Online Appendix
Aggregation and the Gravity Equation

Stephen J. Redding*
Princeton University, NBER and CEPR

David E. Weinstein†
Columbia University and NBER

January 10, 2019

1 Introduction

One of the most successful empirical relationships in international trade 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, since the gravity equation is log linear, whereas aggregation involves summing the level rather than the log level of trade flows. Therefore, Jensen's inequality would appear to imply that if a log-linear gravity equation holds at one level of aggregation, it cannot in general also hold at another level of aggregation. In such circumstances, estimating the gravity equation at another level of aggregation can be interpreted at best as providing a log-linear approximation to the true relationship at this other level of aggregation. This problem is compounded, because there is little consensus in existing theoretical research about the appropriate level of aggregation at which the gravity equation holds. Some models derive this relationship at the aggregate level, while others predict that it holds at the sectoral level, and yet others imply that is holds at an even more disaggregated level below sectors. Mirroring this theoretical ambiguity, researchers have estimated the gravity equation using aggregate, sector and even firm-level data, and find that it provides a reasonable approximation to the data at each of these levels of aggregation.

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. In particular, we use the independence of irrelevant alternatives (IIA) properties of CES to derive an exact Jensen's inequality correction term for aggregation across the nests of utility. Choosing the aggregate economy and sectors as our nests of utility, we estimate gravity equations at both the sectoral and aggregate level, and show how to aggregate exactly from the sectoral to the aggregate level. We decompose the effect of distance on bilateral trade in the aggregate gravity equation into the contribution of a number of different terms from the sectoral gravity equations: (i) origin fixed effects; (ii) destination fixed effects; (iii) distance; (iv) our Jensen's inequality or composition term; and (v) the error term. We show that our composition

*Dept. Economics and WWS, Julis Romo Rabinowitz Building, Princeton, NJ 08544. Tel: 1 609 258 4016. Email: reddings@princeton.edu.
†Dept. Economics, 1022 International Affairs Building (IAB), Mail Code 3308, 420 West 118th Street, New York, NY 10027. Tel: 1 212 854 6880. Email: dew35@columbia.edu.

term makes a quantitatively relevant contribution towards the aggregate effect of distance. Although we choose 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 the gravity equation for international trade, our analysis also goes through for other applications of gravity equations in economics with a nested demand structure, including but not limited to migration, commuting and financial flows.

Our paper is related to a voluminous theoretical literature on the gravity equation in international trade, as recently surveyed in Anderson (2011) and Head and Mayer (2014). In early empirical research following Tinbergen (1962), the gravity equation was found to be empirically successful in explaining observed trade data, but lacked rigorous theoretical microfoundations. More than fifty years later, we have an abundance rather than a scarcity of these theoretical microfoundations, with an entire class of models that are isomorphic in terms of their gravity equation predictions, as emphasized in Arkolakis, Costinot, and Rodriguez-Clare (2012) and Allen, Arkolakis, and Takahashi (2018). This class of models includes neoclassical theories with perfect competition and constant returns to scale (e.g. Deardorff 1998 and Eaton and Kortum 2002), Armington models with differentiation by country of origin (e.g. Anderson and van Wincoop 2003), "new trade" theory models with monopolistic competition and increasing returns to scale (e.g. Krugman 1980), "new new trade" theory models with heterogeneous firms, monopolistic competition and increasing returns to scale (e.g. Melitz 2003 with an untruncated Pareto productivity distribution, as in Chaney 2008 and Arkolakis, Demidova, Klenow, and Rodriguez-Clare 2008), and models of buyer-seller networks (e.g. Chaney 2018). Our main theoretical contribution relative to this literature is to show that a log-linear gravity equation holds exactly at each level of aggregation in a nested CES demand system and to characterize the properties of the error term at each of these levels of aggregation.

Our paper also contributes to an equally-large body of empirical research that has found the gravity equation to provide a good approximation to observed bilateral trade data. Much of this research has estimated the gravity equation using data on aggregate bilateral trade between countries, exploring a whole range of different bilateral trade frictions, including distance, common borders and common currencies along many others (as in Redding and Venables 2004 and the survey by Head and Mayer 2014). Another influential line of research has estimated the gravity equation using more disaggregated data, including sectors, regions within countries, and even firms (e.g. Davis and Weinstein 1999, Head and Ries 2001, Feenstra, Markusen, and Rose 2001, Combes, Lafourcade, and Mayer 2005, Bernard, Redding, and Schott 2011, Berthelon and Freund 2008 and Bas, Mayer, and Thoenig 2017).

