ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Larch, Mario; Norbäck, Pehr-Johan; Urban, Dieter M.

Conference Paper

Globalization and the Distance Puzzle

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2010: Ökonomie der Familie - Session: Trade, Gravity and Distance, No. A4-V3

Provided in Cooperation with:
Verein für Socialpolitik / German Economic Association

Suggested Citation: Larch, Mario; Norbäck, Pehr-Johan; Urban, Dieter M. (2010) : Globalization and the Distance Puzzle, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2010: Ökonomie der Familie - Session: Trade, Gravity and Distance, No. A4-V3, Verein für Socialpolitik, Frankfurt a. M.

This Version is available at:
https://hdl.handle.net/10419/37442

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Globalization and the Distance Puzzle

Mario Larch, Pehr-Johan Norbäck, and Dieter Urban

First draft, 19 February, 2009.
This draft, 30 January, 2010.

Abstract

In a recent meta study, Disdier and Head (2008) summarize 103 empirical studies that find that the (absolute value of the) distance coefficient in a gravity estimation of bilateral exports does not fall over time. Including zero-trade flows is suggested as a solution to this "distance-puzzle" (Felbermayr and Kohler, 2006). However, as shown by Helpman, Melitz, and Rubinstein (2008), HMR henceforth, for cross-section data, including zero-trade flows will rather increase - and not decrease - estimated distance coefficients, due to omitted firm-level heterogeneity. Using industry level data over 28 industries for a time span ranging from 1978 to 2003, we first confirm earlier results that the distance coefficient increases over time using OLS. We then show that the bias from not controlling for firm heterogeneity in OLS rises over time, whereas the non-linear estimation of HMR leads to a declining distance coefficient over time. Our estimates are also shown to capture reductions in trade costs related to distance, since the estimated distance coefficients are strongly correlated with the time pattern in freight costs reported by Hummels (2007).

Keywords: Distance puzzle, gravity estimation, zero trade flows, firm heterogeneity

JEL classifications: F13, F23

Mario Larch: ifo Institute for Economic Research and CESifo, Poschingerstraße 5, 81679 Munich, Germany. E-mail: larch@ifo.de.

Pehr-Johan Norbäck: Research Institute of Industrial Economics (IFN), P.O. Box 5501, SE-114 85 Stockholm, Sweden. E-mail: pehr-johan.norback@ifn.se. Home page: www.ifn.se/pjn.

Dieter Urban: Dept. of Business and Economics, RWTH Aachen University and CESifo, 52062 Aachen, Germany. E-mail: dieter.urban@wiwi.rwth-aachen.de.

1. Introduction

> *"From the telegraph to the Internet, every new communications technology has promised to shrink the distance between people, to increase access to information, and to bring us ever closer to the dream of a perfectly efficient, frictionless global market." (Thomas Friedman, The World is Flat, 2005, p. 204)*

The many facets of globalization like the increased trade in final goods, intermediate inputs and services, or the increased international mobility of capital and labor, are perceived to bring countries closer together, reducing the significance of distance. However, gravity estimations regressing bilateral trade on distance, inter alia, tell us the opposite. Disdier and Head (2008) undertake a meta analysis of the magnitude of the distance coefficient based on 103 empirical studies and find (i) the mean effect of the distance coefficient is about -0.9 across studies, (ii) the negative impact of distance on trade rose around the middle of the century and has remained persistently high since then.[1]

The lack of econometric evidence of distance being less of an impediment for trade over time is puzzling, given the reduction in the indices on transport costs for shipping and air travel (World Trade Report, 2008), mirorring significant productivity increases in logistics over the last decades. It is also puzzling given the large reductions of tariffs due to numerous free trade agreements and trade liberalization rounds through the WTO (see, e.g., Anderson and van Wincoop, 2004).

The main purpose of this paper is to solve this "distance puzzle". In doing so, we draw on a recently developed gravity equation estimator from Helpman, Melitz and Rubinstein (2008), HMR henceforth. They derive a gravity equation for aggregate bilateral trade from a heterogeneous firm model and show that OLS estimates suffer from two sources of bias. First, there is a sample selection bias, because bilateral trade is measured as logarithm and zero values of bilateral trade turn into missing values, yielding a sample selection. If there are unobservable trade barriers, then they are more likely to cause further distant trading partners to have zero trade flows and to drop out of the sample, leaving behind the observations with positive unobservable trade barriers at large distances. As a result, the error term is positively correlated with distance, causing an upward bias in the distance coefficient, which implies too small a distance coefficient in absolute value. Second, there is an omitted variable bias from ignoring that firms are heterogeneous in productivity. If the share of exporting firms is not included as a control in the gravity estimation, then it appears in the regression error, causing a negative correlation between error and distance, because there are less exporters to more distant destinations. Hence, the distance coefficient is downward biased through omitting a control on the share of exporting firms, i.e. the value of the distance coefficient is too large in absolute value. As these two biases work in opposite directions, the overall change of the bias from OLS estimates is ambiguous.

