Mismeasurement of Distance Effects: The Role of Internal Location of Production

Hakan Yilmazkuday*

December 15, 2014

Abstract

The estimated effects of distance in empirical international trade regressions are unrealistically high. Using state-and-sector level U.S. exports data, this paper shows analytically and proves empirically that ignoring the internal location of production (of international exports), which leads to the overestimation of distance effects by about twofold, is a possible explanation. This overestimation is mostly attributed to the mismeasurement of the distance elasticity of trade costs when internal locations of production are ignored. A *corrective distance index* is proposed to avoid such mismeasurements and is shown to work well for the median sector. The results are robust to the consideration of alternative estimation methodologies and data sets.

JEL Classification: F12, F13, F14

Key Words: Corrective Distance Index; Elasticity of Substitution; Distance Elasticity of Trade; State Exports

*Department of Economics, Florida International University, Miami, FL 33199, USA; e-mail: hyilmazk@fiu.edu

1 Introduction

The concept of "trade costs" has been one of the keys to understanding welfare-reducing barriers to international trade. Anderson and van Wincoop (2004) broadly define it by considering its components such as transportation costs (including time-to-ship), policy barriers, information costs, contract enforcement costs, costs associated with the use of different currencies, legal and regulatory costs, and local distribution costs. When it comes to the measurement of these components, though, the data are either limited or nonexistent. To bypass the data problem, the common (and successful) empirical practice in the literature has been to use the geographical location of source and destination countries and thus geographical distance as a proxy to capture the effects of many components of trade costs introduced above. This practice has come at the cost of unrealistically high/overloaded estimated ad-valorem tax equivalents of distance effects, considered under the title of "distance puzzle"; e.g., in their meta-analysis based on 1,466 estimates in the literature, Disdier and Head (2008) have shown that the absolute value of the coefficient in front of (log) distance estimates in gravity-type regressions have a mean of 0.91 and a median of 0.87. While these high distance effects can be investigated under the *magnitude* dimension of the distance puzzle (see Grossman, 1998 or Anderson and van Wincoop, 2004), their persistence over time constitutes the *time* dimension of the puzzle (see Carrere and Schiff, 2005 or Berthelon and Freund, 2008).

In this paper, we focus on the *magnitude* dimension of the distance puzzle. To understand the severity of *magnitude* dimension better, consider the ad-valorem tax equivalents of distance effects. Under the assumption of constant elasticity of substitution (CES) utility functions, the estimated coefficient in front of (log) distance (i.e., the distance elasticity of trade) is the multiplication of the elasticity of substitution and the distance elasticity of trade costs. Following the empirical literature on international trade, if we consider the fact that the elasticity of substitution estimates are as low as 3, the mean/median distance elasticity of trade costs in Disdier and Head (2008) is implied about 0.3, which corresponds to ad-valorem tax equivalents of distance effects as much as 694% when distance (between source and destination) is about 1,000 miles. In the context of the *time* dimension of the puzzle, although the literature has attempted to explain and reduce the severity of these effects through several data sets and methodologies, there are no studies to our knowledge that particularly focus on the *magnitude* dimension of the distance puzzle.[1]

Accordingly, the contribution of this paper is twofold. First, we attempt to understand the *magnitude* of distance effects by considering their possible mismeasurement, which we call the **M**ismeasurement of **D**istance **E**ffects (MDE). Second, we propose a *corrective distance index* that can be used to avoid MDE.

In particular, based on a simple model, we analytically show that the estimated effects of distance would be mismeasured if the internal location of production (of international exports) is ignored in the estimation. The magnitude and the direction of MDE, however, depends on the estimated variables (e.g., source prices), parameters (e.g., elasticity of substitution, distance elasticity of trade costs, taste parameters), and distance data (e.g., the spatial distribution of production). Accordingly, to determine such details empirically, we estimate the implications of our model under two data sets of the U.S. exports at the 3-digit NAICS sector level, one considering the location of production at the state level (i.e., the estimation using state-and-sector level U.S. exports data), the other one ignoring the location of production (i.e., the estimation using only sector-level U.S.

[1] For instance, Estevadeordal et al. (2003) have considered possible increases in marginal costs of transportation with respect to of production, Engel (2002) have focused on the role of nontradables sectors, Felbermayr and Kohler (2006) have taken into account zero-trade observations, Berthelon and Freund (2008) have investigated the role of composition of trade among industries, Head et al. (2009) have included fixed effects in their regressions to account for trading propensities of entrants, and Yotov (2012) has considered the increase in international economic integration relative to the integration of internal markets.

exports data). The results show that the median (across sectors) distance elasticity of trade is estimated around 0.17 with state-and-sector level exports data, while it is around 0.50 when only sector level exports data are used.

In order to depict the role of MDE on the ad-valorem tax equivalents of distance effects, under the assumption of CES utility functions, we further decompose the estimated coefficient in front of distance (i.e., the distance elasticity of trade) into the elasticity of substitution and the distance elasticity of trade costs. Such a decomposition, however, depends on the identification of the elasticity of substitution which requires an additional set of information; e.g., studies such as Harrigan (1993), Hummels (2001), Head and Ries (2001), and Baier and Bergstrand (2001) use additional information on directly observed trade barriers for this identification, studies such as Feenstra (1994) and Eaton and Kortum (2002) use additional information on prices, and Yilmazkuday (2012) uses additional information on markups obtained from production data. Within these options, we follow Yilmazkuday (2012) by using data on state-and-sector level production (for the U.S.) to identify the elasticity of substitution across varieties (each produced in a different U.S. state) of each sector, and by using data on sector level production (for the U.S.) to identify the elasticity of substitution across products of different sectors (produced in the U.S.). In the estimation process, while the former is used to identify the distance elasticity of trade costs when state-and-sector level exports data are used, the latter is used to identify the distance elasticity of trade costs when only sector level exports data are used. This identification strategy is similar to the approach used by Anderson and van Wincoop (2003) who connect CES utility functions to gravity-type estimations; however, this paper is different from theirs, since they use an *ad hoc* measure of the elasticity of substitution for identification, while we estimate it using production-side data. The results show that the median (across sectors) distance elasticity of trade costs is estimated around 0.05 with state-and-sector level exports data, while it is around 0.15 when only sector level exports data are used. In order to have a better idea about the difference between the distance elasticity of trade costs estimates of 0.05 and 0.15, consider the corresponding ad-valorem tax equivalents of distance effects: when distance measure is 1,000 miles, 0.05 corresponds to 41%, and 0.15 corresponds to 182%.

Finally, by considering the appropriate aggregations, we calculate the overall MDE when the internal location of production is ignored. The results show that the distance effects estimated by sector level data are on average about double the distance effects estimated by state-and-sector level data; therefore, distance effects are overestimated when sector level data are employed. These results are robust to the consideration of alternative estimation methodologies and data sets.

When we formally investigate the source of MDE, it is evident that the lion's share belongs to the mismeasurement of the distance elasticity of trade costs and ignoring preferences of individuals in the destination countries (among products produced in different locations within the U.S.). Across sectors, we also show that MDE reduces as the elasticity of substitution (across the products of U.S. states) increases. Therefore, MDE is mostly due to aggregation issues where the underlying micro details are still coming from the internal location of production (i.e., disaggregated data).

However, such disaggregated data are not available all the time. Accordingly, we propose a solution to avoid mismeasurement of distance effects through the estimation of distance elasticity of trade costs. Under certain conditions, we analytically show that the mismeasurement can be avoided by using a *corrective distance index* created by using internal distance measures (i.e., the dispersion of economic activity) and international distance measures (i.e., the remoteness of the source country from the rest of the world). We employ this index and show that it works well to avoid MDE (i.e., the *magnitude* dimension of the distance puzzle) for the median sector. In this context, this paper is complementary to Yotov (2012) who has proposed a solution to the *time* dimension of the distance puzzle considering the international economic integration relative to the integration of internal markets.

The following section introduces a simple model to motivate the empirical investigation where the source of MDE is shown analytically; it also depicts the details of the data set and the empirical methodology. The empirical results are given in Section 3. Section 4 provides guidelines for future studies and show how MDE can be corrected. The last section includes concluding remarks.

2 Methodology

In order to form a simple basis for our empirical framework, we model the exports of U.S. products at the state and sector level to a finite number of importers. In the model, generally speaking, $H^s_{d,g}$ stands for the variable H, where d is the importer country (i.e., the destination), g is the sector (or good), and s is the exporter state (i.e., the source).

In terms of the effects of distance on trade, we would like to show the importance of using the internal location of production for U.S. exports. Since our data for U.S. exports are at the state-and-sector level, we will consider the location of production within the U.S. at the state level for each sector. Therefore, after introducing the preferences of importer countries, we will discuss two possible estimation methodologies for investigating the U.S. exports, one at the state-and-sector level, another at the sector level only. Based on the available data, a researcher may focus on any of these methodologies; however, as we will show in this paper, these two methodologies can result in different estimates of distance effects.

2.1 Preferences of Importers

We assume that the utility maximization problem of the representative agent in destination country d is separable across source countries; hence, we focus on her optimization problem for the U.S. products only. She has the following CES preferences over the products of different sectors (each represented by g) coming from the U.S.:

$$C_d \equiv \left(\sum_g \left(\alpha_{d,g}\right)^{\frac{1}{\gamma}} \left(C_{d,g}\right)^{\frac{\gamma-1}{\gamma}} \right)^{\frac{\gamma}{\gamma-1}}$$

where C_d is the composite index of U.S. products (consisting of $C_{d,g}$'s), $\gamma > 1$ is the elasticity of substitution across the products of U.S. sectors, and $\alpha_{d,g}$ is a destination-sector specific taste parameter. The representative agent also has preferences over the varieties of each sector g, where each variety s is produced in state s in the U.S.:

$$C_{d,g} \equiv \left(\sum_s \left(\beta_{d,g}^s\right)^{\frac{1}{\eta_g}} \left(C_{d,g}^s\right)^{\frac{\eta_g-1}{\eta_g}} \right)^{\frac{\eta_g}{\eta_g-1}}$$

where $C_{d,g}$ is the composite index of sector g consisting of imported products $C_{d,g}^s$'s coming from different states, $\eta_g > 1$ is the elasticity of substitution across products of states for sector g, and $\beta_{d,g}^s$ is a source-destination-sector specific taste parameter.