Great progress has been made in this empirical literature in addressing a number of challenges in estimating the gravity equation, including the presence of zeros in international trade flows (e.g. Santos Silva and Tenreyro 2006), the role of the extensive versus the intensive margins of trade (e.g. Helpman, Melitz, and Rubinstein 2008) and the need to control for changes in multilateral resistance in undertaking counterfactuals for changes in bilateral trade frictions (e.g. Anderson and van Wincoop 2003). One remaining empirical challenge is the choice of the appropriate level of aggregation at which to estimate the gravity equation, as considered in the context of aggregating from regions to countries in Ramondo, Rodríguez-Clare, and Saborio (2016) and Coughin and Novy (2018), and in the context of Ricardian models of trade in Lind and Ramondo (2018). This empirical challenge is an example of the more general

[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 individual firms and products to national trade and welfare using data on the value of trade and unit values.

Modifiable Area Unit Problem (MAUP) in the statistics literature following Gehlke and Biehl (1934) and Fotheringham and Wong (1991), whereby the results of statistical analyses need not be invariant to the scale at which these analyses are undertaken. This empirical challenge is particularly severe for gravity equation estimation, because different theories yield different predictions as to the scale at which the gravity equation holds, including for example the aggregate economy in Eaton and Kortum (2002) and sectors in Costinot, Donaldson, and Komunjer (2012). Our main empirical contribution relative to this literature is to develop a metric for assessing the impact of aggregation on gravity equation estimates.

Finally, as noted above, our results also apply for other applications of the gravity equation in economics with a nested demand structure, including for example migration (e.g. Kennan and Walker 2011), commuting (e.g. Ahlfeldt, Redding, Sturm, and Wolf 2015, Monte, Redding, and Rossi-Hansberg 2018 and Heblich, Redding, and Sturm 2018), and capital flows (e.g. Martin and Rey 2004).

The remainder of the paper is structured as follows. In Section 2, we first develop our main theoretical result that there is an exact Jensen's inequality correction term such that a log-linear gravity equation holds at each level of aggregation in a nested CES demand system. In Section 3, we estimate gravity equations at both the sectoral and aggregate level. We use these estimates to decompose the effect of distance on bilateral trade in the aggregate gravity equation into the contributions of different terms from the sectoral gravity equations. Section 4 concludes.

2 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.

2.1 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 assumed to take the following constant elasticity of substitution (CES) form:

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

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 sector s. Under our assumption that $\Theta_{ds} > 0$, the representative consumer in each country consumes goods from all sectors.

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}) and also takes the CES form:

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

where $\nu_s > 1$ is the elasticity of substitution across countries within sectors; we allow this elasticity of substitution to differ between sectors s; $\theta_{dos} \geq 0$ is the taste of the representative consumer in destination d for the goods produced 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 in positive amounts.

We allow the tastes of the representative consumer for the goods within each sector to have a destination component (φ_{ds}), an origin component (φ_{os}), and a origin-destination component (φ_{ods}) such that:

$$\theta_{dos} = \varphi_{ds}\varphi_{os}\varphi_{ods}. \tag{3}$$

Using the properties of these CES preferences, equilibrium expenditure by destination d on the goods produced by origin o within sector s (x_{dos}) can be written as:

$$x_{dos} = \left(\frac{p_{dos}}{\theta_{dos}}\right)^{1-\nu_s} X_{ds} P_{ds}^{\nu_s - 1}, \tag{4}$$

where p_{dos} is the price in destination d of the goods produced by origin o in sector s; $X_{ds} = \left[\sum_{o \in \Omega_{ds}} x_{dos}\right]$ is total expenditure by destination d on sector s; and P_{ds} is the price index dual to the consumption index in equation (2) for destination d in sector s:

$$P_{ds} = \left[\sum_{o \in \Omega_{ds}} \left(\frac{p_{dos}}{\theta_{dos}}\right)^{1-\nu_s}\right]^{\frac{1}{1-\nu_s}}. \tag{5}$$