[1] This paper provides also a good collection of references for the "distance puzzle". Hence, we dispense here with a discussion of all relevant papers and with providing all references. Rather, we selectively discuss only papers, which are directly related to our paper in their methodology.

Using bilateral country data for the year 1986, HMR find that their estimated distance coefficient represents a drop of roughly by a third as compared to OLS, showing the bias from omitted variables must dominate the sample selection bias. To explore the evolution over time, we apply the HMR estimator on world trade in 28 different manufacturing industries over the time period from 1978 until 2003, estimating a gravity equation for each industry and year separately, running running more than 4000 regressions altogether.

We first find that the HMR-estimator on avarge over all industries does produce distance coefficients which become smaller over time (in absolute value). These estimated coefficients are also shown to capture reductions in trade costs related to distance, since they are strongly correlated with the time pattern in freight costs reported by Hummels (2007) and Brun, Carrère, Guillaumont and de Melo (2005). Applying a Gamma pseudo-maximum-likelihood (GPML) estimator, which was recently suggested by Santos Silva and Tenreyro (2006), albeit augmented by the correction factors of HMR, also gives a very similar time pattern of the estimated distance coefficients, giving additional confidence in our estimates.

Having shown that the HMR estimator does produce distance coefficients which decrease over time, we then compare the outcome with OLS. We first confirm the finding of HMR that OLS produces larger distance coefficients (in absolute value). However, we also show the distance coefficients increase over time. Hence, the distance puzzle arises due to the fact that the bias of OLS increases over time.

To explain the increasing bias of OLS, we derive formally how the bias of the OLS distance coefficient evolves over time if the true data generating process is the HMR model and true trade cost related to distance decrease during globalization through, for instance, improved transport and communication technologies or trade liberalization.

We show that the upward bias through sample selection bias decreases over time. Intuitively, as trade cost decrease, ever less country pairs have zero trade flows and eventually all countries trade with each other. But then the sample selection bias has disappeared. We also show that the downward bias from omitting the number of exporting firms also becomes smaller (in absolute value). Intuitively, at lower trade costs related to a given distance, most firms will export producing a smaller downward bias. Since both biases decrease (in absolute value) globalization has an ambiguous effect on the bias of OLS. However, since our estimates show the bias of OLS increases over time, it must be that upward bias from sample selection decreases less than the decrease of the upward bias from omitting the number of exporting firms. If the HMR model is the data generating process, an implication is then that the distance puzzle actually arises by firm heterogenity becoming relatively less important over time.

For future work, we also suggest a linearization of the HMR estimator, which is easy to implement with standard econometric programs, and show that such a simplified estimator performs just as well. We also show that a Heckman estimator, while deviating from the HMR estimates, also produces a smaller distance coefficient and an increasing difference to the OLS estimates over time. This arises since the Hekman estimator uses a correction factor that is essentially a non-linear but monotonic function of the correction factors of the HMR estimator.

Alternative attempts to solve the distance puzzle stem from Felbermayr and Kohler (2006), using Tobit estimates to take zero trade flows into account.[2] Other studies explain why the substitution elasticity may have been rising over time (Glaeser and Kohlhase, 2004, Krautheim, 2008, Lawless and Whelan, 2007, Berthelon and Freund, 2008), possibly overcompensating the fall in trade costs, which both determine the distance coefficient in theory. Duranton and Storper (2008) provide an alternative model to rationalize rising overall trade costs besides falling transport costs. They assume vertically linked industries in which the quality of inputs is not contractible and where providing a given level of quality to suppliers becomes more costly with distance. Their main finding is that lower transport costs imply that higher quality inputs are traded in equilibrium, and the effect of this higher quality is that trade costs increase. None of them discuss the role of the omitted variable problem of firm heterogeneity in creating an increasing bias over time.

The remainder of the paper is organized as follows. Section 2 derives the gravity equation controlling for zero trade flows and firm-level heterogeneity following HMR in subsection 2.1., whereas we calculate the biases of OLS estimates in subsection 2.2. Section 3 presents our estimation equation in subsection 3.1., describes the data in subsection 3.2., and gives the results in subsection 3.3. The last section concludes.

