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

James E. Anderson
Boston College and NBER

May 29, 2024

Abstract

Intense US-China commercial rivalry is quantified in this paper with novel non-parametric relative resistance sufficient statistics. The accounting method minimizes the demand specification error variance in revealed resistances. China's manufacturing seller incidence falls (seller price rises) 7.6% yearly as China's sales share quadruples over 2000-14. US seller incidence rises 4.1% yearly as US sales share halves. Domestic trade shares closely fit revealed relative resistances with trade elasticity equal to one. Industrial policy pays for itself in suggestive projections. A 10% rise in US 2014 sales share reduces seller incidence 6.0%, exports rise and net benefit is positive.

JEL codes: F10, F14

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

*I am grateful to Danial Lashkari and Arthur Lewbel for 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 various referees especially pushed 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 neglected amplifier of international commercial rivalry.[1] Asymmetric growth of national sales reduces the faster growing seller's incidence of trade frictions. Seller incidence of its lagging rivals rises on average. Since net seller prices in world markets are inversely proportional to the seller incidence of trade frictions, the big revealed seller incidence shifts in manufacturing reported in this paper matter big-time. Quantification is based on a novel non-parametric gravity model accounting. Sufficient statistics for seller incidence and related relative trade frictions are freed from dependence on restrictive parametric specifications and their estimated parameters.

Commercial rivalry in manufacturing between China and the US, 2000-2014 is the focus of the application. Figures 1 and 2 illustrate the close co-movement of revealed inverse seller incidence with sales shares. China's share of world manufacturing sales quadruples while the US share is halved. The association is summarized by yearly average rates of change. Sales share shifts account for a yearly average fall in China's seller incidence of -7.6% and a yearly average rise in US seller incidence of 4.1%. The association of seller incidence with trade shares is analytically derived in the non-parametric gravity accounting model developed below that generates the seller incidences.

[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). This paper's focus is on the consequential commercial rivalry in manufacturing trade between China and the US. 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

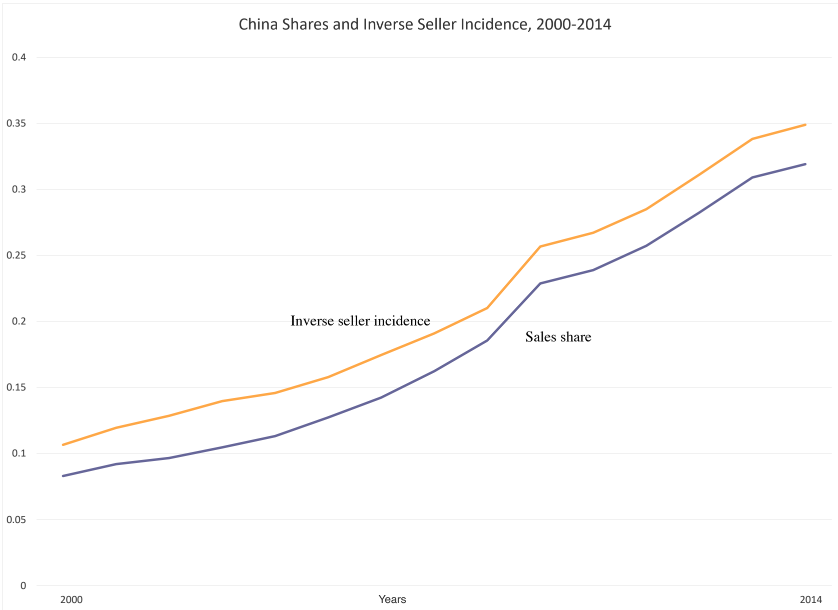
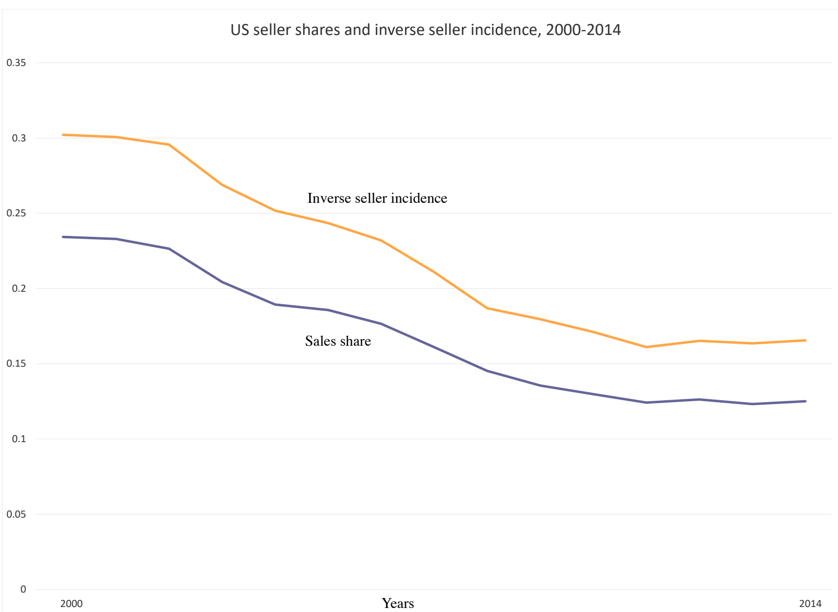


Figure 2: US Decline



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. The World Input Output Database (WIOD) reveals that (i) Much of China's manufacturing growth went to domestic sales, since China's domestic manufacturing sales share rises 2000-2014 in the 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. Rejection of China's admission to the WTO in 2000 (keeping trade policy uncertainty in place) or even escalation with higher tariffs would thus act on a relatively unimportant margin. These facts in the lens of the model suggest that offsetting trade policy sufficient to eliminate the yearly -4.1% fall in seller price (implied by the 4.1% rise in seller's incidence) would have been very costly if not infeasible.

In contrast, the seller incidence shifting mechanism suggests that industrial policy aimed at raising sales share to internalize the incidence shifting externality may have been useful. A rise in sales reduces seller incidence and thus also raises buyer incidence. The latter effect reduces domestic sales and amplifies the net benefit to the industrial policy. A simple counterfactual impact analysis quantified below confirms this suggestion – the 2014 US marginal net benefit is large enough to offset the likely un-modeled costs of the policy.

The net benefit is proportional to the the terms of trade, which is inversely proportional to seller incidence. Quantification of terms of trade effects of supply share changes below reveals that China's terms of trade in manufacturing improve by an average yearly 8.2% while US terms of trade deteriorate by an an average yearly -4.7%. Seller incidence variation thus accounts for much of the terms of trade variation for both countries.[2]

[2]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

A non-parametric compensating variation loss measure of the national interest is developed below. The measure 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 national interest moves with the US terms of trade in manufacturing since a terms of trade improvement reduces the domestic demand share and thus reduces the distance between the domestic and as-if-frictionless shares.[3] 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 in distribution is removed. The average yearly changes in the negative of the loss measure (the gains from trade) are 1.8% for China and −1.5% for the US.

The loss measure is related in Section 3.3.1 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.

The loss measure is to quantify counterfactual industrial policy for the US and China in Section 5.3. Seller incidence reduction suffices to make marginal subsidized sales increases 'pay for themselves'. Thus policy to internalize of the seller incidence externality may be worthwhile. The policy implication is suggestive only, because it omits quantification of plau-

contrary incidence shifting effect when trade is subject to frictions is explained below in Appendix Section 7.1.

[3]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.

sible other costs to the policy.[4] The suggestive benefit of industrial policy further suggests the importance of negotiation to coordinate and constrain such national production subsidy policies.[5] The modeling implies that coordination should be between the small number of big sellers.[6]

The counterfactual seller incidence shifts that drive the welfare effects of industrial policy are based on analytic first order impact elasticities of seller incidence with respect to sales shares. The US and China have large seller incidence elasticities due to their relatively large shares of world manufacturing. In 2014 these are -0.6 for the US and -0.67 for China. The welfare impact measure combines these elasticities amplified by the increase in the share of exports in total sales the terms of trade improvement reduces the domestic share of sales. The amplification of the seller incidence elasticities above for the 2014 US is 6 and for China is 3. US loss from trade frictions falls -3.6% (gains from trade rises 3.6%) while China loss falls -1.95%. The marginal net benefit estimates are far more than sufficient to fully 'pay for' the policy in the crude sense of covering the all else equal resource cost of a marginal share increase.

