Gravity Under Uncertainty

James E. Anderson
Boston College
NBER

Mingyang Wei
Boston College

Yoto V. Yotov
Drexel University
CGPA

March 26, 2026

Abstract

Gravity under uncertainty features *ex ante* efficient arbitrage of bilateral trade. Realizations of trade frictions, effective prices and trade share elasticities with respect to frictions are all uncertain. Observable bilateral trade shares alone are sufficient to infer all realizations in rational expectations *ex post* arbitrage equilibrium. No dubious specifications, parameters or price data are required. Arbitrage error variances are measured. Realized trade elasticities are locally entangled with trade frictions under uncertainty. This paper disentangles them. Congestion uncertainty affects non-price frictions. Larger exporters reduce average realized congestion by delivery arbitrage across more destinations.

The authors have no financial support or conflicts of interest to disclose.

1 Introduction

Trade is carried on by agents acting under uncertainty. Gravity models suppress this reality, despite nods at insurance and instruments for holdup and congestion at choke points. Multiple sources of uncertainty affect all the eventual bilateral payoffs that incentivize traders (arbitrageurs). Arbitrage under uncertainty is incorporated into this paper's gravity model. Recent trade disruptions due to pandemic disease, global recessions, wars and unpredictable trade policy shifts illustrate the need for such a model. History provides ample evidence of even greater uncertainty in the past. The empirical payoff from this paper's gravity model under uncertainty is illustrated by application to manufacturing in the globalization era 2000-2014. Results include novel measures of arbitrage error variances and bilateral trade elasticities that vary across country pairs and time.

Methodological innovation on the preceding gravity model is required. This is because with incomplete markets, even perfectly diversifiable risks have significant real consequences. Shocks in any bilateral pair trade necessarily induce uncertainty in other pairs, an interdependence suppressed in the preceding gravity literature. The non-parametric gravity model of trade in this paper includes the necessarily general equilibrium interdependence of uncertain trade frictions and demand structure.

The workhorse structural gravity model of trade (mostly) assumes certainty about payoffs as well as certainty about one or another of its various parametric specifications. (The brief review of related literature below discusses a partial equilibrium exception that admits uncertainty effects on a bilateral friction.) In counterfactual exercises, the required parameter estimates are thus vulnerable to both specification error and payoff uncertainty. This paper introduces and applies a non-parametric approach to trade as spatial (and inevitably inter-temporal) exchange under uncertainty. Both arbitrageurs and the modeler are uncertain about specification and payoffs.

Uncertain demand is assumed to be generated from a common invertible demand system for goods from all origins in all destinations. Arbitrageurs commit to distribution from each origin to potentially all destinations based on (boundedly) rational expectations about buyer price realizations and trade friction realizations. The difference between anticipations and realizations appears in the interim between order agreement and delivery. Effective realized buyer prices differ from net seller prices by uncertain trade frictions that include non-price shifters of buyer willingness-to-pay.

Ex ante efficient spatial arbitrage applied in this setting yields a tractable gravity model. The application restricts the demand system to the general translog specification without requiring common translog parameters. *Ex post* realized bilateral trade frictions are deduced as residuals of observable realized bilateral trade flows from realized as-if-frictionless trade flows. Their resemblance to productivity residuals suggests

the fruitfulness of reduced form investigation of realized friction determinants.

The key conceptual and quantitative difference from previous gravity models is theory leading to non-parametric measures of arbitrage error variance and its effects. *Ex post* arbitrage error is measured as a deviation of realized frictions from *ex ante* arbitrageur expectations. Wrld and country specific arbitrage error variances are measured from this base. The method of deriving realized frictions disentangles *ex post* bilateral friction variations from variation in local *ex post* bilateral trade elasticities.

The model is applied to quantify uncertainty in arbitrage and some of its effects in the era of globalization. The basis is the WIOD manufacturing trade data from 2000 to 2014. A large (20%) decline in aggregate arbitrage error variance to 2007 is followed by a smaller (10%) increase in 2007 followed by a reversion to the previous minimum variance by 2014. The changes suggest learning and capacity adjustments to the China Shock, disrupted subsequently by the Great Recession. Country level arbitrage error variances mostly decline with the same increase in 2007 due to the Great Recession followed by reversion toward earlier levels. Arbitrage error variance of countries is further decomposed into import and export variances, with interesting patterns discussed below.

The average of local trade elasticities in the results is low compared to the previous gravity literature. Trade elasticity variation across partners has interesting patterns, starting with large bilateral asymmetry in the cross section. For some countries the outward (export) elasticities are larger than the inward (import) elasticities, while for others the reverse is true. Time variation in the trade elasticities and their variances may be further informative about arbitrage efficiency and its consequences.

Non-price trade frictions matter more than observable frictions in explaining trade patterns.[1] The method introduced in this paper re-estimates and extends the reduced form determinants of trade frictions free of trade elasticity assumptions and estimates. Non-price frictions *uncertainty* associated with congestion and delay costs arises due to the temporal aggregation implicit in annual or even monthly trade data. The striking results quantify an important scale effect that favors larger exporters. While congestion and holdup both reduce trade on average, the positive exporter scale effect is associated with uncertainty because larger export origins are better able to arbitrage delays across destinations and have more and likely better outside options when facing holdup. A negative scale effect is found for importers, rationalized by larger importers in the cross section of countries tending to have fewer domestic alternatives or outside options for fulfillment of demand.

The full implications of spatial arbitrage are utilized to obtain the results. Efficient

[1]This regularity is well-established in the gravity literature. Head and Mayer (2014) call the non-price frictions "dark" in a nod to the dark matter and dark energy of cosmology.

spatial arbitrage as in Samuelson (1952) explains the variation of equilibrium trade flows across destinations. Arbitrageurs choose allocations of goods from each origin to active destinations based on expectations of buyer-willingness-to-pay and expectations of trade frictions to each bilateral destination. *Ex ante* efficiency assumes *ex ante* diversification of risk. Arbitrageurs are risk neutral.[2] Buyers' willingness-to-pay *ex post* is generated from duality as in Deaton (1979).

Realized bilateral frictions are interpreted in the model as relative resistances, decomposed as ratios of bilateral resistance to the product of inward and outward multilateral resistance. Multilateral resistances are indexes of all active inward and outward bilateral resistances respectively. In economic terms they measure the equilibrium buyers' and sellers' incidences of all bilateral frictions in the world trade system. The inward multilateral resistance is also interpreted as an effective buyer price index. Bilateral trade elasticities are implicit functions of pairwise relative resistances.

An important feature of the model is that it applies at any level of aggregation of regions, countries and sectors. It admits heterogeneity of buyer types, including intermediate and final demands. In all aggregations, relative resistances are obtained that implicitly aggregate relative resistances across the sources of heterogeneity.

The model extends the preceding gravity literature, reviewed by Head and Mayer (2014). Structural gravity models of trade all assume efficient spatial arbitrage combined with an invertible demand system that represents cost-minimizing buyers behavior. The inverse demand system delivers willingness-to-pay (potential buyer prices) at all destinations. Preceding gravity models assume that arbitrageurs face certain trade frictions and buyer prices. In contrast, arbitrageurs in our model act on rational expectations of uncertain frictions and buyer prices to jointly determine an *ex ante* efficient equilibrium based on anticipated relative resistances. The arbitrageurs are assumed to project the relative resistances based on a state-of-the-art constant trade elasticity (CES) gravity model applied with a trade elasticity estimate. This projection determines their origin-destination distribution shares of trade endowments. The realized buyer trade expenditure shares and the structural model combine to yield *ex post* realizations of relative resistances. The difference between arbitrageurs *ex ante* projections of relative resistances and the *ex post* realizations of relative resistances implied by realizations of trade flows is equal to the realized arbitrage errors.

Anderson (2026) develops the non-parametric gravity model used here under certainty. He notes a country's domestic relative resistance is equal to its terms of trade. Proportional changes in a country's terms of trade weighted by its domestic trade share give changes in a loss measure that is related to the gains from trade measure

[2]The Appendix develops the case of risk aversion with Constant Relative Risk Aversion utility of arbitrageurs. *Ex ante* efficient arbitrage obtains when investment satisfies the conditions of the Diamond stockmarket, as in Helpman and Razin (1978).

of Arkolakis et al. (2012). A natural extension for gravity under uncertainty suggests that changes in the arbitrage error variance of the terms of trade has a useful welfare interpretation under constant relative risk aversion. Development of this intuition is beyond the scope of this paper.

Handley and Limao (2017) innovate a version of gravity under Trade Policy Uncertainty (TPU) of a restricted kind. Their model is based on firms' option value of waiting to invest a fixed (sunk) entry cost to export markets. What is uncertain in their application is whether China will retain access to US MFN tariffs or be reverted to higher preceding US (e.g., Smoot-Hawley) tariffs. The model yields a gradual adjustment pattern as uncertainty resolves. TPU enters their "TPU-augmented gravity" model as a standard bilateral trade cost shifter. Handley and Limao (2022) review a related literature where instruments for trade policy shifts similarly enter the gravity model as standard cost shifters. In contrast, this paper treats uncertainty about levels of frictions from all sources in a general equilibrium arbitrage model where uncertainty about payoffs in any bilateral pair induces uncertainty in all pair frictions.

The demand model of this paper admits heterogeneous buyer costs and heterogeneous tastes. They combine in shifters that multiply prices paid by buyers to yield effective prices. The shifters can include scale effects due to income (final demand) or technology (intermediate demand). The shifters prominently include pricing-to-market markups because efficient arbitrage admits discrimination between markets. Effective price variation across buyers in the cross section also includes quality variation [e.g. Feenstra and Romalis (2014)] on buyers or sellers. The general translog specification that operationalizes the does not require the translog parameters, hence hence heterogeneity in the parameters is admitted and embedded along with the other elements in the revealed relative resistances and effective buyer price indexes.

The static exchange model spatial arbitrage equilibrium takes supplies and qualities as given to simplify an approach modeling the effects of uncertainty. A large related gravity literature [e.g., Eaton and Kortum (2002), Chaney (2008)] embeds the certainty version of gravity in Ricardian production models that select labor allocations to sectors and their efficient distribution to destinations based on labor productivity draws from a known extreme value distribution. The dispersion parameter of the distribution replaces the CES trade elasticity in determining the equilibrium distribution of trade in goods. Anderson and Yotov (2020) model production and trade in a Cobb-Douglas specific factors model of short run gravity. Costly gradual adjustment of bilateral trade capacities toward long run efficient levels is analyzed in an intertemporal and spatial equilibrium model with perfect foresight. The trade elasticity is the product of the CES elasticity and a structurally derived adjustment parameter. Treatment of uncertainty on the lines of this paper points to a production model of resource allocation decisions

with uncertain payoffs that nests a version of the exchange under uncertainty model inside. A starting point is the Helpman and Razin (1978) treatment of *ex ante* efficient resource allocation with uncertain productivity draws. Development of the combined production and trade model under uncertainty is a challenge beyond the scope of this paper.

A byproduct of this paper is revealed effective import price indexes.[3] Observed price index data for traded goods, when available at all, suffer from familiar quality and limited observability deficiencies. The effective price indexes generated in this paper are theory-consistent with the observed trade flow data by construction. The effective price indexes allow for differences in tastes (for final demand) or technologies (for intermediate demand).[4] These properties suggest that effective buyer price indexes generated with the methods of this paper are likely to be useful for a variety of applications where spatial interdependence matters.

2 Non-parametric Spatial Arbitrage Model

The allocation of goods is assumed to nest inside resource allocation decisions previously determined.[5] The arbitrageurs face exogenous (to the arbitrageurs) anticipated bilateral trade 'frictions', the iceberg melting metaphor of Samuelson (1952). Efficient arbitrageurs allocate exogenous anticipated total supplies of tradable goods from each origin to potentially each destination given exogenous rationally anticipated buyer prices. Arbitrageur allocations to destinations generate buyer prices from their willingness-to-pay for quantities supplied by arbitrageurs. The application is to data that combines intermediate and final goods but the exposition focuses on a representative final goods sector.

Buyer willingness-to-pay for given quantities is dual to the more familiar efficient choice of quantities at given prices, reviewed first. Destination j purchases are characterized by a globally common minimum cost (sub-)expenditure function $e(\mathbf{p}^j)u^j$, homogeneous of degree one in effective price vector $\mathbf{p}^j$ with effective price elements $p_{ij}\beta_{ij}$ from each origin i, and (sub-)utility u^j. Non-homothetic effects of u^j and unobservable buyer costs are absorbed in the taste shifters β_{ij} that multiply the buyer prices

[3] Anderson (2026) generates non-parametric sufficient statistics for relative resistances using the *ex post* efficient (no uncertainty) version of the arbitrageurs allocations model. His method requires *observed* price indexes in addition to observed trade flows.

