Aggregation and the Gravity Equation

By Stephen J. Redding and David E. Weinstein*

One of the most successful empirical relationships in economics is the gravity equation, which relates bilateral trade between an origin and destination to bilateral frictions, origin characteristics, and destination characteristics. A key decision for researchers in estimating this relationship is the level of aggregation. While the gravity equation is log linear, aggregation involves summing the level rather than the log level of trade. Therefore, Jensen's Inequality appears to imply that if a log linear gravity equation holds at one level of aggregation, it cannot simultaneously hold at another level of aggregation. In such circumstances, estimating the gravity equation at another level of aggregation at best provides a log linear approximation to the data. This problem is compounded by the absence of a clear theoretical consensus about the appropriate level of aggregation at which to estimate the gravity equation. In line with this theoretical ambiguity, some researchers have estimated this relationship at the aggregate level, while others have estimated it using data on regions, sectors or even firms.

In this paper, we use the nested constant elasticity of substitution (CES) demand system to show that a log linear gravity equation holds exactly at each nest of utility. Using the Independence of Irrelevant Alternatives (IIA) properties of CES, we derive an exact Jensen's Inequality correction term for aggregation across the nests of the utility function. We use this result to decompose the overall effect of distance on bilateral trade in the aggregate gravity equation into the contribution of a number of different terms from sectoral gravity equations: (i) origin fixed effects; (ii) destination fixed effects; (iii) distance; (iv) our Jensen's Inequality correction term; and (v) the error term. These terms vary bilaterally with the set of sectors in which there is positive trade and enter the error term in a conventional aggregate gravity equation. We show that our Jensen's Inequality correction term makes a quantitatively relevant contribution towards the overall effect of distance in the aggregate gravity equation. Although we focus on the aggregate economy and sectors as our two nests of utility, our theoretical results hold for *any* definition and number of nests with the CES demand system. Therefore, our analysis also encompasses, for example, regions and firms as other possible levels of aggregation.[1] Finally, although for brevity we focus on international trade, our analysis also goes through for other gravity applications in economics with a nested CES or nested logit demand structure, including migration, commuting and financial flows.

Our paper is related to a large literature on the gravity equation in international trade, including Anderson and van Wincoop (2003) and Allen, Arkolakis and Takahashi (2018), as recently surveyed in Anderson (2011) and Head and Mayer (2014). Most empirical research has estimated the gravity equation using data on aggregate bilateral trade between countries, as in Eaton and Kortum (2002) and Redding and Venables (2004). However, many studies have instead esti-

* Redding: Princeton University and NBER, Department of Economics and WWS, Princeton, NJ 08544, reddings@princeton.edu. Weinstein: Columbia University and NBER, Department of Economics, 1022 International Affairs Building (IAB), dew35@columbia.edu. The derivations for all results in this paper are reported in further detail in a separate web appendix. We would like to thank Keith Head, Thierry Mayer and Mike Waugh for helpful comments. We are grateful to Lintong Li for excellent research assistance

[1] Whereas we focus on gravity equation estimation for sectoral and aggregate trade, Redding and Weinstein (2018) develop a theoretical framework for aggregating from millions of trade transactions on firms and products to national trade and welfare using data on trade values and quantities.

mated the same gravity relationship using more disaggregated data on sectors, regions and even firms, as in Davis and Weinstein (1999), Head and Ries (2001), Feenstra, Markusen and Rose (2001), Combes, Lafourcade and Mayer (2005), Bernard, Redding and Schott (2011), and Bas, Mayer and Thoenig (2017)).

The remainder of this paper is structured as follows. In Section I, we develop our main theoretical result. In Section II, we estimate gravity equations at both the sectoral and aggregate levels. We use these estimates to decompose the overall effect of distance on bilateral trade in the aggregate gravity equation. Section III concludes.

I. Theoretical Framework

We consider a simple model of international trade across countries and sectors based on differentiation by origin following Armington (1969). Although we choose this formulation for simplicity, our results hold for any international trade model with a nested CES import demand system, including for example the multi-sector Ricardian model of Costinot, Donaldson and Komunjer (2012), a multi-sector version of Krugman (1980), and a multi-sector version of Melitz (2003) with an untruncated Pareto productivity distribution.