In empirical gravity equation estimation, it is common to use datasets in which only data on trade with *foreign* countries is reported. Therefore, we focus on the model's predictions for trade with foreign countries, and use the independence of irrelevant alternatives (IIA) property of CES to partition expenditure in each sector into expenditure on domestic goods and expenditure on foreign goods. In particular, using this property of CES preferences, we have the following equivalent expression for expenditure by destination d on the goods produced by origin $o \neq d$ within sector s (x_{dos}):

$$x_{dos} = \left(\frac{p_{dos}}{\theta_{dos}}\right)^{1-\nu_s} \mathcal{X}_{ds} \mathcal{P}_{ds}^{\nu_s - 1}, \tag{6}$$

where p_{dos} is the price in destination d of the goods produced by origin o in sector s; $\mathcal{X}_{ds} = \left[\sum_{o \in \{\Omega_{ds} : o \neq d\}} x_{dos}\right]$ is total expenditure by destination d on foreign origins $o \neq d$ within sector s; we allow for the possibility that destination d need not import from all foreign origins $o \neq d$ within sector s, such that $\{\Omega_{ds} : o \neq d\} \subseteq \Omega$; and $\mathcal{P}_{ds}$ is a price index for foreign consumption for destination d in sector s that is defined as:

$$\mathcal{P}_{ds} = \left[\sum_{o \in \{\Omega_{ds} : o \neq d\}} \left(\frac{p_{dos}}{\theta_{dos}}\right)^{1-\nu_s}\right]^{\frac{1}{1-\nu_s}}, \tag{7}$$

and the following relationship holds:

$$\mathcal{X}_{ds} = X_{ds}\left(\frac{P_{ds}}{\mathcal{P}_{ds}}\right)^{\nu_s - 1} = \left[\frac{\sum_{o \in \{\Omega_{ds} : o \neq d\}} \left(\frac{p_{dos}}{\theta_{dos}}\right)^{1-\nu_s}}{\sum_{o \in \Omega_{ds}} \left(\frac{p_{dos}}{\theta_{dos}}\right)^{1-\nu_s}}\right] X_{ds}. \tag{8}$$

2.2 Production

The good produced by each country in each sector is supplied under conditions of perfect competition and constant returns to scale. Zero profits implies that the "free on board" (fob) price of each origin's good in each sector is equal to its unit cost of production:

$$p_{os} = \eta_{os}, \tag{9}$$

where η_{os} is a composite measure of unit cost, which could depend on the price of intermediate inputs and the prices of multiple primary factors of production (left implicit here).

Trade between countries 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 $\tau_{dds} = 1$. Under these assumptions, the "cost inclusive of freight" (cif) price in destination d of the good produced by origin o in sector s is:

$$p_{dos} = \tau_{dos} p_{os} = \tau_{dos} \eta_{os}. \tag{10}$$

2.3 Sectoral Gravity

We now show that this multi-sector Armington model implies a log-linear sectoral gravity equation for bilateral trade between destination d and origin o in each sector s for which there is positive trade. Combining CES import demand from equation (6) with the pricing rule from equations (9) and (10), we obtain the following sectoral gravity equation for the value of foreign trade between countries d and $o \neq d$ within sector s:

$$x_{dos} = \left(\frac{\tau_{dos} \eta_{os}}{\theta_{dos}} \right)^{1-\nu_s} \mathcal{X}_{ds} \mathcal{P}_{ds}^{\nu_s - 1}. \tag{11}$$

From this sectoral gravity equation, zero bilateral flows ($x_{dos} = 0$) can arise for two reasons in the model. First, destination d may not import in sector s from origin o if bilateral trade costs are prohibitive ($\tau_{dos} \to \infty$). Second, even if bilateral trade costs are non-prohibitive ($\tau_{dos} < \infty$), destination d need not import in sectors s from origin o if there is no demand for the goods produced by that origin in those sectors ($\theta_{dos} \to 0$).

Taking logarithms in equation (11) for all origin-destination-sector observations for which there is positive trade, this sectoral gravity equation can be written as:

$$\ln x_{dos} = \gamma_{os} + \lambda_{ds} - (\nu_s - 1) \ln \tau_{dos} + u_{dos}, \tag{12}$$

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. The origin fixed effect ($\gamma_{os} = (1 - \nu_s) [\ln \eta_{os} - \ln \varphi_{os}]$) controls for the unit cost of production and the common origin-sector component of tastes across all destinations; the destination fixed effect ($\lambda_{ds} = \ln \mathcal{X}_{ds} + (\nu_s - 1) [\ln \mathcal{P}_{ds} + \ln \varphi_{ds}]$) controls for destination sectoral import expenditure ($\mathcal{X}_{ds}$), the destination sectoral import price index ($\mathcal{P}_{ds}$), and the common destination-sector component of tastes across all origins; and the stochastic error ($u_{dos} = \ln \varphi_{dos}$) captures the idiosyncratic component of tastes (φ_{dos}) that is specific to an individual origin-destination-sector observation, as defined in equation (3).