The optimal condition for expenditure on the products of sector g imported by destination country d from source state s implies that:

$$C_{d,g}^s = \beta_{d,g}^s \left(\frac{P_{d,g}^s}{P_{d,g}} \right)^{-\eta_g} C_{d,g} \tag{1}$$

where P_g^s is the source price per $C_{d,g}^s$ and connected to $P_{d,g}$ by:

$$P_{d,g} \equiv \left(\sum_s \beta_{d,g}^s \left(P_{d,g}^s\right)^{1-\eta_g} \right)^{\frac{1}{1-\eta_g}} \tag{2}$$

where $P_{d,g}$ is the price per $C_{d,g}$. Similarly, the optimal condition for expenditure on the products of sector g imported by destination country d from the U.S. (aggregated across states) implies that:

$$C_{d,g} = \alpha_{d,g} \left(\frac{P_{d,g}}{P_d} \right)^{1-\gamma} C_d \tag{3}$$

where P_d is the price per C_d and connected to $P_{d,g}$ by:

$$P_d \equiv \left(\sum_g \alpha_{d,g} \left(P_{d,g}\right)^{1-\gamma} \right)^{\frac{1}{1-\gamma}}$$

2.2 Estimation at the State-and-Sector Level

When U.S. exports data are available at the state-and-sector level, we can also model the production side at the state level. In particular, suppose that sector g producer in state s maximizes the following profit function:

$$\max_{P_g^s} Y_g^s \left(P_g^s - M_g^s \right)$$

subject to Equation 1 and the following market-clearing condition:

$$P_g^s Y_g^s = \sum_d P_{d,g}^s C_{d,g}^s$$

where Y_g^s is the amount of production, P_g^s is the source price charged, M_g^s is the state-and-sector specific marginal cost of production, and $\sum_d P_{d,g}^s C_{d,g}^s$ is the total international demand. For each sector, the profit maximization results in the following relation between total variable costs and total revenues at the state level:

$$\underbrace{P_g^s Y_g^s}_{\text{Total Revenue}} = \underbrace{\left(\frac{\eta_g}{\eta_g - 1} \right)}_{\text{Gross Markup}} \underbrace{M_g^s Y_g^s}_{\text{Total Variable Costs}} \tag{4}$$

which can be aggregated across states to obtain a similar expression at the U.S. level:

$$\underbrace{\sum_s P_g^s Y_g^s}_{\text{Total Revenue}} = \underbrace{\left(\frac{\eta_g}{\eta_g - 1} \right)}_{\text{Gross Markup}} \underbrace{\sum_s M_g^s Y_g^s}_{\text{Total Variable Costs}} \tag{5}$$

When production-side data are available for total variable costs and total revenues at the state-and-sector level, in order to identify gross markups $\left(\text{i.e., } \frac{\eta_g}{\eta_g - 1}\text{'s}\right)$ as in Yilmazkuday (2012), either Equation 4 (when state-and-sector level data are employed) can be estimated without any constant for each sector, or Equation 5 (when sector level data are employed for the U.S.) can be used (by simply dividing the total revenue by total variable costs at the U.S. level for each sector). Once the gross markups are estimated, the elasticities of substitution η_g's (across products of states for each sector g) can be identified, and, when state-and-sector level data are employed, the corresponding standard errors can be calculated by the Delta method. For robustness, we will consider the implications of both Equations 4 and 5 in our empirical analysis, below.

The source price P_g^s of sector g in state s and the corresponding destination price $P_{d,g}^s$ at destination d are connected to each other through the following iceberg-melting trade costs (to have consistency with the literature that we will compare our empirical results with, below):

$$P_{d,g}^s = P_g^s \left(D_d^s\right)^{\delta_g} (1 + t_d) \tag{6}$$

where D_d^s is the distance between source state s and destination country d, δ_g is the sector-specific distance elasticity of trade costs, and t_d represents destination-specific trade costs (e.g., tariff rates, local distribution costs, etc.).

Using Equations 1 and 6, we can now write an expression for the log source value of U.S. exports at the source-destination-and-sector level as follows:

$$\begin{aligned} \underbrace{\log\left(P_g^s C_{d,g}^s\right)}_{\text{Source Value of Exports}} &= \underbrace{\left(1-\eta_g\right)\log\left(P_g^s\right)}_{\text{Source-and-Sector Fixed Effects}} - \underbrace{\eta_g \delta_g \log\left(D_d^s\right)}_{\text{Sector-Specific Distance Effects}} \\ &\quad + \underbrace{\log\left(\left(P_{d,g}\right)^{\eta_g} C_{d,g}\right) - \eta_g \log\left(1+t_d\right)}_{\text{Destination-and-Sector Fixed Effects}} + \underbrace{\log\left(\beta_{d,g}^s\right)}_{\text{Residuals}} \end{aligned} \tag{7}$$

which can be estimated by a pooled regression using source-and-sector fixed effects, distance interacting with sector-specific dummies, and destination-and-sector fixed effects. In this estimation, the coefficient in front of distance interacting with the sector-g-specific dummy corresponds to the distance elasticity of trade $\eta_g \delta_g$ for sector g. After the estimation, δ_g's can be identified using the already-identified η_g's (through the production-side data), and their standard errors can be calculated by the Delta method when state-and-sector level data for η_g's are employed.

At this stage of the investigation, it is important to emphasize that the distance elasticity of trade costs δ_g is something different than the distance elasticity of trade, which is the coefficient

$\eta_g \delta_g$ in front of distance in Equation 7. We will talk on the implications of this nuance during the empirical investigation, below.

2.3 Estimation at the Sector Level

When U.S. exports data are available at the sector level only, we can model the production side at the U.S. level. In particular, suppose that sector g producer in the U.S. maximizes the following profit function:

$$\max_{P_g} Y_g \left(P_g - M_g\right)$$

subject to Equation 3 and the following market-clearing condition:

$$P_g Y_g = \sum_d P_{d,g} C_{d,g}$$

where Y_g is the amount of production, P_g is the source price charged, M_g is the sector specific marginal cost of production, and $\sum_d P_{d,g} C_{d,g}$ is the total international demand. The profit maximization results in the following relation between total variable costs and total revenues at the sector level within the U.S.:

$$\underbrace{P_g Y_g}_{\text{Total Revenue}} = \underbrace{\left(\frac{\gamma}{\gamma - 1}\right)}_{\text{Gross Markup}} \underbrace{M_g Y_g}_{\text{Total Variable Costs}} \tag{8}$$

which can be aggregated across sectors to obtain a similar expression at the U.S. level (i.e., pooled sample across sectors):

$$\underbrace{\sum_g P_g Y_g}_{\text{Total Revenue}} = \underbrace{\left(\frac{\gamma}{\gamma - 1}\right)}_{\text{Gross Markup}} \underbrace{\sum_g M_g Y_g}_{\text{Total Variable Costs}} \tag{9}$$

Similar to what we have above, when production-side data are available for total variable costs and total revenues at the sector level, in order to identify gross markups $\left(\text{i.e., } \frac{\gamma}{\gamma-1}\text{'s}\right)$, either Equation 8 (when sector level data are employed) can be estimated without any constant, or Equation 9 (when pooled sample across sectors is employed) can be used (by simply dividing the total revenue by total variable costs). Once the gross markups are estimated, the elasticity of substitution γ across U.S. sectors can be identified, and, when sector level data are employed, the corresponding standard errors can be calculated by the Delta method. Again, for robustness, we will consider the implications of both of them in our empirical analysis, below.

The source price P_g of sector g in the U.S. and the corresponding destination price $P_{d,g}$ at destination country d are connected to each other through the following trade costs:

$$P_{d,g} = P_g \left(D_d\right)^{\delta'_g} \left(1 + t_d\right) \tag{10}$$

where D_d is the distance between the U.S. and destination country d (to be compared with its counterpart of D^s_d in the previous subsection), δ'_g is the sector-specific distance elasticity of trade costs, and t_d represents destination-specific trade costs as above. Note that δ'_g is theoretically different from its counterpart δ_g in the previous subsection, because the level of aggregation at which they contribute to the overall distance effects is different; we will work on this connection while depicting the mismeasurement of distance effects, below.

Using Equations 3 and 10, we can now write an expression for the log source value of U.S. exports at the destination-and-sector level as follows:

$$\begin{aligned} \underbrace{\log\left(P_g C_{d,g}\right)}_{\text{Source Value of Exports}} &= \underbrace{(1-\gamma)\log\left(P_g\right)}_{\text{Sector Fixed Effects}} - \underbrace{\gamma\delta'_g \log\left(D_d\right)}_{\text{Sector-Specific Distance Effects}} \\ &\quad + \underbrace{\log\left(\left(P_d\right)^{\gamma} C_d\right) - \gamma\log\left(1+t_d\right)}_{\text{Destination Fixed Effects}} + \underbrace{\log\left(\alpha_{d,g}\right)}_{\text{Residuals}} \end{aligned} \tag{11}$$

where there are no source-fixed effects, because the source is the same (i.e., the U.S.) for all U.S. exports. For the pooled sample, this expression can be estimated using a two-stage estimation process, because, besides having destination fixed effects, we also have destination-specific distance effects due to having only one source country (i.e., the U.S.). Accordingly, in the first stage, source value of U.S. exports are regressed on sector and destination fixed effects. In the second stage, the fitted values of the first stage are regressed on distance interacting with sector-specific dummies to obtain the elasticity of trade estimates $\gamma\delta_g'$'s. After the estimation, δ_g''s can also be identified using the already-identified γ (through the production-side data), and their standard errors can be calculated by the Delta method when sector level data for γ are employed.

