Commercial Rivalry as Seller Incidence Shifting: Non-parametric Accounting of the China Shock*

James E. Anderson
Boston College and NBER

October 18, 2024

Abstract

Intense US-China commercial rivalry is quantified in this paper with novel non-parametric relative resistance sufficient statistics. China's manufacturing seller incidence falls (seller price rises) 8.2% yearly as China's sales share quadruples over 2000-14. US seller incidence rises 6.3% yearly as US sales share halves. A 10% rise in US (China) 2014 sales share reduces seller incidence 10.05% (9.74%) and raises average seller incidence of others. Trade elasticities very close to one fit trade shares to revealed relative resistances. Trade elasticities identified off variation in observable buyer prices or trade costs are biased upward by omitted variation in unobservable buyer frictions.

JEL codes: F10, F14

Keywords: Non-parametric, seller incidence, terms of trade.

*I am grateful to Costas Arkolakis and various referees for great improvements in the paper. Mingyang Wei made helpful comments on the final draft. Danial Lashkari and Arthur Lewbel provided helpful discussions and comments on a first draft. I thank participants at the NBER 2021 Winter ITI and Gravity at Sixty 2022 meetings for helpful comments and encouragement. Arnaud Costinot, Yoto Yotov and other attendees pushed for development and clarification. Previous versions circulated as "Non-parametric Gravity", NBER Working Paper No. 30807, and "Trade Scale Gains with Costly Trade: A Non-parametric Accounting".

Seller incidence shifting is a big amplifier of international commercial rivalry, quantified here for China and the US in manufacturing. A novel non-parametric gravity model is the basis of the accounting. Revealed trade frictions are bigger than those based on previous parametric methods, and shifts in their incidence matter much more.[1] Seller incidence is defined here in the context of equilibrium arbitrage as the equilibrium weighted average proportion of outward bilateral trade frictions to all destinations borne by the seller. (Symmetrically, buyer incidence is the average proportion of inward bilateral frictions from all origins borne by the buyer.) Seller incidence shifting is due to asymmetric growth of national sales that drives reduction in the faster growing seller's incidence of trade frictions. Seller incidence of its lagging rivals rises on average. Net seller prices in world markets are inversely proportional to the seller incidence of trade frictions, so the large seller incidence shifts in manufacturing reported in this paper matter big-time. Sufficient statistics for seller incidence and related relative trade frictions are freed from dependence on restrictive parametric specifications and their estimated parameters.

Figures 1 and 2 illustrate intense commercial rivalry in manufacturing between China and the US, 2000-2014. China's share of world manufacturing sales quadruples while the US share is halved. The seller incidences reported in the Figures are generated from the non-parametric gravity accounting model developed below. The close correlation of revealed inverse seller incidence with sales shares implies sales share shifts account for a yearly average fall in China's seller incidence of -8.2% and a yearly average rise in US seller incidence of 6.3%. The empirical association of seller incidence with trade shares is analytically derived in the non-parametric gravity accounting model developed below.

[1] Subsequent literature has mostly neglected the report of large inter-temporal seller incidence shifting in the inter-regional trade between US states and Canadian provinces by Anderson and Yotov (2010). The methodological differences are more important. The Anderson and Yotov (2010) paper applies parametric constant elasticity gravity. The non-parametric approach applied here frees the implied size of seller incidence changes from dependence on the constant elasticity specification and the validity of its parameter estimate.

Figure 1: China's Rise

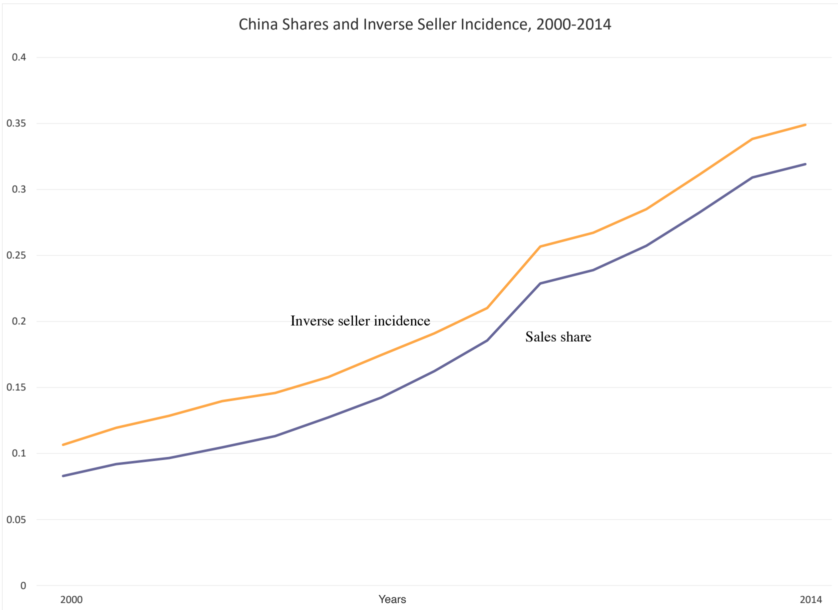
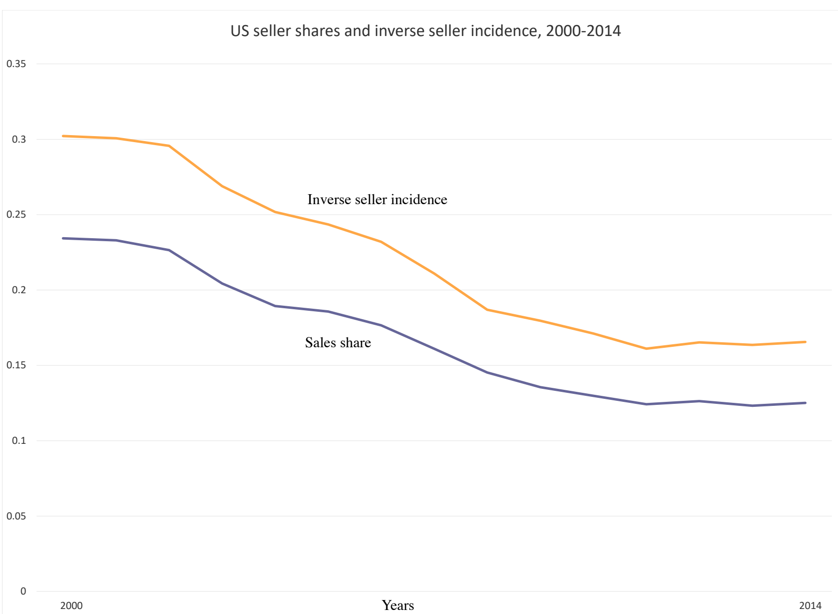


Figure 2: US Decline



The Figures and the analysis behind them counter a common naive opinion among non-economists that much of the China shock impact on US manufacturing could have been avoided by protectionist trade policy. The facts behind the figures sharpen this conclusion. (i) Much of China's manufacturing growth went to domestic sales, since China's domestic manufacturing sales share rises 2000-2014 in the World Input Output Database (WIOD) . (ii) Faster growth by China automatically implies negative share effects for the rest of the world, since shares necessarily sum to one. (iii) The impact effect of the share shifts raises other countries sellers' incidence while China's sellers' incidence falls. (iv) Thus the US faces tougher competition from China in all third party markets while bilateral tariff increases on China's trade with the US affect only part of US imports. Direct US-China tariffs act on a relatively unimportant margin in this context. The lens of the model on these facts suggests that offsetting trade policy by the US on China's exports sufficient to eliminate the yearly -6.3% fall in seller price (implied by the 6.3% rise in seller's incidence) would have been very costly if not infeasible.

Analytic cross-section elasticities of seller incidence with respect to sales shares are derived from the non-parametric model. In 2014 these incidence elasticities are -1.04 for the US and -0.97 for China. The elasticities quantify the intensity of commercial rivalry through seller incidence shifting. Looking over time, the incidence elasticities quantify the impact part of the causal link suggested by the Figures. The cross-section elasticities may be taken to quantify the first order impact effects of counterfactual sales share shifts on seller incidence. Negative incidence elasticities suggest a role for industrial policy to internalize the effect of industry scale on the incidence of the cost of distribution of products. The effect is larger for big sellers, since the analytic elasticities rise in absolute value with sales shares.

Seller incidence shifting resembles external economies of scale in distribution, but the mechanism is fundamentally different. In contrast to scale economies on a single distribution link, general spatial equilibrium implies external scale effects due to resulting shifts in the distribution of sales. A rise in a seller's sales share of world sales will raise its proportion of domestic sales that face relatively lower frictions. This reduces its overall seller incidence, all else equal. Incidence shifting operates even with constant or increasing bilateral trade costs on every link. From this perspective seller incidence shifting is a more pervasive phenomenon than external economies of scale. In the application below, external scale economies may be present, but cancel out in the relative resistances that are the focus of the paper. Thus seller incidence shifting in the distribution of the vector of given supplies is independent of external scale economies and their relationship to cost for inference, projection and policy analysis purposes. The external scale effects in distribution thus complement the scale effects in production that are the focus of Bartelme et al. (2019).

A full treatment of industrial policy is beyond the scope of this paper but a non-parametric basis for sectoral policy evaluation is the compensating variation loss measure of the national interest. The loss measure in this paper is based on the difference between the observed domestic share of sales and the hypothetical domestic share that would obtain in an as-if-frictionless equilibrium. The as-if-frictionless share is observable as the country's share of world manufacturing sales at buyer prices, equal to every destination's expenditure share on the country's goods when the effect of frictions on distribution is removed. The national interest directly moves proportionally with the US terms of trade in manufacturing[2] while indirectly a terms of trade improvement reduces the domestic demand share and thus reduces the distance between the domestic and as-if-frictionless shares. The average yearly changes in the negative of the loss measure (the gains from trade measure) are 1.9% for China and −3.8% for the US.

The sectoral loss measure changes are primarily due to terms of trade changes since the domestic trade shares have relatively small variation. China's terms of trade in manufacturing improve by an average yearly 8.3% while US terms of trade deteriorate by an average

[2]The terms of trade in the presence of trade frictions is defined as the buyer price of domestic products divided by the price index, equal to the utility gain per unit of domestic sales that is reduced to allow efficient reallocation of expenditure to all goods. The sectoral terms of trade here are a part of the economy-wide terms of trade that are the more familiar focus of international trade analysis. The concepts are the same for one good exchange economies.

yearly -5.5%. Seller incidence variation accounts for much of the terms of trade variation for both countries since the sectoral terms of trade is inversely proportional to seller incidence.[3]

The loss measure is related in Appendix Section 8.4 to the well-known gains from trade measure in the Constant Elasticity of Substitution (CES) case, Arkolakis et al. (2012). The CES gains measure is based on the observed domestic expenditure share relative to its hypothetical autarky value equal to one. An equivalent variation real income measure of the gains from trade is given by a power transform of the domestic share where the exponent is the negative inverse of the trade elasticity. For evaluating *ex post* changes, Arkolakis et al. (2012) note that their measure is valid for *external changes only*. In this case, once the loss measure is changed from a difference to the comparable relative form, the two measures are equal provided that trade is balanced (at the sectoral level), changes are foreign only and preferences are CES.

Minimum distance calibration of the CES trade elasticity to fit the variation of log trade shares to the variation of domestic log relative resistances yields a trade elasticity very close to 1. This is significantly lower than previous estimates in the gravity literature. For example, a representative trade elasticity [Simonovska and Waugh (2014)] is 4. The lower elasticity suggests that previous measures of resistance and thus measures of gains from trade are too low, mostly much too low.