If bilateral trade is positive for all pairs of origins and destinations in all sectors ($x_{dos} > 0$ for all d, o, s) and bilateral trade costs (τ_{dos}) are observed, equation (12) can be estimated with a conventional fixed effects or least squares dummy variables (LSDV) estimator. Under the identifying assumption that the stochastic error u_{dos} is orthogonal to observed bilateral trade costs, this fixed effects estimator consistently estimates the sectoral trade elasticity ($\nu_s - 1$). More generally, if bilateral trade costs are not observed, they can be modelled as a function of a number of observed bilateral characteristics and a stochastic error. Finally, if zero bilateral trade flows occur, they could be correlated with bilateral trade costs, and the presence of these zero trade flows can be addressed using for example the Poisson fixed effects estimator of Santos Silva and Tenreyro (2006) or a Heckman selection correction as in Helpman, Melitz, and Rubinstein (2008).

2.4 Aggregate Gravity

We now show that this multi-sector Armington model also implies a log-linear aggregate gravity equation between each destination d and origin o. Aggregate foreign imports in destination d from origin $o \neq d$ are the sum across sectors s of imports from that origin in each sector:

$$\mathcal{X}_{do} = \sum_{s \in \Xi_{do}} x_{dos}, \qquad o \neq d, \tag{13}$$

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 (12) and (13) appear inconsistent with the existence of a log-linear aggregate gravity equation. The sectoral gravity equation (12) is log linear, whereas aggregate trade in equation (13) is the sum of the level rather than the sum of the log level of sectoral trade. Therefore, Jensen's inequality appears to imply that a log-linear gravity equation cannot simultaneously hold at both the sectoral and aggregate level. However, we now use the independence of irrelevant alternatives (IIA) properties of CES to derive an exact Jensen's inequality correction term, such that a log-linear gravity equation holds exactly at both the sectoral and aggregate levels, but the interpretation of the error term in these equations differs across these two different levels of aggregation.

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

$$\mathcal{X}_{do} = \sum_{s \in \Xi_{do}} x_{dos} = \sum_{s \in \Xi_{do}} \frac{x_{dos}}{\mathcal{X}_d} \mathcal{X}_d = \left[\frac{\sum_{s \in \Xi_{do}} x_{dos}}{\sum_{j \in \{\Omega_d : j \neq d\}} \sum_{r \in \Xi_{dj}} x_{djr}} \right] \mathcal{X}_d, \tag{14}$$

where recall that $\Xi_{do} \subseteq \Xi$ is the set of sectors in which destination d has positive imports from origin $o \neq d$; $\{\Omega_d : j \neq d\} \subseteq \Omega$ is the set of foreign origins $j \neq d$ from which destination d imports; and destination d's total imports from all foreign origins are given by $\mathcal{X}_d = \left[\sum_{j \in \{\Omega_d : j \neq d\}} \sum_{r \in \Xi_{dj}} x_{djr} \right]$.

As a second step, we define a measure of the importance of destination d's imports from origin $o \neq d$ in sector s as a share of its imports from that origin across all sectors ($\mathcal{Z}_{dos}$):

$$\mathcal{Z}_{dos} \equiv \frac{x_{dos}}{\sum_{r \in \Xi_{do}} x_{dor}} \qquad \Rightarrow \qquad \left[\sum_{r \in \Xi_{do}} x_{dor} \right] = \frac{x_{dos}}{\mathcal{Z}_{dos}}. \tag{15}$$

This expression must hold for each sector s for which destination d has positive imports from origin $o \neq d$. Therefore, taking logarithms of this relationship, and averaging across all these sectors s with positive imports from origin o to destination d, we obtain:

$$\ln \left[\sum_{r \in \Xi_{do}} x_{dor} \right] = \left[\frac{1}{N^S_{do}} \sum_{s \in \Xi_{do}} \ln \left(\frac{x_{dos}}{\mathcal{Z}_{dos}} \right) \right], \tag{16}$$

where $N^S_{do} = |\Xi_{do}|$ is the number of sectors with positive trade between origin o and destination d.