The counterfactual welfare calculations use the CES model to generate the induced fall in domestic expenditure share as the terms of trade rise due to industrial policy change. Minimum distance calibration of the CES trade elasticity to fit the variation of log trade

[4] The omitted costs include accounting for the market power distortions of big firms, the cost of paying for production subsidies with distortionary taxation and a full general equilibrium quantification of supply cost changes. Quantification of this set of costs is outside the focus of this paper.

[5] The analytic expressions derived from the non-parametric accounting model generally imply that, all else equal, seller incidence of trade frictions (or outward multilateral resistance) is decreasing in own supply share. A negative international externality is implied since shares must sum to one. Thus a rise in one country's share must on average reduce other countries' shares – global crowding out. Decomposition based on this accounting property suggests that the cross effects from China's growth are large. For example the US seller incidence rise is mostly 'accounted for' by the mechanical effect of China's faster than average sales rise on reducing the US share, thereby raising US seller incidence. Assume that the effect on the US is equal to the average effect on all of China's partners. Let s_i denote country i's share and let $\hat{s}_i$ its percentage change. The adding up condition when all of China's partners have the same $\hat{s}_{US}$ percentage change implies $\hat{s}_{US} = -\hat{S}_{CN} s_{CN}/(1 - CN)$. In 2001, the cross effect counts for 39% of the US share fall, while in 2013 it accounts for more than 100%. The net effect arises because the numerator of the US share changes at the same time, amplifying or decreasing the simple accounting effect.

[6] The "large enough" qualification is explained in Section 5.3. It is more involved than the simple intuition that small sellers do not much affect world markets.

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 trade elasticity suggests that previous trade elasticity estimators are biased upward (in absolute value). Section 5.2 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.[7] Thus larger non-pecuniary cost implies lower observed price, and the inferred elasticity must be larger to explain the observed variation in expenditure. Revealed variation of non-pecuniary costs and tastes is an important benefit of relative resistance statistics. The generated data can be used to calibrate minimum distance fits to flexible functional forms suitable for addressing a range of concerns beyond the scope of this paper. For example, the non-homothetic CES is a usefully simple specification that yields constant relative income elasticities. Section 7.3 illustrates calibration for the translog case.

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

[7]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).

analysis purposes. The external scale effects in distribution thus complement the scale effects in production that are the focus of Bartelme et al. (2019).

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 of 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 non-parametric gravity accounting model that results also applies to the set of one factor production models that are observationally equivalent to the endowments model, Arkolakis et al. (2012). The sectoral module focus is consistent with the political economy concerns that drive typical trade policy. For economy-wide analysis, the sector module nests within a class of general equilibrium superstructures.

Section 1 is a non-technical perspective on the non-parametric gravity approach. The analysis begins in Section 2 with unappreciated properties of efficient spatial arbitrage. These properties provide useful intuition when combined with the non-parametric model of buyer willingness-to-pay in Section 3. Multiple goods from a set of origins are distributed to buyers at different destinations who face different equilibrium price vectors for the goods due to trade frictions. Market clearing for each product combines with the budget constraint in each destination to complete the model elements. The intermediate value theorem is applied to derive non-parametric relative resistances from observed shares and buyer price indexes. The unknown intermediate weight of the theorem reflects an unknown demand system specification from the set of invertible demand systems. The expected approximation error variance due to selecting a specification is minimized by the weight equal to 1/2, provided that beliefs are symmetric on the unit interval. The weight equal to 1/2 implies the general translog specification, hence operationality. The resulting accounting mode yields sufficient statistics for relative resistance and seller incidence that do not require the translog parameters. The revealed statistics for China and US manufacturing are reported and discussed in Section 4, followed by projections of the effects of industrial policy in Section 5.

1 Non-parametric Gravity 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 an invertible demand system applicable to all destinations. Parametric gravity adds restrictive parametric demand system specifications. Non-parametric gravity relaxes the third assumption. It delivers relative resistances that are consistent with a wide class of demand system specifications. Remarkably, it consistently aggregates all third party frictions that affect bilateral trade both directly and indirectly via multilateral resistances valid for the wide class of specifications. The consistency property also applies to observed demand systems at any level of aggregation of either sectors or locations.

Each destination faces different effective price vectors (inverse demand vectors) due to bilateral resistances that are equal to trade friction factors that include taste shifters. Observable demand shares differ from the observable world demand shares (all at buyer prices) due to the differences in effective price vectors. Taste differences across buyers act like effective price shifters that are absorbed in the spatial arbitrage context as 'trade frictions'. Moreover, because utilities (or activity levels in the intermediate inputs case) are given in equilibrium, non-homothetic income (activity) effects 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. The shares difference and the invertible common demand system 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.[8] 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.

The best choice of specification to approximate the unknown 'true' invertible demand system is the Toïnqvist approximation: intermediate value weight equal to 1/2. The intermediate weight equal to 1/2 is exactly consistent with the general translog specification. The Toïnqvist approximation is shown to minimize the approximation error variance in relative resistances associated specification choice from the set of invertible demand systems. The projected intermediate share and the projected intermediate price index are known functions of observables. The revealed relative resistances are solved from the resulting exact accounting system.

Non-parametric gravity based on the translog demand system appears to be at the upper limit on extracting information about relative resistances from observed trade within the broader class of invertible demand systems. The translog minimizes the variance of the approximation error associated with choice of specification and is itself understood as a second order approximation to any homothetic demand system generated by cost minimization.

Relative resistance on domestic trade is equal to the actual equilibrium terms of trade relative to the terms of trade that obtain in the as-if-frictionless equilibrium. Sellers' relative incidence, or outward multilateral resistance relative to domestic resistance, is solved as the inverse of the product of domestic relative resistance and the buyer price index. Both the sellers' relative incidence and the relative terms of trade permit comparison of the performance

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

of countries through time, the former for the sellers and the latter for general welfare. The reported results focus on these, but the method generates sufficient statistics for bilateral relative resistance to trade between all active pairs of locations at all time periods.

The cross-section variation of relative resistance with sales shares accounts below for much of the revealed variation in seller incidence. Extrapolation to comparative statics that relate to movement over time is approximated by cross-section partial elasticities that give first order impact effects. These appear to account for a large part of the revealed time series changes. The impact effect method is used to decompose the equilibrium sellers' incidence and terms of trade changes of the US into those due to its own specialization and to other forces including China's growth effect on the US. The impact effect analysis provides a useful perspective on the effects of the China shock on the US.[9]

Future research may also 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 also apply, understanding that the endogenous bilateral frictions include endogenous markups by firms. The general relevance of spatial arbitrage equilibrium and associated seller incidence shifting may be an important part of explaining the dynamics of star firms and failing competitors. 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.[10] t

[9]The full changes could in principle be explained by a full general equilibrium model of the world economy. The first order approximation avoids the data complexities, questionable specifications and parameterizations of such a model.

[10]Calculating the choke value requires projections that apply a parametric demand system along with the observable resistances. In this context, note the difficulties suggested by demand systems where buyers in different locations face different effective price vectors due to taste shifters. As well as the demand parameters, the taste shifters must be parameterized.

2 Efficient Arbitrage

Efficient spatial arbitrage is the foundation of gravity models. Arbitrage imposes a powerful discipline on the equilibrium distribution of goods subject to trade frictions. Previously unappreciated properties of this discipline are set out below. In the class of non-parametric gravity models, a single bilateral relative resistance statistic explains the difference between observed bilateral and as-if-frictionless shares. The properties of efficient arbitrage deepen intuition about why relative resistance suffices.