The lower trade elasticity further suggests that previous trade elasticity estimators are biased upward (in absolute value). Section 6 develops a structural explanation – omitted variable bias. Observable bilateral prices or trade costs vary inversely to the equilibrium bilateral incidence of unobservable bilateral non-pecuniary costs or tastes.[4] Thus larger non-pecuniary cost implies lower observed price, and the inferred elasticity must be larger to

[3]The result that sales expansion of a large exporter improves its terms of trade conflicts with standard intuition based on the immiserizing growth literature that assumed frictionless trade. The intuition for the contrary incidence shifting effect when trade is subject to frictions is explained below in Appendix Section 8.2.

[4]Non-parametric relative resistances are portmanteau residuals that implicitly aggregate across heterogeneous tastes and markups as well as non-price frictions such as delay and uncertainty. Less obviously, relative resistances aggregate across heterogeneous cross effects in demand, products (as in the manufactures application) and locations (as the national markets that aggregate local destinations).

explain the observed variation in expenditure. Revealed relative resistance has much larger variation than observed price or trade cost variation and Section 6 argues that potential endogeneity bias is much attenuated.

Beyond trade elasticity estimation, the omitted variables contained in revealed relative resistances are a rich resource for future investigations. The analogy with Solow productivity residuals and the productivity literature is clear.

Non-parametric gravity as defined here is related to a recent literature extending gravity via non-parametric steps toward more general parametric approximation of demand and supply structures. Closest in spirit is the Adão et al. (2017) non-parametric approach to reduced form spatial equilibrium exchange model where trade is in embodied factors. Both papers assume invertibility of the demand system.

This paper focuses on spatial equilibrium distribution of given sectoral supplies to multiple destinations. The sectoral focus makes the endowments approach to static arbitrage equilibrium natural, but also conveniently avoids the challenge of modeling endogenous supply and demand at the same time. In contrast, the parametric gravity literature extends to a set of one factor production models that are observationally equivalent in gravity equilibrium to the Constant Elasticity of Substitution (CES) endowments model, Arkolakis et al. (2012). Similar to Arkolakis et al. (2012), gains from trade measures in the non-parametric model are based on the observable domestic expenditure share relative to an observable benchmark – autarky in their parametric CES case and domestic sales share of world sales in the non-parametric case. See also Ravikumar and Waugh (2016), who obtain a gains from trade measure in the CES case from the ratio of domestic expenditure share to domestic sales share.

Section 1 is a brief review of the CES gravity model approach to relative resistance measures. Unappreciated properties of efficient spatial arbitrage provide useful intuition here and subsequently in the non-parametric model in Section 3. Section 2 first provides a non-technical perspective on the non-parametric gravity approach. The formal development in

Section 3 derives relative resistances as determined by arbitrage equilibrium in a wide class of invertible demand systems. A specification from this class is required for operationality. The Törnqvist approximation to expenditure shares, the arithmetic average of observed and as-if-frictionless shares, minimizes the approximation error due to deviation of the associate specification from the unknown 'true' specification. Revealed relative resistance statistics under the Törnqvist approximation for China and US manufacturing are reported and discussed in Section 4, followed by discussion of commercial rivalry in Section 5. Implications for trade elasticity estimation are discussed in Section 6.

1 From CES to Non-parametric Gravity

A brief review of CES gravity is a useful starting point. The buyer's effective price p_{ij} in destination j for the good from origin i in arbitrage equilibrium is equal to the product of the net seller price c_i and the friction τ_{ij}. τ_{ij} is a product of trade frictions, unobservable buyer costs and taste shifters, all of which are origin-destination specific. The shipment of quantity x_{ij} from origin i to destination j combines with prices to give demand shares. The demand (expenditure or cost) share $b_{ij} = p_{ij}x_{ij}/\sum_i p_{ij}x_{ij}$ in the CES specification is

$$b_{ij} = \left(\frac{p_{ij}}{P_j}\right)^{-\theta},\ \theta > 0;$$

where the CES price index $P_j = [\sum_i p_{ij}^{-\theta}]^{-1/\theta}$ is solved from the budget constraint $\sum_i b_{ij} = 1$. The same preferences (inclusive of the taste shifters absorbed into effective prices p_{ij}) apply to all destinations.

In arbitrage equilibrium, the supply y_i of goods from each country is distributed with total sales at buyer prices $c_i\Pi_i y_i,\ \forall i$. Here, c_i (unit cost in a simple competitive model) is received by the seller. $\Pi_i = \sum_j (p_{ij}/c_i)x_{ij}/y_i$ is the average friction factor facing seller i, its equilibrium seller incidence of trade frictions, or its outward multilateral resistance. (Appendix 8.1 derives Π_i as a property of efficient spatial arbitrage.) Country i's world sales share is equal to $s_i = c_i\Pi_i y_i / \sum_l c_l \Pi_l y_l$.

Sellers effectively face a world buyer on an as-if-frictionless world market with expenditure shares B_i. In the CES case

$$B_i = (c_i\Pi_i)^{-\theta}$$

where the world price index $[\sum_i (c_i\Pi_i)^{-\theta}]^{-1/\theta} = 1$ implicitly deflates $c_i\Pi_i$. The world budget constraint $\sum_i B_i = 1$ solves for the world price index. With no nominal rigidities in the arbitrage equilibrium, relative prices alone matter. The vector of world prices $\{c_i\Pi_i\}$ is conveniently normalized with $\sum_i c_i\Pi_i y_i / \sum_i y_i = 1$. Thus the as-if-frictionless world price index is equal to one.[5]

The CES gravity model is based on arbitrage equilibrium where world sales share s_i is equal to world buyer purchases share B_i. The bilateral expenditure share b_{ij} relative to its as-if-frictionless share value is given by the reduced form equation

$$\frac{b_{ij}}{B_i} = \left(\frac{p_{ij}/c_i}{\Pi_i P_j}\right)^{-\theta} = \left(\frac{\tau_{ij}}{\Pi_i P_j}\right)^{-\theta}. \qquad (1)$$

The general equilibrium effects of trade frictions are reduced to a power transform of the bilateral relative resistance ratio $\tau_{ij}/\Pi_i P_j = R_{ij}$.

Two properties of arbitrage equilibrium lead to the simple expression on the right hand side of equation (1). First, arbitrage implies that in equilibrium $p_{ij} = c_i\tau_{ij},\ \forall i, j$. Second, less obviously, $c_i\Pi_i$ is the opportunity cost of shifting a unit of i from the world market to some destination j. Willingness-to-pay p_{ij} must cover the extra cost of getting to j, resulting in equilibrium bilateral buyer's incidence τ_{ij}/Π_i. Appendix Section 8.1 shows that the opportunity cost $c_i\Pi_i$ is the Lagrange multiplier associated with the set of market clearing constraints on efficient distribution.

[5] The as-if-frictionless equilibrium is distinct from the counterfactual general equilibrium where all buyers face the same effective price vector. The counterfactual must account for the changes in origin incomes and their effect on each destination's expenditure.

The budget constraint for the world economy implies

$$\sum_j E_j = \sum_i c_i \Pi_i y_i.$$

The normalization of prices applied to the right hand side yields $\sum_i c_i \Pi_i y_i / \sum_i y_i = 1$. The CES expenditure function on the left hand side is $E_j = P_j u^j$ where u^j is the utility of the representative agent. Consistency implies that the CES price indexes must be normalized consistently with the as-if-equilibrium prices

$$\sum_j P_j u^j / \sum_j u^j = 1 = \sum_i c_i \Pi_i y_i / \sum_i y_i.$$

Comparability of relative resistances across countries is assured.

Relative resistance is solved from equation (1) as $R_{ij} = (b_{ij}/s_i)^{-1/\theta}$. The quantitative solution for R_{ij} requires trust in the CES specification and trust in the estimate of the trade elasticity parameter θ. Moreover, the reduction of spatial arbitrage implications to relative resistance $R_{ij} = \tau_{ij}/\Pi_i P_j$ appears to depend on functional form assumptions. Qualms about the restrictiveness of the CES specification and doubts about the accuracy of θ estimates motivate the non-parametric approach.

2 Non-parametric Gravity in Perspective

Gravity models of trade assume that (i) efficient arbitrage governs distribution of supplies where (ii) willingness to pay for goods from all sources is derived from invertible demand systems applicable to all destinations. Parametric gravity adds restrictive parametric demand system specifications. Non-parametric gravity relaxes the third restriction to allow a wide class of invertible demand systems. It delivers relative resistances that aggregate all third party frictions that affect bilateral trade directly and indirectly via multilateral resistances. Consistent aggregation also applies across any level of partners, locations and sectors. An illustrative example is spatial aggregation (across locations, as when regions are aggregated into countries). See Appendix Section 8.3 for development.

Each destination faces different effective price vectors due to bilateral resistances that are equal to trade friction factors that include taste shifters. Because utilities (or activity levels in the intermediate inputs case) are given in equilibrium, non-homothetic income (activity) effects on buyers that act as effective price shifters are similarly absorbed in 'trade frictions'. In the as-if-frictionless equilibrium the observable worldwide sales shares (at buyer prices) from each origin are equal to the hypothetical as-if-frictionless expenditure shares of each destination. As-if-frictionless expenditure shares are associated with the common as-if-frictionless price vector. Observable bilateral demand shares at each destination are assumed to differ from the observable world demand share (evaluated at buyer prices) for goods from each origin due to destination differences in effective price vectors. The shares differences and the invertible common demand system thus imply the difference of actual effective price vectors from the common as-if-frictionless effective price vector. The comparability of observed and as-if-frictionless equilibrium price vectors is achieved with the standard normalization – the observed and as-if-frictionless world buyer price vectors weighted by the origin country endowment shares both sum to one.

The difference in effective price vectors due to difference in shares is accounted for in this paper with an intermediate response 'discrete elasticity' times a discrete percentage change in relative resistance. The 'discrete elasticity' uses the intermediate value theorem.[6] Relative resistances are determined by observable share differences on this reasoning. The intermediate 'discrete elasticity' requires a projected intermediate share, hence a demand specification must be chosen to solve for the implied relative resistance results.

Operational relative resistances require a specification choice. A natural choice of specification to approximate the unknown 'true' invertible demand system is the Törnqvist approximation: the intermediate share is the arithmetic average of the observed and as-if-frictionless

[6] Invertibility justifies the use of the intermediate value theorem.

shares. Given symmetric beliefs about specifications, the Törnqvist approximation is shown to minimize the approximation error variance in relative resistances associated with choice of a specification from the set of invertible demand systems. The Törnqvist approximation is naturally associated with an intermediate value of the price index equal to the geometric average of the observed price index and the normalized value of the as-if-frictionless price index. Revealed relative resistances are solved from the resulting accounting system. The accounting is exact when the Törnqvist approximation is true, associated with the non-homothetic translog specification below. The translog is understood as a second order approximation to any compensated demand system generated by cost minimization.

Non-parametric gravity as defined here appears to be at the upper limit of extracting information about relative resistances from observed trade within the broader class of invertible demand systems. The static sectoral spatial arbitrage equilibrium that results is understood to nest inside a general equilibrium production model that generates the given supply vectors distributed globally by the gravity model. The bilateral frictions are given in the gravity model but similarly understood in perspective to depend on production forces and behavior of non-competitive actors that are outside the model.