[4] The parametric gravity literature recognized the efficient aggregation of heterogeneous tastes in relative resistances [Anderson and van Wincoop (2003)], but imposed common elasticity parameters. The non-parametric model allows unspecified parameters to vary across agents.

[5] This logical modeling simplification is the most natural timing choice and justified by empirical observation for many types of trade. Some buyer-seller contracting will violate the assumption but might usually fit within the perfect foresight case developed below.

p_{ij}. The effective price index is $P_j = e(\mathbf{p}^j)$. The solution to the buyer choice problem is alternatively represented by the dual distance function [Deaton (1979)] where the virtual price vector (effective willingness-to-pay vector) is minimized for a vector of origin-destination supplies from the arbitrageurs.

The realized fixed supplies y_i of origins i are efficiently distributed in arbitrage equilibrium to destinations j.[6] The arbitrageurs face given (in equilibrium) anticipated bilateral 'iceberg' frictions $t_{ij} \geq 1$ that are the product of actual bilateral trade frictions and the β_{ij} taste shifters of the destinations. The sum of sales at anticipated buyer prices is $\sum_j p_{ij}x_{ij}$. Efficient spatial arbitrage by risk neutral profit-maximizing arbitrageurs implies the expected unit value function

$$V = \max_{\{x_{ij}/y_i\}} \mathcal{E}\left[\sum_{i,j} p_{ij}x_{ij}/y_i | \sum_j t_{ij}x_{ij}/y_i = 1\right]; \ \forall i. \tag{1}$$

The expectation operator $\mathcal{E}$ is taken over realizations of buyer prices p_{ij} and frictions t_{ij}. It represents the aggregation of individual actors anticipations in the set of arbitrageurs under the 'wisdom of crowds' hypothesis of rational expectations.

Risk neutrality and the linearity of V justify treating the solution shares $\{x_{ij}/y_i\}$ in certainty equivalent terms with $p^e_{ij} = \mu^e_i t^e_{ij}$ for $x_{ij}/y_i > 0$ and $x_{ij}/y_i = 0$, $\forall p^e_{ij} < \mu^e_i t^e_{ij}{}^e$. The economic interpretation of the Lagrange multiplier μ_i is the anticipated value of $c_i\Pi_i$ where c_i is the anticipated (net) seller price and

$$\Pi^e_i = \sum_j (p_{ij}/c_i)^e x_{ij}/y_i, \tag{2}$$

is the anticipated average equilibrium seller incidence of trade 'frictions' when the set of delivery shares $\{x_{ij}/y_i\}$ is efficiently chosen. Thus Π^e_i is minimized.

Formally, the Lagrangian function derived from (1) is minimized over the choice of the μ^e_is and c_i is dual to efficient supply choice y_i in the outer nest that enfolds (1). Π^e_i is also called the (anticipated) outward multilateral resistance for seller i. The (in principle) observable bilateral trade cost factor is $\tau^e_{ij} = p^e_{ij}/c^e_i$, the anticipated buyer price divided by the anticipated seller price. τ^e_{ij}/Π^e_i is the anticipated bilateral buyer's incidence of trade frictions.

Efficient buyer behavior implies a parallel restriction for the *ex post* price index per unit of utility, or buyer incidence of trade frictions (inward multilateral resistance).

[6]The exposition that follows abstracts from inventory for simplicity, since the application does not use inventory data. Formally, inventory change is both an origin adding to current "supply" and a destination reducing current "supply" to other destinations. The future destination "price" is an expected future period valuation. Active inventory change in the static model follows the familiar (s, S) rule.

The distance function of Deaton (1979) is

$$d(\mathbf{x}^j, u^j) = \min_{\{p_{ij}\}} \sum_i p_{ij} x_{ij} | e(\mathbf{p}^j, u^j) = 1.$$

It is dual to the expenditure function

$$e(\mathbf{p}^j, u^j) = \min_{\{x_{ij}\}} \sum_i p_{ij} x_{ij} | d(\mathbf{x}^j, u^j) = 1.$$

The distance function expresses the minimizing property of the price (willingness-to-pay) index associated with given supplies. This applies to buyers at each destination j, so it applies to the buyers valuation of the world's goods distribution. Impose the common preferences with price-dependent non-homotheticity structure and apply the distance function to the world distribution of goods. The result is:

$$W = \min_{\{p_{ij}\}} [\sum_{i,j} p_{ij} x_{ij} | e(\mathbf{p}^j) u^j / \sum_j u^j = e(\mathbf{p}^*) = 1, \ \forall i, j]. \tag{3}$$

The solution implies the cost minimizing valuation $P_j = e(\mathbf{p}^j)$ of each buyer j's allocation vector of x_{ij}s subject to the common normalization of the price vectors $\{P_j\}$ and the world price vector $e(\mathbf{p}^*)$. These are all the realized buyer prices. The combination of *ex ante* efficient sales (via (1)) and *ex post* efficient purchases (via (3)) implies that the arbitrage equilibrium minimizes both *ex ante* seller incidences $\{\Pi_i\}$ and *ex post* buyer incidences $\{P_j = e(\mathbf{p}^j)\}$.

The non-parametric gravity model assumes a common invertible demand system for all countries and hence the world. Programs (1) and (3) are both well-behaved optimization problems with unique solutions, hence their combination has a unique fixed point. Frictions can be inferred by comparing observed data assumed to reflect (1)-(3) with hypothetical as-if-frictionless equilibrium with relative frictions eliminated and all other exogenous variables held constant. This includes the real incomes $\{u^j\}$. Mild further restrictions on the set of demand systems lead to operational measures of trade frictions.

Sellers in effect ship to a world market, with realized world buyer prices $\{c_i \Pi_i\}$ and realized sales $\{c_i \Pi_i y_i\}$ and sales shares $s_i = c_i \Pi_i y_i / \sum_i c_i \Pi_i y_i$. Shephard's Lemma implies realized world expenditure shares $B_i(\{c_i \Pi_i\})$, equal to the world seller sales share of origin i at buyer prices, s_i. Here $\sum_i B_i(\{c_i \Pi_i\}) = 1$ imposes a normalization on the realized world price vector. Observed expenditure shares are given by $b_{ij} = p_{ij} e_{p_{ij}} / e$, by Shephard's Lemma. In an as-if-frictionless world equilibrium, $b_{ij} = B_i, \ \forall i, j$. All buyers would face the same price vector $\{c_i \Pi_i\} = \mathbf{p}^*$.

The realized (observed) world has equilibrium expenditure shares b_{ij} (which may

be zero) for each country j and each origin i's product. The effective price indexes for each country differ from 1. Once the vector of effective price indexes is normalized to 1, the difference between country j's observed and as-if-frictionless price indexes $P_j - 1$ gives the percentage reduction in expenditure that maintains initial utility at the as-if-frictionless prices. Then the difference between the unit cost functions $P_j - 1 = \sum_i (P_j b_{ij} - B_i)$, from application of Shephard's Lemma. Invertibility justifies application of the intermediate value theorem to the sum. Thus

$$\sum_i (P_j b_{ij} - B_i) = \sum_i b_{ij}(\tilde{\mathbf{p}}^j)\tilde{P}_j(R_{ij} - 1)/\tilde{R}_{ij}, \ \forall j; \tag{4}$$

for some $\lambda_j \in [0, 1]$ and the associated intermediate price vector $\tilde{\mathbf{p}}^{\mathbf{j}} = \lambda_j \mathbf{p}^j + (1 - \lambda_j)\mathbf{p}^*$.

Anderson (2026) shows that the element-by-element differences $P_j b_{ij} - B_i$ in the sums on both sides of equation (4) are the basis for inferring realized relative resistances

$$R_{ij} = \frac{p_{ij}/P_j}{c_i \Pi_i} = \tau_{ij}/\Pi_i P_j.$$

R_{ij} is the proportionate increase in willingness to pay from buyers in j for the increase in frictions over opportunity cost that sellers from i must be paid for service to j. Element-by-element equality is obviously sufficient, but it is also necessary due to the cost minimization property of unit cost function $e(\cdot)$.

Evidently, the signs of the left and right hand sides of the element equations of (4) must agree. Imposing agreement implies a specification that the local trade elasticity is greater or less than zero. The intuitive (standard gravity model) case of a conventionally positive elasticity implies that expenditure shares fall as R_{ij} rises. Then

$$P_j b_{ij} - B_i = -b_{ij}(\tilde{\mathbf{p}}^j)\tilde{P}_j(R_{ij} - 1)/\tilde{R}_{ij}$$

and home bias $b_{jj} > s_j = B_j$ is associated with $R_{jj} < 1$ and $b_{ij} < s_i$ is associated with $R_{ij} > 1$. With a local trade elasticity less than zero, the element equation $P_j b_{ij} - B_i = b_{ij}(\tilde{\mathbf{p}}^j)\bar{P}_j(R_{ij} - 1)/\tilde{R}_{ij}$ is applied.[7]

Ex post inefficient arbitrage arises when realizations of R_{ij} deviate from s_i/b_{ij}. The deviations induce *ex post* realizations of P_j that deviate from 1. The general translog specification makes inference in this case operational. The combination of the large number of parameters (if the number of origins is N, there are $N \times (N-1)/2$ and the large number of bilateral taste shifters implies that this restriction is significantly more general than the translog. Note that the underlying parameters themselves are

[7]Equation (4) holds regardless of the pattern of values of the local trade elasticities, as does the validity of the element-by-element equations. The sign convention procedure applies a plausible economic interpretation to extract information from data where the element equations generate R_{ij}s that differ in sign.

aggregates of heterogeneous agents parameters. The element-by-element version of equation (4) becomes:

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

Equation (5) applies the convention that buyer shares fall with relative resistance, hence low expenditure share b_{ij} is associated with $R_{ij} > 1$. This is a strong regularity in trade data, which justifies the assumption of a positive trade elasticity (elasticity of substitution greater than one).[8]

The right hand side of (5) is the product of a discrete percentage change in relative resistance and what is interpreted as a trade elasticity evaluated at the intermediate value. The left hand side is the discrete percentage change in the buyer price index that results. The equation resembles the Constant Elasticity of Substitution (CES) case in which demand expenditure or cost shares are log-linear in relative prices and are applied to log differences from as-if-frictionless equilibrium:[9]

$$\ln(b_{ij}/s_i - \ln 1) = -\theta(\ln R_{ij} - \ln 1).$$

This resemblance is combined with the mostly close approximation of log points to discrete percentage changes in the application below to generate non-parametric trade elasticities deviating from an average elasticity $bar\theta$ relative to the *ex ante* anticipated CES θ^e elasticity.

Equation (5) is solved for revealed relative resistance as a function of observables:

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

$R_{ij} = 1/(b_{ij}/s_i)$ at $P_j = 1$, the local trade elasticity $\theta = 1$. With P_js close to 1, the shares data may imply that CES approximates well with trade elasticity close to 1. Results in Section 2.1 reveal that implied local trade elasticities vary but extreme low values of b_{ij}/s_i (implying extreme high R_{ij}s are required for substantial variation.

Note that the as-if-frictionless equilibrium used to infer R_{ij}s is *compensated* – utility is held constant at its observed equilibrium values. By construction, the normalized price indexes in (4) are compared to the as-if-frictionless price index based on purchase of supplies from the same origins as the observed equilibrium price indexes. Any gain in utility implied is removed by the assumed compensation and change in the set of

[8]In application to cases where a substitution elasticity less than one applies, the minus sign is removed from the right hand side of (5) and the solution for R_{ij} on the right hand side in (6) is inverted.

[9]The ratio of observed trade shares to as-if-frictionless shares is assumed to be $b_{ij}/s_i = R_{ij}^{-\theta}$. In as-if-frictionless equilibrium the ratio is equal to 1.

goods purchased is similarly assumed away. The compensated equilibrium is essential for the cross-section property of trade elasticities.

Equation (6) generates *ex post* realizations of relative resistances is operational if shares data b_{ij} and as-if-frictionless shares $B_i = s_i$ can be combined with realized effective price indexes P_js. [10]

In general, realized price index data is notoriously suspect. It is limited in coverage or even absent in many cases. The novel method in this paper generates the missing realized effective price index data from the implications of (3). The method combines this *ex post* information with the *ex ante* boundedly rational expectations of arbitrageurs that determine *ex ante* allocations of quantity shares from origins to destinations. The difference between *ex ante* and *ex post* relative resistances is interpreted as arbitrage error.