2. Theory

2.1. Deriving the gravity equation from Helpman, Melitz and Rubinstein (2008)

The HMR model is a multi-country monopolistic competition partial equilibrium model with heterogeneous firms and identical consumers with CES "love-of-variety" utility functions à al Dixit and Stiglitz (1977). Since we have bilateral industry rather than country data of trade, we will add the assumption of multiple sectors in the world economy, each characterized by monopolistic competition. There are N_{ih} firms in a sector h of country i each producing a differentiated variety l. For ease of notation, we will drop the subscript h for sector whenever obvious. With a substitution elasticity between any two varieties $\varepsilon > 1$, the demand $x_{ij}(l)$ for a variety l, consumed in country i and produced in country j is

$$x_{ij}(l) = \frac{p_{ij}(l)^{-\varepsilon}}{P_i^{1-\varepsilon}} \mu_i Y_i, \tag{2.1}$$

where $p_{ij}(l)$ is the associated price, $P_i = \left[\int_{\mathcal{B}_i} p_{ij}(l)^{1-\varepsilon} dl\right]^{\frac{1}{1-\varepsilon}}$ is the price index on the set of $\mathcal{B}_i$ symmetric domestic and imported differentiated goods consumed in country i, Y_i is the income

[2] There is ample evidence from microdata for particular countries that the extensive margin matters. Bernard, Jensen and Schott (2006) use firm-level data and the entry and exit of firms into and out of exporting (extensive margin) from the export volumes of exporting firms (intensive margin). They find that reduction in trade costs may increase industry productivity through changes on the extensive margin. Hummels and Klenow (2005) use disaggregated product-level data to distinguish between the variety dimension (extensive margin) and the quality as well as the quantity dimension (intensive margins). One of their main results is that adverse terms-of-trade effects occur more frequently if growth takes place mainly at the extensive margin. Similarly, Baldwin and Harrigan (2007) use product-level data on bilateral U.S. exports demonstrating that large part of potential export flows are zero, and showing that the incidence of these zero export flows is strongly correlated with distance and importing country size. Hilberry and Hummels (2008) analyze trade at the 5-digit zip codes and decompose the extensive and intensive margins of shipments. Their main finding is that distance reduces aggregate trade values primarily by reducing the number of commodities shipped and the number of establishments shipping. However, the extensive margins are important over very short distances.

in country i and μ_i is the share of income spent on this sector.

A firm l in country j produces one unit of output at cost $c_j a$, where c_j is the minimum cost of a bundle of inputs which is country- and sector specific, and where $a(l) = a$ is a firm-specific input coefficient, where $1/a$ is the firm's productivity. Shipping goods across borders involves iceberg trade costs, which implies that $\tau_{ij} > 1$ units of output needs to be shipped from country j to country $i \neq j$ in order for one unit to arrive. Delivering to home country customers involves no trade costs, i.e. $\tau_{ii} = 1$. Exporting across borders are also associated with country- and sector specific fixed export costs $f_{ij} c_j$, i.e. $f_{ii} = 0$ and $f_{ij} > 0$ for $j \neq i$.

The operating profit from producing a variety l in country j and selling it to country i is then $\pi_{ij}(l) = [p_{ij}(l) - \tau_{ij} a(l) c_{jh}]\, x_{ij}(l) - f_{ij} c_j$. From (2.1), this implies a firm with productivity $1/a$ producing in country j for exports to country i will charge the price:

$$p_{ij}^*(a) = \frac{1}{\alpha} \tau_{ij} a c_j, \tag{2.2}$$

where $\frac{1}{\alpha} = \frac{\varepsilon}{\varepsilon - 1}$ is the standard mark-up. It also follows that domestic consumers are priced at $p_{jj}^*(a) = \frac{1}{\alpha} a c_j$.

The assumption of absence of fixed costs in home sales operations and fixed set-up costs incurred in exporting operations, implies that only a fraction of country $j's$ N_j firms will export to country i. To characterize exporters, define the reduced-form operating profit for country j exporters as $\pi_{ij}^*(a) = \pi_{ij}^*(p_{ij}^*(a))$, or:

$$\pi_{ij}^*(a) = (1/a)^{\varepsilon - 1} (1 - \alpha) \left(\frac{\tau_{ij} c_j}{\alpha P_i} \right)^{1-\varepsilon} \mu_i Y_i - f_{ij} c_j. \tag{2.3}$$

Let firm-productivity $1/a$ be characterized from the cumulative distribution $G(a)$ with density $g(a)$ over the finite support $a \in [a_L, a_H]$. The cut-off productivity for being an exporter to country i based in country j, $1/a_{ij}$, is then determined from the zero-profit condition, $\pi_{ij}^*(a_{ij}) = 0$, or:

$$a_{ij} = \left[(1 - \alpha) \frac{\mu_i Y_i}{f_{ij} c_j} \right]^{\frac{1}{\varepsilon - 1}} \frac{\alpha P_i}{\tau_{ij} c_j}. \tag{2.4}$$

In Figure 1(iii) we show operating profits for a country j firm as a function of the firm-specific input-coefficient a in exporting $\pi_{ij}^*(a)$ and home sales $\pi_{jj}^*(a)$. Operating profits decrease in a, and hence increase in productivity $1/a$, as shown in Figure 1(ii). Thus, firms in country j can only recover export fixed costs and export to country i if their productivity is sufficiently high, $1/a > 1/a_{ij}$. Firms with productivity $1/a < 1/a_{ij}$ will only sell on the domestic market.