The relative resistances implied by the model are *residuals* with a rich structure inside, like Solow productivity residuals but defined for discrete differences. Future research may usefully seek to identify components of bilateral resistance residuals beyond the received gravity literature border policies and a list of proxies, following the strategy of the productivity literature. The model extends to include the treatment of heterogeneous firms, with origins interpreted as firms' locations in product as well as physical space. The concepts of arbitrage equilibrium and seller incidence shifting still apply. The endogenous bilateral frictions may include endogenous markups by firms. Zero demand shares are due to unobservable delivery cost that exceeds the willingness to pay of buyers. Relative resistance exceeds the choke value for these cases.

3 Non-parametric Gravity

The non-parametric approach to gravity retains the key idea from the CES demand structural approach – infer relative resistances from the observable difference between national buyers expenditure shares and the world's buyer expenditure shares. The former reflect actual relative resistances faced by buyers at each destination on shipments from each origin. The latter, equal to the actual world sales shares at buyer prices, are interpreted to reflect the hypothetical 'as-if-frictionless' relative resistances equal to 1.

The common invertible demand system assumption enables aggregation of the countries' expenditure functions into a world expenditure function evaluated at the as-if-frictionless common price vector in compensated hypothetical equilibrium. The assumption also enables aggregation of the expenditure of heterogeneous buyers within countries at compensated equilibrium. Heterogeneity within countries is pushed into the background for simplicity in what follows.

Each country's expenditure function is decomposed by application of Shephard's Lemma to expenditure evaluated at actual and as-if-frictionless buyer prices. The difference between the two decomposed values of expenditure is evaluated in terms of relative resistance differences by application of the intermediate value theorem. The intermediate value theorem applies because the common demand system is assumed to be invertible.

In this setup, relative resistance to bilateral trade is a sufficient statistic that incorporates cross effects of frictions on observable shares as well as own effects. The sufficient statistics are locally solved from the observed shares and prices. The rich set of cross-effects in demand that generally enter into the determination of bilateral expenditure shares may be regarded as implicitly aggregated in the un-modeled share functions. The implicit aggregation takes explicit form in the translog demand case developed in Appendix section 8.6.

The focus on demand structure to provide measures of relative resistances is justified by thinking of efficient distribution as nested inside a full model that efficiently determines the equilibrium supply of products and costs of distribution given the revealed relative resistances. Specifications of structural demand and supply lead to a set of structural gravity models in the literature, all of which could in principle be parameterized by revealed relative resistances. Section 6 does so for the CES endowments gravity model.

3.1 Non-parametric Demand Model

Demand is assumed to be based on cost minimization by buyers. The manufactured goods application below thus nests cost minimization inside an external choice superstructure. The trade data includes both final and intermediate goods, so a common demand structure is forced onto both types of buyers.

x_{ij} is the amount of goods from origin i purchased by destination j buyers facing prices p_{ij} in arbitrage equilibrium. The objective of inferring comparable trade frictions from observed trade patterns dictates the first restriction on buyer behavior.

A1: cost-minimizing buyer behavior is represented by a common invertible demand system.

Berry et al. (2013) provide a 'connected substitutes' class of models that is sufficient for invertibility. Importantly for gravity modeling, connected substitutes allows for both complementarity and zeros in demand.

Matsuyama and Ushchev (2022) provide a homothetic functional form that satisfies the three 'laws of demand'. This escapes unrealistic features of the CES form; e.g. constant markups. But non-homothetic income effects on demand are suggested by the large variation of expenditure shares across countries despite apparently similar trade costs. Non-homothetic CES [Hanoch (1975)] is a functional form that enables income effects but it loses generality in substitution effects. The non-parametric model combines flexibility and non-homotheticity with assumption

A2: non-homothetic income effects are price-dependent only.

The combination of A1 and A2 includes the flexible EASI class of demand systems characterized by Lewbel and Pendakur (2009).

Buyers have per unit effective willingness-to-pay associated with the amounts purchased $\{x_{ij}\}$ based on cost-minimizing selection of amounts. Effective price vectors are given by $\mathbf{p}^j = \{p_{ij}q_{ij}\}$ where p_{ij} is an observable buyer price and $q_{ij}(\cdot, u)$ is a 'taste shifter' function that combines unobservable buyer costs and heterogeneity characteristics with non-homothetic income effects via changes in utility u.[7] q_{ij} ends up absorbed into the trade frictions, as in the CES case where the q_{ij} taste shifters are constants.

Expenditure by buyers is represented by the expenditure function $e(\mathbf{p})u$, homogeneous of degree one and concave in the effective price vector $\mathbf{p}$. (Sub)-utility u at constant $\mathbf{p}$ increases expenditure proportionately for simplicity.[8] Shephard's Lemma implies that the buyer j's expenditure share b_{ij} on each good i is equal to $\partial \ln e / \partial \ln p_{ij}q_{ij}$.[9]

World sales at buyers prices in equilibrium is equal to the sales obtained as if sellers faced a single aggregate buyer with a common effective price vector. The buyer price vector in the as-if-frictionless equilibrium is $\mathbf{p}^*$. The endowment vector $\mathbf{y}$ normalizes the price vector, as in the CES case. World expenditure shares $B_i(\mathbf{p}^*)$ satisfy Shephard's Lemma, and $\sum_i B_i(\cdot) = 1 \Rightarrow P^* = 1$. World sales at buyer prices in the observed equilibrium are equal to world purchases at those prices, with prices normalized consistently with the as-if-frictionless price index:[10]

$$\sum_j P_j u^j / \sum_j u^j = e(\mathbf{p}^*) = 1. \tag{2}$$

[7]Implicit restrictions on the effects of u on q_{ij} are required to preserve invertibility.

[8]The representative buyer simplification of the text is justified by the common invertible demand assumption and A1-A2. Household h in country j has expenditure function $e(\mathbf{p}^{jh})u^{jh}$, with aggregate expenditure

$$\sum_h e(\mathbf{p}^{jh})u^{jh} = [\sum_h e(\mathbf{p}^{jh}) \frac{u^{jh}}{\sum_h u^{jh}}]u^j.$$

Then [using degree one homogeneity of $e(\cdot)$], $\mathbf{p}^j$ solves $e(\mathbf{p}^j) = \sum_h e(\mathbf{p}^{jh}u^{jh}/u^j)$ where $u^j = \sum_h u^{jh}$.

[9]$x_{ij} = \partial e(\mathbf{p}^j)u^j / \partial p_{ij}$ by Shephard's Lemma. Multiply the result by $p_{ij}/e(\cdot)u^j$ to obtain b_{ij}, interpreted as the result in the text.

[10]The adding up condition for all world sales at buyer prices gives the consistent normalization for observed world price indexes:

$$\sum_j \sum_i p_{ij} \frac{x_{ij}}{E_j} E_j = \sum_j E_j \Rightarrow \sum_j P_j \frac{E_j}{\sum_j E_j} = 1.$$

The normalized price indexes for observed and as-if-frictionless equilibria are the basis for non-parametric relative resistance measures. $P_j - 1 = e(\mathbf{p}^j) - e(\mathbf{p}^*)$ is the difference in country j's unit cost of utility at observed and as-if-frictionless prices, a measure of how advantaged or disadvantaged country j is in the distribution of goods relative to other countries. The world adding up condition implies that at constant utility (in compensated equilibrium),

$$\sum_j [e(\mathbf{p}^j) - e(\mathbf{p}^*)] \frac{u^j}{\sum_j u^j} = 0.$$

The square bracket term for each country j on the left hand side equals the difference in j's buyer price function at observed and as-if-frictionless equilibria. The utility share weighted average of the differences is equal to zero. The actual equilibrium utilities are associated with unique price indexes (given invertibility) and thus unique differences.

Use cost minimization (hence Shephard's Lemma) to expand country j's buyer price index difference $e(\mathbf{p}^j) - e(\mathbf{p}^*) = P_j - 1$:[11]

$$P_j - 1 = \sum_i \left[\frac{p_{ij} x_{ij}}{u^j} - \frac{p_i^* x_{ij}^*}{u^j} \right] = \sum_i (P_j b_{ij} - B_i).$$

The intermediate value theorem applied to the difference in unit cost of utility $e(\mathbf{p}^j) - e(\mathbf{p}^*)$ combines with the Shephard's Lemma property of the unit cost function $e(\cdot)$ to yield the equivalent decomposition of $P_j - 1$ given as:

$$\sum_i (P_j b_{ij} - B_i) = \sum_i \tilde{P}_j \tilde{b}_{ij} \frac{p_{ij} - p_i^*}{\lambda_j p_{ij} + (1 - \lambda_j) p_i^*} \quad (3)$$

for some $\lambda_j \in [0, 1]$.[12] The intermediate price index $\tilde{P}_j$ and the intermediate shares $\tilde{b}_{ij}$ are evaluated at the point where the intermediate price vector for j is given by $\tilde{p}_{ij} = \lambda_j p_{ij} + (1 -$

[11] Expenditure $E_j = e(\mathbf{p}^j) u_j$ and $\partial \ln E_j / \partial \ln p_{ij} = b_{ij} = p_{ij} x_{ij} / E_j$.

[12] Equation (3) follows from the univariate intermediate value theorem applied to $e(\tilde{\mathbf{p}}^j)$ where $\tilde{\mathbf{p}}^j = \lambda \mathbf{p}^j + (1 - \lambda) \mathbf{p}^*$ after application of the chain rule. (3) solves for the value of λ such that $u^j [e(\mathbf{p}^j) - e(\mathbf{p}^*)] = u^j de(\tilde{\mathbf{p}}^j)/d\lambda$.

$\lambda_j)p_i^* \; \forall i$. Invertibility is sufficient for the applicability of the intermediate value theorem and also guarantees that the solution is unique.

The ratios on the right hand side of (3) reduce to discrete percentage differences in relative prices p_{ij}/p_i^*. Use $p_i^* = c_i \Pi_i$. It is convenient to scale the vectors of buyer prices by their indexes, normalized P_i and $P^* = 1$. Then $p_{ij}/p_i^* = R_{ij} = \tau_{ij}/\Pi_i P_j$. Then:

$$\frac{p_{ij} - p_i^*}{\lambda_j p_{ij} + (1 - \lambda_j)p_i^*} = \frac{R_{ij} - 1}{\lambda_j R_{ij} + 1 - \lambda_j}.$$

Equation (3) implies that the observable differences on the left hand side are explained by the relative resistances on the right hand side.

3.1.1 Relative Resistance Inference

The relative resistances are determined from the elements of the sum in (3).

Proposition 1

For demand systems satisfying A1-A2, relative resistances are identified from buyer expenditure shares and normalized price indexes in each element of the sum in (3):

$$P_j b_{ij} - B_i = \tilde{P}_j \tilde{b}_{ij} \frac{R_{ij} - 1}{\lambda_j R_{ij} + 1 - \lambda_j}, \; \forall i, j. \tag{4}$$

(4) is obviously sufficient for equality in (3). (4) is also necessary because the cost minimization property of $e(\mathbf{p})$ applies to both sides of (3).

Proof *Suppose the contrary:*

$$P_j b_{ij} - B_i = \tilde{P}_j \tilde{b}_{ij} \frac{R_{ij} - 1}{\lambda_j R_{ij} + 1 - \lambda_j} + \epsilon_{ij}, \; \forall i, j$$

where $\epsilon_{ij} \neq 0$ subject to $\sum_i \epsilon_{ij} = 0, \; \forall j$. Then contrary to cost minimization, there exist changes in the intermediate allocation of expenditure shares $\tilde{b}_{ij}$s at constant price vector $\tilde{\mathbf{p}}^j$ that would lower the price index for given utility. Cost minimization combined with A1-A2 *implies (4) is necessary for (3).* ‖

Proposition 1 confirms the intuition that relative resistance characterizes the implications of spatial arbitrage more widely than in the previously known parametric cases. But it is not operational. The right hand side of equation (4) varies with the unknown true value of $\lambda_j \in [0,1]$ both directly in the percentage change ratios and indirectly due to the effect of λ_j on the discrete elasticity term $\tilde{P}_j \tilde{b}_{ij}$.