Implementation of the model requires an assumption about the boundedly rational expectations of arbitrageurs. The tactical assumption of this paper is to equate bounded rationality with state-of-the-art gravity implementation. Thus arbitrageurs allocate trade shares based on standard gravity projections of relative resistances. The projections use current structural gravity techniques combined with projections of the outputs and expenditures. The 'wisdom of crowds' hypothesis implies that the average of arbitrageurs makes this use of available prior information. The next section implements this setup.

2.1 *Ex Post* Friction and Effective Price Inference

Significant *ex post* arbitrage error in global trade is plausible due to the shocks of globalization and anti-globalization over the past three decades. Assume that the arbitrageurs are risk neutral.[11] Arbitrageurs have (bounded) rational expectations of anticipated *ex post* buyer prices and bilateral frictions. $\mathcal{E}$ is the arbitrageurs' expectation operator. The arbitrageurs commit shipment shares x_{ij}/y_i with expected buyer payment per unit $\mathcal{E}[p_{ij}] = p^e_{ij} = \mathcal{E}[t_{ij}] \times \mathcal{E}[c_i \Pi_i] = t^e_{ij} \times c^e_i \Pi^e_i$. Expected relative resistance R^e_{ij} is given by $\mathcal{E}[p_{ij}]/\mathcal{E}[c_i \Pi_i]$.

The arbitrageurs' expectations and shipments decisions result in realized values of $\{b_{ij},\ s_i,\ P_j\}$ and R_{ij} given by (6) multiplied by an error term ϵ_{ij}. Sellers' expectations are boundedly rational, thus ϵ_{ij} need not satisfy the usual restriction that it be

[10] Anderson (2026) reports *ex post* R_{ij} results for aggregate manufacturing using the WIOD data 2000-14 focused on China and the US. The data allowed *ex post* price index data construction for international trade.

[11] Risk neutrality is plausible for much of international trade under *normal* conditions where risks seem diversifiable and insurance is available. Major disruptions such as pandemics and wars are obvious exceptions that imply non-diversifiable risk. The Appendix models the effect of risk aversion for the case of Constant Relative Risk Aversion. Risk premia alter the revealed relative resistances so as to ordinarily increase international trade and reduce domestic trade – risk neutrality undervalues diversification.

identically independently distributed with expectation equal to 1.

Realized buyer and seller price indexes are given by the solution to the quadratic system of equations $1 = \sum_i b_{ij} R_{ij} P_j c_i \Pi_i, \ \forall j$ and $s_i = \sum_j (E_j/Y) b_{ij} R_{ij} c_i \Pi_i P_j, \ \forall i$; all subject to the normalization $\sum_j (E_j/Y) P_j = 1$. based on $\sum_i s_i c_i \Pi_i = 1$. $\{R_{ij}, P_j, c_i \Pi_i, P_j, s_i\}$ depend on the realizations of the random draws ϵ_{ij}.

Ex ante expected relative resistances of arbitrageurs are generated assuming they use the econometric best practice PPML reduced form regression of bilateral trade flows on origin and destination fixed effects and a set of controls that include direct bilateral cost variables such as tariffs or transport costs or bilateral price observations. With complete datasets that include origin sales and destination expenditure, this combination yields power functions of relative resistance with fitted estimates f_{ij} of size adjusted trade $f_{ij} \equiv b_{ij}/s_i$:

$$(b_{ij}/s_i)^e = f_{ij} = \exp[-\theta^e \sum_h \gamma_h^e \ln z_{ij}^h - \ln \Pi_i^e - \ln P_j^e]. \tag{7}$$

The coefficients $\{\gamma_h^e, \ln \Pi_i^e, \ln P_j^e\}$ of the control variables $\mathbf{z}_{ij}$ are interpreted as the product of the trade elasticity θ^e and the elasticity of relative resistance with respect to the control variable. The CES specification interprets the fitted result as $(R_{ij}^e)^{-\theta^e}$. The 'trade elasticity' θ^e may be estimated from variation of direct trade cost variables to generate the multilateral resistances and thus all components of relative resistance. Given the trade elasticity, R_{ij}^es are generated and used by arbitrageurs to determine their *ex ante* distribution of the supplies $\{x_{ij}/y_i\}$.

From the perspective of the general model, the trade elasticity parameter θ^e in (7) is an approximation to an average of the *ex post* realized local trade elasticities θ_{ij}. In the general *ex post* model, realized relative resistance is

$$\tilde{R}_{ij} = \frac{(\sqrt{P_j} - P_j) b_{ij}/s_i + (1 + \sqrt{P_j})}{(\sqrt{P_j} + P_j) b_{ij}/s_i - (1 - \sqrt{P_j})} = \mathcal{R}(P_j, b_{ij}/s_i).$$

Given P_j, $\mathcal{R}(P_j, b_{ij}/s_i)$ is operational. The inverse trade elasticity $\partial \ln \tilde{R}_{ij}/\partial \ln b_{ij}/s_i$ is a function, not the theoretical CES parameter $-1/\theta$. $\tilde{R}_{ij}$ requires estimates of the P_js. Ex ante P_js are usually missing or unreliable, while treatment of uncertainty requires ex post P_js.

A combination of the standard PPML estimation used by arbitrageurs with the restrictions of the general model generates sufficient statistics for the ex post P_js along with the best fit average value of realized local inverse trade elasticities $\bar{\theta}$. One method is based on fitting $\tilde{R}_{ij}(P_j, \cdot)$ to fitted size-adjusted trade $f_{ij}(\cdot)^{-1/\bar{\theta}}$ by varying $\{P_j, \bar{\theta}\}$. The high degree of non-linearity of this method led to convergence difficulties in practice. The other method exploits the close approximation of the discrete percentage

deviation of R_{ij} from one, $\widehat{R_{ij}} = (R_{ij} - 1)/[(R_{ij} + 1)/2]$, to $\ln R^e_{ij}$ derived using the log-linearity of the CES gravity equation. This method converges smoothly.

Non-parametric R_{ij} is derived from:

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

The rightmost ratio is the discrete percentage deviation of R_{ij} from its as-if-frictionless value equal to one. Solve for the percentage deviation and simplify the resulting left hand side of the equation. The result is

$$\frac{\sqrt{P_j}b_{ij}/s_i - 1/\sqrt{P_j}}{(b_{ij}/s_i + 1)/2} = -\frac{R_{ij} - 1}{(R_{ij} + 1)/2} \equiv -\widehat{R_{ij}}. \tag{8}$$

The left hand side is a function of the share ratio data and the unknown P_j.

The right hand side of (8) is approximated by minus $\ln R_{ij}$, equal to the inverse of the CES gravity equation percentage change in in the share ratio from its frictionless share ratio expressed in log points:

$$(1/\theta^e)[\ln(b_{ij}/s_i)^e - \ln 1] = -(\ln R^e_{ij} - \ln 1).$$

For any particular share ratio and value of the discrete percentage change in relative resistance, a value of $1/\theta^e$ and *ex post* (realized) value of P_j can be found to exactly match the two values of the percentage change in relative resistance. For that value of $1/\theta^e$, the two equations will differ for other bilateral pairs because the general demand specification and its operational translog version generate local trade elasticities different from fixed θ^e assumed to be used by arbitrageurs. Assume for now that the arbitrageurs correctly anticipate the best fit average realized elasticity $\bar{\theta} = \theta^e$ The minimum distance estimator of the best fit average trade elasticity $\bar{\theta}$ and the *ex post* price indexes $\{P_j\}$ is

$$\min_{\{P_j,\bar{\theta}\}} \sum_{i,j} \left[\frac{-2}{b_{ij}/s_i + 1}\left(\sqrt{P_j}b_{ij}/s_i - 1/\sqrt{P_j}\right) + (1/\bar{\theta})\ln f_{ij}(\cdot)\right]^2 | \sum_j \frac{E_j}{Y}P_j = 1. \tag{9}$$

The right hand side of equation (7) is used for $f_{ij}(\cdot)$.

Program (9) minimizes the variance of arbitrage errors expressed by the square bracket term. Specifically, (9) finds the set of *ex post* effective price indexes and *ex post* trade elasticity $\bar{\theta}$ that minimizes the sum of squared arbitrage errors given by the square bracket of (9).

If the true model is CES and arbitrageurs' trade elasticity expectations are exact, $\theta^e = \bar{\theta}$, the square bracket term is equal to the pure arbitrage error plus the discrete

percentage approximation error. The calibrated elasticity $\bar{\theta}$ could be used for counterfactual projections. The approximation error is small except for large R_{ij}. Evidence of large realized R_{ij}s could be taken as evidence against the CES specification.[12] If the arbitrageurs' elasticity expectations are wrong, $\bar{\theta} \neq \theta^e$, the variance is underestimated. If the true model is the general translog, specification error is combined with the approximation error and and error due to the covariation of local θ_{ij} with R_{ij}.

The resulting solution value of the objective function is equal to the variance of the arbitrage errors times the number of observations. In the application, the observation number is the square of the number of countries, equal to $43^2 = 1849$. The linear structure of the sum of squared errors allows for useful decomposition of the overall variance. For example variation of variances by destination or origin may be associated with economically meaningful characteristics of the information problem facing arbitrageurs. The variances measure the arbitrage efficiency loss due to imperfect information. With constant relative risk aversion, the welfare effect on arbitrageurs is the product of the variance and the coefficient of relative risk aversion. The log utility case developed in Appendix 4.1 implies a coefficient of relative risk aversion equal to one.

Program (9) requires $f_{ij}(P_j, \cdot) \geq 0$ for $\ln f_{ij}$ to be defined. The implied negative values of R_{ij} imply negative buyer willingness-to-pay p_{ij}. Computation of (9) drops such cases with the rationale that they represent extreme arbitrage error with no useful information. Buyer price $p_{ij} < 0$ is very unlikely but not actually impossible.[13]

The *ex post* procedure (9) with *ex ante* program (1) embedded thus partially embeds selection. (9) implicitly eliminates bilateral flows for which an incipient $R_{ij} < 0 \Rightarrow p_{ij} < 0$.[14] In contrast, error associated with a missed positive profit opportunity is unobservable.

In practice, the PPML estimator uses a simpler (for estimation purposes) normalization: the German $P_{DEU} = 1$. This implies that all other P_js are normalized by $\sum_{j \neq DEU} P_j E_j/(Y - E_{DEU}) = 1$. This stands in contrast to the standard normalization in general equilibrium exercises, $\sum_j P_j E_j/Y = 1$ applied in (9). The choice of normalization is irrelevant to determination of the expenditure shares $\{b_{ij},\ B_i = s_i\}$ because they are homogeneity of degree zero in relative prices. Thus the PPML estimator is invariant to this choice. Program (9) *is* sensitive to the normalization, but is consistently based on the normalization used for the PPML estimator.

[12] Most revealed R_{ij}s cluster below 10. At $R = 10$, $\widehat{R} = 1.63$ while $\ln R = 2.3$ which exceeds the asymptotic upper bound $\widehat{R} = 2$. In the application described below, $R_{ij} > 1$ are in high enough proportions to cast doubt on the constant elasticity hypothesis.

[13] An example is the notorious Permian Basin case of negative price for natural gas at the source due to excess production meeting fixed pipeline capacity to ship it in the normal fashion.

[14] The percentage of bilateral trade eliminated is low, around 1-2 percent.

The non-parametric approach to gravity has important implications for counterfactual quantification exercises. Projected results are sensitive to the values of the trade elasticities used. For some exercises, the estimate of $\bar{\theta}$ from inverting the average realized trade elasticities may be appropriate. For other exercises, the cross-section variation of local trade elasticities around $\bar{\theta}$ may be important. Local counterfactuals that will be useful for bilateral trade negotiations imply that the bilateral asymmetry of trade elasticities is important. Results below demonstrate that asymmetry is prominent. Another example concerns the extensive margin of trade. Policies aimed at influencing this margin would be usefully guided by fitting translog parameterizations to the revealed R_{ij}s as in Section 4.4.

Elimination of the sectoral dimension eases exposition where it is not needed for clarity.

Tremendous heterogeneity of buyers (within as well as between countries) is allowed because spatial arbitrage buries the heterogeneity inside sufficient statistics for relative resistances. The essential assumptions are that a common invertible demand system governs all buyers and that the as-if-equilibrium comparison is *compensated* – by construction, utility is held constant at its observed equilibrium values. Also by construction, the normalized price indexes generated by (9) are compared to the as-if-frictionless price index based on purchase of supplies from the same origins as the observed equilibrium price indexes. Any gain in utility implied is removed by the assumed compensation and change in the set of goods purchased is similarly assumed away. On the supply side, variation of the net seller prices (c_is) cancels out in relative resistances cancel out c_is, so their cross section variation is irrelevant.