It is again important to emphasize that the distance elasticity of trade costs δ_g' is something different than the distance elasticity of trade, which is this time the coefficient $\gamma\delta_g'$ in front of distance in Equation 11.

2.4 The Mismeasurement of Distance Effects (MDE)

Before the empirical investigation, we would like to analytically show MDE when sector-level (rather than state-and-sector level) U.S. exports data are employed. In order to depict MDE, we focus on the combined version of Equations 2, 6 and 10:

$$P_g \left(D_d\right)^{\delta_g'} \left(1+t_d\right) \equiv \left(\sum_s \beta_{d,g}^s \left(P_g^s \left(D_d^s\right)^{\delta_g} \left(1+t_d\right)\right)^{1-\eta_g} \right)^{\frac{1}{1-\eta_g}}$$

which reduces to the following expression after some manipulation:

$$\left(D_d\right)^{\delta_g'} \equiv \left(\sum_s \beta_{d,g}^s \left(\frac{P_g^s \left(D_d^s\right)^{\delta_g}}{P_g} \right)^{1-\eta_g} \right)^{\frac{1}{1-\eta_g}} \tag{12}$$

where destination-specific trade costs t_d's have been effectively eliminated. This expression depicts that the distance effects when sector level data $\left(\text{i.e., } \left(D_d\right)^{\delta_g'}\text{'s}\right)$ are used are weighted averages of the distance effects when state-and-sector level data $\left(\text{i.e., } \left(D_d^s\right)^{\delta_g}\text{'s}\right)$ are used, where weights are given by the relative source price of a particular state P_g^s with respect to the U.S. average source prices P_g, as well as the taste parameters $\beta_{d,g}^s$'s representing the preferences of destination individuals for the products of particular source states.

In order to further understand the details of Equation 12 analytically, consider the following definition of MDE (in percentage terms) representing as the log difference between the two sides of Equation 12:

$$MDE_0 = \underbrace{\log (D_d)^{\delta'_g}}_{\text{Sector Level Data}} \underbrace{- \log \left(\sum_s \beta^s_{d,g} \left(\frac{P^s_g (D^s_d)^{\delta_g}}{P_g} \right)^{1-\eta_g} \right)^{\frac{1}{1-\eta_g}}}_{\text{State-and-Sector Level Data}} \tag{13}$$

Equation 13 clearly shows that the sector level distance effects may be mismeasured with respect to the aggregated version of state-and-sector level distance effects. The mismeasurement (if any) would be affected by not only the elasticities of substitution η_g's and taste parameters of $\beta^s_{d,g}$'s but also source prices at the U.S. level (i.e., P_g's) versus at the source-state level (i.e., P^s_g's), as well as distance measures at the state level $\left(\text{i.e., } (D^s_d)^{\delta_g}\text{'s}\right)$ versus the U.S. level $\left(\text{i.e., } (D_d)^{\delta'_g}\text{'s}\right)$. Without knowing these variables/parameters, it is hard to talk about the details of MDE; therefore, the determination of MDE requires an empirical investigation at both layers of aggregation.

Before jumping to the empirical investigation, to further understand the details of Equation 13 analytically, consider a special case in which $P^s_g = P_g$ for all s (i.e., source prices are the same across source states); in such a case, MDE_0 would reduce to the following expression:

$$MDE_1 = \underbrace{\log (D_d)^{\delta'_g}}_{\text{Sector Level Data}} \underbrace{- \log \left(\sum_s \beta^s_{d,g} \left((D^s_d)^{\delta_g} \right)^{1-\eta_g} \right)^{\frac{1}{1-\eta_g}}}_{\text{State-and-Sector Level Data}} \tag{14}$$

which is similar to what Berthelon and Freund (2008) have shown regarding the aggregation across goods (compared to aggregation across states in our case) in the context of gravity-type estimations. Therefore, by considering the location of production within the exporter country, our investigation goes one step further compared to their analysis in terms of depicting MDE. If we further assume that $D^s_d = D_d$ for all s (i.e., distance between source states and destination countries are the same with the distance between the U.S. and destination countries, which is an unrealistic assumption, but consider this special case to see the properties of MDE), MDE_1 in Equation 14 would further reduce to the following expression, after also assuming that $\sum_s \beta^s_{d,g} = 1$ (or $\eta_g \to \infty$):

$$MDE_2 = \log\left(D_d\right)^{\delta'_g} - \log\left(D_d\right)^{\delta_g}$$

Finally, on top of the previous assumptions, if $\delta'_g = \delta_g$, then MDE_2 would disappear. Therefore, only in a very special (and unrealistic) case of $P^s_g = P_g$, $D^s_d = D_d$, $\delta'_g = \delta_g$, and $\sum_s \beta^s_{d,g} = 1$ (or $\eta_g \to \infty$) for all s, we can talk about the disappearance of MDE when sector level (rather than state-and-sector level) data are used for U.S. exports.

In order to calculate MDE_0 in Equation 13, we will achieve estimations at both state-and-sector level (Equation 7) and sector level (Equation 11). In particular, estimations of Equations 7 and 11 result in the identification of P^s_g's (by source-and-sector fixed effects), P_g's (by sector fixed effects), δ_g's and δ'_g's (by coefficients in front of distance interacting with sector dummies), and $\beta^s_{d,g}$'s (by residuals). Afterwards, we will investigate the source of MDE by shutting down alternative mechanisms by considering alternative combinations of $P^s_g = P_g$, $D^s_d = D_d$, $\delta'_g = \delta_g$, and $\sum_s \beta^s_{d,g} = 1$ (or $\eta_g \to \infty$). This investigation will also have implications for future studies to understand the source of MDE due to the spatial dispersion of export production (i.e., mostly reflected by $P^s_g = P_g$, $D^s_d = D_d$, $\delta'_g = \delta_g$) versus preferences at the destination countries (reflected by $\sum_s \beta^s_{d,g} = 1$ or $\eta_g \to \infty$).

2.5 Data and Estimation Methodology

In order to consider the location of production (of exports), we use U.S. exports data at the state level for 3-digit NAICS manufacturing sectors/goods. The list of these sectors is given in Table 1. Since the identification of the distance elasticity of trade costs depends on using production-side data according to Equation 4 (or Equation 5, for robustness) and Equation 8 (or Equation 9, for robustness), we use the corresponding production data for the sectors given in Table 1 that are available for the years of 2002 and 2007 in Economic Census Data of U.S. Census Bureau. In particular, under the assumptions of CES utility functions and constant returns to scale, we use the state-and-sector level production data of total variable costs and total revenues in the U.S. to determine/estimate gross markups and thus elasticities of substitution η_g's across varieties of any sector g produced in different states (by Equation 4 or Equation 5) and elasticities of substitution γ across the products of different U.S. sectors (by Equation 8 or Equation 9).[2] In Economic Census Data, the total costs are decomposed into the sum of annual payroll, total cost of materials, and total capital expenditures. We accept annual payroll and total cost of materials as variable costs, however total capital expenditures can be a part of either fixed or variable costs. Hence, for robustness, in order to calculate total variable costs, we consider two alternative approaches. According to the first (second) approach, we consider flexible (fixed) capital, which implies that total capital expenditures are (not) a part of total variable costs. Therefore, according to the first approach, which we accept as our benchmark case, total variable costs are defined as the sum of annual payroll, total cost of materials and total capital expenditures, and according to the second approach, which we accept as the alternative/robustness case, total variable costs are defined as the sum of annual payroll and total cost of materials.

The main source of trade data is the U.S. State-Export Data obtained from the TradeStats Express.[3] The data cover the exports of 50 states of the U.S. and the District of Columbia to 230 countries around the globe.[4] The value of 3-digit NAICS industry-level exports from each U.S. state

[2]It is important to emphasize that under the assumption CES, which implies a pricing strategy of constant markups over marginal costs, the portion of the production that is sold internally within the U.S. is irrelevant in our investigation.

[3]TradeStats Express draws all state export statistics from the Origin of Movement (OM) series compiled by the Foreign Trade Division of the U.S. Census Bureau.

[4]For further information on U.S. trade data, visit http://www.census.gov/foreign-trade/guide/

to the destination countries (measured at the source state) are used. A typical export observation is the value of exports for "Beverages & Tobacco Products" from the state of California to Japan. Although these trade data are available starting from 1999, we will restrict our empirical analysis to the years of 2002 and 2007 to have consistent trade data with our production data. For robustness, we will run separate set of regressions (of Equations 7 and 11) for the years of 2002 and 2007. We consider the same observations (in terms of sectors and destinations) across 2002 and 2007 to have a healthy comparison across years.

Regarding the distance measures, when U.S. exports data at the state-and-sector level are employed (i.e., estimation of Equation 7), great circle distances between each importer country and each exporter state of the U.S. are calculated by using population-weighted latitudes and longitudes of the exporter states and importer countries. Similarly, when U.S. exports data at only the sector level are employed (i.e., estimation of Equation 11), great circle distances between each importer country and the U.S. are calculated using the population-weighted latitudes and longitudes of the importer countries and the U.S..[5]

Following Santos Silva and Tenreyro (2006) or Henderson and Millimet (2008), in order to consider additive versus multiplicative error terms, which have implications mostly on the distribution of estimated taste parameters according to the model in this paper, Equations 7 and 11 are estimated by two different estimation methodologies, namely Ordinary Least Squares (OLS) and Poisson Pseudo-Maximum Likelihood (PPML). As Head and Mayer (forthcoming) show, when OLS and PPML estimates are close to each other, it is an indicator of robustness for the estimated

[5]For further robustness, we also considered alternative bilateral distance indicators in the economic geography database of CEPII (Centre d'e´tudes prospectives et d'informations internationales); see Mayer and Zignago (2011) for the details in this database. The results were virtually the same across different distance measures. Although these results have been skipped to save space, they are available upon request.

parameters.