Observed trade patterns strongly suggest that trade expenditure shares b_{ij} fall with rising trade frictions, associated with observed $b_{jj} > s_j$ and $b_{ij} < s_i, i \neq j$. Consistent with this observation, *impose the sign convention that $P_j b_{ij} - s_i$ varies with $-(R_{ij} - 1)$, $\forall i, j$.*[13] The convention has no effect on (3).

Operationality further requires a value of λ_j, implied by choice of a specification of the demand system. $\lambda_j = 1/2$ (the Törnqvist approximation) implies the general non-homothetic translog specification, and operationalizes (4). Specifically, $\lambda_j = 1/2$ implies $\tilde{b}_{ij} = (b_{ij} + s_i)/2$ and $\tilde{P}_j = \sqrt{P_j} = \exp[\ln(P_j)/2 + \ln(1)/2]$. No translog parameters are needed in (4).

The elements of equation (4) are operationalized with the sign convention and $\lambda_j = 1/2$ as

$$P_j b_{ij} - s_i = -\sqrt{P_j}\, \bar{b}_{ij} \frac{R_{ij} - 1}{(R_{ij} + 1)/2}. \tag{5}$$

Equation (5) can be solved for R_{ij}. If the translog is the true demand model, equation (5) is exact and yields exact non-parametric relative resistance indexes given the absence of measurement error.

Proposition 2 *Revealed relative resistances are given by*

$$R_{ij} = \frac{2\bar{b}_{ij}\sqrt{P_j} - (P_j b_{ij} - s_i)}{2\bar{b}_{ij}\sqrt{P_j} + (P_j b_{ij} - s_i)};\ \forall i, j. \tag{6}$$

Seller incidence Π_j is revealed from inverting $P_j R_{jj}$ where (6) is used for R_{jj}.

[13] An equivalent convention in the CES gravity literature restricts the sign of the trade elasticity parameter: $\theta > 0$. In the non-parametric case the convention implicitly restricts the specification, its parameters and the data. Appendix 8.6 develops a translog model that can illustrate the implications of the convention.

Figure 3: Revealed R_{jj} Logic

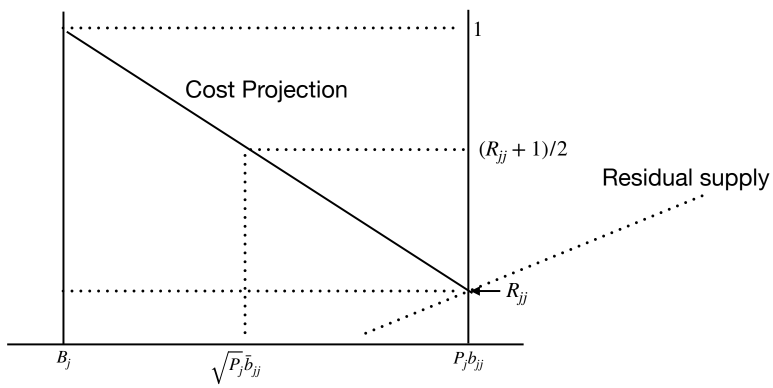


Figure 3 illustrates the logic of equations (4) and (5) and provides insight into how the intermediate value theorem enables non-parametric calculation of relative resistances. The diagram focuses on the case $i = j$. As-if-frictionless demand shares, equal to sales shares at buyer prices, are generated from the common price vector $\mathbf{p}^*$.

The right vertical axis at horizontal coordinate P_jb_{jj} is scaled in relative resistances. On that axis $\tilde{R}_{jj} = \lambda_j R_{jj} + 1 - \lambda_j \in [R_{jj}, 1]$ is an intermediate value point based on equation (4). (The values are projected across to the left vertical axis to indicate association of the values on the horizontal axis with their relative resistances.) The horizontal axis is in units of intermediate domestic friction cost shares $\tilde{P}_j(z)\tilde{b}_{jj}(z) \in [B_i, P_jb_{jj}]$ where z denote a specification choice proxied by an associated specification-specific $\lambda_j(z) \in [0, 1]$. The projection line of intermediate domestic friction cost shares based on equation (4) uses ratios of

$$\frac{P_jb_{jj}}{\tilde{P}_j(z)\tilde{b}_{jj}(z)} = \frac{R_{jj} - 1}{\tilde{R}_{jj}(z)}$$

with negative slope. [The projection line has negative slope, as shown by equation (7) in Section 3.2.]

$\lambda_j = 1/2$ selects the midpoint on the horizontal axis between B_i and $P_j b_{jj}$. $\lambda = 1/2$ implies the general translog specification. The translog specification in turn implies that $\tilde{P}_j \tilde{b}_{ij} = \sqrt{P_j}\, \bar{b}_{jj}$. The intermediate value theorem projects this point to the midpoint on the cost projection line, from which it projects to the right vertical axis at $(R_{jj} + 1)/2$. This is the midpoint between R_{jj} and 1, associated with discrete percentage change $(R_{jj} - 1)/[(R_{jj} + 1)/2]$. The vector of relative resistances is implicitly active in all values of the shares, observed and intermediate. The relative resistances are revealed by the midpoints. No parameters are needed. The location of both the projection line and the residual supply schedule are determined by general equilibrium determination of the full set of relative resistances $\{R_{ij}\}$. Thus the as-if-partial equilibrium picture above applies to all bilateral pairs simultaneously, justifying the solution (6).

Specification choice should be based on a belief that system (4) generates a good approximation to 'true' relative resistances. A cost projection line that linearizes equation (4) relates $\tilde{R}_{jj}$ to an associated $\tilde{P}_j \tilde{b}_{jj}$ in Figure 3. Given its linearity (which depends on the unknowable effect of changes in demand system specification upon $\tilde{b}_{jj}$ and $\tilde{P}_j$), $\lambda_j = 1/2$ minimizes the variance in the approximation error in relative resistance that is associated with specification choice error. Specification choice implies acting on a belief z that $\lambda_j(z),\ z \in [0, 1]$ is true, knowing it may be false. The variance minimizing argument holds for all belief distributions that are symmetric around the mean. Consider the worst case example where all $\lambda(z),\ z \in [0, 1]$ are equally likely to the analyst – i.e., probability densities are represented by the uniform distribution on $[0, 1]$. Let $r_{ij}(\lambda_j(z)) = (R_{ij} - 1)/(\lambda_j R_{ij} + 1 - \lambda_j)$ denote the projected value for any λ_j. The approximation error variance for an arbitrary $\bar{\lambda} \in [0, 1]$ is $V = E[r_{ij}(\lambda_j(z)) - r_{ij}(\bar{\lambda})]^2$ where the expectation is taken over the distribution of z.

Proposition 3

$\lambda_j = 1/2$ *minimizes the approximation error variance for the case where* $r_{ij}(\lambda_j(z))$ *linear* *in z.*

Proof *Choose* $\bar{\lambda}$ *to minimize* V*. This implies (the necessary condition)* $-2E[r_{ij} - r_{ij}(\bar{\lambda})]\partial r_{ij}/\partial\bar{\lambda} = 0 \Rightarrow E[r_{ij}(\lambda_j(z))] = r_{ij}(\bar{\lambda})$. $\bar{\lambda} = 1/2$ *satisfies this condition. The second order condition is also satisfied.* ‖

Figure 3 also gives intuition about sensitivity to approximation error due to the translog restriction, whether or not linearity obtains. The unknown true value λ_j^* on the projection line moves locally around the midpoint. The analytic and quantitative effects of approximation error from deviation from the translog are developed in Section 3.2. Figure 3 for the linear case also suggests why $\lambda_j = 1/2$ minimizes the approximation error variance when beliefs about λ_j (each implicitly associated with a demand system specification that fits the data) are symmetrically distributed on $[0, 1]$.

3.2 Approximation Error

The general case equation for a typical element of the linear decomposition of the change in world expenditure implied by the shift from observed to as-if-frictionless relative prices (5) is

$$P_j b_{ij} - s_i = \tilde{P}_j \tilde{b}_{ij} \frac{R_{ij} - 1}{\lambda_j R_{ij} + 1 - \lambda_j}.$$

The value of R_{ij} that satisfies the equation depends on both the specification and its parameters that yield the intermediate value λ_j and the intermediate price indexes and shares $\tilde{P}_j \tilde{b}_{ij}$ from the observed $P_j b_{ij} - s_i$. In terms of Figure 3, a different specification implies a different λ_j and hence a different point on the projection line. How sensitive is the inferred value to the approximation error when the translog specification is false?

A mechanical answer to the question is provided by local sensitivity analysis of equation (6) at $\lambda_j = 1/2$. The local elasticity of R_{ij} with respect to $\tilde{P}_j \tilde{b}_{ij}$ evaluated at $\lambda_j =$

$1/2$, $\tilde{P}_j \tilde{b}_{ij} = 2\sqrt{P_j}\ \bar{b}_{ij}$ is

$$\frac{\partial \ln R_{ij}}{\partial \ln \tilde{P}_j \tilde{b}_{ij}} = \frac{2\sqrt{P_j}\ \bar{b}_{ij}}{2\sqrt{P_j}\ \bar{b}_{ij} - s_i}(1 - R_{ij}).$$

Combine with the effect of variation in λ_j at $\lambda_j = 1/2$.[14]

$$\frac{\partial \ln R_{jj}}{\partial \ln \lambda_j} = \frac{\partial \ln R_{jj}}{\partial \ln \tilde{P}_j \tilde{b}_{ij}} \frac{\partial \ln \tilde{P}_j\ \tilde{b}_{jj}}{\partial \ln \lambda_j} = \frac{\sqrt{P_j}\ \bar{b}_{jj}}{2\sqrt{P_j}\ \bar{b}_{jj} - s_j}(b_{jj} - s_j)(1 - R_{jj}) \qquad (7)$$

For China and the US in manufacturing 2000-2014, the sensitivity elasticities in equation (7) range over time from 1.72 to 0.36 and 0.37 to 1.03 respectively, falling with rising sales share s_i for China and rising with falling sales share for the US. This implies significant sensitivity to approximation error, larger for small sellers. If the translog itself appears dubious, the non-parametric approach is similarly contaminated.

In perspective, near $\lambda_j = 1/2$, large approximation error requires a specification within the class of invertible demand systems that diverges sufficiently from the translog to be poorly approximated by variation in the translog parameters. The flexibility of the general translog suggests that many alternative specifications may be closely approximated due to a large number of parameters [$N \times (N-1)/2$ where N is the number of countries] that are free to vary subject to the constraints imposed by homogeneity and negative definiteness of the substitution effects matrix.

Measurement error in the data is another important source of errors in the revealed relative resistances that should be faced in future research. Given the translog specification as true, the problem is to estimate the relative resistances from

$$\frac{R_{ij} - 1}{R_{ij} + 1} = \frac{P_j b_{ij} - s_i}{2\sqrt{P_j}\ \bar{b}_{ij}}$$

where on the right hand side P_j, s_i and thus $\bar{b}_{ij}$ are all measured with error that is correlated.

[14] Equation (7) uses $\partial \tilde{P}_j / \partial \lambda_j = (1/2) \partial P_j / \partial \lambda_j = 0$.