Inside the black box of arbitrage is heterogeneity of arbitrageurs and trade services providers along with price shifters and trade frictions. The assumed efficiency of arbitrage implies that revealed relative resistances absorb differences in individual arbitrageur errors and pricing to market markups or markdowns along with the other shifters. Bilateral trade frictions and trade elasticities are reduced form variables that locally combine demand with residual supply schedules in spatial equilibrium.

2.2 Measurement and Specification Errors

Ex ante efficient arbitrage seems plausible for observations aggregated over the yearlong time interval of most gravity applications. Annual production in each origin follows an arrival process as does annual demand in each destination. Shocks that violate orthogonality at daily frequencies are smoothed by error correction over the year. The one year time interval is a compromise justified by resource allocation decisions mostly set at yearly frequencies.

Measurement error of the shares affects the method in both the initial PPML gravity

estimation stage assumed to be the basis for arbitrageurs' expectations and the *ex post* realization stage (9). For the initial stage, measurement errors are part of the error term in equation (8), assumed to be an independent random variable. In program (9), the realized shares have measurement errors that enter nonlinearly in equation (8) that determines $\widehat{R_{ij}}$. The share measurement errors thus do not have multiplicative errors with mean one that are zeroed out in the log differences equal to realized measured arbitrage error $\widehat{R_{ij}} - (\bar{\theta})/\theta^e)\hat{f}_{ij}$. This casts much doubt on the convenient assumption that realized log resistance arbitrage errors are random independent errors with zero mean. The violation results from the bounded rationality of arbitrageurs.

Bounded rational expectations are formed by arbitrageurs. Their expectations of $\ln R_{ij}$ are based on (7). In the absence of controls for measurement error, this is the best that can be done. The minimum distance calibration of (**??**) still has attractive features, but probability theoretic interpretations are not valid.

Finally, there is specification error of three kinds, considered in order of increasing importance. First, least important, the approximation error difference in discrete and log percentage changes in R is small except for the instances where R_{ij} is large. From the point of view that the general translog is the 'true' model, the approximation error is part of the arbitrage error due to specification error

Second, the arbitrageurs may not have boundedly rational expectations or our PPML gravity model assumption of their expectations may be in error. In the absence of specific alternatives, there is no way to evaluate this critique.

Finally, the general translog approximation may not adequately capture the variation of the demand system. The associated error can be approximated using $\tilde{P}_j \approx \exp[\lambda \ln P_j + (1-\lambda)\ln 1] = P_j^\lambda$ and $\tilde{b}_{ij} \approx \lambda b_{ij} + (1-\lambda)s_i$.

The discrete percentage change in R_{ij} is $(R_{ij}-1)/(\lambda R_{ij} + 1 - \lambda)$. Evaluation of sensitivity of R_{ij} calculated from (6) to deviation of λ from $1/2$ is investigated by replacing $\sqrt{P_j}$ with P_j^λ and $\bar{b}_{ij}$ with $\lambda b_{ij} + (1-\lambda)s_i$. Exploration of this case with numerical methods shows low sensitivity around $\lambda = 1/2$ except when b_{ij}/s_i is small. In perspective both the technical result and common sense suggests that cases with small ratios of b_{ij}/s_i are imprecise source of information about R_{ij}.

3 Results

New material to be reported here. The application starts with modeling the arbitrageurs' anticipated relative resistances that will guide their distribution of the endowments. *Ex ante* efficient arbitrageurs infer *ex ante* relative resistances from a fitted standard PPML gravity equation of size-adjusted trade flows. Since there is nothing original in this stage, we skip reporting results. The arbitrageurs obtain their antici-

pated relative resistances by inversion the fitted equation using an estimate of the CES trade elasticity parameter. We equate their anticipated trade elasticity with the second stage calibrated trade elasticity that minimizes the variance of the arbitrage errors.

The anticipated relative resistances determine the arbitrageurs' share allocations x_{ij}/y_i. The predetermined share allocations to efficient cost-minimizing buyers determine their *ex post* willingness-to-pay dependent on the realizations of shocks to trade frictions and non-price buyer costs. The combined *ex post* spatial arbitrage equilibrium distribution is directly associated with realized relative resistances expressed in equation (6) as a function of realized effective buyer price P_j and realized size-adjusted trade shares b_{ij}/s_i. The realized buyer prices are not observed, but the minimization of buyer cost property allows them to be inferred. Stage two does this in program (9) where the inference includes the trade elasticity assumed to be used by arbitageurs in calculating their anticipations of relative resistance. The minimized value of the objective function in (9) is interpreted as an arbitrage error variance.

The last stage re-examines the determinants of relative resistances as in the first stage, but based on the relative resistances from stage two. Here we report results with an emphasis on the novel scale effects we uncover. Larger exporters benefit from greater opportunities to arbitrage congestion shocks across more destinations and have more and likely better outside options when facing holdup. In contrast, no scale effects are observed for importers, explained by fewer opportunities to arbitrage deliveries and fewer outside options vs. holdup.

The model and results are focused on the static equilibrium of spatial arbitrage, so results are reported for a yearly sequence of cross sections of bilateral trade. For some purposes the yearly results are pooled.

3.1 Arbitrage Error Variance

The standard PPML gravity estimation of the combined effect of controls on the fitted value of size-adjusted trade gives $(b_{ij}/s_i)^e = \hat{f}_{ij}$. The fitted share ratio is interpreted in the CES framework as equal to $(R^e_{ij})^{-1theta^e}$. $\hat{f}_{ij}$ is fitted to standard gravity controls. The arbitrageurs are presumed to infer anticipated log point relative resistances from the log of the fitted PPML model as $\ln R^e_{ij} = -(1/\theta^e) \ln \hat{f}_{ij}$ by assuming the CES specification and a fixed trade elasticity estimate θ^e. The boundedly rational expectations of the arbitrageurs are assumed to include using this state of the art model, with the CES elasticity inferred 'efficiently' to minimize the variance of arbitrage errors. Program (9) yields the results below.

The value of the minimized objective function divided by the number of observations is equal to the arbitrage error variance for the world. Its component country variances report the arbitrage error variance as it affects each country. The country variance for

each country i is calculated as the sum of the terms in minimized (9) containing i as origin or destination divided by the number of observations, 43+42=85. The variances can be further broken into country export and import arbitrage error variances. In this case the number of observations is 84 due to the exlusion of domestic trade. The figures below report the world variances by year and a few country variances. (There are 43 countries, so the number of observations for the world variance is $43 \times 43 = 1849$.

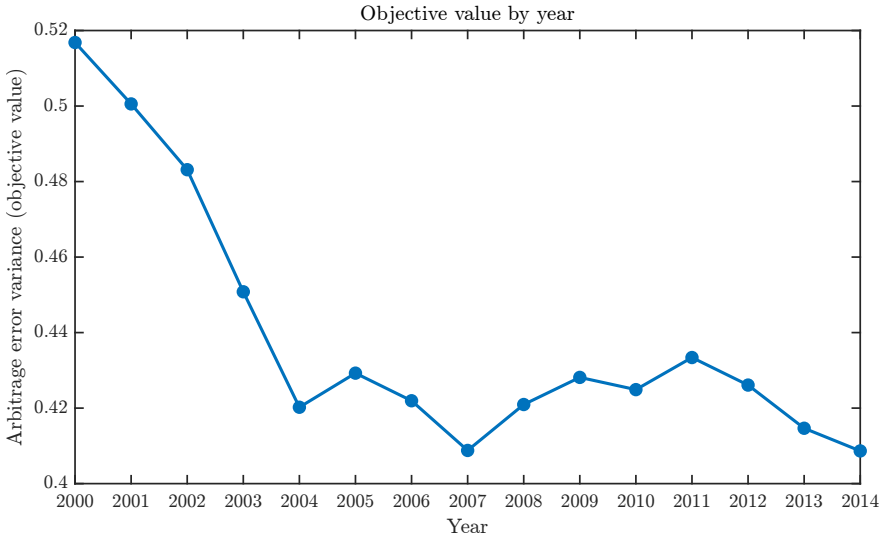


Figure 1: Yearly Arbitrage Error Variance

Figure 1 reveals a significant reduction in world arbitrage error variance over the period of globalization, around 20%. This progress was interrupted by a 2% increase associated with the Great Recession of 2007-08, followed by reversion by 2014 to the previous low error variance in 2007.

The selected country variances mostly reveal a similar pattern of shifts over time but there are interesting exceptions that suggest tentative explanations. The most prominent exception is China, which reveals rising variance both as exporter and importer. A tentative explanation is that China's much faster than average growth of manufacturing output drove much more than average expansion of the extensive margin of trade. New and new-ish markets plausibly are associated with higher rates of learning about bilateral trade frictions. The US in contrast conforms most closely to the world variance evolution. The figures also suggest that reduced form investigation of the panel of arbitrage error variances would be useful in future work focused on explanations using measures of investment in learning as well as policy and other disruptions beyond the Great Recession.

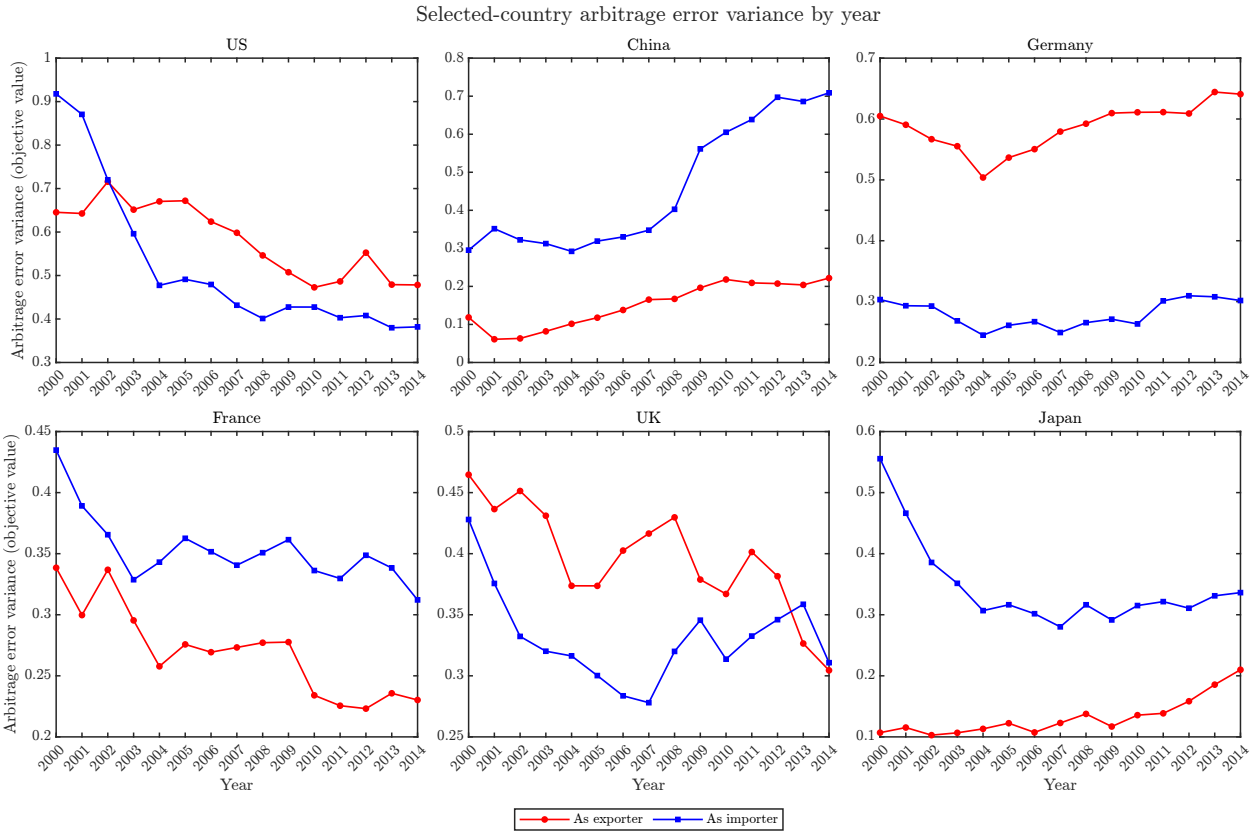


Figure 2: Country-level Variance

3.2 Trade Elasticity Variation

An implication of the model is local trade elasticity variation across time and across pairs at a point in time and over time. The cross section differences are important for applied trade policy work where local elasticities may be applied to project the effects of small changes in policies when affected part of the economy is small enough to ignore full general equilibrium resource reallocation and income distribution effects. An interesting pattern is that variation across pairs exhibits bilateral directional asymmetry for almost all countries. We illustrate with some examples.