3 Estimation Results

We estimate η_g's by Equation 4 using state-and-sector level production data in the U.S., γ by Equation 8 using (pooled-across-states version of) sector level production data in the U.S., $\eta_g \delta_g$'s by Equation 7 using data on state-and-sector level U.S. exports and distance, and $\gamma \delta_g'$'s by Equation 11 using data on sector level U.S. exports and distance. The estimation results are given in the following subsections. For robustness, we also estimate/identify η_g's by Equation 5 using sector level production data in the U.S. and γ by Equation 9 using pooled (across sectors) production data in the U.S..

3.1 Elasticity of Substitution

The estimates of elasticities of substitution (i.e., η_g's and γ) in the benchmark case (i.e., flexible capital) are introduced in Table 2 for the years of 2002 and 2007. When state-level production data are used, as is evident on the left panel of Table 2, all η_g estimates are significant at the 1% level; the highest η_g measures of 10.39 and 6.85 belong to "Petroleum and Coal Products", mostly due to the homogenous nature of the goods produced in this sector. In contrast, "Beverages and Tobacco Products" have the lowest η_g measures of 1.80 and 1.78. Similarly, the estimates of elasticities of substitution γ across the products of different sectors produced in the U.S. are also given in Table 2. These estimates (of η_g's and γ) are in line with mostly the lower bound of the estimates in the corresponding literature; estimates of Hummels (2001) range between 4.79 and 8.26; the estimates of Head and Ries (2001) range between 7.9 and 11.4; the estimate of Baier and Bergstrand (2001) is about 6.4; Harrigan's (1996) estimates range from 5 to 10; Feenstra's (1994) estimates range from

3 to 8.4; the estimate by Eaton and Kortum (2002) is about 9.28; the estimates by Romalis (2007) range between 6.2 and 10.9; the (mean) estimates of Broda and Weinstein (2006) range between 4 and 17.3; estimates of Simonovska and Waugh (2011) range between 3.47 and 5.42; and the median estimate of Yilmazkuday (2012) is about 3.01.

One important detail is that η_g estimates (slightly) change across years. Therefore, between 2002 and 2007, when one would like to decompose the distance elasticity of trade $\eta_g \delta_g$ into the elasticity of substitution η_g versus the distance elasticity of trade costs δ_g, the main effect may be coming from either η_g's or δ_g's; we will investigate this in detail, below. For each sector, having high R-squared values is important to show the validity of assuming common markups across states for each sector.

As a robustness measure, the elasticity of substitution estimates are very similar when production data at the U.S. level are used on the right panel of Table 2. For further robustness, we also consider the case of fixed capital, for which results are given in Table 3. As is evident, considering fixed capital has reduced the overall value of total variable costs that results in higher markups and thus lower elasticities of substitution. Nevertheless, across Table 2 and Table 3, the difference between the estimates is only about 10% on average, and the ranking of elasticities is mostly stable across sectors. These additional results support the robustness of our elasticity of substitution estimates.

3.2 Distance Elasticity of Trade

The distance elasticities of trade $\eta_g \delta_g$ at the sector level are first estimated (according to Equation 7) for the year of 2002. The results, which are all significant at the 1% level, are depicted in Table 4, where we distinguish between the estimates obtained by using state-and-sector level data (on the left panel) versus sector level data (on the right panel) as well as between estimation methodologies of OLS versus PPML. As is evident, all $\eta_g \delta_g$ estimates are lower with a median (across sectors) value of 0.17 when state-and-sector level data are employed, showing that $\gamma\delta'_g$'s are overestimated with a median (across sectors) value of about 0.50 (on average across OLS and PPML) when only sector level trade data are employed. Therefore, on average, the distance elasticity of trade estimates obtained by sector level data are about three times the estimates obtained by state-and-sector level data. If one thinks that this difference is specific to the year of 2002, she is wrong, because the same difference shows up also for the year of 2007 as shown in Table 5. Compared to the average (across studies) estimates in the literature, our distance elasticity of trade estimates are relatively lower, because, according to the excellent meta-analysis by Disdier and Head (2008) based on 1,466 estimates in the literature, mean distance elasticity of trade estimate is 0.91 and the median is 0.87, although the minimum estimate is as low as 0.04.[6]

3.3 Distance Elasticity of Trade Costs

In order to calculate MDE_0 in Equation 13, we need to identify the distance elasticity of trade costs δ_g's and δ'_g's, which we achieve by using the already estimated η_g's, γ's, $\eta_g\delta_g$'s, and $\gamma\delta'_g$'s. The estimated values of δ_g's and δ'_g's are given in Table 4 for the year of 2002. As is evident, independent of the estimation methodology, δ'_g estimates are higher than δ_g estimates about two to three times, even after considering their standard errors calculated by the Delta method. This violates one of the restrictive assumptions (i.e., $\delta_g = \delta'_g$) that we considered while depicting the analytical properties of MDE in Equation 13, above. The results are the same for the year of 2007 given in Table 5. Therefore, there is evidence for possible mismeasurement of distance effects

[6] As Disdier and Head (2008) nicely puts, any difference between empirical studies may be due to sampling error (chance errors in estimating a population parameter arising from the finite sample drawn from that population), "structural" heterogeneity (differences in parameters across subpopulations of the data), or "sampling" error (differences in statistical technique lead to different estimates).

according to the estimation results given in Tables 4-5. Using these estimates, we will calculate the exact amount of MDE in the next subsection; before that, we will compare our distance elasticity of trade costs estimates with the literature.

Compared to the existing literature, independent of the estimation methodology, our δ_g estimates obtained by state-and-sector level data are significantly lower with a median of about 0.05 (for both 2002 and 2007). In particular, Hummels (2001), Limao and Venables (2001), and Anderson and Wincoop (2004) all have δ_g estimates of around 0.3. When we consider our δ_g' estimates obtained by sector level data, they are very close to each other across sectors and are on average about 0.15 (across years and/or estimation methodologies); 0.15 is still lower compared to the literature. To understand the implications of these estimates, we will also consider their ad-valorem tax equivalents. For example, when distance measure is 1,000 miles, 0.05 corresponds to 41% $(= 100 \times (1000^{0.05} - 1))$, 0.15 corresponds to 182% $(= 100 \times (1000^{0.15} - 1))$, and 0.3 corresponds to 694% $(= 100 \times (1000^{0.3} - 1))$. Since we have several different distance measures and sectors in our data set, having a complete analysis is possible only through a formal investigation with the comparison of distance effects across appropriate levels of aggregation, which we achieve next.

3.4 The Source of MDE

Having estimates for P_g^s's, P_g's, δ_g's, δ_g''s, η_g's, and $\beta_{d,g}^s$'s, together with the data for D_d^s and D_d, we can now calculate MDE_0 given by Equation 13. MDE_0 estimates for the years of 2002 and 2007 are given in the first columns of Table 6 and Table 7, respectively, where we depict the average percentage deviation (divided by 100) across destination countries for each sector. As is evident, the distance effects estimated by sector level data are on average about double the distance effects estimated by state-and-sector level data. Therefore, our results in fact show one source of MDE that leads to higher estimates of distance effects that seem to be puzzling in the literature that has been discussed in detail during the introduction section, above.

In Tables 6-7, the highest MDE belongs to "Beverages and Tobacco Products", while the lowest MDE belongs to "Petroleum and Coal Products". Interestingly, these are the sectors with the lowest and the highest elasticity of substitution, respectively, according to Table 2 and Table 3. Therefore, there seems to be a negative relation between MDE and the elasticity of substitution. This relationship is further depicted for all sectors in Figure 1, where the correlation coefficient between MDE and log elasticity of substitution is about -0.78 (-0.85) for the year of 2002 (2007). Hence, as the products across states get more substitutable for each other, MDE reduces.[7] Nevertheless, as we had shown above, the elasticity of substitution (i.e., η_g being away from ∞) is only one source of MDE; we rather need a formal investigation to depict the contribution of each factor, as we achieve next.

Recall that we had shown analytically that only in a very special case of $P_g^s = P_g$, $D_d^s = D_d$, $\delta_g' = \delta_g$, and $\sum_s \beta_{d,g}^s = 1$ (or $\eta_g \to \infty$), MDE_0 given by Equation 13 would disappear; we empirically confirm this in the last columns of Tables 6-7. Therefore, by considering each of these restrictions, we can shut down particular mechanisms to have an idea about the source of MDE due to differences between P_g^s and P_g, between D_d^s and D_d, between δ_g' and δ_g, and between $\sum_s \beta_{d,g}^s$ and 1 (or η_g and ∞). We follow such an approach in the remaining columns of Table 6 and Table 7 for the years of 2002 and 2007, respectively, where each considered mechanism contributes to MDE with different magnitudes. As is evident, the special cases of $\delta_g' = \delta_g$ or $\sum_s \beta_{d,g}^s = 1$ (or $\eta_g \to \infty$) reduce MDE by about half; therefore, the lion's share for explaining MDE belongs to the mismeasurement of

[7] We also ran a regression of MDE on the log elasticity of substutition and the squared log elasticity of substutition, including a constant. For the year of 2002 (2007), the coefficient in front of log elasticity of substitution is significantly estimated as -11.48 (-16.57), while the coefficient in front of squared log elasticity of substitution is significantly estimated as 2.98 (4.86), with an R-squared value of 0.90 (0.92). The fitted lines for these two regressions are given in Figure 1. Therefore, across sectors, there is nonlinear relationship between MDE and the elasticity of substitution.

distance elasticity trade costs or of preferences.[8] The special case of $P_g^s = P_g$ also depicts significant differences with respect to the benchmark case, meaning that the price dispersion across states has also contributed to MDE. However, in the special case of $D_d^s = D_d$, MDE remains almost the same, meaning that the distance measures used (i.e., either state or nation specific) do not contribute much to MDE, which makes sense, because, from a global perspective, the state-specific distance measures are relatively close to each other; e.g., the log distance between New York State and China is about 8.87 while the log distance between California State and China is about 8.79. These results are also supported by the alternative combinations of the special cases considered in the rest of Table 6 and Table 7.