A. *Preferences*

The world economy consists of a number of countries indexed by $d, o \in \Omega$, where we use d as a mnemonic for destination and o as a mnemonic for origin. The preferences of the representative consumer in each destination are defined over consumption indexes (C_{ds}) for a number of sectors indexed by $s \in \Xi$, where we use s as a mnemonic for sector. The utility function is:

$$U_d = \left[\sum_{s \in \Xi} \left(\Theta_{ds} C_{ds} \right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}, \tag{1}$$

where $\sigma > 1$ is the elasticity of substitution between sectors and $\Theta_{ds} > 0$ is the taste of the representative consumer in destination d for the output of sector s.

The consumption index for destination d in sector s (C_{ds}) is defined over the consumption of the output of each origin o within that sector (c_{dos}):

$$C_{ds} = \left[\sum_{o \in \Omega_{ds}} \left(\theta_{dos} c_{dos} \right)^{\frac{\nu_s-1}{\nu_s}} \right]^{\frac{\nu_s}{\nu_s-1}}, \tag{2}$$

where $\nu_s > 1$ is the elasticity of substitution across countries within sectors; we allow this elasticity to differ across sectors s; $\theta_{dos} \geq 0$ is the taste of the representative consumer in destination d for the goods supplied by origin o within sector s; and $\Omega_{ds} \subseteq \Omega$ is the set of origins from which destination d consumes goods in sector s.

Goods are produced under conditions of perfect competition and constant returns to scale using a composite factor with unit cost η_{os} in sector s in origin o. Trade is subject to iceberg variable trade costs, such that $\tau_{dos} > 1$ units of a good must be shipped from origin o to destination $d \neq o$ in order for one unit to arrive, where $\tau_{dds} = 1$. As a result, the "cost inclusive of freight" (cif) price in destination d of the good produced by origin o in sector s is:

$$p_{dos} = \tau_{dos} p_{os} = \tau_{dos} \eta_{os}. \tag{3}$$

Using these equilibrium prices and CES preferences, we can write the import expenditure of destination d on goods in sector s from a foreign origin $o \neq d$ as:

$$x_{dos} = \left(\frac{\tau_{dos} \eta_{os}}{\theta_{dos}} \right)^{1-\nu_s} \mathcal{X}_{ds} \mathcal{P}_{ds}^{\nu_s - 1}, \tag{4}$$

where $\mathcal{X}_{ds} = \left[\sum_{o \in \{\Omega_{ds}: o \neq d\}} x_{dos} \right]$ is total expenditure on foreign imports within sector s; we allow destination d to have zero imports from some origins $o \neq d$ within sector s, such that $\{\Omega_{ds} : o \neq d\} \subseteq \Omega$; we rationalize these zeros in terms of either zero tastes ($\theta_{dos} \to 0$) or infinite trade costs ($\tau_{dos} \to \infty$); and $\mathcal{P}_{ds}$ is the price index for foreign imports defined as:

$$\mathcal{P}_{ds} = \left[\sum_{o \in \{\Omega_{ds}: o \neq d\}} \left(\frac{p_{dos}}{\theta_{dos}} \right)^{1-\nu_s} \right]^{\frac{1}{1-\nu_s}}. \tag{5}$$

B. Sectoral Gravity

We first show that this multi-sector Armington model implies a log linear gravity equation for sectoral trade. Taking logarithms in equation (4) for pairs with positive trade, we obtain:

$$\ln x_{dos} = \gamma_{os} + \lambda_{ds} - (\nu_s - 1) \ln \tau_{dos} + u_{dos}, \tag{6}$$

where γ_{os} is a fixed effect for origin o in sector s; λ_{ds} is a fixed effect for destination d in sector s; and u_{dos} is a stochastic error.

C. Aggregate Gravity

We next show that this multi-sector Armington model also implies a log linear gravity equation for aggregate trade. As a first step, note that aggregate foreign imports in destination d from origin $o \neq d$ are the sum of imports across sectors s:

$$\mathcal{X}_{do} = \sum_{s \in \Xi_{do}} x_{dos}, \qquad o \neq d, \tag{7}$$

where $\Xi_{do} \subseteq \Xi$ is the set of sectors in which destination d has positive imports from origin $o \neq d$.

At first sight, equations (6) and (7) appear inconsistent with a log linear aggregate gravity equation. Although the sectoral gravity equation (6) is log linear, aggregate trade in equation (7) is the sum of the level rather than the log level of sectoral trade. However, we now derive an exact Jensen's Inequality correction term, which enables us to write aggregate bilateral trade in a log linear form.