Progress depends on imposing strong but plausible restrictions on the correlation structure, informed by knowledge about the construction of the data.

3.3 Gains from Trade and Terms of Trade

The buyers' loss per unit of utility of country j due to heterogeneity of frictions is equal to $P_j - 1$. $P_j - 1$ is also interpreted as the average percentage incidence of normalized frictions borne by buyers in j. From the social welfare point of view, country j is both buyer and seller. Its gains on domestic sales as seller are offset by the loss to the country's domestic good buyers. The loss per unit of utility of country j due to *cross-border* trade frictions is given by $L_j = \sum_{i \neq j}(P_j b_{ij} - s_i)$. A compact form of L_j manipulates $P_j - 1 = \sum_i (P_j b_{ij} - s_i)$ to yield

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

The loss $L_j = P_j b_{jj} - s_j$ on the right hand side is due to frictions on the left hand side of the equation (both on average and due to cross-border imports). L_j reduces the loss to a measure based on domestic sales.

The economics behind accounting loss measure L_j are revealed by a decomposition using equation (5) for $i = j$.

Lemma

$$L_j = -\sqrt{P}_j \bar{b}_{jj} \frac{R_{jj} - 1}{(R_{jj} + 1)/2}. \tag{8}$$

The loss falls as $\bar{b}_{jj}$ falls, hence b_{jj} falls toward s_j. This means the country is becoming more open to trade. The loss also falls as the terms of trade R_{jj} rises, holding all else equal in the cross-section of countries. [$R_{jj} < 1$ (almost) universally.] As R_{jj} approaches one, cross-border trade bears the average cost of frictions and loss $L_j \to 0$.

Ex post changes in loss can be non-parametrically evaluated with the percentage change in loss relative to as-if-frictionless trade $L_{j,t} - L_{j,t-1}$ where L_j at each time period is given by (8). Equation (8) in changes incorporates changes in s_j. Thus it reflects changes in specialization due to terms of trade changes along with any other supply side forces at work. As a measure of the change in the exchange gain at the sectoral level, it excludes specialization gains or other sources of real income change. Note also that the formula in principle incorporates the effects of changes in both the intensive and extensive margins of trade. Equation (8) is useful for non-parametric *ex post* evaluation of change in arbitrage gains from trade in a single sector. The application below uses (8) to quantify the differing welfare effects of globalization on manufacturing in China and the US.

Non-parametric loss measure (8) builds on the well-known Arkolakis et al. (2012) demonstration that the observable domestic share b_{ii} variable is negatively related to the gains from trade, requiring only a trade elasticity to quantify gains from trade changes that result from foreign changes. Non-parametric (8) is a compensating variation measure, in contrast to the equivalent variation real income measure of Arkolakis et al. (2012). An offsetting advantage of the non-parametric loss measure (8) is its ability to include changes in domestic frictions and endowments as well as foreign ones. This is crucial for applications to large national sales share changes, as in the manufacturing trade of the US and China during the globalization era, 2000-14. See Appendix Section 8.4 for a detailed comparison.

In wider perspective, economic gravity characterized by (8) and (6) re-connects to physical gravity in the two body case. The attractive force of trade is the gains from trade. A country's terms of trade is interpreted as the inverse square of its economic distance to and from the world market, and its exchange gains from trade are locally proportional to the inverse square of its economic distance to and from the world market. The terms of trade interpretation as the inverse square of economic distance follows from interpreting the numerator and denominator of R_{jj} as squares of geometric means. The denominator $\Pi_i P_j$ of R_{jj} is the square of the geometric mean of inward and outward multilateral resistances, the natural average of a product. Similarly the numerator τ_{ij} is understood as (an index of) the square of the geometric mean of inward and outward resistances on shipments between domestic locations.[15]

4 Application to China and US Manufacturing Trade

The application quantifies changes in manufacturing terms of trade and gains from trade for China and the US over the period 2000-2014. The China and US cases highlight the value of a non-parametric approach to gravity because big general equilibrium propagation effects are implied by their large shares of world manufacturing and the large changes in these shares over time. Moreover, manufacturing itself is an exceptionally tradable set of products. Thus multilateral resistance changes are likely to be important. In contrast, the constant trade elasticities models of structural gravity practice are likely to significantly mislead in quantifying the evolution of relative resistance and seller incidence.

Data are drawn from the World Input-Output Database. Non-parametric gravity measure (6) uses usually high quality observations on value of production and trade combined with observed buyer price P_j data that is subject to the usual problems of price comparison indexes. The step from the preceding theory of relative resistance to practice depends on consistent data for purchases at buyers prices in all destinations from all origins along with buyer price data that is consistent. Price indexes from the WIOD are consistently associated with the production and expenditure flows. They are constructed from the intermediate input buyer prices in the database. See the online replication folder for details.

The adding up condition on bilateral shares to world market shares, implies that the normalization of the price indexes is $\sum_j E_j P_j / \sum_j E_j = 1$.[16] Thus the observed price indexes $\hat{P}_j$ are deflated to form the normalized $P_j = \hat{P}_j \sum_j E_j / \sum_j E_j \hat{P}_j$. In the application below, normalized P_j for manufacturing is lower than 1 for China and nearly constant. P_j for the US rises about 10% from below 1 to above 1 over the 2000-14 period.

[15] Appendix section 8.3 deals with internal economic distance formally.

[16] The adding up condition 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.

Suspicion about the quality of the buyer price indexes P_j is justified, but the present application is arguably robust for two reasons. First, the US and China have highly diversified manufacturing so the sample size of observed component prices is large. Second, the focus on annual change in the normalized indexes shows only small variation around 1, so their movement has small influence on the calculations.

The terms of trade are given by relative resistance on domestic trade, R_{ij} for $i = j$. The calculation of relative resistance R_{jj} applies equation (6). This sufficient statistic appears superior to standard measures of the terms of trade based on the well known deficiencies of price indexes. Price comparison is mostly based on unit values and their associated measurement error, while incomplete coverage for exports is especially salient for the exports of diversified economies. Less obviously but perhaps more important, prices do not contain unobserved user costs that vary across users and product types.

Non-parametric measures of changes in exchange gains from trade and terms of trade for China and the US reported presented at the outset are discussed below. Treatment of final demand and intermediate input demand separately is essential impossible for familiar reasons, so the cost function $e(\mathbf{p}^j)$ is assumed to be identical for both uses. The buyers side price indexes of the theory are thus the intermediate input price indexes of the WIOD.[17]

Non-parametric sufficient statistics for percentage changes in gains from trade rand terms of trade relative to as-if-frictionless trade are summarized below with average annual percentage rates of change. The discrete percentage change in gains is $2(L_j^1 - L_j^0)/(L_j^1 + L_j^0)$ for any years 0 and 1 where equation (8) is applied to calculate L_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 (6) for the case $i = j$.

The application reveals that US manufacturing experienced a 3.8% annual average fall in gains from trade relative to as-if-frictionless trade from 2000 to 2014. This was accompanied

[17]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.

by a 5.5% annual average fall 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. China's gains from trade relative to as-if-frictionless trade rose an annual average 1.9%, accompanied by an annual average 8.3% rise 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 gains measures incorporate the effect of a rise in s_j on $\bar{b}_{jj}$ that increases loss L_j. The rise in s_j directly raises $\bar{b}_{jj} = (b_{jj} + s_j)/2$, offset by the indirect effect whereby the rise in s_j raises R_{jj} and thus reduces b_{jj}. Thus the gains % changes are lower in absolute value than the terms of trade changes for both China and the US. The association of terms of trade and gains from trade is analyzed in terms of the model below in SectionSpecification.

The seller incidence measure Π_j/τ_{jj} is obtained from solving $R_{jj} = \tau_{jj}/(\Pi_j P_j)$. Recall that the seller net price c_j varies inversely to relative seller incidence Π_j/τ_{jj}. The yearly average percentage changes are -8.2% for China and 6.3% for the US. Thus the terms of trade movement of both countries is mostly explained by the global effects of shifts in the sellers' incidence of trade frictions Π_j – sellers' incidence falls as sales shares rise. Terms of trade $R_{jj} = \tau_{jj}/\Pi_j P_j$ component P_j plays a subsidiary role. In the US case with mature internal distribution infrastructure, internal distribution frictions τ_{jj} presumably do not change much, while P_j rises slightly only about 10% over 2000-2014. Almost all the change in R_{jj} is due to a rise in Π_j. In China's case, τ_{jj} presumably falls as internal infrastructure dramatically improves while P_j is almost constant. The decline in China's Π_j/τ_{jj} implies an equal rise in c_j but attributing all the change to a fall in Π_j overestimates its role. Both cases point to the dominant role and large effects of seller incidence shifting.

Two caveats about interpretation need emphasis. First, the gains from trade and terms of trade statistics are for single sectors, only a part of of the national economies. In particular, a full national accounting would relate the changes in manufacturing sales shares to the alternative uses of the national resources in the rest of the economy along with changes in sectoral terms of trade for other sectors. Second, the aggregation of sub-sectors into all of manufacturing conceals the effects of compositional change on relative resistances. Keeping these limitations in mind, the lens of the model still provides a sharp interpretation.

5 Commercial Rivalry Implications

Revealed relative resistances vary with seller size in the static equilibrium cross section of origin destination trade pairs. The lens of the model reveals that, all else equal, equation (6) implies that relative resistance R_{ij} is increasing in s_i; $\forall i, j$:

$$\frac{\partial \ln R_{ij}}{\partial \ln s_i} = \frac{s_i}{2\bar{b}_{ij}\sqrt{P_j} - (P_j b_{ij} - s_i)}[1 + R_{ij} + \sqrt{P_j}(1 - R_{ij})] > 0. \tag{9}$$

This intuitive sharp result implies that, larger countries in the cross section have lower outward multilateral resistance Π_i – seller incidence shifting. Lower Π_i raises $R_{ij} = \tau_{ij}/\Pi_i P_j$; $\forall i, j$. The indirect effect is amplified on average by a fall in $s_j,\ j \neq i$ due to $\sum_i s_i = 1$.

Equation (9) applies to the impact effect of supply share *differences* on relative resistance in the cross section. Full general equilibrium comparative statics combine impact effect (9) with knock-on changes to P_j and the bilateral frictions τ_{ij} that blur this quantification but the intuition is likely hold. Thus the positive sign of (9) helps explain the time series results showing perfect positive correlation of inverse seller incidence $1/\Pi_i$ and sales shares s_i for China and the US in Figures 1 and 2 . The pattern arises because seller incidence changes dominate the movement of $R_{ii} = \tau_{ii}/\Pi_i P_i$ in the data.

The non-parametric terms of trade elasticity with respect to sales size is given by equation (9) for the domestic case $i = j$. Plug the observed and inferred data in to (9) where j is the US or China for a given year. Note that the terms of trade elasticity is increasing in s_j, as is the cross effect on other sellers. The externality is thus quantitatively significant mostly for large sellers.

The US 2014 terms of trade elasticity with respect to the share is equal to 1.04. China's

2014 terms of trade elasticity with respect to its share is 0.97. $\partial \ln R_{ii}/\partial \ln s_i \approx -1$. The results and their approximate constancy over time are consistent with a CES trade elasticity $\theta \approx 1$. This implies that $B_i = (c_i\Pi_i)^{-\theta} = s_i$ induces approximately a 1% fall in $c_i\Pi_i$, $\forall i$ for every 1% rise in s_i. The consistency alternatively implies that given a CES specification, the Törnqvist approximation $\lambda_i = 1/2\ \forall i$ generates non-parametric relative resistances that are accurate.