Figure 3 displays the mean local inverse trade elasticity by year where the mean is taken first for each country and year and then the 95% confidence interval of country means is displayed around the mean over counries indicated by a dot. The line connecting the dots displays the same time variation pattern as the arbitrage error Figure 1.

Figure 4 displays the bilateral asymmetry of local elasticities. The yearly elasticities are pooled by country and the mean over time is calculated for each pair. The scatter plot of directional bilateral trade elasticities is displayed for the selected countries. The dashed 45 degree line indicates bilateral symmetry with the domestic trade elasticity necessarily on the dashed line. The strikingly flat bar for Germany (Deutschland) arises because the normalization is by Germany's effective price index. This implies that $P_{DEU} = 1$, which implies that its inward elasticity =1 by construction. This in turn implies that countries with P_j near 1 will tend to have rather horizontal scatter plots. Further away may indicate more vertical scatter plots. But many more sources of asymmetry may be at work. The patterns invite further investigation.

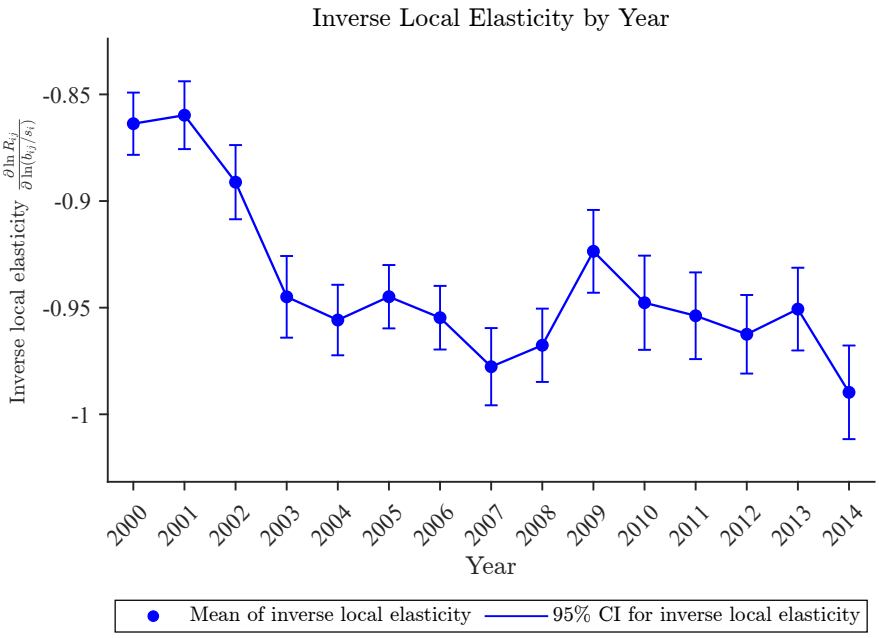


Figure 3: Inverse Local Trade Elasticity across Years

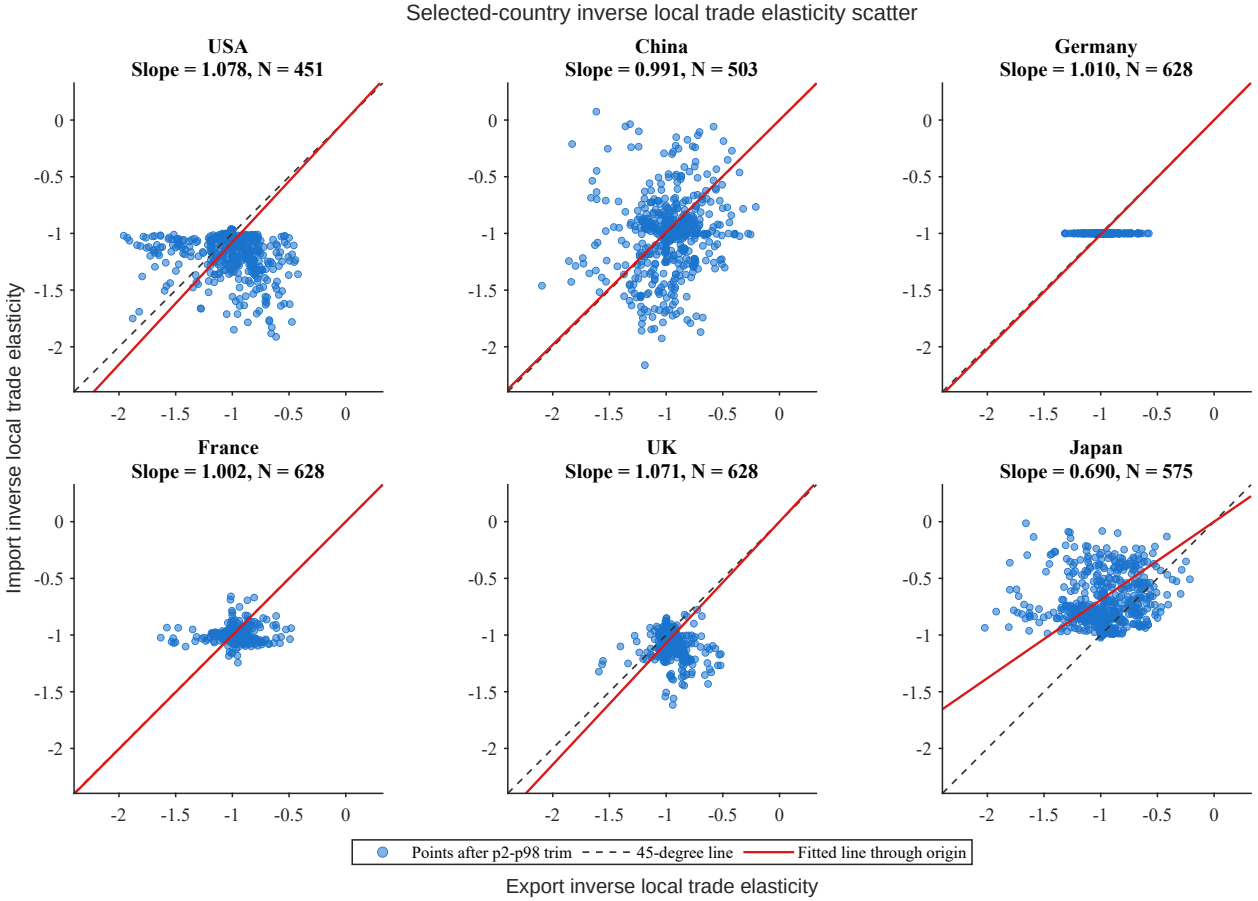


Figure 4: Bilateral Directional Asymmetry of Inverse Local Elasticities, Selected Countries

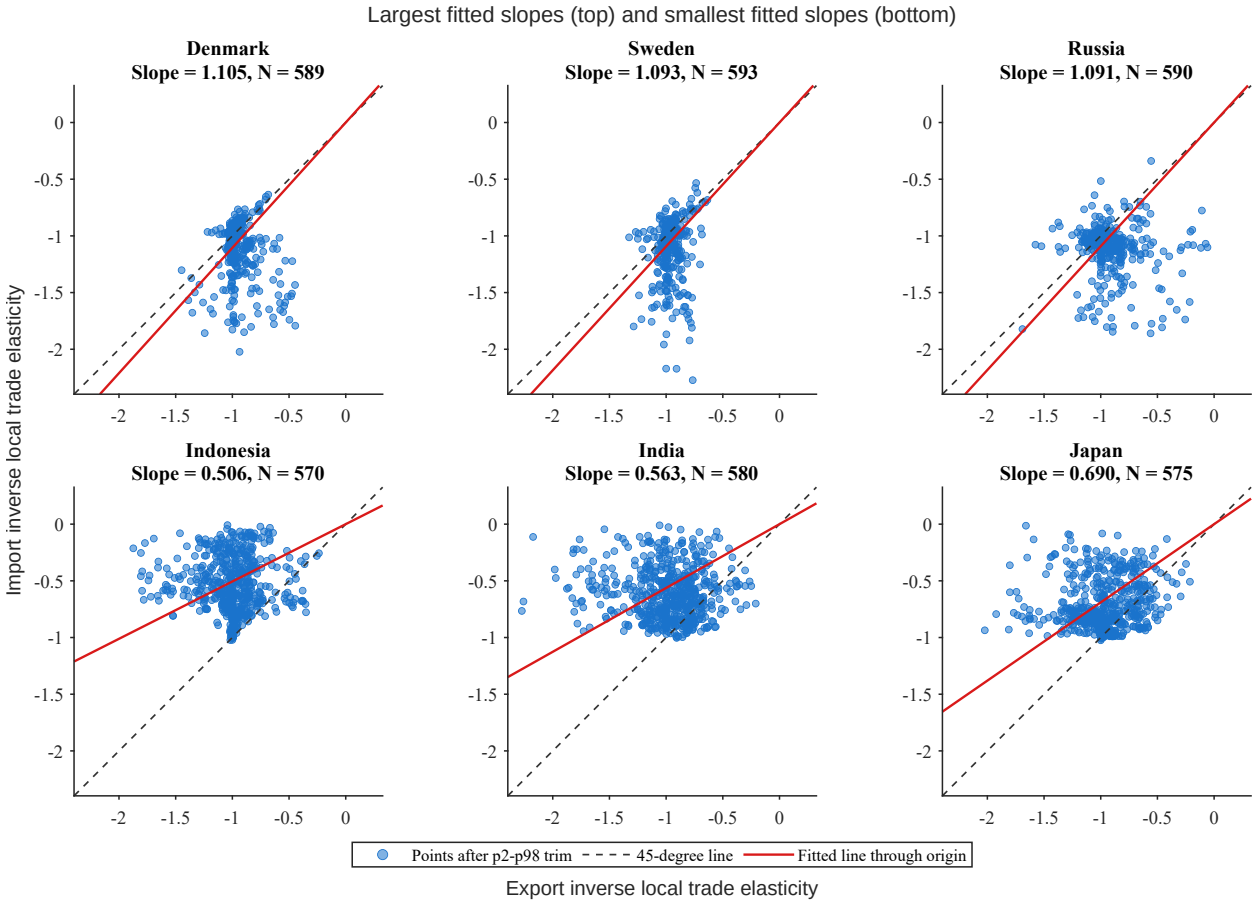


Figure 5: Bilateral Directional Asymmetry of Inverse Local Elasticities, Strongest Countries

3.3 Determinants of Revealed Relative Resistance

The investigation of revealed relative resistance determinants begins with descriptive statistics in Table 1. The full sample of observations of manufacturing bilateral trade for 43 countries over 14 years ends up with some deletions due to negative R_{ij} arising in program (9). The second row of the table reports a small number of further deletions that arise due to the inverse local trade elasticities associated with solutions to program (9) being unstable with respect to computations repeated with different starting points. These are associated mechanically with low size adjusted trade ratios b_{ij}/s_i and effective price P_j close to one. Low ratios are also associated with $\mathcal{R}_{ij}(P_j, b_{ij}/s_i)$ being close to switching signs and suggest possible error such that arbitrage may be close to earning negative profits.

Table 1: Summary Statistics of $\widehat{R}_{ij}$ under Alternative Trimming Rules

Sample	N	Deleted	% Deleted	Mean	Median	$s.d.$
Full sample	27,247	488	1.76	11.85	3.13	154.65
Inverse elasticity trim	26,975	760	2.74	8.72	3.11	28.90

Notes: The sample is constructed from aggregate manufacturing data based on the WIOD dataset. The theoretical number of observations is 43 origins×43 destinations×15 years, totaling 27,735 observations. The full sample includes all observations with $\widehat{R}_{ij} > 0$. The inverse local trade elasticity trim excludes observations in the top and bottom 0.5 percent of the distribution of inverse local trade elasticity. Reported deletion counts and percentages are computed relative to the theoretical sample size.

The full distribution of the trimmed sample of $\ln R_{ij}$ is shown in Figure 6. It reveals a puzzling drop in density near the center of the distribution.