The arbitrageurs' problem is to efficiently allocate a given total supply y_i to multiple destinations. For expositional simplicity, arbitrageurs are competitive. (Monopolistic behavior is analyzed later.) Arbitrageurs take as given the willingness to pay at each destination p_{ij}, not internalizing the fall in p_{ij} as delivered product x_{ij} increases. In any single link, $\bar{t}_{ij}$ is the metaphorical iceberg melting trade friction and c_i is the unit seller cost that must be covered before delivery. The collective arbitrageurs' problem is modeled as:

$$\max_{\{x_{ij}\}} \sum_j p_{ij} x_{ij} | \sum_j \bar{t}_{ij} x_{ij} \leq y_i. \tag{1}$$

The Lagrange multiplier μ_i on the constraint is interpreted as the opportunity cost of serving any particular market j with another unit of good i.

The first order conditions imply $p_{ij} = \mu_i \bar{t}_{ij},\ \forall x_{ij} > 0; p_{il} < \mu_i \bar{t}_{il}\ \forall x_{il} = 0$. The opportunity cost is decomposed economically below as $c_i \Pi_i$, the product of the net seller cost c_i and the average seller incidence of frictions cost Π_i. Π_i is also called outward multilateral resistance in the gravity literature.

Willingness to pay p_{ij} relative to opportunity cost is $p_{ij}/c_i\Pi_i$. In arbitrage equilibrium with zero profits $p_{ij}/c_i\Pi_i = \bar{t}_{ij}$. The left hand side is interpreted as the buyer's incidence of bilateral trade costs. $p_{ij}/c_i = \Pi_i \bar{t}_{ij}$ is the full equilibrium cost of trade frictions.

In what follows below, it is convenient to work in terms of relative buyer prices p_{ij}/P_j where P_j is the buyer price index. In the equilibrium allocation $p_{ij}/P_j = c_i \Pi_i \bar{t}_{ij}$. Divide by

$c_i\Pi_i$ and simplify to $R_{ij} = \tau_{ij}/\Pi_i P_j = \bar{t}_{ij}$ where $\tau_{ij} = p_{ij}/c_i$. Then relative resistance

$$R_{ij} \equiv \frac{\tau_{ij}}{\Pi_i P_j}$$

characterizes arbitrage equilibrium. All elements of R_{ij} are evidently endogenous general equilibrium variables, as is the bilateral friction $\bar{t}_{ij}$ in (2) when interpreted in general equilibrium.

Summing the first order conditions implies $\sum_j p_{ij}x_{ij} = \mu_i \sum_j \bar{t}_{ij}x_{ij} = \mu_i y_i$. Replace μ_i with $c_i\Pi_i$ and solve:

$$\Pi_i = \sum_j \frac{p_{ij}x_{ij}}{c_i y_i}. \tag{2}$$

Equation (2) gives Π_i as the ratio of sales at buyer prices to sales at seller prices. Thus Π_i is interpreted as the average seller incidence of trade friction costs.

The competitive arbitrage assumption can be relaxed to allow efficient monopoly markups on bilateral links. Efficiency means that however the friction τ_{ij} is determined, the equilibrium allocation of the x_{ij}s at the equilibrium τ_{ij}s is efficient as defined in equation (1). Thus the τ_{ij}s in equilibrium are interpreted as containing the monopoly markups. The equilibrium τ_{ij} also may vary endogenously with bilateral trade volume due to fixed bilateral link capacity or fixed bilateral link entry cost. The efficient arbitrage property moreover implies that a wide range of heterogeneous buyer behavior may be absorbed into equilibrium τ_{ij}s, and thus implicitly aggregated into the revealed relative resistances below. The theme of implicit aggregation into revealed relative resistance keeps playing through the remainder of the paper.

Efficient arbitrage nests within a superstructure that determines a equilibrium set of buyer willingness to pay buyer prices $\{p_{ij}\}$ in destinations j that are served, given the amounts of a products y_i. The superstructure links shipments x_{ij} back to buyer prices $\{p_{ij}\}$ in destinations served. Some also add restrictive constant elasticity superstructure that determines the amounts of a set of possible products that are actually produced in each location i. Non-parametric gravity as developed here imposes minimal restriction on the demand superstructure and assumes that supply is given.[11]

Maximization problem (1) implies that a rise in supply constraint y_i raises sales at given buyer prices p_{ij} by $c_i\Pi_i$ while for larger changes it incipiently reduces the Lagrange multiplier $\mu_i = c_i\Pi_i$. A central quantitative concern of the application below to discrete changes is how much of the fall in $c_i\Pi_i$ fall is absorbed by a fall in sellers incidence Π_i. The pictures above show how dramatic the measured incidence shifting effect is for manufacturing by China and the US, 2000-2014.

Accounting for discrete changes in supply and their effects requires a demand model to impute the price changes to $\{p_{ij}\}$. Foreshadowing the analysis, contrast the no trade case with one destination and a vertical supply schedule with the multiple destinations case of trade. In the no trade case, the incidence of friction τ is borne by the seller. Changes in supply move the vertical supply schedule along a downward sloping demand schedule. At constant τ, the change in $c\Pi$ falls entirely on the seller's price c whereas a fall in distribution friction τ implies an equal fall in Π. With multiple destinations, in contrast, shifts in upward sloping residual supply move along downward sloping demand in each bilateral market. In the formal model that follows, the heterogeneity of frictions τ_{ij} across destinations j makes for rich interaction of supply shifts with the behavior of inverse demand schedules p_{ij}.

Intuition for the results depicted in the introduction is suggested by the pattern of international trade in manufactures (and almost all other products) – home bias, local sales shares are larger than the product's global sales share. The pattern suggests $\tau_{ii} < \tau_{ij}, \forall j \neq i$. Then the distribution of a discrete rise in y_i will tend to disproportionately favor local sales, hence x_{ii}/y_i will rise. At constant p_{ij}s this implies a fall in Π_i. Of course, the p_{ij}s will change. Accounting for the changes requires the formal model below.

[11]This strategy avoids apparent limitations such as independence of irrelevant alternatives in demand structure, while the given supply assumption avoids an equivalent limitation to supply structure due to the assumption of heterogeneous firm or sector productivities being independent draws from an identical probability distribution.

3 Non-parametric Gravity Modeling

A brief review of CES gravity is a useful starting point. The buyer's effective price p_{ij} in destination j for good i in arbitrage equilibrium is equal to the product of the net seller price c_i and the friction τ_{ij}, itself a product of trade frictions and taste shifters, both of which are origin-destination specific in principle. The demand share $b_{ij} = p_{ij}x_{ij}/\sum_i p_{ij}x_{ij}$ in the CES demand share 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 implied by the budget constraint $\sum_i b_{ij} = 1$.

In arbitrage equilibrium, the supply y_i of goods from each country is efficiently distributed as in Section 2 with total sales at buyer prices $c_i\Pi_i y_i$, $\forall i$. Each country's sales to the world have shares $s_i = c_i\Pi_i y_i/\sum_l c_l\Pi_l y_l$. The same preferences (up to the influence of the taste shifters absorbed into τ_{ij}s) apply to all destinations, so it is as if sellers faced a single buyer on an as-if-frictionless world market with shares $B_i = (c_i\Pi_i y_i)^{-\theta}$. Here the world budget constraint $\sum_i B_i = 1$ is used to normalize the as-if-frictionless world price index to one. The relative resistances are all equal to one since all buyers in the world face the same price vector. The CES model implies

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

The budget constraint for the world economy implies

$$\sum_{i,j} s_i \left(\frac{\tau_{ij}}{\Pi_i P_j}\right)^{-\theta} = 1.$$

The constraint imposes a normalization on the set of actual price indexes P_j and implies that the normalized prices measure overall buyer incidence of trade costs for each country. The normalization also implies that the world set of relative resistances is normalized. Comparability of relative resistances across countries is thus assured.