Trade between country i and j can now be characterized as follows: From (2.1) and (2.2), the export revenue for a country j firm is $p_{ij}^*(l) x_{ij}^*(l) = (1/a)^{\varepsilon - 1} \left(\frac{\tau_{ij} c_j}{\alpha P_i} \right)^{1-\varepsilon} \mu_i Y_i$. As shown in Figure 1(i), the number of firms endowed with a productivity $1/a$ is $N_j g(a)$. Hence, the aggregate imports of county i from country j, M_{ij}, are:

$$M_{ij} = \int_{a_L}^{a_{ij}} (1/a)^{\varepsilon - 1} \left(\frac{\tau_{ij} c_j}{\alpha P_i} \right)^{1-\varepsilon} \mu_i Y_i N_j g(a) da. \tag{2.5}$$

Define V_{ij} as a term indicating firm-heterogeneity:

$$V_{ij} = \begin{cases} \int_{a_L}^{a_{ij}} a^{1-\varepsilon} g(a) da & \text{for} \quad 1/a \geq 1/a_{ij}, \\ 0, & \text{otherwise.} \end{cases} \tag{2.6}$$

Combining (2.5) and (2.6), imports from country j to country i can be written:

$$M_{ij} = \left(\frac{c_j \tau_{ij}}{\alpha P_i} \right)^{1-\varepsilon} \mu_i Y_i N_j V_{ij}. \tag{2.7}$$

To obtain an estimation equation from (2.7), HMR proceed in several steps. First, they specify trade cost to include an error component as follows:

$$\tau_{ij}^{\varepsilon-1} = D_{ij}^{\gamma} \exp^{-u_{ij}}, \tag{2.8}$$

where D_{ij} is the distance between source and destination country, γ is the trade cost per unit of distance, which may vary across industries, and u_{ij} is a trade barrier unobservable to econometricians but observable to the firms. The latter is assumed to be log normally distributed such that $u_{ij} \sim N\left(0, \sigma_u^2\right)$.[3]

Second, they assume a Pareto distribution $G(a) = a^k/(a_H^k - a_L^k)$, where $k > (\varepsilon - 1)$ is the shape parameter and show that V_{ij} in (2.6) takes the form $V_{ij} = \psi W_{ij}$, where:

$$W_{ij} = \max\left\{ \left(\frac{a_{ij}}{a_L} \right)^{k-\varepsilon+1} - 1, 0 \right\} \tag{2.9}$$

and $\psi = \frac{k a_L^{k-\varepsilon+1}}{(k-\varepsilon+1)(a_H^k - a_L^k)}$ is a constant. Note that trade is observed whenever $W_{ij} > 0$, that is, if $a_{ij} > a_L$, implying that the cut-off productivity is smaller than the productivity of the most efficient firm in the industry, $1/a_{ij} < 1/a_L$. From Figure 1(i), note that the number of exporting firms $\int_{a_{Lh}}^{a_{ij}} N_j g(a) da$ is increasing in a_{ij}. Importantly, from (2.9) it then follows that the number of country j exporters is increasing in W_{ij}, since W_{ij} is increasing in a_{ij}.[4] Note also that a reduction of trade cost γ increases thus the number of exporting firms, since it lowers the threshold productivity level $1/a_{ij}$.

Inserting (2.8) and (2.9) into (2.7) and taking logs, HMR obtain the gravity estimation equation in logs:

$$m_{ij} = \beta_0 + \lambda_j + \chi_i - \gamma d_{ij} + \omega_{ij} + u_{ij}, \tag{2.10}$$

where importer country-fixed effects χ_i contain $\chi_i = (\varepsilon - 1)p_i + \ln \mu_i + y_i$, and exporter country fixed effects λ_j contain $\lambda_j = -(\varepsilon - 1) \ln c_j + n_j$ and the number of exporters are captured by

$$\omega_{ij} = \ln \left[\left(\frac{a_{ij}}{a_L} \right)^{k-\varepsilon+1} - 1 \right]. \tag{2.11}$$

[3] Choosing $\tau_{ij}^{\varepsilon-1}$ rather than τ_{ij} to be related to distance D_{ij} will ensure that the distance coefficient will exclusively measure trade costs rather than being a function of trade cost and the substitution elasticity of demand.

[4] As illustrated in Figure 1, note that positive trade $M_{ij} > 0$ requires $a_{ij} > a_L$. But then $W_{ij} = \max\left\{ \left(\frac{a_{ij}}{a_L} \right)^{k-\varepsilon+1} - 1, 0 \right\} = \left(\frac{a_{ij}}{a_L} \right)^{k-\varepsilon+1} - 1$. Thus, we have $\omega_{ij} \equiv \ln W_{ij} = \ln \left[\left(\frac{a_{ij}}{a_L} \right)^{k-\varepsilon+1} - 1 \right]$.

Note that the term ω_{ij} is the only one anew in the gravity equation compared to Anderson and van Wincoop (2003).