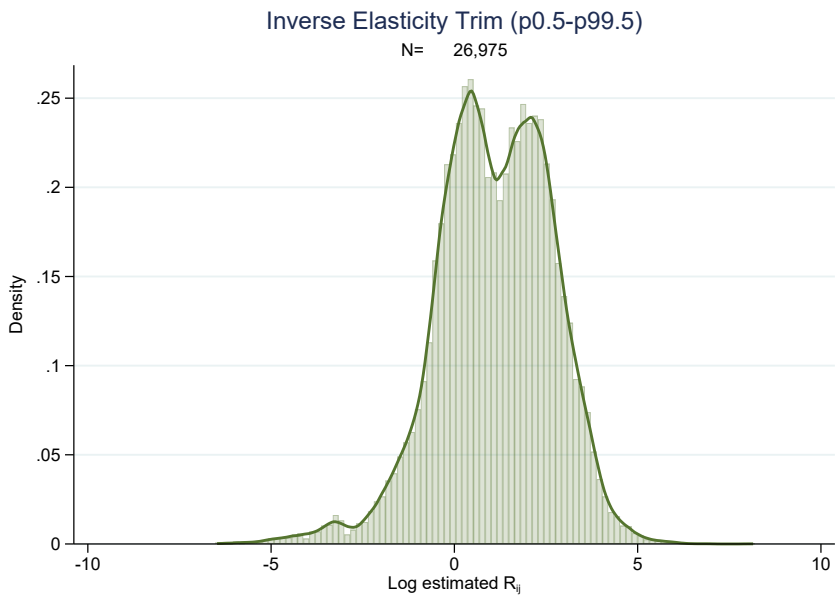


Figure 6: Distribution of Estimated $\ln R_{ij}$ under Inverse Local Trade Elasticity Trim

We apply this data in a reduced form regression of $\ln R_{ij}$s on a set of standard controls for bilateral resistance $\hat{\tau}_{ij,t} = \beta \cdot \mathbf{Z}_{ij}$:

$$\ln R_{ij,t} = \chi_{i,t} + \mu_{j,t} + \hat{\tau}_{ij,t} + \epsilon_{ij,t}$$

Then model possible scale effects that are absorbed in the origin-time and destination-time fixed effects or remain in the error term $\epsilon_{ij,t}$. The scale effects are modeled with $(1 - \kappa_{ij}) \times (\gamma \ln IM_{j,t} + \phi \ln EX_{i,t})$ to augment the previous version of $\hat{\tau}_{ij,t}$:

$$\ln R_{ij,t} = \chi_{i,t} + \mu_{j,t} + \beta \cdot \mathbf{Z}_{ij_t} + (1 - \kappa_{ij}) \times (\gamma \ln IM_{j,t} + \phi \ln EX_{i,t}) + \upsilon_{ij,t}. \qquad (10)$$

Here κ_{ij} is the Kronecker delta and γ and ϕ are the import and export border scale elasticities with respect to import and export volumes $IM_{j,t}$ and $EX_{i,t}$ respectively. (A subtlety is that re-export trade should be included in the inward and outward trade flows, on the grounds that its mass does affect the scale costs inward and outward.)

Scale effects in (10) are identified off both partner and time variation in imports and exports *relative to domestic trade*. The estimated import scale effect is positive, resistance is increased with volume. We interpret this to mean that larger importers face worse congestion and holdup at border entry points because they tend to have fewer domestic alternatives to substitute during delays and thus also fewer and thus less favorable outside options in the face of holdup. In contrast, the export scale effect is negative. We interpret this to mean that larger exporters have more alternatives to

arbitrage congestion at border exit points in the interim between order confirmation and delivery. This is because they typically have larger domestic markets and more alternative export destinations. The same reasoning applies to holdup: larger exporters have more alternative outside options in renegotiation when faced with holdup. More discussion is provided following the presentation of results.

In principle, the model of equation (10) is better applied to sectoral data, where the import and export volume data should be aggregated across all sectors. This reduces the collinearity of bilateral trade with the aggregate volumes in (4). The model can alternatively be run with uniform sector-time γ and ϕ, with time varying uniform sector γ_t, ϕ_t or as a series of cross-section regressions to allow observation of the time path of the sequence of $\{\gamma_t^k, \phi_t^k\}$. The sequence of time varying scale effects may reveal capacity adjustment, but it is unmodeled. The structure can be fine tuned by allowing for groups of countries to have different scale elasticities. The suggested regressions reflect a crude notion of how scale effects at the border might operate.

To make equation (10) operational, we rely on the voluminous empirical gravity literature to proxy for the trade frictions in vector $\mathbf{Z}_{ijt}^k$. Specifically, we employ three sets of trade cost variables. First, we employ 'standard' proxies for time-invariant bilateral trade frictions that are used most often in the gravity literature, including the log of bilateral distance (LN_Dist_{ij}) between two trading partners, and indicators for contiguous borders ($CNTG_{ij}$), the presence of colonial relationships ($CLNY_{ij}$), and common official language ($LANG_{ij}$).[15] Second, we add a series of time-varying bilateral policy covariates, including the log of tariffs (LN_Tariff), membership in Customs Unions ($CU_{ij,t}$), membership in Economic Integration Agreements ($EIA_{ij,t}$), membership in the World Trade Organization ($WTO_{ij,t}$), and membership in Currency Unions ($CURR_{ij,t}$).[16] Finally, we use a vector of time-varying border indicators, which take a value of one for international RRs and are equal to zero for domestic RRs for each year in our sample. The estimates on these dummy variables capture the impact of common (de-)globalization trends that have affected the international relative to the domestic RRs, e.g., improvements in communication, transportation, global recessions, etc. To further proxy for domestic trade costs, we also use a variable for domestic distance ($LN_Dist_Intra_{ij}$).[17]

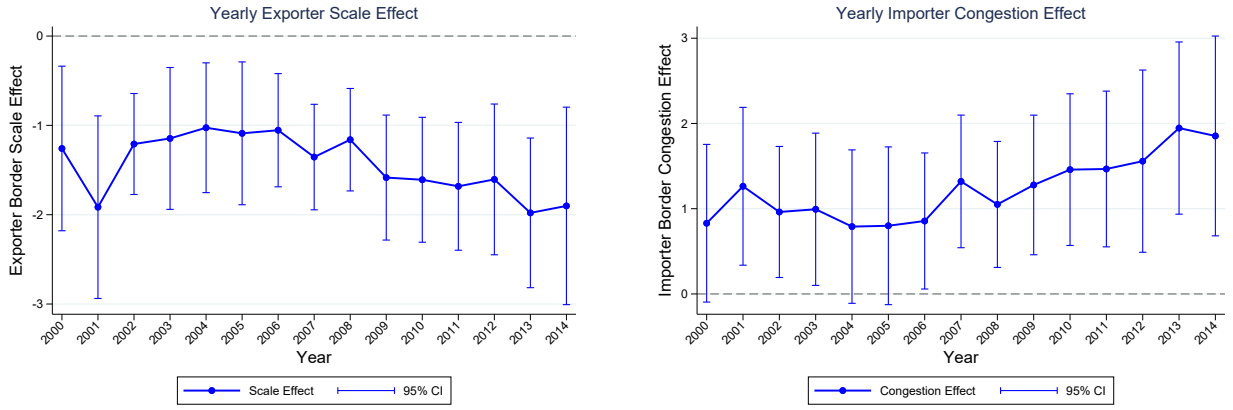
Our estimates appear in Table 2, and we develop the analysis in three steps. Our

[15] Data on the standard time-invariant bilateral covariates comes from the *Dynamic Gravity Database (DGD)* of the U.S. International Trade Commission (Gurevich and Herman, 2018).

[16] LN_Tariff is defined as $\ln(1 + tariff)$, where $tariff$ is the *ad-valorem* tariff, and the data on tariffs comes from Caliendo and Parro (2015). The data on CUs, and EIAs, is from Mario Larch's RTA database (Egger and Larch, 2008), and these variables are mutually exclusive by definition. The data on membership in currency unions and the WTO if from the DGD database.

[17] Consistent with our measure of international distance, domestic distance is constructed as population-weighted average of the great circle distance between the cities within a country.

benchmark results, with the full set of gravity variables, are presented in column (1). Column (2) introduces country-pair-sector fixed effects. Finally, we introduce the scale effects in column (3). The results in column (1) are consistent with corresponding estimates from the existing parametric gravity literature. Specifically, all estimates of the effects of the 'standard' time-invariant bilateral gravity variables are statistically significant, and they suggest that larger distances (both domestic and international) translate into higher relative resistance, while sharing common borders, common language, and colonial relationships are all associated with lower relative resistance. In terms of sign and statistical significance, these estimates are consistent and comparable with corresponding results from the parametric gravity literature. Thus, they reinforce the validity of the relative resistance.



(a) Yearly Exporter Scale Effect

(b) Yearly Importer Congestion Effect

Figure 7: Relative Resistance and Scale/Congestion effects

Turning to the estimates on the policy variables, we see that most of them are as expected. Specifically, our results suggest that higher tariffs are associated with higher relative resistance, while customs unions, economic integration agreements, and WTO membership decrease the relative resistance. The sign of the estimate currency unions is opposite to our expectations. A possible explanation for this counterintuitive result is that the policy variables in our setting are potentially endogenous, and we will revisit this result in the next specification. Before that, we note that all estimates on the time-varying border effects are large, positive, and statistically significant, as expected, i.e., international resistance is significantly larger than domestic resistance. Finally, the time-varying border estimates are decreasing over time, which reflects common globalization trends, i.e., falling international relative do domestic resistance during the period of investigation.

The results in column (2) of Table 2 are obtained with country-pair-sector fixed effects, which control for all possible time-invariant bilateral trade frictions at the

Table 2: The Determinants of the Relative Resistance

	(1) Benchmark	(2) Pair FE	(3) Scale Effects
Log distance	1.197***		
	(0.043)		
Log intra-national distance	0.566***		
	(0.212)		
Contiguity	-0.438***		
	(0.095)		
Common language	-0.086		
	(0.094)		
Colonial relationship	-0.417***		
	(0.114)		
Log tariff	0.206***	0.187***	0.183***
	(0.039)	(0.022)	(0.022)
Customs union	-0.126	-0.120	-0.113
	(0.091)	(0.076)	(0.076)
Economic integration agreement	-0.087	-0.077**	-0.071*
	(0.069)	(0.037)	(0.037)
WTO membership	-0.033	-0.045	0.029
	(0.197)	(0.176)	(0.134)
Common currency	0.185***	-0.011	-0.011
	(0.051)	(0.044)	(0.044)
$(1-\kappa_{ij})$Log Exports$_{i,t}$			-1.510***
			(0.327)
$(1-\kappa_{ij})$Log Imports$_{j,t}$			1.070**
			(0.438)
Pair FE	No	Yes	Yes
Exporter-year FE	Yes	Yes	Yes
Importer-year FE	Yes	Yes	Yes
Cross border-year FE	Yes	Yes	Yes
Scale controls	No	No	Yes
Observations	26,975	26,969	26,969
Adjusted R^2	0.806	0.955	0.955

Notes: The sample is restricted to observations with inverse local elasticity trim described in Table 1. The dependent variable is $\ln \widehat{R}_{ij}$. κ_{ij} is the Kronecker delta where $\kappa_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$. Column (3) additionally includes cross-border exporter and cross-border importer scale controls. The inverse local elasticity trim removes some observations, which can generate singleton pairs; these are dropped when exporter-importer pair fixed effects are included, explaining the smaller number of observations in columns (2) and (3). Standard errors clustered at the exporter-importer pair level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Robustness to Alternative Specifications of Scale and Congestion Effects

	(1) No BRDR	(2) No Pair FE	(3) No Pair FE+No BRDR
Log distance		1.219***	1.220***
		(0.043)	(0.043)
Log intra-national distance		-0.057	-0.049
		(0.161)	(0.162)
Contiguity		-0.421***	-0.419***
		(0.096)	(0.096)
Common language		-0.079	-0.080
		(0.095)	(0.095)
Colonial relationship		-0.408***	-0.407***
		(0.116)	(0.116)
Log tariff	0.183***	0.200***	0.197***
	(0.022)	(0.038)	(0.038)
Customs union	-0.113	-0.101	-0.101
	(0.076)	(0.091)	(0.091)
Economic integration agreement	-0.071*	-0.037	-0.034
	(0.037)	(0.067)	(0.068)
WTO membership	0.008	0.445***	0.518***
	(0.137)	(0.169)	(0.186)
Common currency	-0.012	0.200***	0.198***
	(0.044)	(0.050)	(0.050)
$(1-\kappa_{ij})$Log Exports$_{i,t}$	-1.502***	-0.994**	-1.043**
	(0.337)	(0.429)	(0.440)
$(1-\kappa_{ij})$Log Imports$_{j,t}$	0.930***	0.434	0.534
	(0.327)	(0.471)	(0.461)
Pair FE	Yes	No	No
Exporter-year FE	Yes	Yes	Yes
Importer-year FE	Yes	Yes	Yes
Cross border-year FE	No	Yes	No
Observations	26,969	26,975	26,975
Adjusted R^2	0.955	0.810	0.810