Relative resistance is solved as $R_{ij} = (b_{ij}/s_i)^{-1/\theta}$. The ratio b_{ij}/B_i suggests the role of trade frictions in general, with home bias observed in $b_{jj}/s_j > 1$ and $b_{ij} < s_i$; $\forall i \neq j$. The quantitative solution for R_{ij} requires trust in the CES specification and trust in the estimate of the trade elasticity parameter θ. Qualms about the restrictiveness of the CES specification and doubts about the accuracy of θ estimates motivate the non-parametric approach.

The key idea from the structural approach is retained – infer relative resistances from the difference between the buyer's expenditure shares pattern facing the actual trade frictions and the expenditure shares pattern the same buyer would hypothetically have in an 'as-if-frictionless' world. The expenditure shares of the as-if-frictionless equilibrium are observable because they are equal to the actual world sales shares at buyer prices. The difference in each country's expenditure function from the expenditure function it would face in the as-if-frictionless world equilibrium is decomposed by application of the intermediate value theorem and Shephard's Lemma. The intermediate value theorem applies because the common demand system is 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 rich set of cross-effects in demand that enter into the determination of bilateral expenditure shares may be regarded as implicitly aggregated into bilateral relative resistances $\overline{R_{ij}} = \bar{\tau}_{ij}/\bar{\Pi}_i P_j$ where the bars denote implicit aggregation. The implicit aggregation takes explicit form in the translog structure developed in Appendix section 7.3.

3.1 Demand Model

x_{ij} is the amount of goods from origin i purchased by destination j buyers in arbitrage equilibrium. The set of prices $\{p_{ij}\}$ give the buyers per unit willingness to pay associated with the set of amounts purchased $\{x_{ij}\}$. The equilibrium value of buyers' willingness to pay p_{ij} for a marginal increase in x_{ij} is based on cost-minimizing selection of amounts. The application below focuses on the manufacturing sub-set of goods, implying that the choice problem separably nests inside an external choice superstructure. Manufacturing includes deliveries to both final consumers and intermediate input buyers, hence necessarily imposes the same structure on buyers. The exposition focuses on the final consumer, exploiting the common cost minimization foundation of expenditure and cost functions.

Expenditure by buyers is represented by the expenditure function $e(\mathbf{p}, u)$, homogeneous of degree one and concave in the price vector $\mathbf{p}$ and increasing in (sub)-utility u. Shephard's Lemma ($\partial e/\partial p_i = x_i$) implies that the buyer's expenditure share b_i on each good i is equal to $\partial \ln e/\partial \ln p_i$. Restrictions on $e(\cdot)$ are imposed in steps to reach operationality.

The first restriction is that all the effects of utility (real income) variation and other sources of heterogeneity in tastes operate as effective price shifters on the expenditure function $e(\cdot)$. Specifically, $e(\mathbf{p}, u) = e(\{p_i q_i(u)\})u$ is concave and homogeneous of degree one in effective prices $\mathbf{p} = \{p_i q_i(u)\}$. The taste shifters $q_i(u)$ include income effects via dependence on real income u. Thus price-dependent taste shifters may vary across buyers as real income u changes. Location j specific effective prices are p_{ij}, including taste shifters q_{ij}. Shephard's Lemma applied to this expenditure function implies origin-destination expenditure shares $b_{ij} \equiv p_{ij}x_{ij}/E_j = \partial \ln e(\mathbf{p}^j)/\partial \ln(p_{ij})$. The vector of shares $\{b_i(\mathbf{p}^j)\}$ is independent of utility u^j for given equilibrium taste shifters $\{q_i(u^j)\}$.[12]

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 $\mathbf{p}^*$. Endowments are constant and so are real incomes $\{u^j\}$. World expenditure remains equal to world sales, or $\sum_j [e(\mathbf{p}^j) - \sum_j e(\mathbf{p}^*)]u^j = 0$. World expenditure shares $B_i(\mathbf{p}^*)$ satisfy Shephard's Lemma, and $\sum_i B_i(\cdot) = 1 \Rightarrow P^* = 1$.

The setup implies a key normalization on observed buyer price indexes to be used in applications. The adding up condition for all world sales at buyer prices gives the normalization

[12] Consistency with the axioms of revealed preference requires restrictions on the $\{q_i(u)\}$ that are of no concern here. The Constant Relative Income Elasticities class of expenditure functions generated by the non-homothetic CES specification is an example.

for world price indexes:

$$\sum_j \sum_i p_{ij} \frac{x_{ij}}{E_j} E_j = \sum_j E_j \Rightarrow \sum_j P_j \frac{E_j}{\sum_j E_j} = 1. \tag{3}$$

Thus the normalization of observed price indexes (3) is consistent with normalization of the as-if-frictionless price index: $\sum_j P_j u^j / \sum_j u^j = e(\mathbf{p}^*) = 1$.

$P_j - 1 = e(\mathbf{p}^j) - e(\mathbf{p}^*)$ is the difference in country j's unit cost of utility at observed and as-if-frictionless prices. Use Shephard's Lemma to expand this difference:

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

The intermediate value theorem applied to the unit cost of utility implies that this difference is equal to

$$\sum_i (P_j b_{ij} - B_i) = \sum_i \tilde{P}_j \tilde{b}_{ij} \frac{p_{ij} - p_i^*}{\lambda_j p_{ij} + (1 - \lambda_j) p_i^*} \tag{4}$$

for some $\lambda_j \in [0, 1]$.

3.1.1 Relative Resistance Inference

Relative resistance is inferred from each element of the sum in (4) by applying equality element-by-element:

$$P_j b_{ij} - B_i = \tilde{P}_j \tilde{b}_{ij} \frac{p_{ij} - p_i^*}{\lambda_j p_{ij} + (1 - \lambda_j) p_i^*}. \tag{5}$$

Divide numerator and denominator of the ratio on the right hand side by $p_i^* = c_i \Pi_i$ to yield $(R_{ij} - 1)/(\lambda_j R_{ij} + 1 - \lambda_j)$. The right hand side of equation (5) varies with the unknown true value of $\lambda_j \in [0, 1]$ both directly in the percentage change terms and indirectly due to its implications for the discrete elasticity term $\tilde{P}_j \tilde{b}_{ij}$. A choice of λ_j is effectively a choice of a specification based on a belief that it is a good approximation.

A specification choice is equivalent to acting on a belief z that $\lambda_j(z),\ z \in [0, 1]$ is true, knowing it may be false. Suppose initially all $\lambda(z),\ z \in [0, 1]$ are equally likely – i.e., probabilities 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 $V(r_{ij}) = E[r_{ij}(\lambda_j(z)) - r_{ij}(\bar{\lambda})]^2$ where the expectation is taken over the distribution of z.

Proposition 1

$\lambda_j = 1/2$ minimizes the approximation error variance.

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.* $\|$

$\lambda_j = 1/2$ wonderfully also implies the general translog specification with zero expected approximation error. The variance minimizing argument extends to include all belief distributions that are symmetric around the mean.[13] $\lambda_j = 1/2$ (the Törnqvist approximation) implies $\tilde{b}_{ij} = (b_{ij} + s_i)/2$ and $\tilde{P}_j = \sqrt{P_j}$. The translog parameters are not needed to reveal relative resistances. If the translog is the true demand model, it is exact and equation (5) with $\lambda_j = 1/2$ yields operational exact non-parametric relative resistance indexes, given the absence of measurement error.

The elements given by equation (5) are equated to their intermediate values with $\lambda_j = 1/2$ as

$$P_j b_{ij} - s_i = \sqrt{P_j}\ \bar{b}_{ij}\frac{R_{ij} - 1}{(R_{ij} + 1)/2}. \tag{6}$$

Equation (6) can be solved for R_{ij}.

Proposition 2 *Revealed relative resistances are given by*

$$R_{ij} = \frac{2\bar{b}_{ij}\sqrt{P_j} - (P_j b_{ij} - s_i)}{2\bar{b}_{ij}\sqrt{P_j} + (P_j b_{ij} - s_i)};\ \forall i, j. \tag{7}$$

Seller incidence Π_j is revealed from inverting $P_j R_{jj}$ where (7) is used for R_{jj}.