The cross effect of Chinese sales share on US terms of trade comes through its necessary effect on reducing the average sales shares of all other sellers. Assume that the effect on the US share is equal to the average effect on the rest of the world. (This estimate is likely downward biased.) Then the requirement that shares sum to one implies

$$\frac{s_{CN}}{1 - s_{CN}}\hat{s}_{CN} = -\sum_{j \neq CN} \frac{s_j}{\sum_{j \neq CN} s_j}\hat{s}_j.$$

Using China's 2014 share of 31.9% implies that a 10% rise in China's sales share reduces the average non-China sales share by 4.68%. The reduced US sales share times the US terms of trade elasticity of 1.04 reduces the US manufacturing terms of trade by 4.87%. The large negative externality is due to China's large size in world manufacturing. The same calculation for the US effect on China uses the US 2014 share of 12.5%. A 10% rise in US sales share reduces the average rest of world share by 1.43%. The 2014 Chinese terms of trade elasticity of 0.97 implies that China's terms of trade fall 1.39% on the assumption that China's sales share falls at the rest of world average rate.

The own effects of US and China share changes on their terms of trade can be decomposed relative to other forces based on the local elasticity estimates for 2014. The attribution overstates China's own effect contribution because it uses the most recent of the annual elasticity calculations – $\partial \ln R_{ij}/\partial \ln s_i$ is increasing in s_i, and China's share and its resistance elasticity rises over time. (The US case is more complicated because while the US share falls over time, the effect on its resistance elasticity is offset by the effect of a rise in its price index.) The combination of the two cases brackets the implication that the own effect due to the local terms of trade elasticity (9) accounts for more than half of the observed terms of trade movement.

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 fall in US terms of trade is 2.9%, slightly more than half of the 5.5% fall in the estimated results. The own effect of China's 10.2% average annual rise in sales share implies that it accounts for 6.83 percentage points of the annual 8.3 percentage point rise in its terms of trade.

Approximation error affects the terms of trade elasticity $\partial \ln R_{ij}/\partial \ln s_i$ reported and used above. Perspective on its reported value based on equation (9) is provided by comparison to the upper and lower bound local change cases generated by setting $\lambda_j = 1$ and $\lambda_j = 0$ in equation (5). Both cases reduce the ratio on the right hand side of equation (9) to $s_j/s_j = 1$. At $\lambda_j = 1$ the terms of trade elasticity for the US in 2014 is equal to 1.16 versus its calculated value 1.04, while at $\lambda_j = 0$ the terms of trade elasticity is equal to 0. (For China the corresponding terms of trade elasticities are 1.37 at $\lambda_j = 1$ versus its calculated value 0.97 while the elasticity is equal to 0 at $\lambda_j = 0$.) The range of the revealed terms of trade elasticities is comfortably far above zero, consistent with the observed equilibrium of trade that is very far from as-if-frictionless equilibrium. This is a smell test on the adequacy of equation (9) and its association with the revealed relative resistance statistics.

6 Trade Elasticity Inference

Projection of counterfactuals requires parametric modeling. The simplicity of CES and its wide use in the gravity literature suggest calibration of the CES parameter θ that best fits the observed equilibrium trade expenditure shares to the relative resistance statistics generated from (6). A tightly fitted θ close to 1 is the result. This is much lower than the range of estimates in the previous literature.

The difference is explained here by omitted variable bias in previous estimates of the trade elasticity. The standard method identifies the trade elasticity off the variation of observable bilateral buyer prices or other observable price shifters such as tariffs or transport costs. Revealed relative resistances capture variation in unobservable 'taste shifter frictions' that is omitted in previous estimation. Negative correlation of observable bilateral prices and trade costs with unobservable bilateral frictions is plausibly explained in the arbitrage model by incomplete buyer incidence of unobservable frictions.

The minimum distance calibration selects the value of θ that minimizes the variance of local elasticities calibrated for each observation to exactly fit the revealed relative resistances to the observed relative shares. Two interpretations of the minimum distance calibrator are possible. In the first, the general translog specification that generates the statistics is treated as true. In the second, neither specification is treated as true but the method averages results from a widely used CES model and a model widely interpreted as a good approximation to a flexible general functional form. Under either interpretation, a tight fit suggests that CES and its parameterization is a good local approximation to the underlying arbitrage equilibrium process.

The buyers' equilibrium expenditure share with CES demand in the theoretical gravity equation is

$$b_{ij} = s_i(\tau_{ij}/\Pi_i P_j)^{-\theta} = s_i(R_{ij}^{CES})^{-\theta}. \tag{10}$$

Invert (10) to isolate the unobservable R_{ij}^{CES} on the left hand side and then take logs. The result is

$$\ln R_{ij}^{CES} = -(1/\theta)[\ln b_{ij} - \ln s_i]. \tag{11}$$

The revealed non-parametric $\ln R_{ij}$ is given by the log of (6). The minimum distance CES parameter is the CES trade elasticity (inverse) that minimizes the sum of squared residuals

η_{ij}^2 from the cross-section 'regression' equation:

$$\ln R_{ij} = (-1/\theta)[\ln b_{ij} - \ln s_i] + \ln \eta_{ij}. \tag{12}$$

Here $\ln \eta_{ij}$ represents the effect of specification error (interpretable as the difference between the true translog local elasticity and the CES parameter) as well as measurement error. (Inability to treat final and intermediate demand systems separately introduces further specification error.)

Equation (12) extends to a panel setting, adding the time subscript t. The minimum distance CES elasticity estimated from panel data solves

$$\min_{\theta} \sum_{i,j,t} \ln \eta_{ij,t}^2. \tag{13}$$

Minimization serves to both to minimize the average difference of the CES representation from the translog specification used to generate relative resistance and to average out the effect of pure orthogonal measurement error.

The application uses the terms of trade and domestic shares for the US and China, 2000-2014. Thus time variation in $\ln R_{ii,t}$ is fitted to the time variation in $\ln b_{ii,t} - \ln s_{i,t}$. Procedure (13) yields a tightly estimated θ equal to 1.04 with standard deviation 0.07 in the US subsample, and 1.05 with standard deviation 0.04 in the China sub-sample. Appendix 8.5 discusses elasticity estimation further. The CES specification choice is natural for its simplicity and connection to the large empirical gravity literature.[18]

The calibration method and results contrast with standard econometric estimation of θ based on the CES gravity specification treated as true. The econometric estimator in the literature seeks the best fit unbiased estimate of the elasticity parameter θ. From the econometric perspective, 'regression' (12) yields a biased estimate of the trade elasticity. The error term $\ln \eta_{ij}$ cannot be orthogonal to the regressor $\ln(b_{ij}/s_i)$ because b_{ij} and s_i both

[18]See Appendix Section 8.6 for a sketch of fitting the translog model to relative resistances.

determine R_{ij} given by (6) and appear on the right hand side of (12).

Resolution of the tension between the two perspectives is beyond the scope of this paper but considerations below favor the calibration approach for use in projections and counterfactuals. On the one hand, omitted variable bias plausibly makes previous econometric θ estimates too high. On the other hand, from the econometric perspective, the standard method avoids the endogeneity bias associated with the calibration method. On consideration, endogeneity bias in the R_{ij}s appears to be a lesser problem.

The variation of manufacturing R_{ij} is driven mainly by the variation of seller incidence Π_i which itself is due to cross section variation in aggregates sales shares s_i. Classic identification analysis thus implies that bilateral residual supply functions are shifted along downward sloping bilateral demand curve functions to identify the trade elasticity by variation in Π_i.[19] Indirect influence of aggregate sales variation on destination buyer incidence P_j could lead to endogeneity bias, but normalized manufacturing P_j has small variation for the China and US. Finally, potential endogeneity bias in τ_{ij} is common to both revealed τ_{ij} and the observable bilateral prices or trade cost components used in the standard econometric approach. A further advantage of the calibration method is efficiency – the standard method identifies θ off the small variation of observable buyer price elements while the calibration method identifes θ off the much larger variation of relative resistances R_{ij}.

The results of this paper suggest that standard trade elasticity estimates are substantially biased upward, while the calibration method suffers only attenuated endogenety bias. The combination suggests picking the minimum distance calibration method even when relying on the CES structure for projections.

7 Conclusion

The model has implications for future applications in multiple areas. Evaluation of industrial policy was foreshadowed above. A few others are discussed below.

[19]Specifically, the reservation supply price equal to R_{ij} shifts along the demand curve.

Application of (6) to the WIOD will yield a rich panel of sectoral bilateral relative resistances. The residual accounting property of the R_{ij}s resembles the Solow productivity residual, differing in discrete rather than local differences. The productivity literature may provide clues for explaining relative resistance variation.

Trade elasticity inference from observable prices or trade costs appears significantly upward biased by omitted variables captured in relative resistances. Gains from trade changes in associated counterfactual exercises are thus significantly downward biased. Parameter inference from variation of relative resistances with sparsely parameterized specifications of CES or translog seems useful. Appropriate specification and inference of parameters to be used in model projections is a deep intellectual challenge.

The preceding gravity literature suggests two more ambitious extensions – to endogenous supply and to discrete choice of migration or investment. The parallels for endogenous supply are straightforward – the GDP deflator plays the role of the price index $e(\mathbf{p})$. Functional form restriction to the translog produces non-parametric relative resistance sufficient statistics. The main difficulty is the plausibility of invertibility. The extensive Heckscher-Ohlin-Vanek literature provides compelling counter-examples.

Discrete choice models are more promising targets for extension. Extreme value distribution theory in the literature has extended the effectively CES initial structure to a nested CES structure. The literature has accepted efficient arbitrage and the adding up constraints on demand and supply sides that combine in the trade model of the text. Is there a generalization akin to non-parametric gravity for discrete choice?

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.

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. and Penglong Zhang, "Latent Exports: Almost Ideal Gravity and Zeros," 2022.

_ **and Yoto V. Yotov**, "The Changing Incidence of Geography," *American Economic Review*, 2010, *100* (5), 2157–2186.

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

Bartelme, Dominick G., Arnaud Costinot, Dave Donaldson, and Andres Rodriguez-Clare, "The Textbook Case for Industrial Policy: Theory Meets Data," *NBER Working Paper No. 26193*, 2019.

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

Chaney, Thomas, "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, 2008, *98* (4), 1707–1721.

Hanoch, Giora, "Production and Demand Models with Direct or Indirect Implicit Additivity," *Econometrica*, 1975, *43* (3), 395–419.

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

Lewbel, Arthur and Krishna Pendakur, "Tricks with Hicks: The EASI Demand System," *American Economic Review*, 2009, *99* (3), 827–863.

Matsuyama, Kiminori and Philip Ushchev, "Destabilizing Effects of Market Size in the Dynamics of Innovation," *Journal of Econmic Theory*, 2022, *200.*

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

Ravikumar, B. and Michael E. Waugh, "Measuring Openness to Trade," *Journal of Economic Dynamics and Control*, 2016, *72.*

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

8 Appendix

8.1 Efficient Spatial Arbitrage

The set of buyer prices p_{ij} for each origin's product i at destination j are taken as given to the arbitrageurs. The arbitrageurs distribute each origin i's supply y_i to potentially all destinations j with x_{ij} received on payment of $p_{ij}x_{ij}$. Shipments are subject to given iceberg melting frictions $t_{ij} \geq 1$ such that $t_{ij}x_{ij}$ is required for receipt of x_{ij}.