As a second step, we rewrite destination d's aggregate imports from origin $o \neq d$ ($\mathcal{X}_{do}$) as the sum across sectors of the share of these imports in its total foreign import expenditure multiplied by total foreign import expenditure ($\mathcal{X}_d$):

$$\mathcal{X}_{do} = \left[\frac{\sum_{s \in \Xi_{do}} x_{dos}}{\sum_{j \in \{\Omega_d : j \neq d\}} \sum_{r \in \Xi_{dj}} x_{djr}} \right] \mathcal{X}_d. \tag{8}$$

As a third step, we define two measures of the importance of destination d's imports from foreign origin $o \neq d$ in sector s. The first is relative to total imports from foreign origin ($\mathcal{Z}_{dos}$) and the second is relative to total imports from all foreign origins ($\mathcal{Y}_{dos}$):

$$\mathcal{Z}_{dos} \equiv \frac{x_{dos}}{\sum_{r \in \Xi_{do}} x_{dor}}. \tag{9}$$

$$\mathcal{Y}_{dos} \equiv \frac{x_{dos}}{\sum_{j \in \{\Omega_d : j \neq d\}} \sum_{r \in \Xi_{dj}} x_{djr}}. \tag{10}$$

Using the denominators in these two definitions, destination d's aggregate imports from origin o in equation (8) can be rewritten in the following log linear form:

$$\ln \mathcal{X}_{do} = \Gamma_{do} + \Lambda_{do} - T_{do} + J_{do} + U_{do}, \tag{11}$$

where Γ_{do} is an average of the origin-sector fixed effects (γ_{os}); Λ_{do} is an average of the destination-sector fixed effects (λ_{ds}); T_{do} captures the average effect of sectoral bilateral trade costs ($(\nu_s - 1) \ln \tau_{dos}$); J_{do} is our Jensen's Inequality correction term, which includes $\mathcal{Z}_{dos}$ and $\mathcal{Y}_{dos}$, and controls for the difference between the mean of the logs and the log of the means; U_{do} is an average of the sectoral error terms (u_{dos}). These averages are taken across sectors with positive trade, and hence vary bilaterally, as shown in the web appendix.

Absorbing the bilateral variation in the components Γ_{do}, Λ_{do}, J_{do} and U_{do} into the error term, we can re-write equation (11) as a conventional aggregate gravity equation:

$$\ln \mathcal{X}_{do} = \eta_o^{\mathcal{X}} + \mu_d^{\mathcal{X}} - V^{\mathcal{X}} \ln \tau_{do} + w_{do}^{\mathcal{X}}, \tag{12}$$

where $\eta_o^{\mathcal{X}}$ is an origin fixed effect; $\mu_d^{\mathcal{X}}$ is a destination fixed effect; τ_{do} is an aggregate measure of bilateral trade costs; $V^{\mathcal{X}}$ is the coefficient on this aggregate trade cost measure; and $w_{do}^{\mathcal{X}}$ is the transformed error term, which includes all bilateral variation not captured in the aggregate trade cost measure, as defined in the web appendix.

As well as estimating the aggregate gravity equation for overall bilateral trade in equation (12), we can also use the log linear form of equation (11) to estimate aggregate gravity equations for each bilateral component of overall trade (Γ_{do}, Λ_{do}, T_{do}, U_{do}):

$$
\begin{aligned}
\Gamma_{do} &= \eta_o^{\Gamma} + \mu_d^{\Gamma} - V^{\Gamma}\tau_{do} + w_{do}^{\Gamma}, \\
\Lambda_{do} &= \eta_o^{\Lambda} + \mu_d^{\Lambda} - V^{\Lambda}\tau_{do} + w_{do}^{\Lambda}, \\
-T_{do} &= \eta_o^{T} + \mu_d^{T} - V^{T}\tau_{do} + w_{do}^{T}, \\
J_{do} &= \eta_o^{J} + \mu_d^{J} - V^{J}\tau_{do} + w_{do}^{J}, \\
U_{do} &= \eta_o^{U} + \mu_d^{U} - V^{U}\tau_{do} + w_{do}^{U},
\end{aligned} \tag{13}
$$

where we can compute Γ_{do}, Λ_{do}, $-T_{do}$, J_{do}, and U_{do} from estimates of sectoral gravity equations and observed data.

Estimating equations (12) and (13) using ordinary least squares (OLS), the estimated coefficient on bilateral trade costs for aggregate trade ($V^{\mathcal{X}}$) is the sum of those for each component (V^{Γ}, V^{Λ}, V^{T}, V^{J}, V^{U}). Therefore, the relative magnitude of these estimated coefficients reveals the extent to which the effect of trade costs on aggregate bilateral trade ($V^{\mathcal{X}}$) captures the direct effect of these trade costs on sectoral bilateral trade (V^{T}) versus indirect effects through changes in the composition of sectors with different origin fixed effects (V^{Γ}), destination fixed effects (V^{Λ}), import shares (V^{J}), and error terms (V^{U}).