[13] The analysis is closely related to price index theory but the latter does not have a similar variance minimizing property. This is because R_{ij} is linear in the choice of belief while the price index being projected is non-linear in the choice of belief.

Figure 3: Revealed R_{jj} Logic

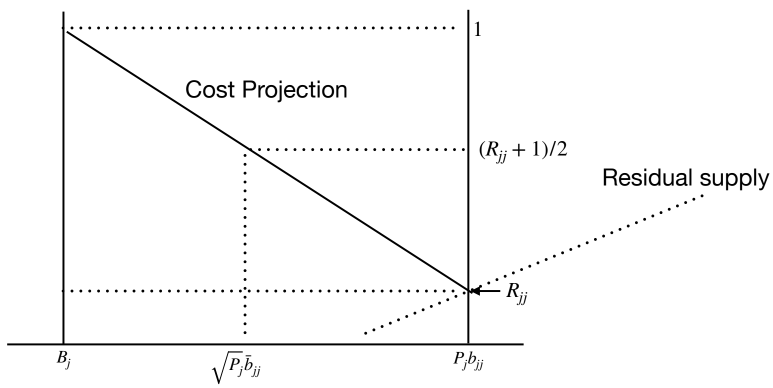


Figure 3 illustrates the logic of equations (5) and (6) and provides insight into how the intermediate value theorem enables non-parametric calculation of relative resistances. The diagram focuses on the case $i = j$. As-if-frictionless demand shares, equal to sales shares at buyer prices, are generated from the common price vector $\mathbf{p}^*$.

The right vertical axis at horizontal coordinate $P_j b_{jj}$ 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 (5). (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 $\tilde{P}_j(z)\tilde{b}_{jj}(z) \in [B_i, P_j b_{jj}]$ 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 (5) uses ratios of

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

with slope -1.

The translog specification $\lambda_j = 1/2$ selects the midpoint on the horizontal axis between B_i and $P_j b_{jj}$ with associated value $\sqrt{P_j}\,\bar{b}_{jj}$. 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 (7).

Figure 3 also gives intuition about sensitivity to approximation error due to the translog restriction. The unkown 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 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 $[0,1]$.

Revealed relative resistances vary with seller size. All else equal, equation (7) implies that relative resistance R_{ij} is increasing in s_i; $\forall i,j$:

$$\frac{\partial \ln R_{ij}}{\partial \ln s_i} = \frac{s_i}{2\bar{b}_{ij}\sqrt{P_j} - (P_j b_{ij} - s_i)}(1+R_{ij}) > 0. \tag{8}$$

This intuitive sharp result implies that in the cross section, larger countries 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 (8) also applies to the impact effect of supply share *changes* on relative resistance. Thus the positive sign of (8) helps explain the results displayed in the graphs above showing perfect positive correlation of inverse seller incidence $1/\Pi_i$ and sales shares s_i for China and the US. This is because seller incidence changes dominate the movement of $R_{ii} = \tau_{ii}/\Pi_i P_i$ in the data. Full general equilibrium comparative statics combine impact effect (8) with knock-on changes to P_j and the bilateral frictions τ_{ij} that blur this quantification but the intuition is likely hold.

The application focuses on the case of domestic trade, $i = j$. Relative resistance here is recognized as the terms of trade for the manufacturing sector. Standard measures of the terms of trade have well known deficiencies. 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 importantly prices do not contain unobserved user costs, costs that vary across users and product types. Non-parametric gravity measure (7) 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. The applications below in Section 4 assume the accuracy of the WIOD data.

Appendix Section 7.3 shows how translog parameters may be calibrated from the relationship of observed shares and revealed relative resistances for use in projections. Alternatively, a CES approximation can be calibrated using $\ln(b_{ij}/s_i) = -\theta_i \ln R_{ij}$ as a check on how closely the familiar CES restriction comes. In the application below, the minimum distance estimates of θ_i for China and the US fit extremely well and yield values close to 1.

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 the translog aggregation of own and cross effects 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.

Non-parametric gravity has a natural treatment of the zeros problem that arises with CES gravity. Relative resistance associated with zero bilateral flows is unknown but exceeds the value that chokes off trade. Thus net willingness to pay p_{ij}/τ_{ij} is less than the value that would cover the excess of bilateral shipment cost relative to its opportunity cost $c_i\Pi_i$. Relative resistance on positive flows can be measured because willingness to pay equals the excess of bilateral shipment cost relative to opportunity cost.

The cross-section variation in the number of active links (variation in the pattern of zeros) is thus appropriately accounted for by the relative resistances. In panel settings, each cross section is a static equilibrium and variation in the active links (entry or exit) is due to shifts in the relationship of net willingness to pay to opportunity cost. The implication is that a simple calibration of translog share structure on the positive shares alone is not subject to selection bias. The alternative applied below in Section 5.2 is to calibrate a local CES trade elasticity on the positive shares alone. It closely approximates a restricted translog structure developed below.

When zeros switch off or on over time, relative resistances remain useful as measures of the distorting effect of frictions. But an important consequence is that domestic relative resistance interpreted as the terms of trade no longer reliably links change in domestic resistance to compensated real income change. A fall in domestic relative resistance could indicate a welfare improving ability to pay for new expensive products while a rise could indicate the reverse force.[14]

[14] A properly adjusted terms of trade to associate with compensated real income change requires a model that links real income to the price-dependent income effect shifters. Such a model can remove the income effects from the τ_{ij}s. It is beyond the scope of this paper.

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 (6) 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 $P_j b_{ij} - s_i$. Why choose the translog specification over any other specification that may be false. 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 (7) at $\lambda_j = 1/2$. The partial 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 $\lambda_j = 1/2$.[15]

$$\frac{\partial \ln R_{jj}}{\partial \ln \lambda_j} = \frac{\partial \ln R_{jj}}{\partial \ln \tilde{P}_j \tilde{b}_{ij}} \frac{\partial \ln \tilde{P}_j\ \tilde{b}_{jj}}{\partial \ln \lambda_j} = \frac{\sqrt{P_j}\ \bar{b}_{jj}}{2\sqrt{P_j}\ \bar{b}_{jj} - s_j}(b_{jj} - s_j)(1 - R_{jj}) \tag{9}$$

For China and the US in manufacturing 2000-2014, the sensitivity elasticities in equation (9) 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, the general translog

[15] Equation (9) uses $\partial \tilde{P}_j / \partial \lambda_j = (1/2) \partial P_j / \partial \lambda_j = 0$.

has 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. 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.

Approximation error also affects the terms of trade elasticity $\partial \ln R_{ij}/\partial \ln s_i$. Its quantification is crucial for the evaluation of industrial policy in the calculations below. Perspective on its value in equation (8) 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 (6). Both cases reduce the ratio on the right hand side of equation (8) 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 0.60, 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 0.67 while the elasticity is equal to 0 at $\lambda_j = 0$.) It is plausible to assume that the terms of trade elasticity is monotonically increasing in λ_j, but this is only guaranteed with a regularity condition on $\tilde{b}_{jj}\tilde{P}_j$. Nevertheless, the location of the terms of trade elasticities as comfortably in the middle of their ranges provides a perspective check on the adequacy of equation (8) and its association with the revealed relative resistance statistics.

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. 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 rearranging $P_j - 1 = \sum_i (P_j b_{ij} - s_i)$ to yield

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

The loss $L_j \equiv P_j b_{jj} - s_j$ is due to frictions on the right hand side (both on average and due to cross-border imports) but is simplified on the left hand side to a measure based on domestic sales. Using equation (6) for $i = j$ and $\epsilon_{jj} = 0$:

$$L_j = -\sqrt{P}_j \bar{b}_{jj} \frac{R_{jj} - 1}{(R_{jj} + 1)/2}, \tag{10}$$

where the right hand side uses the domestic trade case of equation (6). Note that $R_{jj} < 1$ (almost) universally. Loss measure (10) is a (negative) measure of gains from trade. The loss falls as the terms of trade R_{jj} rises, holding all else equal in the cross-section of countries. As R_{jj} rises toward one, relative loss $L_j \to 0$, cross-border trade bears the average cost of frictions. As $\bar{b}_{jj}$ rises, the loss rises, reflecting the volume effect of relatively low R_{jj} that reduces b_{jj} and thus $\bar{b}_{jj}$.

