large changes in world manufacturing trade shares of China and the US, 2000-2014. It may encourage use of CES structures for developed countries trade with highly aggregated trade partners, as in the back-of-the-envelope exercise. A restriction to time series variation of domestic expenditures shares is justified in the wider non-parametric context because all third party relationships are aggregated into the single relative resistance for domestic relative to global trade.

In sharp contrast, extension of estimator (21) to fit the entire bilateral trade panel (44 times 44 countries over 15 years) reveals a specification problem [(i)]. More than 20% of calculated $R_{ij,t}$s are negative, with numerous examples for almost all exporting countries and years. For purposes of the preceding back-of-envelope evaluation of US industrial policy, it is plausible [problem (iii)] to regard these errors in bilateral trade as unimportant to the basic approach to the problem.

Aggregation bias is the most obvious explanation for the negative R_{ij}s that is consistent with the non-parametric gravity model being valid. Manufacturing is a highly aggregated set of sectors with very large compositional differences across countries. The composition differences become relatively unimportant when the purpose of the exercise depends on explaining variation in the exchange of one country's aggregate manufacturing with the world's aggregate manufacturing.

In contrast, the purpose of the exercise could be a full general equilibrium evaluation of industrial policy. In that case the entire network of multilateral trade needs to be accurately modeled, the nonsensical (in economic terms) negative R_{ij} reject the aggregation of data. Disaggregation into multiple manufacturing sectors could reduce the proportion of negative $R_{ij,t}$s, and justify sectorally differing trade elasticities in a nested CES specification. The implied exercise includes action on the extensive margins of trade (new destinations for existing products) and production (new products).

For non-nested specifications, further research is needed to answer questions raised in this paper. Two model selection issues are suggested in the simple CES context alone. First, the CES manufacturing sub-sectors must be connected with an upper level CES structure. Beyond reducing or eliminating negative non-parametric R_{ij}s, what is the most appropriate nested CES specification? Second, the problem of zeros and dropped data points suggests an alternative specification that combines the simple CES model with a model that treats selection into trade. The standard econometric response to selection bias is just this.[32] From the non-parametric perspective, consideration of the purpose of the counterfactual exercise may often suggest not treating selection. A typical purpose is to evaluate the aggregate welfare effects of trade friction changes. Since the effects of friction changes on welfare are proportional to the share of trade in national production or expenditure, more accurate treatment of the variation of small trade flows above zero is relatively unimportant. The least inaccurate trade elasticity to fit $\ln R_{ij}$ for positive trade observations to the simple CES model (20) may dominate (for purposes of accuracy of counterfactual welfare effects) the least inaccurate CES trade elasticity conditional on a selection equation that is also presumed false. When is this the right choice?

[32]Fixed costs of export can explain zeros in the CES model. Subject to finding an instrument for fixed cost that plausibly does not affect variable cost, a standard procedure can be applied that removes selection bias from the trade elasticity, Helpman et al. (2008).

4 Supply Side Extensions and Specialization

The supply side of the world economy prominently includes the demand for inputs. The cost function aggregated across sectors and countries formally resembles the aggregation of household expenditure functions across households and countries in Section 1.3.[33]

The specialization gains from trade are based on the reallocation of outputs. Specialization gains 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.

A theoretical obstacle is that the connected substitutes property of Berry et al. (2013)) must be taken to apply to the maximum value profit or GDP functions. For profit functions applied to sectors, connected substitutes is no more restrictive than it is for buyer expenditure functions. For GDP functions, in contrast, connected substitutes 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

[33]A practical obstacle to a detailed parallel treatment of final and intermediate inputs is the well-known dubious quality of data on imported intermediate inputs. First, the division of imports into final and intermediate uses is rather arbitrary. Second, 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. Most of the measurement error issues raised by this can be considered as absorbed in the various shifters assumed active in the aggregate manufacturing demand systems assumed in the application Section **??**.

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.

5 Conclusion

Economic gravity describes the static equilibrium of bilateral trade between N^2 pairs of regions where $N > 2$ is an integer. The attractive force is profit maximizing arbitrage drawn by the gains from trade between locationally separated supplies and demands by 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.

Arbitrage equilibrium reduces the distribution of goods to a set of two body relationships that 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 cleanly characterizes 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.

Mild restrictions are provided under which non-parametric sufficient statistics for gains from trade and terms of trade are calculated using the manufacturing trade data of the WIOD, 2000-2014. Non-parametric sufficient statistics for relative resistances for bilateral pairs are also derived.

Counterfactual calculations require parametric representations of gravity. The non-parametric relative resistances of this paper are the basis for estimating least inaccurate fit trade elasticity parameters for use in the projections. These differ in method and magnitude from standard econometric estimates.

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.

Agnosteva, Delina E., James E. Anderson, and Yoto V. Yotov, "Intra-national Trade Costs: Assaying Regional Frictions," *European Economic Review*, 2019, *112*, 32–50.

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 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.

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.

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein, "Trading Partners and Trading Volumes," *Quarterly Journal of Economics*, 2008, *123* (2), 441–487.

Mayer, Thierry and Keith Head, "Illusory Border Effects: Distance Mismeasurement Inflates Estimates of Home Bias in Trade," 2002, (2002-01).