As a third step, we define a measure of the importance of destination d's imports from country $o \neq d$ in sector s as a share of its imports across all foreign origins and sectors ($\mathcal{Y}_{dos}$):

$$\mathcal{Y}_{dos} \equiv \frac{x_{dos}}{\sum_{j \in \{\Omega_d : j \neq d\}} \sum_{r \in \Xi_{dj}} x_{djr}} \qquad \Rightarrow \qquad \left[\sum_{j \in \{\Omega_d : j \neq d\}} \sum_{r \in \Xi_{dj}} x_{djr} \right] = \frac{x_{dos}}{\mathcal{Y}_{dos}}. \tag{17}$$

Again this expression must hold for each sector s for which destination d has positive imports from origin o. Hence, taking logarithms in this relationship, and averaging across all origins o and sectors s with positive imports to destination d, we obtain:

$$\ln\left[\sum_{j\in\{\Omega_d:j\neq d\}}\sum_{r\in\Xi_{dj}}x_{djr}\right]=\left[\frac{1}{N_d^O}\sum_{o\in\{\Omega_d:o\neq d\}}\frac{1}{N_{do}^S}\sum_{s\in\Xi_{do}}\ln\left(\frac{x_{dos}}{\mathcal{Y}_{dos}}\right)\right], \tag{18}$$

where $N_d^O=|\{\Omega_d:o\neq d\}|$ is the number of origins o with positive trade for destination d.

In a fourth step, we take logarithms in equation (14) for destination d's imports from origin $o\neq d$ ($\mathcal{X}_{do}$) and use equations (16) and (18) to substitute for the two summation terms, which yields:

$$\ln\mathcal{X}_{do}=\left[\frac{1}{N_{do}^S}\sum_{s\in\Xi_{do}}\ln\left(\frac{x_{dos}}{\mathcal{Z}_{dos}}\right)\right]-\left[\frac{1}{N_d^O}\sum_{o\in\{\Omega_d:o\neq d\}}\frac{1}{N_{do}^S}\sum_{s\in\Xi_{do}}\ln\left(\frac{x_{dos}}{\mathcal{Y}_{dos}}\right)\right]+\ln\mathcal{X}_d. \tag{19}$$

Using our sectoral gravity equation (12) to substitute for x_{dos} in equation (19), we obtain the following log-linear expression for aggregate bilateral trade between origin o and destination d:

$$\ln\mathcal{X}_{do}=\Gamma_{do}+\Lambda_{do}-T_{do}+J_{do}+U_{do}, \tag{20}$$

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 frictions ($(\nu_s-1)\ln\tau_{dos}$); J_{do} is our Jensen's inequality or composition 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; and U_{do} is an average of the sectoral error terms (u_{dos}).

Each of these averages is taken across the sectors with positive trade between origin o and destination d and hence varies bilaterally. In particular, Γ_{do} equals the average origin-sector fixed effect (γ_{os}) across the set of sectors with positive trade between destination d and origin o minus the average origin-sector fixed sector (γ_{os}) across all origins and sectors for that destination:

$$\Gamma_{do}\equiv\left[\frac{1}{N_{do}^S}\sum_{s\in\Xi_{do}}\gamma_{os}-\frac{1}{N_d^O}\sum_{o\in\{\Omega_d:o\neq d\}}\frac{1}{N_{do}^S}\sum_{s\in\Xi_{do}}\gamma_{os}\right]. \tag{21}$$

Each of the other averages is defined analogously. Therefore, Λ_{do} equals the average destination-sector fixed effect (λ_{ds}) across the set of sectors with positive trade between between destination d and origin o minus the average destination-sector fixed sector (λ_{ds}) across all origins and sectors for that destination:

$$\Lambda_{do}\equiv\left[\frac{1}{N_{do}^S}\sum_{s\in\Xi_{do}}\lambda_{ds}-\frac{1}{N_d^O}\sum_{o\in\{\Omega_d:o\neq d\}}\frac{1}{N_{do}^S}\sum_{s\in\Xi_{do}}\lambda_{ds}\right]. \tag{22}$$

Similarly, T_{do} captures the average effect of sectoral trade costs ($(\nu_s-1)\ln\tau_{dos}$) across the set of sectors with positive trade between destination d and origin o minus the average effect of these sectoral trade costs ($(\nu_s-1)\ln\tau_{dos}$) across all origins and sectors for that destination:

$$T_{do}\equiv\left[\frac{1}{N_{do}^S}\sum_{s\in\Xi_{do}}(\nu_s-1)\ln\tau_{dos}-\frac{1}{N_d^O}\sum_{o\in\{\Omega_d:o\neq d\}}\frac{1}{N_{do}^S}\sum_{s\in\Xi_{do}}(\nu_s-1)\ln\tau_{dos}\right]. \tag{23}$$