Loss measure (10) is an analog to the sectoral rate of effective protection defined in the spatial arbitrage setting.

In wider perspective, economic gravity characterized by (10) and (7) pleasingly 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 inverse square of distance interpretation of terms of trade follows from interpreting the denominator of R_{jj} as the square of the geometric mean of inward and outward multilateral resistances while the numerator of R_{jj} is understood as (an index of) the square of the geometric mean of inward and outward resistances on shipments between domestic locations.[16]

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 (10). Equation (10) 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 (10) is useful for non-parametric *ex post* evaluation of change in arbitrage gains from trade in a single sector. The application below uses (10) to quantify the differing welfare effects of globalization on manufacturing in China and the US.

3.3.1 Relationship to CES Gains Measure

Non-parametric loss measure (10) 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. Loss measure (10) differs in ways that are clarified below, with convergence in a special case.

In contrast to Arkolakis et al. (2012), the non-parametric loss measure (10) allows for changes in domestic frictions and endowments as well as foreign ones. This is a crucial advantage when national sales share changes are large, as in the applications below to man-

[16] Appendix section 7.2 deals with internal economic distance formally.

ufacturing trade and the gains from trade changes of the US and China during the globalization era, 2000-14. A potential disadvantage is that non-parametric (10) is a compensating variation 'real gains' measure, in contrast to the parametric equivalent variation real income measure (11) below.

Arkolakis et al. (2012) show that under the CES demand specification, the observed domestic share b_{jj} and the hypothetical autarky share $b_{jj}^A = 1$ 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_j^A. The gains from trade relative to autarky are measured by

$$G_j = b_{jj}^{-1/\theta} = s_j^{-1/\theta} R_{jj}, \tag{11}$$

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 $b_{jj,t}^A = b_{jj,t-1}^A = 1$. In relative form,

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

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 (11). 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:

$$\frac{RL_{j,t}}{RL_{j,t-1}} = \frac{P_{j,t}}{P_{j,t-1}} \left(\frac{R_{jj,t}}{R_{jj,t-1}}\right)^{-\theta} \tag{12}$$

Apply the $-1/\theta$ power transform to equation (12) that converts loss to gains and completes the comparability to (11). The result is a gains expression in proportional change terms that approaches the Arkolakis et al. (2012) expression above as $s_{j,t}/s_{j,t-1}$ approaches the ratio of normalized price indexes $P_{j,t}/P_{j,t-1}$. 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 CES preferences, the two are equal.[17]

While the quantitative *measures* are the same in this limiting case, the relative gains proportional changes differ in interpretation. (11) is an equivalent variation measure while (12) is a compensating variation loss measure. For the former, the no domestic changes assumption means that u_j^A does not change. For the latter, as-if-frictionless equilibrium utility changes over time, $u^*_{j,t} \neq u^*_{j,t-1}$. Calculations are simplified with the alternative compensating variation measure of income needed to maintain actual utility in each point in time when hypothetically shifting to the as-if-frictionless equilibrium.

The loss measure is more generally useful in allowing for domestic changes and in applying to a much wider class of demand systems without need of a parameter estimate.

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. Data are drawn from the World Input-Output Database. The China and US cases highlight the value of a non-parametric approach

[17] The balanced trade requirement implies that total sales share in the world is equal to total expenditure share in the world. CES preferences are homothetic so there is no income effect on price indexes over time.

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. The parametric constant trade elasticities models of structural gravity practice may significantly mislead in quantifying the evolution of relative resistance and seller incidence in this context.

The calculation of relative resistance R_{jj} applies equation (7). 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. Price indexes from the WIOD are consistently associated with the production and expenditure flows. Treatment of final demand and intermediate input demand separately is suspect for familiar reasons, so the cost function $e(\mathbf{p}^j)$ is assumed to be identical for both uses. The buyers side price indexes of the theory are thus the intermediate input price indexes of the WIOD.[18]

The adding up condition on bilateral shares to world market shares, implies that the normalization of the price indexes is $\sum_j E_j P_j / \sum_j E_j = 1$.[19] Thus the observed price indexes $\hat{P}_j$ are deflated to form the normalized $P_j = \hat{P}_j / \sum_j E_j \hat{P}_j$. 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.

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 (10) 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

[18]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.

[19]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.

(7) for the case $i = j$.

The application reveals that US manufacturing experienced a 1.5% annual average fall in gains from trade relative to as-if-frictionless trade from 2000 to 2014. This was accompanied by a 4.7% 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. [See equations (10) and (8) and the discussion following the latter.] China's gains from trade relative to as-if-frictionless trade rose an annual average 1.8%, accompanied by an annual average 8.2% 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 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 -7.6% for China and 4.1% 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 implied decline in Π_j/τ_{jj} is implies an equal rise in c_j but this over-estimates the role of the fall in Π_j. 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 Industrial Policy Implications

Changes in world sales shares drive changes in the terms of trade and gains from trade in the lens of the non-parametric gravity model. Big effects measured above for manufacturing of China and the US 2000-2014 are plausibly due in part to China's industrial policy. The US Inflation Reduction Act of 2022 is an industrial policy response. The results reported above suggest that industrial policy to some extent pays for itself by improving the terms of trade. This suggestion is sharpened and quantified with a focus on seller incidence shifting, shown below to be the main force driving terms of trade changes.

The terms of trade effects of policy-induced own supply shifts uses non-parametric equation (8). Results are reported below in Section 5.1. Gains effect quantification requires a parametric model to project the buyers domestic share change in response to the resulting terms of trade change. The CES share model is chosen to represent this response for simplicity and to connect to the previous gravity literature. Locally, the trade elasticity can be calibrated to each country's domestic share for each year, with the local translog elasticity evaluated at that point equal to a CES elasticity. The non-parametric point of view privileges neither translog nor CES, suggesting that a minimum distance average of the point estimates is used in Section 5.1 to generate a representative trade elasticity θ. The estimated elasticity is applied to loss measure (10) quantify the implications for the gains measure reported in Section 5.2.

5.1 Terms of Trade Elasticity Estimates

The non-parametric terms of trade elasticity with respect to sales size is given by equation (8) for the domestic case $i = j$.[20] Thus

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

The local non-parametric elasticity of R_{jj} with respect to s_j is calculated by plugging into the equation the observed and inferred data, 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 0.60. A 10% rise in sales share implies a 6% terms of trade improvement. Assuming a CES trade elasticity $\theta = 1$ for the world expenditure share would imply that $c_i\Pi_i$ falls 10% for every 10% increase in sales share. China's 2014 terms of trade elasticity with respect to its share reveals an elasticity equal to 0.67, so a 10% rise in its share (from 31.9% to 35%) induces a 6.7% rise in its terms of trade.

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 is likely a downward biased estimate.) 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 0.60 reduces the US manufacturing terms of trade by 2.81%. The

[20] This is an all else equal measure. Accounting for system interaction effects requires a full general equilibrium approach that is far beyond the aim of this paper.

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 0.67 implies that China's terms of trade fall 0.96% 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 and understates the US 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 rises over time while the US share falls over time. The combination of the two cases brackets the implication that the own effect due to the local terms of trade elasticity (8) 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%, a bit more than half of the 4.7% 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 8.2 percentage point rise in its terms of trade.

5.2 CES Trade Elasticity

Projection of counterfactuals requires parametric modeling. The simplicity of CES and its wide use in the parametric gravity literature both suggest the CES demand model to project the gains from trade effects of a change in industrial policy below in Section 5.3.