Estimation of (2.10) is hampered by two problems. First, it is estimated only on data with positive trade flows, since the dependent variable, the log of trade volume m_{ij}, is not defined for zero import values $M_{ij} = 0$. Second, there is an omitted variable problem through ω_{ij}, which captures the number of exporting firms in country j, an information typically not available for gravity estimations on a world trade dataset.[5]

HMR note that both problems are related to the extensive margin of trade. Rearranging (2.4), they then define an auxiliary variable Z_{ij}:

$$\begin{aligned} \left(\frac{a_{ij}}{a_L}\right)^{\varepsilon-1} &= \frac{(1-\alpha)\left(\frac{\tau_{ij}c_j a_L}{\alpha P_i}\right)^{1-\varepsilon}\mu_i Y_i}{c_j f_{ij}} \\ &\equiv Z_{ij}. \end{aligned} \tag{2.12}$$

As illustrated in Figure 1(i), changes in Z_{ij} will indicate both, changes in the number of exporting firms through the cut-off a_{ij}, as well as zero trade links. From Figure 1(i), we note that there will be no exports if $a_{ij} < a_L$, which is equivalent to the condition $Z_{ij} < 1$. In contrast, trade is observed when $a_{ij} > a_L$, which implies $Z_{ij} > 1$.

Now, we can express the omitted variable ω_{ij} by inserting equation (2.12) into (2.11)

$$\omega_{ij} = \ln\left[e^{\tilde{\delta} z_{ij}} - 1\right], \tag{2.13}$$

where $\tilde{\delta} = (k - \varepsilon + 1) / (\varepsilon - 1)$ and $z_{ij} = \ln Z_{ij}$.The estimation strategy of HMR in the following steps is to obtain an estimate of the expected value of the omitted variable ω_{ij} by estimating an expected value of the auxiliary variable z_{ij}.

To obtain another estimation equation, another error term is introduced by decomposing the fixed trade costs f_{ij} as follows:

$$f_{ij} = \exp^{(\phi_{EX,j} + \phi_{IM,i} + \kappa\phi_{ij} - \nu_{ij})}, \tag{2.14}$$

where $\nu_{ij} \sim N\left(0, \sigma_\nu^2\right)$ and $\phi_{EX,j}$ is a measure of fixed export cost common across all export destinations, $\phi_{IM,i}$ a fixed trade barrier imposed by the importing country on all exporters, and ϕ_{ij} is an observed measure of any additional country-pair specific fixed trade costs.[6]

Taking logs of Z_{ij} in (2.12) and using (2.8) and (2.14), HMR obtain an equation for a latent variable $z_{ij} \equiv \ln Z_{ij}$:

$$\begin{aligned} z_{ij} &= \gamma_0 + \xi_j + \zeta_i - \gamma d_{ij} - \kappa\phi_{ij} + \eta_{ij} \\ &= E\left[z_{ij} \left| d_{ij,}\xi_j, \zeta_i, \phi_{ij}\right.\right] + \eta_{ij}, \end{aligned} \tag{2.15}$$

[5] Flam and Nordström (2008) have recently included a proxy variable for ω_{ij}, which is available for Swedish exports. However, they did not estimate the distance coefficient over time, which is the focus of this paper.

[6] ϕ_{ij} takes the role of an instrument in the empirical implementation and is assumed to be statistically independent of distance and country dummies. This assumption is not innocuous. Dropping this assumption may potentially cause additional bias of unknown direction, which is not taken into consideration in HMR.

where $\eta_{ij} = u_{ij} + \nu_{ij}$, $\zeta_i = (\varepsilon - 1) p_i + y_i + \ln \mu_i - \phi_{IM,i}$ is an importer fixed effect, $\xi_j = -\varepsilon \ln c_j + \phi_{EX,j}$ is an exporter fixed effect and σ_η^2 is the variance of η_{ij}. While the latent variable z_{ij} cannot be observed, one can observe if trade takes place. Thus, an indicator variable $T_{ij} = 1[z_{ij} > 0]$ can be defined from which one obtains an estimation equation for the probability of strictly positive exports:

$$\begin{aligned} \Pr\left(T_{ij} = 1 \left| d_{ij}, \xi_j^*, \zeta_i^*, \phi_{ij}\right.\right) &= \Phi\left(z_{ij}^* > 0\right) \\ &= \Phi\left(\gamma_0^* + \xi_j^* + \zeta_i^* - \gamma^* d_{ij} - \kappa^* \phi_{ij} > -\eta_{ij}\right) \\ &= E\left[z_{ij}^* \left| d_{ij}, \xi_j^*, \zeta_i^*, \phi_{ij}\right.\right], \end{aligned} \tag{2.16}$$

where $\Phi(.)$ is the cumulative distribution function of the unit normal distribution and every starred coefficient represents the original coefficient divided by σ_η.[7]

One can now estimate in a first stage (2.16) by a probit estimation. Inverting the predicted probability from (2.16) yields an estimate of the underlying latent variable $\hat{z}_{ij}^*$.