Notes: The sample is restricted to observations with inverse local elasticity trim described in Table 1. κ_{ij} is the Kronecker delta where $\kappa_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$. The dependent variable is $\ln \widehat{R}_{ij}$. All columns include the exporter and importer scale controls, $\log \text{Exports}_{i,t}$ and $\log \text{Imports}_{j,t}$, to assess the robustness of the scale/congestion effects. Column (1) excludes cross-border-year dummies relative to the pair fixed effects specification(Column (3) in Table 2). Column (2) adds the scale controls to the benchmark specification (Column (1) in Table 2) while retaining the cross-border-year dummies. Column (3) removes the cross-border-year dummies from column (2). Time-invariant bilateral controls are absorbed in column (1) because exporter-importer pair fixed effects are included. Standard errors clustered at the exporter-importer pair level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Cross-Border-Year Dummy Estimates

	(1) Benchmark	(2) Pair FE	(3) Scale Effects
Other estimates		See Table 2	
Cross_Border_2000	-1.605		
	(1.358)		
Cross_Border_2001	-1.873	-0.197**	-0.153*
	(1.293)	(0.091)	(0.087)
Cross_Border_2002	-1.900	-0.391***	-0.405***
	(1.366)	(0.150)	(0.122)
Cross_Border_2003	-1.829	-0.525***	-0.438***
	(1.361)	(0.157)	(0.137)
Cross_Border_2004	-1.633	-0.445***	-0.241
	(1.382)	(0.150)	(0.169)
Cross_Border_2005	-1.535	-0.485***	-0.244
	(1.396)	(0.134)	(0.196)
Cross_Border_2006	-1.504	-0.525***	-0.253
	(1.361)	(0.154)	(0.217)
Cross_Border_2007	-1.613	-0.714***	-0.414
	(1.356)	(0.165)	(0.268)
Cross_Border_2008	-1.758	-0.688***	-0.312
	(1.374)	(0.171)	(0.283)
Cross_Border_2009	-1.604	-0.408**	-0.142
	(1.397)	(0.164)	(0.229)
Cross_Border_2010	-1.799	-0.664***	-0.332
	(1.411)	(0.189)	(0.262)
Cross_Border_2011	-1.738	-0.743***	-0.290
	(1.399)	(0.190)	(0.297)
Cross_Border_2012	-1.800	-0.835***	-0.433
	(1.395)	(0.203)	(0.278)
Cross_Border_2013	-1.848	-0.793***	-0.409
	(1.396)	(0.192)	(0.294)
Cross_Border_2014	-1.817	-0.915***	-0.505
	(1.370)	(0.214)	(0.324)
Pair FE	No	Yes	Yes
Exporter-year FE	Yes	Yes	Yes
Importer-year FE	Yes	Yes	Yes
Observations	26,975	26,969	26,969
Adjusted R^2	0.806	0.955	0.955

Notes: The sample is restricted to observations with inverse local elasticity trim described in Table 1. The dependent variable is $\ln \widehat{R}_{ij}$. All non-border-year estimates are omitted here and reported in Table 2. Column (1) reports the full set of cross-border-year dummy estimates from the benchmark specification. In columns (2) and (3), Cross_Border_2000 is omitted as the reference category once the exporter-importer pair fixed effects specification is imposed, so only Cross_Border_2001–Cross_Border_2014 are reported. Standard errors clustered at the exporter-importer pair level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

country-pair-sector level. Before we interpret the results, we note the following. First, we can no longer obtain estimates of the effects of the standard time-invariant gravity variables, as those are perfectly collinear with (and therefore fully controlled for and absorbed by) the country-pair(-sector) fixed effects. Second, we can no longer obtain estimates on all time-varying border indicators. Due to perfect collinearity in the presence of the country-pair(-sector) fixed effects, we can identify $T-1$ of the time-varying border estimates, where $T=15$ is the number of years in our sample. Without implications for any estimates other then those of the time-varying border dummies, we select to drop the border for 2000, the first year in the sample. Thus, all other time-varying border estimates will be relative to the dropped border effect in 2000.

A comparison between the estimates from columns (1) and (2) revels four differences. First, the policy estimates in column (2) are smaller in magnitude. Second, the estimate on WTO membership is no longer statistically significant. A possible explanation for this result is that, in the presence of the country-pair fixed effects, the WTO estimate is identified of very few observations, i.e., due to new members that joined during the period 2000-2014.[18] Third, the estimate on currency unions is negative and it is statistically significant, as expected. This change can be due to the fact that the country-pair fixed effects have mitigated endogeneity concerns (Baier and Bergstrand, 2007) or because the currency unions that were formed during the period of investigation had stronger impact on trade. Finally, the estimates on the border dummies are all negative and increasing (in absolute value) over time. As discussed earlier, this is because they are obtained as deviations from the dropped border estimate for 2000. As before, our border estimates capture common globalization trends.

The results in column (3) are obtained after introducing the scale effects ($\ln IM_{j,t}$ and $\ln EX_{i,t}$) to our specification. Three main findings stand out from column (3). First, focusing on the policy estimates, the estimate on the customs unions is no longer statistically significant, while the estimate on the WTO becomes positive and statistically significant, which is counterintuitive. Second, the border estimates become mostly positive and increasing in magnitude over time, i.e., the trend in the border/globalization effects is reversed and, once we explicitly account for possible scale effects, our estimates capture de-globalization during the period 2000-2014. Finally, we obtain large and statistically significant scale effects on the exporter side, while the corresponding effects on the importer side are much smaller and not statistically significant.

The positive scale effect for exports can be explained by spatial arbitrage opportunities available to exporting firms that are not available to importing firms due to

[18]The only new members in our sample that joined the WTO during the period of investigation are China (2001), Lithuania (2001), Taiwan (2002), and Russia (2012).

the timing of routing decisions. Trade statistics are temporal aggregates of a stream of transactions, typically a year in the gravity literature. A contract between buyer and seller sets a price to deliver a product within an interval typically specified with an upper bound, subject to possible penalties for default. The bargained price is a weighted average of the outside options of buyer and seller. A seller typically holds multiple contracts for delivery within an interval for any particular contract while also negotiating and/or entering new contracts with other buyers. The multiple choices for deliveries open for the seller the opportunity for intertemporal-spatial arbitrage. Deliveries are diverted from lower value to higher value buyers, subject to the upper bound or the default penalty. (Think of Uber ride buyers being dropped by Uber ride sellers at a busy airport.) Buyers can in principle engage in similar behavior with multiple sellers to arbitrage shipment delays or shift to alternative options in the face of holdup but in practice their sourcing is far less diversified than sellers' sales. The effect of this disparity in the cross section of importers rationalizes the positive effect of scale on cost because larger imports tend to be associated with smaller domestic alternatives.

In the cross section of aggregated trade flows for a longer period such as a year, larger exporters typically have more diversified sales, selling to more buyers within destinations, to more destinations and to larger domestic markets. So their arbitrage opportunities are better than for smaller exporters. The scale effect operates at the firm level for all sorts of intertemporal variation of buyers willingness-to-pay, but it has an external scale effect when the variation is due to congestion at border-crossing points. It is worthwhile to examine whether the scale effect is larger for time periods with more congestion and in the cross section at locations with more congestion.

This story relies on buyer heterogeneity and the associated heterogeneity of contractual prices. The theoretical model is flexible enough to accommodate this buyer-seller heterogeneity because the relative resistance statistic is a sufficient statistic consistently aggregated over heterogeneous buyers.

4 Conclusion

This paper introduces and applies a model of spatial arbitrage of trade in goods under uncertainty. The novel method is its main contribution. The application illustrates its usefulness with some examples that suggest many questions for future research both theoretical and empirical.

On the theory front, integrated treatment of production and trade under uncertainty is important and may be feasible with *ex ante* efficiency imposed by applying the Diamond stock market model as in Helpman and Razin (1978). The intertemporal effects of uncertainty on investment via the option value of waiting as in Handley and

Limao (2017) bring in dynamic considerations.

On the empirical front, the arbitrageurs and their actions are a black box that micro-economic investigations can fruitfully open and inform. Data and modeling of what trade intermediaries actually do may advance knowledge. Congestion and holdup are suggested as explanations for important empirical regularities in the application of this paper. Micro models that feature costly responses to them are available.

References

Anderson, James E., "Commercial Rivalry as Seller Incidence Shifting: Non-parametric Accounting of the China Shock," *Journal of International Economics*, 2026, *159*.

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

_ **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.

Baier, Scott L. and Jeffrey H. Bergstrand, "Do Free Trade Agreements Actually Increase Members' International Trade?," *Journal of International Economics*, 2007, *71* (1), 72–95.

Caliendo, Lorenzo and Fernando Parro, "Estimates of the Trade and Welfare Effects of NAFTA," *Review of Economic Studies*, 2015, *82* (1), 1–44.

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

Deaton, Angus, "The Distance Function in Consumer Behaviour with Applications to Index Numbers and Optimal Taxation," *The Review of Economic Studies*, 1979, *46* (3), 391–405.

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

Egger, Peter and Mario Larch, "Interdependent Preferential Trade Agreement Memberships: An Empirical Analysis," *Journal of International Economics*, 2008, *76* (2), 384–399.

Feenstra, Robert C. and John Romalis, "International Prices and Endogenous Quality," *The Quarterly Journal of Economics*, 2014, *129* (2), 477–527.

Gurevich, Tamara and Peter Herman, "The Dynamic Gravity Dataset: 1948-2016," 2018. USITC Working Paper 2018-02-A.

Handley, Kyle and Nuno Limao, "Policy Uncertainty, Trade, and Welfare: Theory and Evidence for China and the United States," *American Economic Review*, 2017, *107*, 2731–83.

_ **and** _ , "Trade Policy Uncertainty," *Annual Review of Economics*, 2022, *14* (1), 363–395.

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

Helpman, Elhanan and Assaf Razin, *A Theory of International Trade Under Uncertainty*, Academic Press, 1978.

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

Samuelson, Paul A., "Spatial Price Equilibrium and Linear Programming," *American Economic Review*, June 1952, *42* (3), 283–303.

Online Appendix

4.1 Risk Aversion

The implications of risk averse behavior by arbitrageurs are illustrated by assuming that arbitrageurs maximize the log of profits. (This is the case where the coefficient of relative risk aversion is equal to one.) The arbitrageurs program is:

$$\max_{\{x_{ij}/y_i\}} \mathcal{E}\left[\ln\left(\sum_i\sum_j p_{ij}x_{ij}/y_i\right) | \sum_j t_{ij}x_{ij}/y_i = 1,\ \forall i\right].$$

The first order conditions are

$$\mathcal{E}\left[\frac{p_{ij}}{\sum_{i,j} p_{ij}x_{ij}/y_i}\right] = \mu_i \mathcal{E}[t_{ij}],\ \forall i,j.$$

The Lagrange multiplier $\mu_i = c_i^e \Pi_i^e$, the anticipated opportunity cost of of serving any destination j from origin i.

Based on results in Section 2.1, the normalization of world price indexes implies that the denominator in the square bracket term on the left is equal to one as the numerator is equal to the relative realized price p_{ij}/P_j. Then the first order condition is equivalent to

$$\mathcal{E}[R_{ij}P_j] = \mathcal{E}[t_{ij}].$$

The algebra of covariance implies

$$Cov[R_{ij}, P_j] + \mathcal{E}[R_{ij}]\mathcal{E}[P_j] = \mathcal{E}[t_{ij}].$$

Risk neutrality as in the main text eliminates the role of the covariance.

The contribution of risk averse behavior implies a percentage change in relative resistance equal to the covariance between R_{ij} and P_j given by the properties of equation (6). Here, variation of realized R_{ij} is induced by variation in realized b_{ij}/s_i, which in turn induces variation in realized P_j. The partial elasticity of R_{ij} with respect to P_j is ordinarily negative for $R_{ij} > 1$ (hence most imports, $i \neq j$) and positive for $R_{ij} < 1$ (hence most domestic trade $i = j$). Thus risk aversion implies an increase in imports (x_{ij}/y_i rises) and reduction in domestic sales relative to risk neutrality (x_{ii}/y_i falls), and similarly for increases in riskiness.

Development and application of this model is beyond the scope of this paper.