This section generates the trade elasticity as CES parameter θ that best quantifies the relationship of equilibrium trade expenditure shares to the non-parametric relative resistance statistics generated from (7). The results imply a tightly fitted trade elasticity very close to 1, much lower than the range in the previous literature. The minimum distance calibrator ('estimator') method here is contrasted below to the standard econometric method. The standard method uses the variation of bilateral buyer prices or other observable price shifters used to identify the trade elasticity rather than the variation in revealed relative resistance. Thus it omits variation in the unobservable 'taste shifters' in revealed relative resistances. The difference in results is explained by negative correlation of observable with unobservable bilateral frictions. The lens of the model is applied below to explain the negative correlation.

The minimum distance of the set of CES implied relative resistances from the revealed relative resistances is 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. This method will be called minimum distance calibration, though it is alternatively an estimation method comparable to standard econometric methods of estimating a CES trade elasticity. 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 they average results from a widely used good-fit model and a model widely interpreted as a good approximation to a flexible general functional form. From either perspective, the minimum distance estimate or calibration is intuitively appealing.

The buyers' expenditure share with CES demand is given by $b_{ij} = (c_i\tau_{ij}/P_j)^{-\theta}, \theta > 0$. The spatial equilibrium distribution is given by the standard gravity equation

$$b_{ij} = s_i(\tau_{ij}/\Pi_i P_j)^{-\theta} = s_i(R_{ij})^{-\theta}. \tag{13}$$

The relationship of (13) to the unobservable R_{ij}^{CES} is given by first inverting (13) to isolate R_{ij}^{CES} on the left hand side:

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

and then taking logs. The result is

$$\ln R_{ij}^{CES} = -(1/\theta)[\ln b_{ij} - \ln s_i]. \tag{14}$$

Given the value of the trade elasticity θ, the right hand side of (14) generates the log of relative resistance R_{ij}^{CES} implied by the log distance of b_{ij} from as-if-frictionless expenditure share s_i.

In the non-parametric case, observable $\ln R_{ij}$ is given by the log of (7). The non-parametric approach to CES parameter fitting finds the best-fit CES trade elasticity (inverse) that minimizes the sum of squared residuals η_{ij}^2 from the cross-section 'regression' equation:

$$\ln R_{ij} = (-1/\theta)[\ln b_{ij} - \ln s_i] + \ln \eta_{ij}. \tag{15}$$

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.) The fitted θ is the average of the calibrated values that solve equation (15) for each observation with $\ln \eta_{ij}$ set equal to zero.

In contrast, the CES gravity literature treats the CES specification as true and the estimator seeks the best fit unbiased estimate of the parameter θ. From the econometric perspective, 'regression' (15) 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 determine R_{ij} given by (7) and appear on the right hand side of (15). Viewed from the non-parametric perspective, endogeneity bias is a feature, not the bug it is from the econometric perspective.

For use in projections and counterfactuals and given that a specification (CES here) is chosen, the best prediction of out of sample value of trade elasticity θ is desired. With no downward omitted variable bias, the standard method remains the choice because it presumably avoids endogeneity bias. The results of this paper demonstrate omitted variable variation with substantial downward bias effect on the trade elasticity. My choice here is to pick the lesser of two elasticities as likely to be the lesser of the two biases. Attenuation of the endogeneity bias is feasible with time differencing the relationship between relative resistances $R_{ij,t}$ and the share terms $P_{j,t}b_{ij,t} - s_{i,t}$, especially with long time differences.

Equation (15) extends to a panel setting, adding the time subscript t. The minimum distance CES elasticity estimated from panel data solves

$$\min_{\theta} \sum_{i,j,t} \ln \eta^2_{ij,t}. \tag{16}$$

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 (16) yields a tightly estimated θ equal to 1.03 with standard deviation 0.04 in the US subsample, and 1.04 with standard deviation 0.04 in the China sub-sample. The adjusted R^2 is .93 in both cases.

Extension of the estimator (16) 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 7.3. 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]

[21]The large panel suggests measurement error associated with small trade flow shares $b_{ij,t}$ (many on

The large difference in trade elasticity estimates requires explanation. The chief difference in application is that the econometric best fit trade elasticity is identified off variation in tariffs or other directly observed trade costs or bilateral prices [Simonovska and Waugh (2014)], while the non-parametric approach fits the trade elasticity to the (much larger and potentially more informative) variation in non-parametric relative resistance statistics $\ln R_{ij}$. Both methods control for variation in the multilateral resistances, so the difference lies in the implied bilateral frictions. The equilibrium bilateral friction is τ_{ij}. Standard methods use variation in an observable component such as tariffs or transportation costs, or variation of τ_{ij} inferred from variation in observable bilateral prices p_{ij}. Revealed relative resistance contains sources of variation in bilateral resistance such as taste shifters that are unobservable and hence omitted from the standard approach that uses observables such as p_{ij} and tariffs or transport costs. The much lower minimum distance trade elasticity estimates than the standard estimates is explained by omitted variable bias if the taste shifters are negatively correlated with the observable variation in bilateral prices or costs. The model focus on incidence shifting suggests a structural explanation for negative correlation. Increases in destination j's unmeasured buyer cost or distaste for seller i's good raises the bilateral seller incidence τ_{ij}/Π_i of a given bilateral friction τ_{ij}, hence it reduces the equilibrium bilateral price p_{ij}. Transport costs similarly fall, assuming that they rise endengously with bilateral volume, as in Anderson and Yotov (2020). Political economy suggests a similar endogeneity for tariffs.

While omitted variable bias can explain the lower trade elasticity estimate, a less comfortable alternative explanation is aggregation composition effects. Non-parametric gravity implies implicit aggregation in the relative resistances in theoretical equation (5) and opera-

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 (7) 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. The theory could be wrong due to approximation error. Either way, 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 (7), 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.

tional equation (6). Manufacturing is highly aggregated, and it has large sectoral composition differences across countries. Where these differences are important, they affect elasticity estimates and the fit of the CES model in ways that are outside the CES model and may lie outside the translog model. Disaggregation is the appropriate treatment for this problem.[22] Aggregation 'bias' is arguably not important for the manufacturing trade of the US and China, where large diversified economies sell to and purchase from the world with many types of manufactures.

Other possible reasons for the difference are also relevant. All methods are subject to measurement error in the trade and production data, but the non-parametric method additionally relies on buyer price indexes subject to error. The CES specification controls for the multilateral resistances in fitting the elasticity, but the revealed relative resistances are contaminated with the errors in the price indexes in complex ways that can affect the estimate.

The zeros problem (some bilateral trade shares $b_{ij} = 0$) from the econometric perspective suggests that regression (15) may yield trade elasticities subject to selection bias.[23] The efficient arbitrage properties of gravity noted in Section 2 imply that the opportunity cost of delivering a unit of i, $c_i\Pi_i$ that is embedded in all active relative resistances is also the cutoff value of net delivered price in the arbitrage equilibrium. This is true whether there are fixed export costs or not. A Tobit estimator of the translog structure would be appropriate if the error term could be thought of as orthogonal. See [Anderson and Zhang (2022)] for Almost Ideal gravity Tobit estimation and projections of entry or exit that feature both fixed costs and choke prices.[24]

[22] The implied exercise for panel data includes action on the extensive margins of trade (new destinations for existing products) and production (new products by some countries). Developing a useful treatment is beyond the scope of this paper.

[23] If the CES demand specification with $\theta > 0$ is true, it implies that there must be fixed bilateral trade cost to explain observed $b_{ij} = 0$. Thus positive trade volume must be large enough to cover fixed cost, so a selection equation is needed, [Helpman et al. (2008)].

[24] That model features entry or exit by heterogeneous firms and approximates selection. The present case of selection of heterogeneous buyers implies discrete choice by a mass of heterogeneous buyers will select entry or exit of each product in the set. The aggregate purchase varies with the mass who select to buy. Presumably a similar approximation would apply. In the CES case, the combination of heterogeneous firm

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 (15). 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.