Efficient arbitrage is characterized by:

$$\max_{\{x_{ij}\}} \sum_j p_{ij}x_{ij} | \sum_j t_{ij}x_{ij} \leq y_i,\ foralli. \tag{14}$$

The first order conditions are: $p_{ij} = \mu_i t_{ij},\ \forall x_{ij} > 0; p_{il} < \mu_i t_{il},\ \forall x_{il} = 0$. Here the Lagrange multiplier μ_i on the adding up constraint gives the opportunity cost of dx_{ij}. The economic interpretation of the opportunity cost is $\mu_i = c_i \Pi_i$, given net seller cost c_i times average seller incidence of frictions cost Π_i.

The adding up constraints in maximization problem (14 imply that seller incidence is

$$\Pi_i = \sum_j \frac{p_{ij}}{c_i} \frac{x_{ij}}{y_i}.$$

Thus seller incidence is defined directly from efficient spatial arbitrage, independent of the process defining the buyer willingness to pay p_{ij} or the supply process generating y_i or its cost c_i.

Converting buyer willingness to pay into relative buyer prices as in the text and using $p_{ij}/c_i = \tau_{ij}$, spatial arbitrage defines relative resistance $R_{ij} = \tau_{ij}/\Pi_i P_j$. All the elements of R_{ij} are endogenous in equilibrium, to be pinned down with more structure.

8.2 Terms of Trade Rise with Supply: Intuition

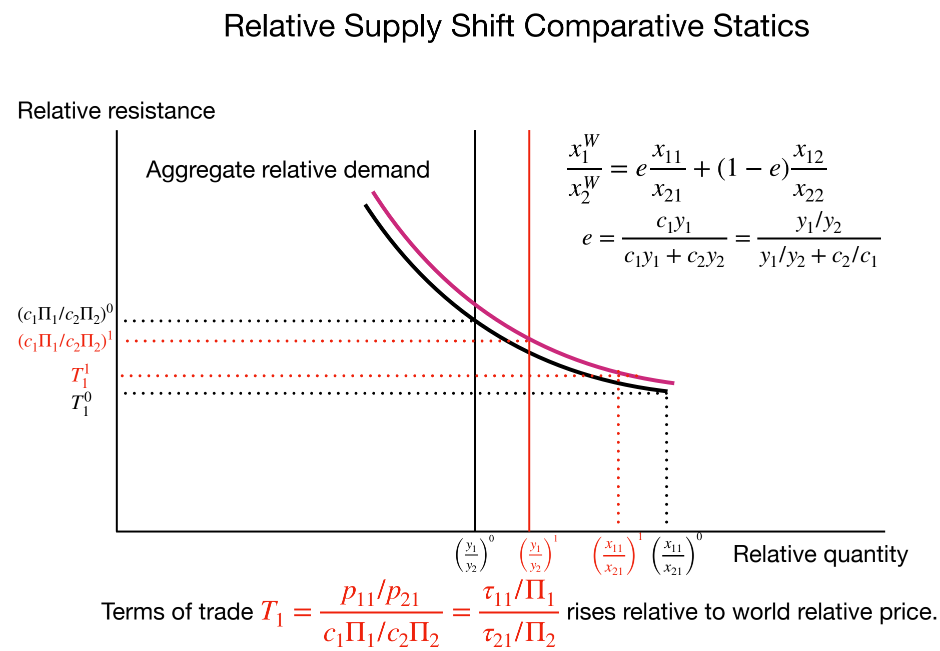
The headline result that China's terms of trade improve when China's manufacturing production rises faster than the US is contrary to intuition based on frictionless exchange. In the simple two good two country model, when relative supply of China's good rises, downward sloping relative demand for China's good implies that the world average buyer price of China's good must fall relative to the numéraire. This is true even when there are frictions present. But in the presence of frictions, China's internal buyer's relative price of its own good must rise, freeing increased sales to exchange for relatively cheaper foreign goods. The diagram below illustrates.

Two countries exchange their endowments denoted y_1, y_2 for countries 1 and 2. Demand for the two goods is generated by buyer expenditure minimization based on homothetic preferences that are identical up to country-product specific taste shifters that favor the local good. Taste shifters and distribution frictions combine in friction factors on domestic sales $\tau_{11} > 1, \tau_{22} > 1$ and foreign sales $\tau_{12} > \tau_{11}$ and $\tau_{21} > \tau_{22}$ where the order of subscripts denotes the origin-destination direction of trade. Competitive traders generate a spatial arbitrage equilibrium.

The 'world' market clears with relative world price of good 1 $c_1\Pi_1/c_2\Pi_2$ where $c_i,\ i = 1, 2$ is the net seller price and $\Pi_i,\ i = 1, 2 > 1$ is a trade weighted average of two outward frictions, the average sellers' incidence of trade frictions. The buyer relative prices are p_{11}/p_{21} for country 1 and p_{12}/p_{22} for country 2. Arbitrage equilibrium implies $p_{ij} = c_i\tau_{ij};\ \forall i, j$. The arbitrageur's opportunity cost on the sale of good i to any j is $c_i\Pi_i$. Then $\tau_{ij}/\Pi_j = c_i\tau_{ij}/c_i\Pi_i$ is the equilibrium premium or discount factor that buyer j is willing to pay to obtain good i. The variation of p_{11}/p_{21} relative to opportunity cost is equal to the terms of trade T_1. World relative demand for good 1 in equilibrium x_1^W/x_2^W must equal relative supply y_1/y_2, associated with world relative buyer price $c_1\Pi_1/c_2\Pi_2$. Trade frictions drive local relative buyer prices away from the world relative price.

The diagram focuses on country 1 and the effect of a growth in its relative size. Relative demand is downward sloping due to the substitution effect. e is country 1's share of world income, also equal to its share of world expenditure under the assumption of balanced trade. (Balanced trade is a harmless simplification since a rise in y_1/y_2 is highly correlated with a rise in e when trade is not balanced.) Equilibrium relative demand in the world market is generated by the intersection of downward sloping relative demand with vertical endowment ratio y_1/y_2. The assumed pattern of trade frictions implies that equilibrium $x_{11}/x_{21} > y_1/y_2 > x_{12}/x_{22}$. Equilibrium is associated with terms of trade T_1 for country 1.

The diagram illustrates the effect of a rise in y_1/y_2 on country 1's terms of trade. The vertical relative endowment line shifts to the right by a given percentage $\alpha > 0$. Relative size e rises by $\hat{e} < \alpha$. The result is a shift of the relative world demand schedule to the right that is less than the shift in the relative supply line. World relative price $c1\Pi_1/c_2/\Pi_2$ falls, while T_1 rises. Assuming for simplicity that the underlying τ_{ij} frictions are constant, country 1's terms of trade rise as its relative size increases because (i) it buys more than the world average amount of its own lower friction good and (ii) its relative expenditure size increase raises the weight on the lower friction good in its seller incidence average Π_1. Country 2 experiences a relative size decrease that acts in the opposite direction, raising its sellers incidence and reducing its terms of trade.



It is useful to consider the 'as-if-frictionless' equilibrium case where $\tau_{ij} = \tau_i\tau_j;\ \forall i, j$. Then $\tau_{11}/\tau_{21} = \tau_1/\tau_2 = \tau_{12}/\tau_{22} = \Pi_1/\Pi_2$ and the world relative price becomes $(c_i/c_2)(\tau_1/\tau_2) = p_{11}/p_{21} = p_{12}/p_{22}$. Incidence shifting obtains with asymmetry of frictions, most importantly the asymmetry between internal and cross-border frictions.

The logic of seller incidence shifting in the diagrammatic analysis basically carries over to the generalization in the text to many countries and its quantification focused on the effects of differential growth of China and the US on their seller incidence and terms of trade.

8.3 SpatialAggregation

Non-parametric gravity equation (5) provides a useful interpretation of the relationship between gravity applications across many varieties of aggregation. In application to aggregated sectors such as manufacturing, it is useful to note that implicit aggregation applies straightforwardly across products as well as origins, expanding to include aggregation across product-origin and product-destination categories. Less obviously in terms of notation, the same treatment extends to the aggregation of true physical locations within origin and destination aggregates. All the detail is compressed by implicit aggregation into bilateral relative resistances. The structural parametric interpretation of aggregation that is implicit in the accounting system is a guide to future work that drills into decomposing the causes of variation in the relative resistances.

The focus in this Appendix is on 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. The existing literature does not treat aggregation of frictions between city pairs not related to distance and not uniformly associated with international borders. Section 8.3.1 lays out a general treatment. Section 8.3.2 treats aggregation of internal distances in the context of infrastructure that may imply directionally asymmetric internal economic distances (e.g. up vs. down hill between locations).

8.3.1 General Logic

The general non-parametric logic of spatial aggregation of frictions is nested within the logic of (5). 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 (5) 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 $j \in J$:

$$P_j b_{Ij} - Y_i/Y = 2\sqrt{P_j}\bar{b}_{Ij} \sum_i \frac{\bar{b}_{ij}}{\bar{b}_{Ij}} \frac{R_{ij}-1}{R_{ij}+1},$$

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:

$$b_{IJ} \sum_{j \in J} \frac{b_{Ij}}{b_{IJ}} P_j - Y_I/Y = \bar{b}_{IJ} \sum_{j \in J} \frac{\bar{b}_{Ij}}{\bar{b}_{IJ}} 2\sqrt{P_j} \bar{b}_{Ij} \sum_{i \in I} \frac{\bar{b}_{ij}}{\bar{b}_{Ij}} \frac{R_{ij} - 1}{R_{ij} + 1}. \tag{15}$$

The double sum on the right hand side of (15) is interpreted as the weighted average of the effect of the granular relative resistances on observable bilateral trade between I and J,

$$\bar{b}_{IJ} 2\sqrt{P}_J \frac{R_{IJ} - 1}{R_{IJ} + 1}.$$

This interpretation is approximately consistent (i.e. consistent linear aggregation is approached) under the general translog assumption.

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.

8.3.2 Internal Distance

Industrial policy often includes infrastructure measures that reduce internal distance. In contrast, the applied gravity literature often sets internal distance to unity everywhere. The practice is justified for many purposes but can conceal variation that is important for some purposes.[20] The simplification of frictionless internal distance 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 internal frictions are absorbed in the multilateral resistances.

Variation of internal distance resolves a spatial units puzzle. Gravity applies to spatial

[20] For example, 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).

arbitrage between units of any chosen size (countries, regions, commuting zones, ...). The natural asymmetries of directional distance are geometrically averaged in internal distances $\tau_{ii} = \sqrt{\tau_{lk}\tau_{kl}},\ \forall (k,l) \in i$ for the chosen unit size i. This procedure is without consequence for characterizing spatial arbitrage between the units of the chosen size. However, small unit sizes are associated with smaller τ_{ii}, hence larger R_{ii}, contributing to a regularity observed in CES gravity model applications.

Variation in internal distance also helps explain the apparent wide variation in "openness to trade" measures across similar sized regions. Relative resistance R_{ii} is an inverse measure of open-ness that is comparable across countries in the cross section and over time, and defined for here the wide class of non-parametric gravity models. Variation in internal frictions may be as important or more important than cross-border frictions in explaining the variation in open-ness and its consequences for real incomes.

8.4 Relationship to CES Gains Measure

Arkolakis et al. (2012) show that under the CES demand specification, the observed domestic share b_{jj} and the hypothetical autarky share $b^A_{jj} = 1$ are sufficient statistics that in combination with the trade elasticity θ can quantify the gains from trade as a proportional real income rise in utility u_j relative to autarky utility u^A_j. The gains from trade relative to autarky are measured by

$$G_j = b_{jj}^{-1/\theta} = s_j^{-1/\theta} R_{jj},$$

where relative internal resistance R_{jj} is the terms of trade of country j.