Defining $\delta = \sigma_\eta \frac{k-\varepsilon+1}{\varepsilon-1}$, HMR then show that $\hat{\omega}_{ij}^* \equiv \ln\left\{\exp\left[\delta\left(\hat{z}_{ij}^* + \hat{\bar{\eta}}_{ij}^*\right)\right] - 1\right\}$ is a consistent estimate for $E[\omega_{ij}|., z_{ij}^* > 0]$,[8] where $\hat{\bar{\eta}}_{ij}^* = \phi(\hat{z}_{ij}^*)/\Phi(\hat{z}_{ij}^*)$ is the inverse Mills ratio from the first stage probit estimation, which itself is well-known to be a consistent estimate of $E[u_{ij}|., z_{ij}^* > 0]$.[9] Inserting these terms into (2.10) HMR show that consistent estimation of the gravity model requires estimating the following specification:

$$m_{ij} = \beta_0 + \lambda_j + \chi_i - \gamma d_{ij} + \ln\left\{\exp\left[\delta\left(\hat{z}_{ij}^* + \hat{\bar{\eta}}_{ij}^*\right)\right] - 1\right\} + \beta_{u\eta}\hat{\bar{\eta}}_{ij}^* + e_{ij}, \tag{2.17}$$

where $\beta_{u\eta} \equiv corr(u_{ij}, \eta_{ij})(\sigma_u/\sigma_\eta)$. The term $\ln\left\{\exp\left[\delta\left(\hat{z}_{ij}^* + \hat{\bar{\eta}}_{ij}^*\right)\right] - 1\right\}$ corrects for the omitted variable ω_{ij} in the presence of sample selection[10] and $\beta_{u\eta}\hat{\bar{\eta}}_{ij}^*$ is the well-known correction of the error term u_{ij} in the presence of sample selection. As a result, e_{ij} is an i.i.d. error term satisfying $E[e_{ij}|., T_{ij} = 1] = 0$. Therefore, one can estimate (2.17) using NLS and obtain a consistent estimate of the distance coefficient $-\gamma$, having the structural interpretation of a trade cost per unit of distance.

2.2. The distance puzzle

The last decades of increased globalization have led to a more integrated world economy. World Trade Report (2008) describes this process emerging from improved transport and better communication technologies lowering trade costs between countries. When estimating gravity equations of trade, one would therefore expect that distance should become less of an obstacle to

[7] As in every discrete choice model, the level and scale can be arbitrarily chosen, i.e., the model has to be properly normalized. We normalize by dividing through σ_η, following HMR. This leads the error term $\eta_{ijl}^* = \eta_{ijl}/\sigma_\eta$ to be distributed unit normal.

[8] Santos Silva and Tenreyro (2008) note that this is not an unbiased estimate because of Jensen's inequality. However, it is a consistent estimate because of Slutsky's Theorem. To see this note $p\lim \ln\{\exp[\delta z_{ij}] - 1\} = \ln\{\exp[\delta p\lim z_{ij}] - 1\}$. But $p\lim z_{ij} = p\lim\left(E\left[z_{ij}^* \left| d_{ij}, \xi_j^*, \zeta_i^*, \phi_{ij}\right.\right] + u_{ij}\right) = E\left[z_{ij}^* \left| d_{ij}, \xi_j^*, \zeta_i^*, \phi_{ij}\right.\right] + E[u_{ij}|., z_{ij}^* > 0]$. A consistent estimate of the first term stems from (2.16), and a consistent estimate of the latter term $E[u_{ij}|., z_{ij}^* > 0] = \phi(\hat{z}_{ij}^*)/\Phi(\hat{z}_{ij}^*)$ as shown in Heckman (1979).

[9] This term is also known as Heckman's lambda (see Heckman, 1979).

[10] In the absence of sample selection bias but in the presence of the omitted variable bias, the correction term would simplify to $\ln\left\{\exp\left(\delta\hat{z}_{ij}^*\right) - 1\right\}$, since $p\lim\hat{\bar{\eta}}_{ij}^* = E[u_{ij}|., z_{ij}^* > 0] = 0$ in this case.

trade over time. Thus, globalization should decrease the absolute value of the distance coefficient $|-\gamma|$ in estimation equations like (2.10). As noted in the introduction, empirical work based on OLS regressions rather find the opposite - the negative impact of distance on trade seems to increase over time.

In this section, we use the HMR model to examine this "distance puzzle". We first make use of the result that estimating the specification (2.17) will estimate the distance coefficient consistently. We then examine to what extent OLS estimation of (2.10) is biased and in what direction this bias goes. Finally, we examine how the bias of OLS is affected by globalization, which reduces trade costs associated with geographical distance. In Section 3 we take these predictions to the data and examine if our estimates can explain the "distance puzzle".