4.2 Demand Heterogeneity

The scale effects directly connected to non-price bilateral trade frictions in the preceding section are complemented in this section with non-price friction effects associated with heterogeneity (across destinations) in the composition of buyers. Given the nature of international trade data, final and intermediate demand cannot be disentangled. In final demand, heterogeneity is due to non-homothetic preferences. In intermediate demand heterogeneity is due to non-homothetic technology but the resulting non-homothetic cost functions are formally equivalent to the non-homothetic expenditure functions for final demand buyers. The formal treatment focuses on final demand, the difference reappearing in the implications for application.[19]

The effect on revealed relative resistances of differences across origins in average real income per capita are absorbed in origin fixed effects (and similarly for differences across origins in average firm size). The concern addressed in this section is the effect of cross-origin heterogeneity in the distributions of real income or firm sizes.

The expenditure function of an agent is generally $E(\mathbf{p}, u)$, with price vector $\mathbf{p}$ and utility u. $E(\cdot)$ is concave and homogeneous of degree one in $\mathbf{p}$ due to cost minimization. This note specializes $E(\cdot)$ to $E = e(\mathbf{p})u$ with the *effective* price elements p_i of $\mathbf{p}$ equal to observable prices $\bar{p}_i$ times exogenous taste shifters for each i times potential taste shifter functions of utility u for each i. The setup here is a member of the EASI demand system class of Lewbel and Pendakur (2009).

The combined price shifter functions must be restricted for consistency. Insight into sufficient restrictions follows from differentiation of $E(\cdot)$ totally with respect to u. Simplify using Shephard's Lemma to form

$$\frac{E_u u}{E} - 1 = \sum_i b_i \frac{p_{iu} u}{p_i},$$

where b_i is the expenditure share on good i and $p_{iu}u/p_i = \varepsilon_i$ is the elasticity of effective price i with respect to real income u. Homotheticity implies both sides of the equation are equal to zero, where a sufficient condition for the right hand side to be zero is $\varepsilon_i = 0;\ \forall i$. For this study we parameterize the ε_is; then restrict the price-dependent income effects such that $\sum_i \varepsilon_i = 0$, homotheticity in effective prices prevails.

[19]The literature has focused on non-homotheticity with respect to goods classes (i.e. sectors). The variation of total expenditure across the goods classes at given prices is related to the variation of income, hence within-sector variation by origin of varieties is assumed away by homothetic CES sectoral preferences. In this setup, income variation plays no role in the sector-origin price shifters of the non-parametric analysis. In broader perspective, homotheticity seems more appropriate for higher levels of aggregation, whereas non-homotheticity is a bigger concern for narrowly defined categories such as national origin sectors.

The expenditure (cost) function is defined by

$$E(\mathbf{p}, u) = e(\{\bar{p}_i u^{\epsilon_i - \bar{\epsilon}}\})u^{\bar{\epsilon}}. \tag{11}$$

The multi-household version of a country's expenditure function extends (11) with households denoted h.

$$\sum_h e(\mathbf{p}^h, u^h) = \sum_h e(\{\bar{p}_i (u^h)^{\epsilon_i - \bar{\epsilon}}\})(u^h)^{\bar{\epsilon}}$$

Shephard's Lemma yields the expenditure shares $b_i = \partial \ln E / \partial \ln p_{ij}$. $\{\epsilon_i\}$ is a set of shift parameters and $\bar{\epsilon} = \sum_i b_i \epsilon_i$ is implicitly defined so that $e(\cdot)$ is concave and homogeneous of degree one in its effective price arguments.[20]

The share function dependence on real income is given by

$$\frac{b_{iu} u}{b_i} = \epsilon_i - \bar{\epsilon}. \tag{12}$$

The ratio of shares b_i/b_l at constant prices varies with real income u with elasticity $\epsilon_i - \epsilon_l$.

Arbitrage equilibrium expenditure shares of countries vary with income in the cross section according to (12). Allowing for income distribution, the shares also vary with *changes* in the real income distribution. Applied to revealed relative resistances, this setup implies that In the one consumer case for each country, the variation in u^j is absorbed in the buyer multilateral resistances. The multi-household extension implies a cross-country variation in effective prices due to heterogeneity of income effect shifters that is not absorbed into the buyer multilateral resistances. This observation suggests a reduced form model to capture the effect of non-homotheticity using measures of real income inequality of the origin countries.

The Gini coefficient is taken to represent the heterogeneity of the distribution of income. The effect of non-homotheticity via heterogeneity in income distribution on the expenditure shares is constrained by the adding up condition, the sum of shares must equal one. Thus in the data the non-homotheticity effect is observed in cross-border shares relative to the domestic share. The implied regression to investigate

[20]There is no closed form for the indirect utility function. It is implicitly defined by the expenditure function so long as the expenditure function is concave and homogeneous of degree one in the effective price vector, and increasing in utility. Expenditure (cost) function (11) is novel (as far as I have found), in that the effective price vector enters with no restrictions other than concavity and homogeneity of degree one. It is a member of a set of demand systems characterized by Lewbel and Pendakur (2009) that similarly are associated with implicit indirect and direct utility functions. In practice, since relative resistances are exact when the expenditure function is translog, the expenditure function is a version of the non-homothetic translog.

non-homotheticity is based on

$$R^k_{ij,t} = \chi^k_{i,t} + \mu^k_{j,t} + (1-\kappa_{ij})\Gamma^k_i \ln G_{j,t} + \ln \hat{\tau}^k_{ij,t} + \omega^k_{ij,t}. \tag{13}$$

Here $G_{j,t}$ is the Gini coefficient for origin j at time t, Γ^k_i is a demand parameter specific to goods from origin i subject to $\sum_k b^k_{ij}\Gamma^k_i = 0$, $\tau_{ij,t}$ is a pair fixed effect and $\omega^k_{ij,t}$ is a random error term.

Combine the non-homothetic specification (13) with the scale effect specification and impose the theoretical constraint $P_j R_{ij} = \tau_{ij}/\Pi_j$ on the generated data. Then the full regression is

$$\ln(P^k_{j,t}R^k_{ij,t}) = \chi^k_{i,t}+(1-\kappa_{ij})\Gamma^k_i \ln G_{j,t}+\ln \hat{\tau}^k_{ij,t}+(1-\kappa_{ij})\times(\gamma^k \ln IM_{j,t}+\phi^k \ln EX_{i,t})+\omega^k_{ij,t}. \tag{14}$$

The analogous treatment of intermediate demand replaces the Gini coefficient (of inequality in buyer real incomes) with the Herfindahl index $H_{j,t}$ of the size distribution of firms.

$$\ln(P^k_{j,t}R^k_{ij,t}) = \chi^k_{i,t}+(1-\kappa_{ij})\Gamma^k_i \ln H_{j,t}+\ln \hat{\tau}^k_{ij,t}+(1-\kappa_{ij})\times(\gamma^k \ln IM_{j,t}+\phi^k \ln EX_{i,t})+\omega^k_{ij,t}. \tag{15}$$

Finally we combine the two buyer heterogeneity sources in the same regression.

4.3 Intuition

A simple intuition for how the heterogeneity composition effects could work uses the goalpost diagram plotting x_{ij}/y_i against buyer price and residual supply against seller price. Consider final goods. For cases where income elasticity of demand > 1, a rise in Gini (in the cross section of destinations) implies an upward shift in demand, cet. par. For income elasticity < 1 the demand shift is downward. Across time, with smooth rise in inequality, the sign in each sector is predicted by income elasticity minus 1.

Similarly, for intermediate demand the HH effect on the demand shift (in the cross section of destinations) might be upward under the hypothesis that bigger firms buy more sophisticated inputs and downward under the hypothesis that bigger firms have better outside options and thus buy at llower prices. But here the story must shift to include changes in markups. Across time with smooth rise in HH, the sign in each sector in the regression tells us which case obtains.

A structural interpretation on these lines applies the translog specification developed next in Section 4.4. The share equation b_{ij} has an intercept parameter that is shifted (across destinations) by the composition effects. For two destinations with the same observed share, the shift in the intercept due to heterogeneity must generate a

compensating shift in R_{ij} to maintain equality.

The regression coefficients in (15) are interpreted as averages of the implied local effects of the deep determinants on b_{ij} and hence through (**??**) on the revealed R_{ij}

At the very least, our results indicate that composition effects correlated with changing inequality of the size distributions of buyers and of sellers have a significant effect on revealed relative resistance.

4.4 Projection Modeling with Translog

Counterfactual projections are essential in applied trade policy modeling. They requires parametric models. The derivation of non-parametric relative resistances suggests choosing parametric approximations based either on the general translog used to generate the R_{ij}s or a CES variant that may fit the generated data well and has the virtues of simplicity and familiarity. The methods of the text generate the CES average trade elasticity parameter. Here we consider the translog case.

The translog form implies a semi-parametric aggregation procedure for projecting realized relative resistance effects on *equilibrium* b_{ij} for each bilateral link. The general translog price index is

$$\ln P_j = \sum_l \alpha_l \ln p_{lj} + \frac{1}{2} \sum_{l,m} \ln p_{lj} \gamma_{lm} \ln p_{mj},$$

where $\alpha_l \in [0,1]$, $\sum_l \alpha_l = 1$ and the matrix Γ of the γ_{lm}s is negative semi-definite. A tractable[21] special case is $\gamma_{lm} = \gamma_l \gamma_m$, $\forall l \neq m$; $\gamma_{ll} = \gamma_l(1-\gamma_l)$ where $\gamma_l \in [0,1]$ and $\sum_l \gamma_l = 1$. In this case Shephard's Lemma implies that the share equation is:

$$b_{ij} = \alpha_i - \gamma_i \ln(p_{ij}/\bar{p}_j),$$

where $\ln \bar{p}_j = \sum_l \gamma_l \ln p_{lj}$. The log price index for the special case is $\ln P_j = \sum_i \alpha_i \ln p_{ij} - \sum_i \gamma_i (\ln p_{ij})^2$. The revealed relative resistances are generated by (6) and are consistent with the translog share structural model given the special case.

The translog parameters $\{\gamma_i\}$ can be estimated using the information in the revealed R_{ij}s and the reduced form regressions as follows. First, the parameters α_i are theoretically equal to observed s_i. Second, use $p_{ij} = c_i \Pi_i P_j R_{ij} = c_i \tau_{ij}$ and manipulate the share equation to yield:

$$\frac{b_{ij}}{s_i} - 1 = -\frac{\gamma_i}{s_i}(\ln p_{ij} - FE_j) = -FE_i - \frac{\gamma_i}{s_i}(\ln \tau_{ij} - FE_j) \tag{16}$$

where the right hand side is the deterministic part of the regression equation to be

[21] This is due to familiar dimensionality constraints on being able to identify the many components of Γ.

fitted. The theoretical interpretation of the fixed effects is $FE_i = c_i\gamma_i/s_i$ and $FE_j = \bar{p}_j$.

4.5 Omissions

The reduced form analysis of realized relative resistance omit treatment of two forces that affect trade frictions: time variation and market power. Here we discuss and defend the omissions.

Time variation of *bilateral* capacity and its effect on trade costs is presumably important. It may be possible to look for such trade scale effects with the use of bilateral volume variation, but more structure is required. Anderson and Yotov (2020) model the rising bilateral cost effect with a CES demand model combined with Cobb-Douglas bilateral shipment technology subject to short run fixed bilateral capacity and hence rising shipment costs. The rising shipment cost combines with the demand elasticity in a composite trade elasticity. The effect of changes in the bilateral fixed capacities over time with perfect foresight absorbs all the action of short run elasticities approaching long run elasticity. The exchange model lacks a resource allocation structure to found modeling. Adding the resource allocation treatment of capacity decisions faces the challenge of modeling dynamics and uncertainty.

Market power of buyers and/or sellers is likely to be significant in bilateral trade. In the reduced form analysis of relative resistance determinants there is no attempt to quantify the role of market power separately from other factors affected by the same controls.

4.6 Dynamics

The static model of preceding sections suggest that trade frictions will change over time due to exogenous forces. More important for modeling purposes, trade frictions change due to endogenous reactions to these and other secular supply and demand forces. The endogenous changes are taken to imply adjustment toward constrained efficient trade frictions subject to adjustment costs. Dynamic models of short run gravity with adjustment to long run equilibrium such as Anderson and Yotov (2020) are based on pair-specific capacities that adjust over time.

The revealed relative resistances incorporate the effects of the adjusted bilateral capacities, so the minimum distance calibration of a trade elasticity from the set of revealed local non-parametric trade elasticities is a long run elasticity. The empirical results for the constant elasticity case reveal tightly fitted results close to 1. The tight fit over the cross section argues that the constant elasticity approximation is close. The low variability of the time series estimates, over time for a given pair and over the times series of cross section estimates suggest that the dynamic adjustment of bilateral

capacities is close to being efficient.