4 Guideline for Future Studies

The results for the source of MDE have implications for future studies. For example, it is by now evident that estimation of the distance elasticity of trade costs (i.e., δ_g's versus δ_g''s), together with preferences (i.e., $\beta_{d,g}^s$'s), is the key in understanding MDE. Although it is obvious that using more disaggregated data (e.g., state-and-sector level U.S. exports data in this paper) is a solution for the mismeasurement, what can an empirical researcher do when such data are not available? This section proposes a solution to this problem under certain conditions.

Accordingly, consider Equation 12, divide both sides of it by $(D_d)^{\delta_g}$, and take the summation

[8] It is important to emphasize that preferences measured by residuals may also be capturing any type of measurement errors, including the errors in trade costs. Therefore, if there are in fact measurement errors, the special case of $\sum_s \beta_{d,g}^s = 1$, which is the mainly used assumption in the literature, also provides insight regarding how much MDE would be reduced in the absence of such errors. However, since we do not have any information regarding these errors, we cannot say anything further.

across destination countries to obtain (after some manipulation):

$$\delta_g' - \delta_g \equiv \sum_d \left(\frac{\log \left(\sum_s \beta_{d,g}^s \left(\frac{P_g^s}{P_g} \left(\frac{D_d^s}{D_d} \right)^{\delta_g} \right)^{1-\eta_g} \right)^{\frac{1}{1-\eta_g}}}{N \log \left(D_d \right)} \right)$$

where N is the number of destination countries. Since the special case of $D_d^s = D_d$ does not have a significant effect on MDE (according to Table 6 and Table 7), one can employ this special case as an assumption to obtain:

$$\delta_g' - \delta_g \equiv \sum_d \left(\frac{\log \left(\sum_s \beta_{d,g}^s \left(\frac{P_g^s}{P_g} \right)^{1-\eta_g} \right)^{\frac{1}{1-\eta_g}}}{N \log D_d} \right)$$

which shows that the mismeasurement of the distance elasticity of trade costs is due to the source-price dispersion across states (i.e., P_g^s/P_g) as well as the remoteness of the U.S. from the rest of the world (measured by D_d's). Since P_g^s's are not known in the absence of disaggregated (state-level) data, as an approximation, one can use the following arbitrage condition to connect state-level source prices P_g^s's to the national source prices P_g's measured at a central location of the nation (e.g., at the capital city or the population-weighted center of the country):

$$P_g^s = P_g \left(D^s \right)^{\delta_g'}$$

where D^s represents the distance between the central location of the nation and the source state s. This arbitrage condition literally says that any sector g producer in state s is indifferent between producing its product and reselling a composite index of the same sector purchased from the central location of the nation, subject to internal (multiplicative) trade costs of $\left(D^s \right)^{\delta_g'}$, similar to Equation 10 that depicts the trade costs for the composite index. Assuming that the producer always chooses to produce its product out of this indifferency, it is implied that the mismeasurement of the distance elasticity of trade costs can be avoided by using a corrective distance index:

$$\delta_g = \delta'_g \underbrace{\left(1 - \frac{\log\left(D^{US}\right)}{R^{US}}\right)}_{\text{Corrective Distance Index}} \tag{15}$$

where D^{US} represents an index of the internal distance in the U.S. and is given by:

$$D^{US} = \left(\sum_s \beta^s_{d,g} \left(D^s\right)^{\delta'_g\left(1-\eta_g\right)}\right)^{\frac{1}{\delta'_g\left(1-\eta_g\right)}}$$

and R^{US} represents an index of the remoteness of the U.S. from the rest of the world and is given by:

$$R^{US} = \left(\sum_d \left(N \log D_d\right)^{-1}\right)^{-1}$$

As is evident by Equation 15, δ_g is always lower than δ'_g as long as D^{US} and R^{US} are positive. This is consistent with the estimation results in Table 4 and Table 5. For any given δ'_g, it is also evident that δ_g gets lower as internal distance gets higher (i.e., as the country gets more dispersed) or as remoteness of the country gets lower (i.e., as the country gets closer to other countries).[9]

How can one calculate the corrective distance index in Equation 15 in the absence of disaggregated (e.g., state-level) data? In such a case, since $\beta^s_{d,g}$'s and η_g's are not known, one can simply proxy D^{US} by the internal distance measure provided by economic geography database of CEPII (Centre d'e´tudes prospectives et d'informations internationales), documented by Mayer and Zignago (2011), which is a population/agglomeration weighted average of distances within a country. R^{US} can also be easily obtained by using the distance measures between the U.S. and the destination countries (i.e., D_d's). Finally, δ'_g's are already known by the aggregate-level (i.e., sector-level) estimations when sector-level international trade data are available.

[9]This implication is also in line with Yotov (2012) who has proposed a solution for the time dimension of the distance puzzle by showing that the effect of international distance relative to the effect of internal distance is falling over time.

Accordingly, we borrow the internal distance measure of $D^{US} = 261.72$ from CEPII (i.e., the measure called *distwces* that is consistent with gravity models of bilateral trade flows, as shown by Mayer and Zignago, 2011); we calculate $R^{US} = 9.03$ using data on the distance D_d between the U.S. and the destination countries; and, we use $\delta'_g = 0.14$ as the average of the OLS and PPML median sector-level estimates from Table 4 and Table 5. It is implied that:

$$\delta_g = \delta'_g \left(1 - \frac{\log\left(D^{US}\right)}{R^{US}}\right) = 0.14\left(1 - \frac{\log(261.72)}{9.03}\right) = 0.05$$

which is exactly the median state-and-sector level median estimate that we have for δ_g in Table 4 and Table 5. Hence, our proposed solution works well for the median sector in our sample.

For all sectors, it is also implied that:

$$\frac{\delta_g}{\delta'_g} = \left(1 - \frac{\log\left(D^{US}\right)}{R^{US}}\right) = 0.38$$

However, across sectors, according to Table 4 and Table 5, this ratio ranges between 0.11 and 0.50 when OLS is used, and it ranges between 0.14 and 0.69 when PPML is used as an estimation methodology; hence, the measures that we have used in our calculations have not captured the heterogeneity across sectors. In this context, it is important to emphasize that the internal distance measure D^{US} borrowed from CEPII considers the agglomeration that is on average across all sectors of a country; therefore, it works well only for the median sector. However, one better needs an sector-level measure of internal distance, possibly by using the sector-level agglomeration within countries, in order to have better estimates of δ_g by using sector-level data.

5 Conclusion

Estimations for the ad-valorem tax equivalents of distance effects are unrealistically high in magnitude and persistent over time in the literature, the so-called *distance puzzle*. Focusing on the *magnitude* dimension of the distance puzzle, this paper has shown analytically and proved empirically that the distance effects are overestimated, at least partly, due to ignoring the internal location of production of international exports that leads to the mismeasurement of distance. Using data on U.S. exports at the state-and-sector level, it has been shown that the mismeasurement of distance results in overestimation of distance effects by about twofold. The results are robust to the consideration of alternative estimation methodologies and data sets.

The mismeasurement of distance effects are found to be mostly due to the mismeasurement of distance elasticity of trade costs. In the presence of disaggregated data that consider the internal location of production, the obvious solution is to employ such data to avoid any mismeasurement. In the absence of disaggregated data, this paper proposes a *corrective distance index* created by using internal distance measures (i.e., the dispersion of economic activity) and international distance measures (i.e., the remoteness of the source country from the rest of the world). We employ this index on our data and show that it works well to avoid any mismeasurement for the median sector.

The paper is not without caveats, though. For instance, considering production locations of exports at the state level (rather than, say, at the plant-level) may be a restrictive approach, since, within each state, there is also a spatial dispersion of production locations. The same idea holds also for any other layer of aggregation; e.g., the mode of transportation. Nevertheless, the simple and clean message of this paper can be understood best by considering exactly the opposite case: If distance elasticity estimates are mismeasured even when state-level data are used, the true mismeasurement may be much higher when the exact spatial locations of production and exports would be considered. Moreover, to avoid the mismeasurement of distance effects at the sectoral level, one better needs internal distance measures created by using the agglomeration of sectors at the disaggregated level. In sum, with the corresponding spatial data, an investigation using production, trade, and location information would shed more light on the mismeasurement of distance effects.

References

[1] Anderson, J.E., Wincoop, E.V., 2003. Gravity with gravitas: a solution to the border puzzle. The American Economic Review 93 (1), 170–192.

[2] Anderson, J.E., Wincoop, E.V., 2004. Trade costs. Journal of Economic Literature 42, 691–751.

[3] Baier, S.L., Bergstrand, J.H., 2001. The growth of world trade: tariffs, transport costs and income similarity. Journal of International Economics 53 (1), 1–27.

[4] Berthelon, M and Freund, C. 2008. On the conservation of distance in international trade. Journal of International Economics 75, 310–320.

[5] Broda, C., Weinstein, D.E., 2006. Globalization and the gains from variety. Quarterly Journal of Economics 121 (2), 541–585.

[6] Carrere, C. and Schiff, M. 2005. On the geography of trade: distance is alive and well. Revue Economique 56, 1249–1274.

[7] Disdier, A.C. and Head, K. 2008. The puzzling persistence of the distance effect on bilateral trade. The Review of Economics and Statistics 90(1), 37–41.

[8] Eaton, J., Kortum, S., 2002. Technology, geography and trade. Econometrica 70 (5), 1741–1779.

[9] Engel, Charles. 2002. "Comment on Anderson and van Wincoop" in Brookings Trade Forum 2001. Susan Collins and Dani Rodrik, eds. Washington: Brookings Inst.