2.2.1. The Bias of OLS

Let us start to examine the properties of an OLS estimate of the distance coefficient, $-\hat{\gamma}^{OLS}$ from estimating gravity equation (2.10) without sample selection correction and when ignoring the omitted variable for firm heterogeneity ω_{ij}.

To gain an intuition on these two biases and their direction, we look at them first in turn before considering them simultaneously, beginning with the sample selection bias. This can be illustrated in Figure 2, which contains on the horizontal axis distance d_{ij} and on the vertical axis imports m_{ij} in the top and the error term u_{ij} in the bottom panel. We depict by circles imports to country i from countries $j = 1, 2, 3, 4, 5$. The missing trade link selection equation (2.16) reveals that missing observations are the more likely the larger is distance and the smaller is the error term u_{ij}. For this reason, we draw potential imports between countries i and j=4 and j=5 such that distance is large and the error terms u_{i4} and u_{i5} negative, causing these two observations to drop out of the sample, which we indicate by hollow circles in the top panel. Since the negative shocks u_{i4} and u_{i5} are not only contained in the selection equation but also in the gravity equation (2.10), the imports that drop out occur not only at large distance but also at unusually low values of imports.[11] The non-missing imports at large distances, indicated by filled circles, are those with positive values of u_{ij}. But then the error term in the outcome equation u_{ij} and distance d_{ij} are positively correlated, which is shown by the positive slope of the $E[u_{ij}|d_{ij}]$ schedule in the bottom panel. Fitting OLS to the remaining three strictly positive import data from j=1,2,3 causes a steeper line indicated as OLS in the top panel as compared to the dashed line that fits OLS to the population including the data points j=4 and j=5. Such a twist in the slope occurs whenever the error terms in outcome and selection equation are positively correlated.

Turning next to the omitted variable bias, we need first understand how the omitted variable ω_{ij} is correlated with distance d_{ij}. This can easily be seen by inserting (2.15) into (2.13) and taking the expected value conditional on distance d_{ij} and the other control variables $z_0 \equiv$

[11] Note that we have drawn negative values of m_{ij}. Of course, negative values of m_{ij} can never exist, but are generated by the gravity equation (2.10), since shocks are by assumption normally distributed on a range from $-\infty$ to $+\infty$. However, whenever m_{ij} is negative, it is not observed, which is the same as if it does not exist.

$\gamma_0 + \zeta_i + \xi_j - \kappa\phi_{ij}$ to obtain

$$E[\omega_{ij} | d_{ij}, z_0] = \int_{\eta_{ij}} f_{\eta_{ij}}(\eta_{ij}) \ln \left[e^{\delta(z_o - \gamma d_{ij} + \eta_{ij})} - 1 \right] d\eta_{ij} \equiv \Omega(z_o, d_{ij}), \tag{2.18}$$

where $f_{\eta_{ij}}(\eta_{ij})$ is the marginal distribution function of η_{ij}, and we assumed that η_{ij} is conditionally independent of d_{ij} and z_0, which excludes by assumption the selection bias to separate the omitted variable bias from the sample selection bias conceptually.[12] Excluding conceptually the selection bias implies also to assume that $e^{\delta(z_o - \gamma d_{ij} + \eta_{ij})} > 1$ to ensure that there are no missing observations. Moreover, $\Omega(z_o, d_{ij})$ is the non-linear conditional expectation function, the shape of which is easy to analyse. Taking the derivative of (2.18) with respect to distance d_{ij}, we obtain

$$\frac{\partial E[\omega_{ij} | d_{ij}, z_0]}{\partial d_{ij}} = -\gamma\delta \int_{\eta_{ij}} f_{\eta_{ij}}(\eta_{ij}) \frac{e^{\delta(z_o - \gamma d_{ij} + \eta_{ij})}}{e^{\delta(z_o - \gamma d_{ij} + \eta_{ij})} - 1} d\eta_{ij} < 0.$$

Hence, there is a negative correlation between ω_{ij} and d_{ij}, because the share of exporting firms becomes smaller the larger is distance. The conditional expectation function is depicted in the bottom panel of Figure 2. If ω_{ij} was known, the influence of ω_{ij} would be deducted from the import values m_{ij}, indicated by crosses ($\times$) in the top panel, where the difference between m_{ij} and $m_{ij} - \omega_{ij}$ can be read off from the bottom panel, an example of which is given for ω_{i1}. An OLS estimator controlling for ω_{ij} and does not suffer from sample selection indicated by "HMR" fits thus the crosses ($\times$) rather than the circles which yields a flatter line than the dashed line. Overall, it is indeterminate whether the OLS line is flatter than the HMR line. In anticipation of our empirical results, we have drawn it such that the OLS line is steeper than the HMR line, which implies that the omitted variable bias dominates the sample selection bias in levels.