II. Data and Empirical Results

In our empirical analysis, we use the BACI CEPII world trade database, which reports the bilateral value of trade by Harmonized System (HS) 6-digit product, origin and destination. To abstract from considerations that are specific to the agricultural sector, we focus on mining and manufacturing products (HS 2-digit sectors 16-96), excluding arms and ammunition (HS 2-digit sector 93). We model bilateral trade costs as a constant elasticity function of bilateral distance between the most-populated cities of each origin and destination. We allow this elasticity of bilateral trade costs with respect to bilateral distance to differ across sectors. We report results using bilateral trade data for 2012, but find similar results for other years.

We begin by estimating both an aggregate gravity equation and gravity equations for each sector. We do so for a range of different definitions of sectors, including HS 1-digit, HS 2-digit, HS 3-digit and HS 4-digit categories. As we include exporter and importer fixed effects in our gravity equations, we drop exporter-sector cells with less than 3 importers and importer-sector cells with less than 3 exporters, which results in slightly different samples of exporters and importers for each definition of sector.

As a first step, we sum bilateral trade flows across sectors, and estimate the aggregate gravity equation (12) for each of our samples. As reported at the bottom of Table 1 (row (vi)), we estimate a similar aggregate distance coefficient across these four samples. We find a elasticity of aggregate trade with respect to bilateral distance of around −1.65, which is in line with existing studies, and is statistically significant at conventional critical values.

As a second step, we estimate separate gravity equations for each sector for our alternative definitions of sectors. We find substantial heterogeneity in the estimated distance coefficients across sectors. These estimated distance coefficients range from −1.9011 to −1.2794 using 1-digit sectors, −1.9428 to −0.8692 using 2-digit sectors, −1.9480 to −0.7242 using 3-digit sectors, and −2.0576 to 1.5683 using 4-digit sectors. By itself, this heterogeneity in estimated distance coefficients across sectors suggests that the average distance coefficient will vary across origin-destination pairs with the set of sectors in which there is positive trade. We find that the extent of these differences in average distance coefficients generally increases as we move from less to more disaggregated definitions of sectors. For example, using 4-digit sectors, the unweighted average distance coefficient varies across origin-destination pairs from −1.3995 at the 10th percentile to −1.0970 at the 90th percentile, and the trade-weighted average distance coefficient ranges from −1.5012 to −0.9885 between these same percentiles.

As a third and final step, we compute each of the components of aggregate bilateral trade in equation (11), and estimate separate gravity equations for each component, as in equation (13) above. In rows (i)-(v) of Table 1, we report the estimated distance coefficient for each component for alternative definitions of sectors (across the columns). The sum of the coefficients across these components in rows (i)-(v) equals the coefficient for aggregate bilateral trade in row (vi).

Perhaps unsurprisingly, we find that much of the effect of distance on aggregate trade (row (vi)) occurs through the average effect of distance on sectoral trade (row (iii)), with the difference between these two estimates becoming larger as we consider more disaggregated definitions of sectors. In line with Alchian-Allen type selection effects, we find positive correlations with distance for the origin-sector fixed effects (row (i)), the destination-sector fixed effects (row (ii)) and the error term (row (v)). Therefore, trade flows over longer distances are a selected sample with superior supply-side, demand-side and/or match-specific characteristics. Finally, we find a substantial negative and statistically significant coefficient on our Jensen's Inequality correction term (row (iv)), which ranges from -0.5188 using 1-digit sectors to -1.2846 using 4-digit sectors. Therefore, as we consider trade over longer versus shorter distances, we find substantial differences in the distribution of imports across sectors. These differences in turn imply a substantial discrepancy between the mean of log imports and the log of mean imports, highlighting the relevance of the choice of the level aggregation at which to estimate the gravity equation.

III. Conclusions

Although the gravity equation is one of the most successful empirical relationships in economics, existing research provides relatively little guidance as to the appropriate level of aggregation at which to estimate this relationship. In this paper, we make two main contributions to this question.

First, we derive an exact Jensen's Inequality correction term for the nested CES demand structure, such that a log linear gravity equation holds exactly for each nest of utility. Second, we use this result to decompose the effect of distance on bilateral trade in the aggregate gravity equation into the contribution of a number of different terms from gravity equations estimated at a more disaggregated level: (i) origin fixed effects; (ii) destination fixed effects; (iii) distance; (iv) our Jensen's Inequality correction term; and (v) the error term.