[10] Estevadeordal, Antoni; Brian Frantz and Alan M. Taylor. 2003. The Rise and Fall of World Trade, 1870–1939. Quartarly Journal of Economics 68(2), 359–408.

[11] Feenstra, R.C., 1994. New product varieties and the measurement of international prices. The American Economic Review 84 (1), 157–177.

[12] Felbermayr, G.J. and Kohler, W. 2006. Exploring the intensive and extensive margins of world trade. Review of World Economics 112: 642–674.

[13] Grossman, Gene. 1998. "Comment on Deardorff" in The Regionalization of the World Economy. J.A. Frankel, ed. Chicago: U. Chicago Press, pp. 33–57.

[14] Harrigan, J., 1996. Openness to trade in manufactures in the OECD. Journal of International Economics 40, 23–39.

[15] Head, K. and T. Mayer, forthcoming, "Gravity Equations: Workhorse, toolkit, cookbook," in Gopinath, Helpman, Rogoff eds. Handbook of International Economics Volume 4 eds.

[16] Head, K., Ries, J., 2001. Increasing returns versus national product differentiation as an explanation for the pattern of U.S.–Canada trade. The American Economic Review 91 (4), 858–876.

[17] Head, K., Mayer, T. and Ries, J. 2009 How remote is the offshoring threat? European Economic Review 53: 429–444.

[18] Henderson, D.J. and Millimet, D.L., 2008. Is gravity linear? Journal of Applied Econometrics, 23(2), 137-172.

[19] Hummels, D., (2001), "Toward a geography of trade costs", mimeo.

[20] Limao, N., Venables, A.J., 2001. Infrastructure, geographical disadvantage, transport costs and trade. World Bank Economic Review 15 (3), 451–479.

[21] Mayer, T. and Zignago, S. 2011. Notes on CEPII's distances measures: the GeoDist database. CEPII Working Paper 2011-25.

[22] Romalis, J., 2007. NAFTA's and CUSFTA's impact on international trade. The Review of Economics and Statistics 89 (3), 416–435.

[23] Santos Silva, J. M. C. and Tenreyro, S., 2006. The Log of Gravity. The Review of Economics and Statistics, 88(4), 641-658.

[24] Simonovska, I., Waugh, M.E., 2011. The Elasticity of Trade: Estimates and Evidence, NBER Working Paper No: 16796.

[25] Yilmazkuday, H. 2012. Understanding Interstate Trade Patterns. Journal of International Economics 86, 158-166.

[26] Yotov, Y. 2012. A simple solution to the distance puzzle in international trade. Economics Letters. 117: 794–798.

Table 1 - NAICS Manufacturing Sectors

NAICS Code	Definition
311	Food Manufacturing Products
312	Beverages & Tobacco Products
313	Textiles & Fabrics
314	Textile Mill Products
315	Apparel & Accessories
316	Leather & Allied Products
321	Wood Products
322	Paper
323	Printing, Publishing And Similar Products
324	Petroleum & Coal Products
325	Chemicals
326	Plastics & Rubber Products
327	Nonmetallic Mineral Products
331	Primary Metal Manufacturing
332	Fabricated Metal Products
333	Machinery, Except Electrical
334	Computer & Electronic Products
335	Electrical Equipment, Appliances & Components
336	Transportation Equipment
337	Furniture & Fixtures
339	Miscellaneous Manufactured Commodities

Table 2 – Elasticity of Substitution Estimates when Capital is Flexible (Benchmark Case)

NAICS Code	Using Production Data at the State-Level: 2002 Estimate (S.E.)	2002 R-Sqd.	2007 Estimate (S.E.)	2007 R-Sqd.	Using Production Data at the U.S. Level (For Robustness): 2002	2007
η_g for 311	3.16* (0.03)	0.984	3.43* (0.03)	0.985	3.13	3.38
η_g for 312	1.80* (0.07)	0.694	1.78* (0.09)	0.668	1.88	1.85
η_g for 313	4.81* (0.02)	0.999	4.00* (0.04)	0.995	4.48	3.71
η_g for 314	3.64* (0.01)	0.999	2.89* (0.02)	0.998	3.75	3.22
η_g for 315	3.44* (0.02)	0.998	3.25* (0.02)	0.999	3.26	3.29
η_g for 316	2.99* (0.06)	0.980	3.78* (0.06)	0.990	3.67	3.78
η_g for 321	5.49* (0.03)	0.997	5.04* (0.03)	0.995	5.37	4.91
η_g for 322	3.12* (0.03)	0.982	3.19* (0.03)	0.983	3.21	3.24
η_g for 323	3.24* (0.01)	0.999	3.10* (0.01)	0.999	3.23	3.12
η_g for 324	10.39* (0.03)	1.000	6.85* (0.04)	0.998	10.08	6.83
η_g for 325	2.81* (0.06)	0.925	3.17* (0.06)	0.961	2.41	2.50
η_g for 326	3.35* (0.01)	0.998	3.61* (0.01)	0.997	3.34	3.62
η_g for 327	2.99* (0.02)	0.993	2.95* (0.01)	0.995	3.02	2.90
η_g for 331	4.46* (0.04)	0.993	4.35* (0.03)	0.996	4.42	4.48
η_g for 332	3.37* (0.01)	0.999	3.27* (0.01)	0.998	3.36	3.27
η_g for 333	3.53* (0.04)	0.981	3.40* (0.02)	0.996	3.45	3.49
η_g for 334	3.03* (0.03)	0.988	2.75* (0.03)	0.982	2.85	2.62
η_g for 335	3.24* (0.02)	0.994	3.32* (0.03)	0.985	3.22	3.29
η_g for 336	4.26* (0.03)	0.995	4.28* (0.03)	0.993	4.12	4.14
η_g for 337	3.23* (0.02)	0.995	3.04* (0.03)	0.986	3.20	3.08
η_g for 339	2.59* (0.02)	0.988	2.44* (0.02)	0.988	2.64	2.42
γ for Pooled Sample	3.39* (0.06)	0.972	3.60* (0.08)	0.967	3.29	3.36

Notes: * indicates significance at the 1% level. S.E. stands for standard errors that have been calculated using the Delta method. Elasticity estimates are exactly identified when production data at the U.S. level are used at the right panel; hence, there are no standard errors or R-squared values to depict in this case.

Table 3 – Elasticity of Substitution Estimates when Capital is Fixed (For Robustness)

	Using Production Data at the State-Level				Using Production Data at the U.S. Level (For Robustness)	
	2002		2007			
NAICS Code	Estimate (S.E.)	R-Sqd.	Estimate (S.E.)	R-Sqd.	2002	2007
η_g for 311	2.94* (0.03)	0.982	3.20* (0.03)	0.983	2.91	3.14
η_g for 312	1.69* (0.06)	0.715	1.69* (0.07)	0.689	1.77	1.76
η_g for 313	4.17* (0.01)	0.999	3.60* (0.03)	0.995	3.94	3.36
η_g for 314	3.53* (0.01)	0.999	2.76* (0.01)	0.998	3.55	3.04
η_g for 315	3.33* (0.02)	0.996	3.14* (0.01)	1.000	3.15	3.17
η_g for 316	3.73* (0.08)	0.957	3.52* (0.06)	0.980	3.51	3.44
η_g for 321	4.83* (0.02)	0.997	4.38* (0.02)	0.996	4.69	4.28
η_g for 322	2.78* (0.02)	0.984	2.86* (0.02)	0.982	2.84	2.89
η_g for 323	2.88* (0.01)	0.999	2.72* (0.01)	0.999	2.86	2.74
η_g for 324	7.65* (0.03)	0.999	5.74* (0.03)	0.999	7.29	5.68
η_g for 325	2.52* (0.05)	0.924	2.89* (0.05)	0.957	2.20	2.31
η_g for 326	2.94* (0.01)	0.999	3.20* (0.01)	0.997	2.92	3.20
η_g for 327	2.60* (0.01)	0.997	2.50* (0.01)	0.994	2.59	2.46
η_g for 331	3.92* (0.02)	0.995	3.91* (0.03)	0.995	3.89	3.97
η_g for 332	3.04* (0.01)	0.999	2.98* (0.01)	0.998	3.03	2.98
η_g for 333	3.22* (0.03)	0.984	3.16* (0.02)	0.996	3.16	3.23
η_g for 334	2.68* (0.02)	0.988	2.48* (0.03)	0.974	2.57	2.33
η_g for 335	2.98* (0.02)	0.995	3.10* (0.03)	0.985	2.97	3.07
η_g for 336	3.86* (0.03)	0.995	3.93* (0.03)	0.993	3.72	3.81
η_g for 337	3.02* (0.01)	0.996	2.89* (0.02)	0.988	2.98	2.91
η_g for 339	2.40* (0.02)	0.986	2.27* (0.02)	0.985	2.43	2.25
γ for Pooled Sample	3.08* (0.06)	0.982	3.28* (0.07)	0.983	2.98	3.06

Notes: * indicates significance at the 1% level. S.E. stands for standard errors that have been calculated using the Delta method. Elasticity estimates are exactly identified when production data at the U.S. level are used at the right panel; hence, there are no standard errors or R-squared values to depict in this case.