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

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

6 Appendix

6.1 SpatialAggregation

Non-parametric gravity equation (9) provides a useful interpretation of the relationship between gravity applications across many varieties of spatial aggregation. In practice, gravity is widely used for trade between cities, regions and countries and sometimes commuting zones. How may we understand relative resistances based on views at varying focal lengths?

Aggregation of locations necessarily implies spatial aggregation of frictions. Mayer and Head (2002) address the aggregation of frictions related to distance. Their solution in the CES gravity context uses city-pair distance aggregation with population weights. Population weights proxy economic mass weights with the useful virtue of plausible exogeneity to contemporaneous trade flows. Aggregation of frictions between city pairs not related to distance and not uniformly associated with international borders are untreated in the existing literature.

The general non-parametric logic of spatial aggregation of frictions is nested within the logic of (9). Define the primary set S of the granular locations as origins $i \in S$ and destinations $j \in S$, with aggregation into distinct subsets $i \in I$ and $j \in J$. Linear aggregation of (9) describes the aggregate relationship between aggregate origin I and aggregate destination J. First add over $i \in I$ to give aggregate location I's relation to granular locations $j \in J$:

$$P_j b_{Ij} - Y_i/Y = \tilde{b}_{Ij} \sum_i \frac{\tilde{b}_{ij}}{\tilde{b}_{Ij}} \frac{R_{ij} - 1}{\lambda_j R_{ij} + 1 - \lambda_j},$$

where $b_{Ij} \equiv \sum_{i \in I} b_{ij}$ and similarly for $\tilde{b}_{Ij}$. Then add the result above over $j \in J$ to give:

$$b_{IJ} \sum_{j \in J} \frac{b_{Ij}}{b_{IJ}} P_j - Y_I/Y = \tilde{b}_{IJ} \sum_{j \in J} \frac{\tilde{b}_{Ij}}{\tilde{b}_{IJ}} \tilde{P}_j \tilde{b}_{Ij} \sum_{i \in I} \frac{\tilde{b}_{ij}}{\tilde{b}_{Ij}} \frac{R_{ij} - 1}{\lambda_j R_{ij} + 1 - \lambda_j}. \tag{22}$$

The double sum on the right hand side of (22) is interpreted as the weighted average of the effect of the granular relative resistances on observable bilateral trade between I and J,

$$\tilde{b}_{IJ}\tilde{P}_J\frac{R_{IJ}-1}{\lambda_J R_{IJ}+1-\lambda_J}.$$

This interpretation is approximately consistent (i.e. consistent linear aggregation is approached) under conditions given below in Section 1.4.

All the linear aggregation analysis above applies straightforwardly to aggregation across goods. In contrast to spatial aggregation, trade flow data is sufficient to permit disaggregated non-parametric gravity measurement.

6.2 Intermediate Goods

The methods of the text extend easily to multiple distinct sectors (products), as noted above for final goods. In particular, Propositions 1 and 2 hold for demand systems for any number of products k purchased from any number of origins i. Specifically, Proposition 2 becomes:

$$P_j b_{ij}^k - Y_i^k/Y = -\bar{b}_{ij}^k(P_j+1)\frac{R_{ij}^k-1}{R_{ij}^k+1},\ \forall i,j,k.$$

Note that P_j is the price index in j for all goods prices $\{p_{ij}^k\}$. Note also that the equation allows for $Y_i^k=0$ for some sectors k, reflecting specialization as in the real world.

For intermediate input demand systems, the formal equivalent of the expenditure function is the cost function of the using producers, with the utility of the buyer replaced by the real output of the users. Allowing for different real outputs y_i^k, the cost function for sector k purchasing inputs from sectors l in origins o is $C(\{p_{oi}^l\})y_i^k$ and the aggregate cost for origin i is $C_i=\sum_k C(\{p_{oi}^l\})y_i^k$. In the cross section the real outputs y_i^k are given. The aggregate input demand system by i for inputs l from origins i is given by Shephard's Lemma. Market clearing conditions and adding up conditions apply as with final goods, leading to equivalent expressions to Proposition 2:

$$C_i b^l_{oi} - Y^l_o/Y = -\bar{b}^l_{oi}(C_i + 1)\frac{R^l_{oi} - 1}{R^l_{oi} + 1},\ \forall o, i, l.$$

In practice, sectoral trade data is sorted into intermediate and final use with rather arbitrary and often implausible assumptions due to lack of information. Most gravity applications thus lump them together for purposes of estimation and projection.

6.3 GDP Function Approach

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 $\mathbf{p}^j_y$ and input buyer prices $\mathbf{p}^j_m$. 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.[34] Appendix section 6.3 has more details.

Panel data changes in non-parametric specialization gains from trade and terms of trade measures analogous to (21) 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 (21) is implausibly restrictive, since it is associated with specific factors and mobile labor allocation based on identical Cobb-Douglas production functions (Anderson (2011b)).

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

[34] 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.

$\mathbf{y}^i, \mathbf{m}^i, \mathbf{v}^i \in \mathcal{T}(\mathbf{y}^i, \mathbf{m}^i, \mathbf{v}^i)$ 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 $R^W = \sum_i R^i = \sum_i r(\mathbf{p}_y^i, \mathbf{p}_m^i)f(\mathbf{v}^i)$.