Second, using the aggregate economy and sectors as our two nests of utility, we show that sectoral composition makes a quantitatively relevant contribution to the overall effect of bilateral distance on bilateral trade in the aggregate gravity equation.

REFERENCES

Allen, Treb, Costas Arkolakis, and Yuta Takahashi. 2018. "Universal Gravity." *Journal of Political Economy.* forthcoming.

Anderson, James. 2011. "The Gravity Model." *Annual Review of Economics*, 3: 133–60.

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

Armington, Paul S. 1969. "A Theory of Demand for Products Distinguished by Place of Production." *IMF Staff Papers*, 16: 159–178.

Bas, Maria, Thierry Mayer, and Mathias Thoenig. 2017. "From Micro to Macro: Demand, Supply, and Heterogeneity in the Trade Elasticity." *Journal of International Economics*, 108: 1–19.

Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott. 2011. "Multi-product Firms and Trade Liberalization." *Quarterly Journal of Economics*, 126(3): 1271–1318.

Combes, Pierre-Philippe, Miren Lafourcade, and Thierry Mayer. 2005. "The Trade-Creating Effects of Business and Social Networks: Evidence from France." *Journal of International Economics*, 66: 1–29.

Costinot, Arnaud, Dave Donaldson, and Ivana Komunjer. 2012. "What

Table 1—Decomposition of the Distance Effect in the Aggregate Gravity Equation

	(1)	(2)	(3)	(4)
	HS1	HS2	HS3	HS4
(i) Origin fixed effect	0.1639***	0.2732***	0.2739***	0.3146***
	(0.0052)	(0.0066)	(0.0069)	(0.0077)
(ii) Destination fixed effect	0.0472***	0.0861***	0.0841***	0.0915***
	(0.0025)	(0.0034)	(0.0033)	(0.0040)
(iii) Distance	-1.5704***	-1.5389***	-1.4352***	-1.1873***
	(0.0044)	(0.0066)	(0.0067)	(0.0098)
(iv) Composition term	-0.5188***	-0.9138***	-0.9873***	-1.2846***
	(0.0146)	(0.0167)	(0.0177)	(0.0181)
(v) Error term	0.2275***	0.4396***	0.4128***	0.4084***
	(0.0151)	(0.0132)	(0.0130)	(0.0135)
(vi) Aggregate	-1.6505***	-1.6538***	-1.6517***	-1.6574***
	(0.0195)	(0.0196)	(0.0196)	(0.0196)
Observations	23,597	23,379	23,192	22,417

Note: Gravity equation estimates of aggregate bilateral trade from equation (12) (row (vi)) and the components of aggregate bilateral trade from equation (13) (rows (i)-(v)) using the CEPII BACI trade database. Coefficients in rows (i)-(v) sum to the coefficient in row (vi). Columns correspond to different definitions of sectors. Heteroskedasticity robust standard errors in parentheses.

Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas." *Review of Economic Studies*, 79(2): 581–608.

Davis, Donald R., and David E. Weinstein. 1999. "Economic Geography and Regional Production Structure: An Empirical Investigation." *European Economic Review*, 43: 379–407.

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

Feenstra, Robert C., James R. Markusen, and Andrew K. Rose. 2001. "Using the Gravity Equation to Differentiate among Alternative Theories of Trade." *Canadian Journal of Economics*, 34(2): 430–447.

Head, Keith, and John Ries. 2001. "Increasing Returns versus National Product Differentiation as an Explanation for the Pattern of U.S.-Canada Trade." *American Economic Review*, 91: 858–876.

Head, Keith, and Thierry Mayer. 2014. "Gravity Equations: Workhorse, Toolkit, and Cookbook." In *Handbook of International Economics*. Vol. 4, , ed. Elhanan Helpman, Gita Gopinath and Kenneth Rogoff, 131–196. Amsterdam:Elsevier, North Holland.

Krugman, Paul. 1980. "Scale Economies, Product Differentiation, and the Pattern of Trade." *American Economic Review*, 70(5): 950–959.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica*, 71: 1695–1725.

Redding, Stephen J., and Anthony J. Venables. 2004. "Economic Geography and International Inequality." *Journal of International Economics*, 62(1): 53–82.

Redding, Stephen J., and David E. Weinstein. 2018. "Accounting for Trade Patterns." Princeton University, mimeograph.