Table 4 - Distance Effects obtained for 2002

	Estimation using State-and-Sector Level Data				Estimation using Sector Level Data			
	OLS	PPML	OLS	PPML	OLS	PPML	OLS	PPML
NAICS Code	$\eta_g \delta_g$ (S.E.)	$\eta_g \delta_g$ (S.E.)	δ_g (S.E.)	δ_g (S.E.)	$\gamma \delta'_g$ (S.E.)	$\gamma \delta'_g$ (S.E.)	δ'_g (S.E.)	δ'_g (S.E.)
311	0.18*(0.02)	0.18*(0.03)	0.06*(0.00)	0.06*(0.01)	0.52*(0.04)	0.41*(0.03)	0.16*(0.01)	0.12*(0.01)
312	0.16* (0.03)	0.16*(0.06)	0.08*(0.01)	0.09*(0.03)	0.64*(0.01)	0.51*(0.01)	0.19*(0.00)	0.15*(0.00)
313	0.19* (0.02)	0.19*(0.05)	0.04*(0.00)	0.04*(0.01)	0.59*(0.01)	0.44*(0.01)	0.18*(0.00)	0.13*(0.00)
314	0.13* (0.02)	0.14*(0.05)	0.04*(0.01)	0.04*(0.01)	0.60*(0.01)	0.46*(0.01)	0.18*(0.00)	0.14*(0.00)
315	0.13* (0.02)	0.15*(0.05)	0.04*(0.01)	0.04*(0.01)	0.61*(0.01)	0.47*(0.01)	0.19*(0.00)	0.14*(0.00)
316	0.13* (0.02)	0.13*(0.05)	0.04*(0.01)	0.04*(0.01)	0.61*(0.01)	0.46*(0.01)	0.18*(0.00)	0.14*(0.00)
321	0.22* (0.02)	0.23*(0.05)	0.04*(0.00)	0.04*(0.01)	0.60*(0.01)	0.45*(0.01)	0.18*(0.00)	0.14*(0.00)
322	0.21* (0.02)	0.21*(0.04)	0.06*(0.01)	0.06*(0.01)	0.57*(0.01)	0.43*(0.01)	0.17*(0.00)	0.13*(0.00)
323	0.17* (0.02)	0.17*(0.04)	0.05*(0.01)	0.05*(0.01)	0.57*(0.01)	0.44*(0.01)	0.17*(0.00)	0.13*(0.00)
324	0.19* (0.03)	0.17*(0.07)	0.02*(0.00)	0.02*(0.01)	0.62*(0.01)	0.48*(0.01)	0.19*(0.00)	0.14*(0.00)
325	0.16* (0.01)	0.16*(0.03)	0.07*(0.01)	0.07*(0.01)	0.52*(0.01)	0.40*(0.01)	0.16*(0.00)	0.12*(0.00)
326	0.19* (0.02)	0.19*(0.04)	0.06*(0.00)	0.06*(0.01)	0.55*(0.01)	0.42*(0.01)	0.17*(0.00)	0.13*(0.00)
327	0.17* (0.02)	0.17*(0.04)	0.06*(0.01)	0.06*(0.01)	0.58*(0.01)	0.44*(0.01)	0.18*(0.00)	0.13*(0.00)
331	0.23* (0.02)	0.22*(0.05)	0.05*(0.00)	0.05*(0.01)	0.59*(0.01)	0.43*(0.01)	0.18*(0.00)	0.13*(0.00)
332	0.17* (0.02)	0.17*(0.04)	0.05*(0.00)	0.05*(0.01)	0.54*(0.01)	0.42*(0.01)	0.17*(0.00)	0.13*(0.00)
333	0.15* (0.01)	0.15*(0.03)	0.04*(0.00)	0.04*(0.01)	0.48*(0.01)	0.38*(0.01)	0.15*(0.00)	0.12*(0.00)
334	0.15* (0.01)	0.15*(0.03)	0.05*(0.00)	0.05*(0.01)	0.47*(0.01)	0.38*(0.01)	0.14*(0.00)	0.12*(0.00)
335	0.13* (0.02)	0.14*(0.04)	0.04*(0.00)	0.04*(0.01)	0.54*(0.01)	0.42*(0.01)	0.16*(0.00)	0.13*(0.00)
336	0.20* (0.01)	0.20*(0.03)	0.05*(0.00)	0.05*(0.01)	0.51*(0.01)	0.40*(0.01)	0.15*(0.00)	0.12*(0.00)
337	0.16* (0.02)	0.16*(0.04)	0.05*(0.01)	0.05*(0.01)	0.60*(0.01)	0.45*(0.01)	0.18*(0.00)	0.14*(0.00)
339	0.10* (0.01)	0.11*(0.03)	0.04*(0.01)	0.04*(0.01)	0.53*(0.01)	0.41*(0.01)	0.16*(0.00)	0.12*(0.00)
MEDIAN	0.17* (0.02)	0.17*(0.04)	0.05*(0.00)	0.05*(0.01)	0.57*(0.01)	0.43*(0.01)	0.17*(0.00)	0.13*(0.00)
R-Squared	0.69	0.73	-	-	0.88	0.91	-	-

Notes: * indicates significance at the 1% level. S.E. stands for standard error. The standard errors of δ_g and δ'_g have been calculated using the Delta method. The sample size is 60,517 (3,819) in all regressions using state-and-sector level (sector-level) data.

Table 5 - Distance Effects obtained for 2007

	Estimation using State-and-Sector Level Data				Estimation using Sector Level Data			
	OLS	PPML	OLS	PPML	OLS	PPML	OLS	PPML
NAICS Code	$\eta_g\delta_g$ (S.E.)	$\eta_g\delta_g$ (S.E.)	δ_g (S.E.)	δ_g (S.E.)	$\gamma\delta'_g$ (S.E.)	$\gamma\delta'_g$ (S.E.)	δ'_g (S.E.)	δ'_g (S.E.)
311	0.19*(0.01)	0.19*(0.03)	0.06*(0.00)	0.06*(0.01)	0.47*(0.04)	0.35*(0.03)	0.14*(0.01)	0.10*(0.01)
312	0.16*(0.02)	0.17*(0.06)	0.09*(0.01)	0.09*(0.03)	0.59*(0.01)	0.45*(0.01)	0.18*(0.00)	0.13*(0.00)
313	0.15*(0.02)	0.15*(0.05)	0.04*(0.01)	0.04*(0.01)	0.55*(0.01)	0.38*(0.01)	0.16*(0.00)	0.11*(0.00)
314	0.13*(0.02)	0.13*(0.05)	0.04*(0.01)	0.04*(0.01)	0.54*(0.01)	0.39*(0.01)	0.16*(0.00)	0.12*(0.00)
315	0.15*(0.02)	0.16*(0.05)	0.05*(0.01)	0.05*(0.01)	0.55*(0.01)	0.40*(0.01)	0.16*(0.00)	0.12*(0.00)
316	0.10*(0.02)	0.11*(0.05)	0.03*(0.01)	0.03*(0.01)	0.55*(0.01)	0.40*(0.01)	0.16*(0.00)	0.12*(0.00)
321	0.23*(0.02)	0.24*(0.05)	0.05*(0.00)	0.05*(0.01)	0.54*(0.01)	0.39*(0.01)	0.16*(0.00)	0.11*(0.00)
322	0.20*(0.02)	0.20*(0.04)	0.06*(0.01)	0.06*(0.01)	0.52*(0.01)	0.37*(0.01)	0.16*(0.00)	0.11*(0.00)
323	0.14*(0.02)	0.15*(0.04)	0.05*(0.01)	0.05*(0.01)	0.53*(0.01)	0.38*(0.01)	0.16*(0.00)	0.11*(0.00)
324	0.22*(0.03)	0.21*(0.07)	0.03*(0.00)	0.03*(0.01)	0.57*(0.01)	0.41*(0.01)	0.17*(0.00)	0.12*(0.00)
325	0.15*(0.01)	0.15*(0.03)	0.06*(0.01)	0.06*(0.01)	0.46*(0.01)	0.34*(0.01)	0.14*(0.00)	0.10*(0.00)
326	0.21*(0.02)	0.20*(0.04)	0.06*(0.00)	0.06*(0.01)	0.48*(0.01)	0.35*(0.01)	0.14*(0.00)	0.11*(0.00)
327	0.19*(0.02)	0.19*(0.04)	0.06*(0.01)	0.06*(0.01)	0.54*(0.01)	0.37*(0.01)	0.16*(0.00)	0.11*(0.00)
331	0.20*(0.02)	0.20*(0.04)	0.05*(0.00)	0.04*(0.01)	0.52*(0.01)	0.36*(0.01)	0.15*(0.00)	0.11*(0.00)
332	0.17*(0.02)	0.17*(0.04)	0.05*(0.00)	0.05*(0.01)	0.48*(0.01)	0.34*(0.01)	0.14*(0.00)	0.10*(0.00)
333	0.16*(0.01)	0.15*(0.03)	0.04*(0.00)	0.04*(0.01)	0.42*(0.01)	0.31*(0.01)	0.13*(0.00)	0.09*(0.00)
334	0.15*(0.01)	0.15*(0.03)	0.06*(0.00)	0.06*(0.01)	0.42*(0.01)	0.32*(0.01)	0.12*(0.00)	0.09*(0.00)
335	0.13*(0.01)	0.14*(0.03)	0.04*(0.00)	0.04*(0.01)	0.47*(0.01)	0.35*(0.01)	0.14*(0.00)	0.10*(0.00)
336	0.18*(0.01)	0.18*(0.03)	0.04*(0.00)	0.04*(0.01)	0.44*(0.01)	0.33*(0.01)	0.13*(0.00)	0.10*(0.00)
337	0.18*(0.02)	0.18*(0.04)	0.06*(0.01)	0.06*(0.01)	0.54*(0.01)	0.39*(0.01)	0.16*(0.00)	0.12*(0.00)
339	0.12*(0.01)	0.12*(0.03)	0.05*(0.01)	0.05*(0.01)	0.47*(0.01)	0.35*(0.01)	0.14*(0.00)	0.10*(0.00)
MEDIAN	0.16*(0.02)	0.17*(0.04)	0.05*(0.01)	0.05*(0.01)	0.52*(0.01)	0.37*(0.01)	0.16*(0.00)	0.11*(0.00)
R-Squared	0.70	0.74	-	-	0.86	0.90	-	-

Notes: * indicates significance at the 1% level. S.E. stands for standard error. The standard errors of δ_g and δ'_g have been calculated using the Delta method. The sample size is 60,517 (3,819) in all regressions using state-and-sector level (sector-level) data.