In contrast, our Jensen's inequality or composition term (J_{do}) depends on the average across sectors of the log import share $\ln\mathcal{Z}_{dos}$ for origin o and the average across origins and sectors of the log import share $\ln\mathcal{Y}_{dos}$:

$$J_{do}\equiv\ln\mathcal{X}_d+\left[\frac{1}{N_d^O}\sum_{o\in\{\Omega_d:o\neq d\}}\frac{1}{N_{do}^S}\sum_{s\in\Xi_{do}}\ln\mathcal{Y}_{dos}-\frac{1}{N_{do}^S}\sum_{s\in\Xi_{do}}\ln\mathcal{Z}_{dos}\right]. \tag{24}$$

Finally, U_{do} equals the average error term (u_{dos}) across the set of sectors with positive trade between destination d and origin o minus the average error term (u_{dos}) across all origins and sectors for that destination:

$$U_{do} \equiv \left[\frac{1}{N_{do}^{S}} \sum_{s \in \Xi_{do}} u_{dos} - \frac{1}{N_{d}^{O}} \sum_{o \in \{\Omega_d : o \neq d\}} \frac{1}{N_{do}^{S}} \sum_{s \in \Xi_{do}} u_{dos} \right]. \tag{25}$$

Therefore, in each case, a destination's trade with a given origin depends on the average value of a variable with that origin relative to its average value across all origins.

In a fifth and final step, we absorb the bilateral variation in the components Γ_{do}, Λ_{do}, J_{do} and U_{do} into the error term of our log-linear expression for aggregate bilateral trade, such that equation (20) can be re-written as the following conventional aggregate gravity equation:

$$\ln \mathcal{X}_{do} = \eta_{o}^{\mathcal{X}} + \mu_{d}^{\mathcal{X}} - V^{\mathcal{X}} \ln \tau_{do} + w_{do}^{\mathcal{X}}. \tag{26}$$

In this specification, we define the origin fixed effect ($\eta_{o}^{\mathcal{X}}$) as the average across destinations of the bilateral component Γ_{do} from equation (21):

$$\eta_{o}^{\mathcal{X}} = \frac{1}{N_{o}^{D}} \sum_{d \in \{\Omega_o : d \neq o\}} \Gamma_{do}, \tag{27}$$

which only varies by origin o. Similarly, we define the destination fixed effect ($\mu_{d}^{\mathcal{X}}$) as the average across origins of Λ_{do} from equation (22):

$$\mu_{d}^{X} = \frac{1}{N_{d}^{O}} \sum_{o \in \{\Omega_d : o \neq d\}} \Lambda_{do}, \tag{28}$$

which only varies by destination d.

Additionally, we include an aggregate measure of bilateral trade costs (τ_{do}) together with a constant coefficient $V^{\mathcal{X}}$ on this aggregate measure in equation (26). In our empirical work below, we proxy both sectoral and aggregate bilateral trade costs using bilateral distance, which takes a common value across all sectors, although we allow the coefficient on this variable to vary across sectors in our sectoral gravity equations.

Both equations (20) and (26) hold simultaneously, because the error term $w_{do}^{\mathcal{X}}$ includes (i) bilateral variation in Γ_{do}; (ii) bilateral variation in Λ_{do}; (iii) heterogeneity in the average effect of sectoral trade costs ($(\nu_s - 1) \ln \tau_{dos}$) across origin-destination pairs depending on the set of sectors with positive trade; (iv) the Jensen's inequality or composition term; and (v) the error term U_{do}:

$$w_{do}^{\mathcal{X}} = \left(\Gamma_{do} - \frac{1}{N_{o}^{D}} \sum_{d \in \{\Omega_o : d \neq o\}} \Gamma_{do} \right) + \left(\Lambda_{do} - \frac{1}{N_{d}^{O}} \sum_{o \in \{\Omega_d : o \neq d\}} \Lambda_{do} \right) - \left(T_{do} - V^{\mathcal{X}} \ln \tau_{do} \right) + J_{do} + U_{do}. \tag{29}$$

In general, the properties of this composite error term ($w_{do}^{\mathcal{X}}$) can be quite different from those of the average sectoral error term (U_{do}). Therefore, even if this average sectoral error term (U_{do}) is orthogonal to the true measure of bilateral trade costs (T_{do}) in equation (20), there is no necessary reason why this composite error term ($w_{do}^{\mathcal{X}}$) should be orthogonal to the aggregate measure of trade costs ($\ln \tau_{do}$) in equation (26).