Ex post changes in the gains from trade due to foreign changes only can be evaluated from changes in $b_{jj,t}/b_{jj,t-1}$ since $b^A_{jj,t} = b^A_{jj,t-1} = 1$. In relative form,

$$\frac{G_{j,t}}{G_{j,t-1}} = \left(\frac{s_{j,t}}{s_{j,t-1}}\right)^{-1/\theta} \frac{R_{jj,t}}{R_{jj,t-1}}. \tag{16}$$

Here, the supply shares change because the relative net seller prices change due to the foreign changes in supply and/or trade frictions. The first ratio on the right hand side adjusts the domestic demand share to an intermediate value to appropriately weight the second term, the proportionate terms of trade change.

The loss measure relative to as-if-frictionless trade is first put into relative terms for comparison with (16). The result is:

$$\frac{L_j}{s_j} + 1 = \frac{P_j b_{jj}}{s_j} = RL_j.$$

Ex post evaluation in relative form comparable to $G_{j,t}/G_{j,t-1}$ yields:

$$\left(\frac{RL_{j,t}}{RL_{j,t-1}}\right)^{-1/\theta} = \left(\frac{P_{j,t}}{P_{j,t-1}}\right)^{-1/\theta} \frac{R_{jj,t}}{R_{jj,t-1}} \tag{17}$$

The right hand sides of (16) and (17) apply different weights to the proportional change in the terms of trade. Given the no domestic changes condition, $s_{j,t}/s_{j,t-1}$ is the proportional change in sales of country j at normalized buyer prices while $P_{j,t}/P_{j,t-1}$ is the proportional change in j's nomalized buyer price. With balanced trade and the normalized price indexes, $s_j/P_j = u^j$. Take the ratio of (16) to (17) and note that the right hand side ratio simplifies to

$$\left(\frac{u^{j,t}}{u^{j,t-1}}\right)^{-1/\theta}.$$

When the real income effects are negligible, the two measures converge.

Generally the two measures applied to *ex post* comparisons must diverge. This because (16) is an equivalent variation measure of utility change while (17) is a measure of the change (over time) in income needed to maintain actual utility at each point in time. For the equivalent variation measure, the no domestic changes assumption means that autarky real income u_j^A does not change. For the compensation measure, as-if-frictionless equilibrium utility generally changes over time. The requirements to calculate as-if-frictionless equilibrium utility at each point in time include a parametric model and its parameters. The compensating variation approach of (17) avoids the requirements. More significant, (17) allows for domestic changes and yields non-parametric sufficient statistics that are valid for a wide class of demand systems.

8.5 CES Trade Elasticity Notes

Extension of the estimator (13) to fit the entire bilateral trade panel (44 times 44 countries over 15 years) gives a tightly estimated θ that is slightly larger at 1.1, with adjusted $R^2 = .46$. The CES specification still comes quite close to the data, understanding that the specification does less well with the huge variation of bilateral flows in the cross section as well as over time. (Presumably, allowance for origin-specific trade elasticities would improve the fit substantially, as justified by the translog structure in Section 8.6. Investigation of the full panel is deferred to future work.) The difference between the full panel and the time series estimate for the US and China terms of trade alone is surprisingly small. This and the very small time variation of yearly calibrated θs for the US and China suggests they may be close to a long run elasticity.[21]

The non-parametric approach generally comes at the cost of inability to make probability statements about the results. The minimum distance technique permits statistical inference only if the residuals equal to $\ln \eta_{ij,t}$ evaluated at $\hat{\theta}$ are random. Even with standard statistical inference not applicable, 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 (12).

[21] The large panel suggests measurement error associated with small trade flow shares $b_{ij,t}$ (many on the order of e^{-06}). Such cases are associated with calculated negative $R_{ij,t}$s and constitute over 20% of observations, with numerous examples for almost all exporting countries and years. Equation (6) is decreasing in b_{ij}, and as b_{ij} falls the denominator of the formula falls to zero (where R_{ij} is undefined), beyond which the calculated $R_{ij} < 0$. The theory suggests that the observed small b_{ij} is a reporting error, true demand should be choked off. Such observations are uninformative about relative resistance. Note that at $b_{ij} = 0$, $R_{ij} = (\sqrt{P_j} + 1)/(\sqrt{P_j} - 1)$ from equation (6), uninformative about relative resistance. The appropriate treatment is dropping the observation since its corresponding unobservable relative resistance is a choke price rather than an informative relative resistance. Negative R_{ij} is associated with $P_j < 1$ at $b_{ij} = 0$ and also with small positive b_{ij} associated with small enough values of P_j. This suggests that measurement error in P_js plays a role in economically meaningless revealed $R_{ij} < 0$.

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 calibrator (estimator). Information methods such as AIC might then be applied for model selection between CES and non-homothetic CES and translog.

A full treatment is postponed to future work.

8.6 Translog Gravity

The general translog gravity case provides perspective on the non-parametric model above, especially its implicit aggregation of general cross effects. It also gives perspective on the CES case applied for the industrial policy counterfactual below in Section 5.

The translog expenditure share b_{ij} is given by

$$b_{ij} = \alpha_i - \sum_l \gamma_{il} \ln(c_l \tau_{lj}/P_j) = \alpha_i - \bar{\gamma}_i \ln(\bar{p}_{ij}/P_j) \tag{18}$$

where where $\bar{\gamma}_i = \sum_l \gamma_{il}$ and $\ln \bar{p}_{ij} = \ln \bar{c}_i + \ln \bar{\tau}_{ij} = \sum_l (\gamma_{il}/\bar{\gamma}_i)(\ln c_l + \ln \tau_{lj})$. Homogeneity of degree one and concavity require that the the parameters are constrained such that $\alpha_i \geq 0$; $\sum_i \alpha_i = 1$ and the matrix of the $-\gamma_{ij}$s is negative definite. Importantly, net complementarity ($\gamma_{ij} < 0, i \neq j$) is allowed. Admitting complementarity alleviates intuitive unease about its absence from standard parametric gravity models.

Projection of counterfactual changes in trade frictions or industrial policy requires the full set of translog parameters. The translog form implies a semi-parametric implicit aggregation procedure for projecting relative resistance effects on *equilibrium* b_{ij} for each bilateral link:

$$b_{ij} = \alpha_i - \bar{\gamma}_i \ln \overline{R_{ij}},$$

where $\ln \overline{R_{ij}} = \sum_l (\gamma_{il}/\bar{\gamma}_i) \ln R_{lj}$. Let N denote the number of countries. The $N \times (N-1)/2$ parameters $\{\gamma_{lj}\}$ can be identified from panel data on the N^2 shares and inferred R_{ij}s. A more tractable special case is $\gamma_{lj} = \gamma_l \gamma_j,\ \forall l \neq j;\ \gamma_{jj} = \gamma_j(1 - \gamma_j)$ where $\gamma_l \in [0,1]$ and

$\sum_l \gamma_l = 1$. In this case $\bar{\gamma}_i = \gamma_i$ and $\ln \overline{R_{ij}} = \sum_l \gamma_l \ln R_{il}$. The $2N$ parameters γ_l can be fitted from the N^2 equations $b_{ij} = \alpha_i - \gamma_i \sum_l \gamma_l \ln R_{lj}$.

Equation (18) requires amended notation to explicitly account for the variation in active links. At a point in time (suppressing the time notation), the set A_i of active links across destinations l is active links $l \in A_i$: $b_{ij} = \alpha_i - \sum_{l \in A_i} \gamma_{il} \ln(c_l \tau_{lj}/P_j) = \alpha_i - \bar{\gamma}_{ij} \ln(\bar{p}_{ij}/P_j)$. For inactive links l, $b_{il} = 0$ and efficient arbitrage implies that

$$p_{il}/P_l < c_i \Pi_i \tau_{ij} \Rightarrow \frac{p_{il}/P_l}{\tau_{il}} < c_i \Pi_i.$$

8.7 Industrial Policy Implications

Incidence shifting suggests that industrial policy may partially 'pay for itself' via improved terms of trade implied by (9). Also, the volume effect of terms of trade improvements (the rise in buyer relative price shifts more sales to foreign markets) may amplify the benefit. A simple impact accounting for industrial policy combines the two effects on the loss measure (8) where the share b_{jj}'s response to the change in R_{jj} is given by CES trade elasticity θ. The amplification is large for China and the US.

Differentiate loss measure $L_j = P_j b_{jj} - s_j$ holding P_j constant:

$$\frac{dL_j}{ds_j} = -1 + P_j \frac{db_{jj}}{ds_j} = -1 + P_j \frac{b_{jj}}{s_j} \frac{d \ln b_{jj}}{d \ln R_{jj}} \frac{d \ln R_{jj}}{d \ln s_j}.$$

The -1 on the right hand side is a loss reduction that is offset by the resource cost of obtaining it ds_j. Thus the net effect is the second term on the right. The CES specification $\Rightarrow d \ln b_{jj}/d \ln R_{jj} = -\theta$ and $b_{jj}/s_j = R_{jj}^{-\theta}$. Apply this to calculate the net gain

$$1 + \frac{dL_j}{ds_j} = -\theta P_j R_{jj}^{-\theta} \frac{\partial \ln R_{jj}}{\partial \ln s_j}. \qquad (19)$$

Use equation (9) for $\partial \ln R_{jj}/\partial \ln s_j$ in equation (19).

The net benefit of industrial policy at the margin for China and the US is calculated with equation (19). Combine the 2014 trade data with estimated θ from (13) and estimated terms of trade elasticity from (9). The average fitted trade elasticity and the 2014 calibrated trade elasticity are close but both are used along with the 2014 terms of trade elasticity for each country. In all cases there is a very substantial surplus. For 2014 China the net benefit (reduction in loss) is -2.84 with the average θ and -2.82 with the 2014 θ. For the 2014 US the net reductions in loss are -6.75 and -6.3 respectively. In 2000 the ranking is reversed. For 2000 China the net reduction in loss is -8.92 with the average θ and -9.25 with the 2000 θ. For 2000 US, the net reduction in loss is $-3.$ with the average θ and -3.22 with the 2000 θ.

The reason for the reversal in loss reduction rates for China and the US 2000 and 2014 follows from equations (19) and (9). Equation (9) is one the one hand increasing in s_j, bigger countries have more terms of trade power. On the other hand equation (9) is decreasing in terms of trade R_{jj}. Rising terms of trade also directly reduce the benefit of loss reduction directly in equation (19). The latter effect dominates in the cases of the US and China.

A full evaluation of industrial policy must set the incidence shifting benefit against unmeasured social costs such as rising marginal cost of supply in general equilibrium along with the marginal cost of public funds and other sources of distortionary loss. For example, large firms dominate international trade and may well be internalizing much of the seller incidence shifting externality. The pricing-to-market distortion is absorbed in the bilateral resistances τ_{ij}s, with sales increases presumably increasing the markups and then the τ_{ij}s. A proper evaluation of industrial policy requires data and analysis far beyond the scope of this paper.

A further qualification follows from the model. The offset reduction in loss increases with the seller's size due to its positive effect on $\partial \ln R_{jj} / \partial \ln s_j$ given by (9). Seller incidence shifting is a much weaker motive for industrial policy by smaller suppliers. The offset loss reduction also rises with θ, so higher elasticity products have a stronger case for industrial policy. And finally, the benefit of loss reduction falls as terms of trade are already good, which will vary widely across countries for geographic and basic economic as well as policy reasons.