5.3 Industrial Policy Implications

Incidence shifting suggests that industrial policy may partially 'pay for itself' via improved terms of trade implied by (8). 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 (10) 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

entry where firms draw productivity from the Pareto distribution leads to a tractable closed form model, Chaney (2008).

$\Rightarrow d\ln b_{jj}/d\ln R_{jj} = -\theta$. Apply this to calculate the net gain

$$1 + \frac{dL_j}{ds_j} = -\theta P_j \frac{b_{jj}}{s_j}\frac{d\ln R_{jj}}{d\ln s_j}. \tag{17}$$

Substitute $s_j R_{jj}^{-\theta}$ for b_{jj} everywhere in equation (17) and in equation (8) used for $d\ln R_{jj}/d\ln s_j$ in equation (17). After simplification the result is:

$$1 + \frac{dL_j}{ds_j} = -\theta P_j \frac{R_{jj}^{-\theta} + 1}{1 + \sqrt{P_j} - R_{jj}^{-\theta}(P_j - \sqrt{P_j})}.$$

The denominator is negative for $R_{jj}^{-\theta} > (1 + \sqrt{P_j})/(P_j - \sqrt{P_j})$, implying that $R_{jj} < 1$ is sufficiently small. Since R_{jj} falls with s_j by equation (7), sufficiently small countries lose from policy that raises sales share. Countries with sufficiently large shares will reduce the loss measure by marginal increases in shares, a net benefit.

The net benefit of industrial policy at the margin for China and the US is calculated with (17). Combine the 2014 trade data with estimated θ from (16) and estimated terms of trade elasticity from (8). 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 -1.94 with the average θ and -1.95 with the 2014 θ. For the 2014 US the net reductions in loss are -3.88 and -3.61 respectively. In 2000 the ranking is reversed. For 2000 China the net reduction in loss is -5.042 with the average θ and -5.230 with the 2000 θ. For 2000 US, the net reduction in loss is -2.032 with the average θ and -2.186 with the 2000 θ.

Loss reduction rates for China and the US reverse in ranking size between 2000 and 2014. The reason for the reversal follows from equation (17) in its simplified form. Given that s_j is sufficiently large that the loss is reduced by sales increases, the loss reduction from a marginal increase in sales falls in absolute value as sales share increases. The sales share changes in combination are large enough to reverse the ranking over time as China's share rises over the time period and the US share falls. (The decreasing rate of change suggests movement toward an interior optimal sales share.)

The large marginal net benefit of seller incidence shifting illustrates the importance of this mechanism, but 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 (8). Seller incidence shifting is a much weaker motive for industrial policy by smaller suppliers and contra-indicated for sufficiently small sellers. The offset loss reduction also rises with θ, so higher elasticity products have a stronger case for industrial policy.

6 Conclusion

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

The application focuses on domestic expenditure shares and the relative resistance for domestic trade, the terms of trade. The method also produces a large panel of bilateral relative resistances that can be used to dig deeper into the causes of their variation. Their portmanteau property as residuals is shared with the Solow productivity residual, differing in residual accounting for discrete rather than local differences in trade frictions. The productivity literature may provide clues for the relative resistance investigation.

The results of this paper suggest that the trade elasticities inferred in the received parametric gravity literature are significantly upward biased by the effect of omitted variables. This means that the gains from trade changes in associated counterfactual exercises are significantly downward biased. Exploration of sparsely parameterized extenstions of CES or translog seemse useful. The appropriate specification and inference of parameters from residually derived relative resistances to be used in model projections is a deep intellectual challenge. This paper stands at the edge of the forest.

Spatial aggregation (of origin and destination locations at varying sizes) is a feature of all gravity applications. An approach to consistent aggregation is sketched below in Appendix Section 7.2. The model is developed for final demand systems for goods, but it also straightforwardly applies to demand systems for intermediate inputs.

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.

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

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.

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

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

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

7 Appendix

7.1 Terms of Trade Rise with Supply: Intuition

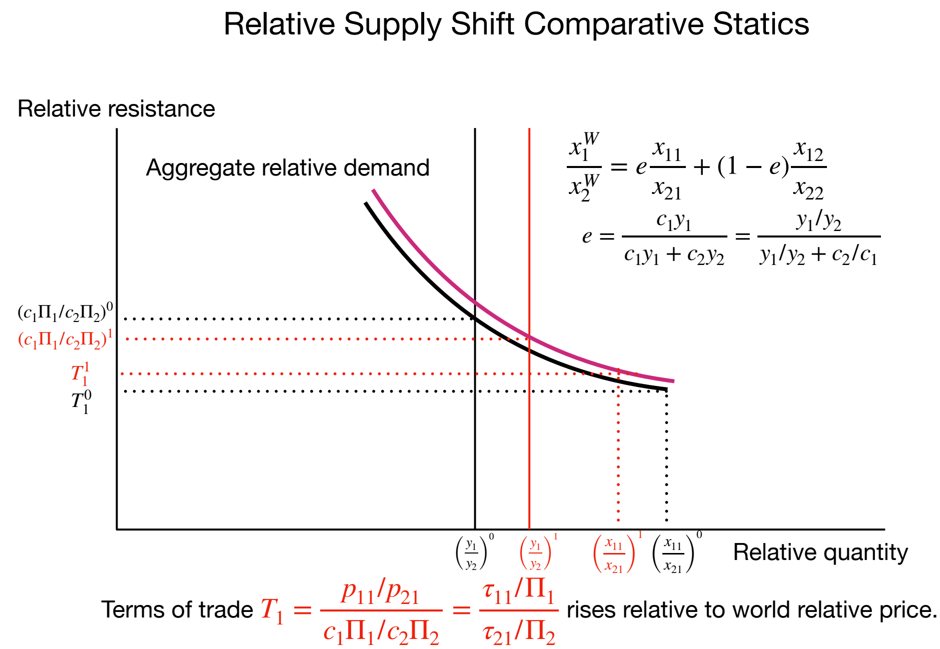
The headline result that China's terms of trade improve when China's manufacturing production rises faster than the US is contrary to intuition based on frictionless exchange. In the simple two good two country model, when relative supply of China's good rises, downward sloping relative demand for China's 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, the 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.

7.2 SpatialAggregation

Non-parametric gravity equation (6) provides a useful interpretation of the relationship between gravity applications across many varieties of spatial aggregation. In practice, gravity is widely used for trade between cities, regions and countries and sometimes commuting zones. How may we understand relative resistances based on views at varying focal lengths?

Aggregation of locations necessarily implies spatial aggregation of frictions. Mayer and Head (2002) address the aggregation of frictions related to distance. Their solution in the

CES gravity context uses city-pair distance aggregation with population weights. Population weights proxy economic mass weights with the useful virtue of plausible exogeneity to contemporaneous trade flows. The existing literature does not treat aggregation of frictions between city pairs not related to distance and not uniformly associated with international borders. Section 7.2.1 lays out a general treatment. Section 7.2.2 treats aggregation of internal distances in the context of infrastructure that may internal distances asymmetrically.

7.2.1 General Logic

The general non-parametric logic of spatial aggregation of frictions is nested within the logic of (6). 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 (6) 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:

$$b_{IJ} \sum_{j \in J} \frac{b_{Ij}}{b_{IJ}} P_j - Y_I/Y = \bar{b}_{IJ} \sum_{j \in J} \frac{\bar{b}_{Ij}}{\bar{b}_{IJ}} 2\sqrt{P_j}\bar{b}_{Ij} \sum_{i \in I} \frac{\bar{b}_{ij}}{\bar{b}_{Ij}} \frac{R_{ij}-1}{R_{ij}+1}. \tag{18}$$

The double sum on the right hand side of (18) 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.

7.2.2 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.[25] 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 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. See the aggregation discussion above for details.

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.

[25] 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).

7.3 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

$$b_{ij} = \alpha_i - \sum_l \gamma_{il} \ln(c_l \tau_{lj}/P_j) = \alpha_i - \bar{\gamma}_i \ln(\bar{p}_{ij}/P_j) \tag{19}$$

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 (19) 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

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