We are thus in a position to establish our main theoretical result:

Proposition 1 *In a nested CES demand system, both sectoral and aggregate bilateral trade flows can be expressed as gravity equations that are log linear in origin characteristics, destination characteristics, a measure of bilateral trade costs, and a stochastic error.*

Proof. The proposition follows immediately from equations (12) and (26). ■

Therefore, although Jensen's inequality implies that the log of a sum is not equal to the sum of the logs, there exists an exact Jensen's inequality correction term for the nested CES demand system, such that bilateral trade flows at each level of aggregation can be expressed as a log-linear function of origin characteristics, destination characteristics, bilateral trade frictions, and a stochastic error.

2.5 Components of Aggregate Gravity

Returning to our log-linear expression for aggregate bilateral trade between origin o and destination d in equation (20), we can use the additive separability of this relationship in the bilateral components of aggregate trade (Γ_{do}, Λ_{do}, T_{do}, J_{do}, U_{do}) to provide further evidence on the mechanisms through which bilateral trade costs affect aggregate bilateral trade flows. In particular, as well as estimating an aggregate gravity equation (26) for aggregate bilateral trade ($\mathcal{X}_{do}$), we can estimate aggregate gravity equations for each of its components (Γ_{do}, Λ_{do}, T_{do}, U_{do}):

$$\begin{aligned}
\Gamma_{do} &= \eta_o^{\Gamma} + \mu_d^{\Gamma} - V^{\Gamma}\tau_{do} + w_{do}^{\Gamma}, \\
\Lambda_{do} &= \eta_o^{\Lambda} + \mu_d^{\Lambda} - V^{\Lambda}\tau_{do} + w_{do}^{\Lambda}, \\
-T_{do} &= \eta_o^{T} + \mu_d^{T} - V^{T}\tau_{do} + w_{do}^{T}, \\
J_{do} &= \eta_o^{J} + \mu_d^{J} - V^{J}\tau_{do} + w_{do}^{J}, \\
U_{do} &= \eta_o^{U} + \mu_d^{U} - V^{U}\tau_{do} + w_{do}^{U}.
\end{aligned} \tag{30}$$

where we can compute Γ_{do}, Λ_{do}, $-T_{do}$, J_{do}, and U_{do} from estimates of the sectoral gravity equations (12) using the observed data on bilateral trade.

Estimating equations (26) and (30) 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 bilateral 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 bilateral 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}).

3 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 (26) 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 (Γ_{do}, Λ_{do}, T_{do}, J_{do}, U_{do}) in equation (20), and estimate separate gravity equations for each component, as in equation (30) 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 for each component across 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)). Nonetheless, we find a substantial negative and statistically significant coefficient on our composition term (row (iv)), which ranges from -0.5188 using 1-digit sectors to -1.2846 using 4-digit sectors. We also find positive and statistically significant correlations with distance for the origin-sector fixed effects (row (i)), the destination-sector fixed effects (row (ii)), and the error term (row (v)). This pattern of results is consistent with Alchian-Allen type effects, in which trade relationships over longer distances are a selected sample of relationships with superior characteristics. The net effect of all of these forces is that the ratio of the aggregate to the sectoral distance coefficients ranges from 1.05 to 1.40, depending on the level of aggregation, which highlights the importance of compositional differences for aggregate trade over long versus short distances.

4 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

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 (26) (row (vi)) and the components of aggregate bilateral trade from equation (30) (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.

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 or composition 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 international trade in the aggregate gravity equation.

References

Ahlfeldt, G., S. Redding, D. Sturm, and N. Wolf (2015): "The Economics of Density: Evidence from the Berlin Wall," *Econometrica*, 83(6), 2127–2189.

Allen, T., C. Arkolakis, and Y. Takahashi (2018): "Universal Gravity," *Journal of Political Economy*, forthcoming.

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

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

Arkolakis, C., A. Costinot, and A. Rodriguez-Clare (2012): "New Trade Models, Same Old Gains," *American Economic Review*, 102(1), 94–130.