Table 6 - Source of Bias for 2002

		Bias in the Special Case of:							
	Bias	$P_g^s = P_g$	$D_d^s = D_d$	$\delta_g' = \delta_g$	$\sum \beta_{d,g}^s = 1$ or $\eta \to \infty$				
NAICS Code	(Benchmark)	(1)	(2)	(3)	(4)	(2)&(3)	(1)&(4)	(1)&(2)&(3)	(1)&(2)&(3)&(4)
311	2.17	2.05	2.14	1.26	1.06	1.23	0.94	1.11	0.00
312	6.37	2.71	6.33	5.40	4.68	5.36	1.02	1.70	0.00
313	1.60	1.80	1.58	0.38	1.05	0.36	1.25	0.56	0.00
314	2.17	2.01	2.15	0.86	1.49	0.84	1.33	0.68	0.00
315	2.56	2.10	2.54	1.26	1.78	1.24	1.32	0.78	0.00
316	2.22	2.03	2.20	0.89	1.55	0.87	1.35	0.68	0.00
321	1.34	1.74	1.31	0.08	0.87	0.06	1.28	0.46	0.00
322	2.37	1.98	2.33	1.39	1.40	1.35	1.01	0.97	0.00
323	2.37	2.04	2.35	1.27	1.46	1.25	1.13	0.92	0.00
324	0.88	1.70	0.87	-0.63	0.70	-0.64	1.52	0.18	0.00
325	3.63	2.65	3.60	2.82	1.84	2.78	0.85	1.80	0.00
326	2.05	2.00	2.02	1.06	1.07	1.03	1.02	0.98	0.00
327	2.69	2.20	2.66	1.60	1.61	1.57	1.12	1.08	0.00
331	1.63	1.77	1.60	0.50	1.01	0.47	1.15	0.62	0.00
332	2.11	2.10	2.09	1.07	1.09	1.05	1.07	1.03	0.00
333	1.54	2.15	1.52	0.63	0.33	0.60	0.94	1.21	0.00
334	2.20	2.49	2.18	1.38	0.56	1.35	0.85	1.64	0.00
335	2.35	2.24	2.33	1.24	1.24	1.22	1.13	1.11	0.00
336	1.31	1.84	1.28	0.35	0.45	0.33	0.98	0.86	0.00
337	2.64	2.14	2.61	1.45	1.72	1.42	1.22	0.92	0.00
339	3.46	2.65	3.44	2.35	1.94	2.33	1.12	1.52	0.00
MEDIAN	2.20	2.05	2.18	1.24	1.24	1.22	1.12	0.97	0.00

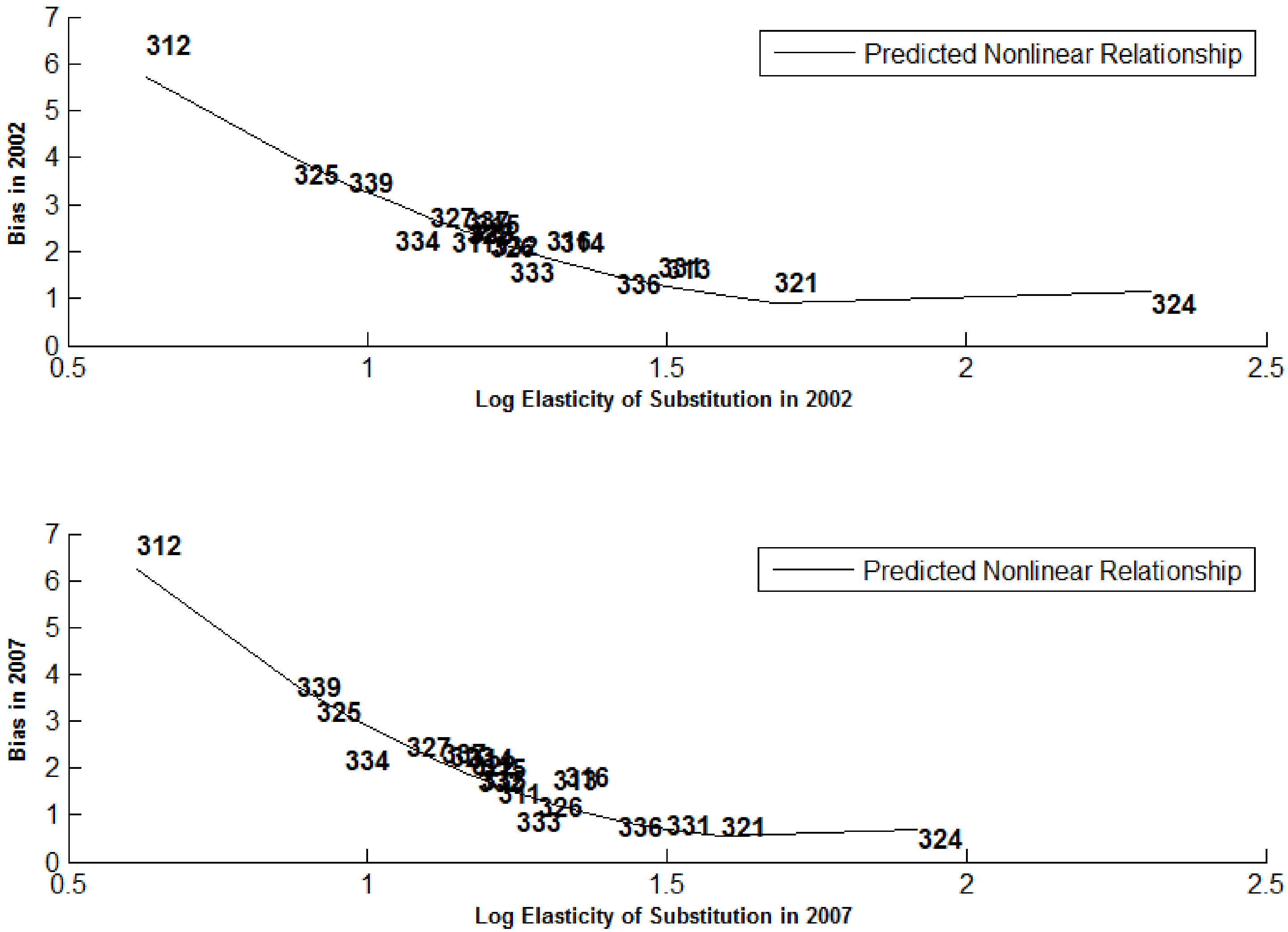
Notes: The bias represents the average percentage deviation (divided by 100) of the OLS-estimated distance effects obtained by sector level data from the aggregated version of the OLS-estimated distance effects obtained by state-and-sector level data.

Table 7 - Source of Bias for 2007

		Bias in the Special Case of:							
	Bias	$P_g^s = P_g$	$D_d^s = D_d$	$\delta_g' = \delta_g$	$\sum \beta_{d,g}^s = 1$ or $\eta \to \infty$				
NAICS Code	(Benchmark)	(1)	(2)	(3)	(4)	(2)&(3)	(1)&(4)	(1)& (2)&(3)	(1)&(2)&(3)&(4)
311	1.45	1.78	1.42	0.69	0.46	0.66	0.79	0.99	0.00
312	6.74	2.61	6.69	5.94	4.98	5.89	0.85	1.76	0.00
313	1.73	1.84	1.70	0.63	1.01	0.60	1.12	0.71	0.00
314	2.19	1.95	2.16	1.11	1.34	1.08	1.10	0.84	0.00
315	1.98	1.85	1.96	0.92	1.22	0.90	1.08	0.77	0.00
316	1.78	1.89	1.77	0.56	1.14	0.55	1.24	0.65	0.00
321	0.74	1.56	0.71	-0.27	0.22	-0.30	1.04	0.52	0.00
322	2.04	1.83	2.01	1.20	1.09	1.16	0.88	0.95	0.00
323	2.24	1.99	2.21	1.23	1.27	1.21	1.03	0.96	0.00
324	0.47	1.53	0.45	-0.75	0.18	-0.77	1.24	0.29	0.00
325	3.18	2.40	3.15	2.50	1.49	2.46	0.72	1.69	0.00
326	1.16	1.68	1.13	0.38	0.28	0.35	0.81	0.88	0.00
327	2.44	2.04	2.41	1.59	1.30	1.55	0.89	1.15	0.00
331	0.75	1.61	0.73	-0.22	0.14	-0.24	1.00	0.61	0.00
332	1.68	1.89	1.65	0.88	0.61	0.85	0.83	1.07	0.00
333	0.83	1.94	0.80	0.10	-0.36	0.07	0.75	1.19	0.00
334	2.15	2.50	2.12	1.55	0.28	1.52	0.63	1.87	0.00
335	1.74	2.00	1.72	0.84	0.66	0.82	0.92	1.08	0.00
336	0.72	1.66	0.70	-0.06	-0.13	-0.08	0.81	0.85	0.00
337	2.29	1.92	2.26	1.38	1.32	1.35	0.94	0.97	0.00
339	3.71	2.60	3.68	2.89	1.95	2.86	0.85	1.76	0.00
MEDIAN	1.45	1.78	1.42	0.69	0.46	0.66	0.79	0.99	0.00

Notes: The bias represents the average percentage deviation (divided by 100) of the OLS-estimated distance effects obtained by sector level data from the aggregated version of the OLS-estimated distance effects obtained by state-and-sector level data.

Figure 1 - The Bias versus the Elasticity of Substitution across Varieties



Notes: The correlation coefficient for 2002 (2007) is -0.78 (-0.85). For each year, the predicted nonlinear relationship has been obtained by running a regression of the bias on the log elasticity of substitution and the squared log elasticity of substitution, including a constant. For the year of 2002 (2007), the coefficient in front of log elasticity of substitution is significantly estimated as -11.48 (-16.57), while the coefficient in front of squared log elasticity of substitution is significantly estimated as 2.98 (4.86), with an R-squared value of 0.90 (0.92).