Arkolakis, C., S. Demidova, P. Klenow, and A. Rodriguez-Clare (2008): "Endogenous Variety and the Gains from Trade," *American Economic Review*, 98(4), 444–450.

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

BAS, M., T. MAYER, AND M. THOENIG (2017): "From Micro to Macro: Demand, Supply, and Heterogeneity in the Trade Elasticity," *Journal of International Economics*, 108, 1–19.

BERNARD, A. B., S. J. REDDING, AND P. K. SCHOTT (2011): "Multi-product Firms and Trade Liberalization," *Quarterly Journal of Economics*, 126(3), 1271–1318.

BERTHELON, M., AND C. FREUND (2008): "On the Conservation of Distance in International Trade," *Journal of International Economics*, 75(2), 310–320.

CHANEY, T. (2008): "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, 98(4), 1707–21.

——— (2018): "The Gravity Equation in International Trade: An Explanation," *Journal of Political Economy*, 126(1), 150–177.

COMBES, P.-P., M. LAFOURCADE, AND T. MAYER (2005): "The Trade-Creating Effects of Business and Social Networks: Evidence from France," *Journal of International Economics*, 66, 1–29.

COSTINOT, A., D. DONALDSON, AND I. KOMUNJER (2012): "What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas," *Review of Economic Studies*, 79(2), 581–608.

COUGHIN, C. C., AND D. NOVY (2018): "Estimating Border Effects: The Impact of Spatial Aggregation," *CEPR Discussion Paper*, 11226.

DAVIS, D. R., AND D. E. WEINSTEIN (1999): "Economic Geography and Regional Production Structure: An Empirical Investigation," *European Economic Review*, 43, 379–407.

DEARDORFF, A. (1998): "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?," in *The Regionalization of the World Economy*, ed. by J. A. Frankel, chap. 1, pp. 7–32. University of Chicago Press, Chicago.

EATON, J., AND S. KORTUM (2002): "Technology, Geography, and Trade," *Econometrica*, 70(5), 1741–1779.

FEENSTRA, R. C., J. R. MARKUSEN, AND A. K. ROSE (2001): "Using the Gravity Equation to Differentiate among Alternative Theories of Trade," *Canadian Journal of Economics*, 34(2), 430–447.

FOTHERINGHAM, A. S., AND D. W. S. WONG (1991): "The Modifiable Areal Unit Problem in Multivariate Statistical Analysis," *Environment and Planning A*, 23, 1025–1044.

GEHLKE, C. E., AND K. BIEHL (1934): "Certain Effect of Grouping Upon the Size of the Correlation Coefficient in Census Tract Material," *Journal of the American Statistical Association*, 29, 169–170.

HEAD, K., AND T. MAYER (2014): "Gravity Equations: Workhorse, Toolkit, and Cookbook," in *Handbook of International Economics*, ed. by E. Helpman, G. Gopinath, and K. Rogoff, vol. 4, pp. 131–196. Elsevier, North Holland, Amsterdam.

HEAD, K., AND J. 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.

Heblich, S., S. Redding, and D. Sturm (2018): "The Making of the Modern Metropolis: Evidence from London," Princeton University, mimeograph.

Helpman, E., M. J. Melitz, and Y. Rubinstein (2008): "Estimating Trade Flows: Trading Partners and Trading Volumes," *Quarterly Journal of Economics*, 123(2), 441–487.

Kennan, J., and J. R. Walker (2011): "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 79(1), 211–251.

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

Lind, N., and N. Ramondo (2018): "Trade with Correlation," *NBER Working Paper*, 24380.

Martin, P., and H. Rey (2004): "Financial Super-markets: Size Matters for Asset Trade," *Journal of International Economics*, 64, 335–361.

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

Monte, F., S. Redding, and E. Rossi-Hansberg (2018): "Commuting, Migration and Local Employment Elasticities," *American Economic Review*, forthcoming.

Ramondo, N., A. Rodríguez-Clare, and M. Saborio (2016): "Trade, Domestic Frictions, and Scale Effects," *American Economic Review*, 106(10), 3159–3184.

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

Redding, S. J., and D. E. Weinstein (2018): "Accounting for Trade Patterns," Princeton University, mimeograph.

Santos Silva, J., and S. Tenreyro (2006): "The Log of Gravity," *Review of Economics and Statistics*, 88(4), 641–658.

Tinbergen, J. (1962): *Shaping the World Economy: Suggestions for an International Economic Policy*. Twentieth-Century Fund, New York.