Now, we consider both biases simultaneously, taking additionally into account the interaction of the two biases. To do this, we need to draw on an approximation of (2.13),

$$\omega_{ij} \approx \delta z_{ij},$$

where $\delta = \partial\omega_{ij}/\partial z_{ij}$ evaluated at the mean of z_{ij}.

We then have the following proposition.

Proposition 1. ***(HMR)*** *When assuming that the HMR model is the data generating process, then the OLS estimate of $-\gamma$ in (2.10) may be (asymptotically) up- or downward biased, depending on whether the omitted variable bias from the share of exporting firms or the sample selection bias due to omission of zero trade flows dominates, respectively.*

[12] To see how one obtains this equation, note that by definition of a conditional expected value $E[\omega_{ij} | d_{ij}, z_0] = \int_{\omega_{ij}} \tilde{f}(\omega_{ij} | d_{ij}, z_0) \omega_{ij} d\omega_{ij}$, where $\tilde{f}(\omega_{ij} | d_{ij}, z_0)$ is the conditional distribution of ω_{ij}. According to Greene (2009), (B-51), this can be written as $E[\omega_{ij} | d_{ij}, z_0] = \int_{\eta_{ij}} f(\eta_{ij} | d_{ij}, z_0) \omega_{ij} d\eta_{ij}$ with $f(\eta_{ij} | d_{ij}, z_0)$ the conditional distribution of η_{ij}. If we then assume that η_{ij} is conditionally independent of d_{ij} and z_0, then we obtain from (B-60) $f(\eta_{ij} | d_{ij}, z_0) = f_{\eta_{ij}}(\eta_{ij})$, where $f_{\eta_{ij}}(\eta_{ij})$ is the marginal probability density (see B-45). Inserting this relation above, we obtain: $E[\omega_{ij} | d_{ij}, z_0] = \int_{\eta_{ij}} f_{\eta_{ij}}(\eta_{ij}) \omega_{ij} d\eta_{ij}$. Inserting into this relation (2.15) and (2.13) yields (2.18).

We derive in the Appendix the simultaneous bias term, which is approximately given by the following simple expression:

$$Bias(-\hat{\gamma}^{OLS}) = -\gamma\delta + \left[\delta + \Xi\beta_{u\eta}\right]\bar{\eta}_{ij} \gtreqless 0, \tag{2.19}$$

where $\Xi = \sum_i \sum_j d_{ij} / \sum_i \sum_j (d_{ij})^2$.

Thus, the term $-\delta\gamma < 0$ in (2.19) represents a downward bias in OLS from not controlling for the number of exporting firms, and the last two terms measure an upward bias from sample selection in OLS when omitting zero trade flows, as $\beta_{u\eta}$, $\bar{\eta}_{ij}$ and Ξ are positive.

In any case, the HMR estimator inserts non-linear correction factors into the OLS estimator to control simultaneously for the sample selection problem and the omitted variable bias as shown in the previous section. However, what would be the bias of a Heckman estimator designed to control for sample selection only if HMR is the true data generating process? The somewhat surprising answer is given in Proposition 2.

Proposition 2. *When assuming that the HMR model is the data generating process, then the Heckman estimate of $-\gamma$ in (2.10) is asymptotically unbiased.*

Proof: See Appendix B.

To gain a simple intuition of this proposition, note that a two-stage Heckman estimator is an OLS estimator using the inverse Mills ratio $\phi\left(\hat{z}^*_{ij}\right)/\Phi\left(\hat{z}^*_{ij}\right)$ as an additional explanatory variable, which is a monotonic transformation of $\hat{z}^*_{ij}$.***footnote???*** The HMR estimator, which is asymptotically unbiased, uses instead a correction factor that is also a monotonic transformation of $\hat{z}^*_{ij}$ (****to be shown in two lines???****). As a result, the two different correction factors are highly correlated in practice and pick up the same type of biases.

2.2.2. Globalization

How would the bias of OLS evolve over time when globalization reduces the impact of distance from the development of new and better communication and transport technologies. Make the following assumption:

Assumption Increased globalization implies $\frac{\partial\gamma}{\partial t} < 0$

We then have the following proposition:

Proposition 3. *When assuming that the HMR model is the data generating process, both the upward bias from sample selection from omitting zero trade flows and the downward bias from omitting the number of exporting decreases from globalization.*

The change in the bias of distance coefficient $\frac{\partial \mathrm{Bias}(-\hat{\gamma}^{OLS})}{\partial t}$ can again be understood intuitively, looking at the two sources of bias separately. Beginning with the change of the sample selection bias over time, we notice first that the bias depends on how the slope of $E[u_{ij}|d_{ij}]$ changes when γ changes over time. To understand this we need to first look at how the selection process is influenced by a reduction in γ. An observation is missing whenever $z^*_{ij} < 0$

according to (2.16). Obviously, a reduction in trade cost γ increases z_{ij}^* $\left(\partial z_{ij}^*/\partial\gamma > 0\right)$, where some missing trade links turn positive. Importantly, a reduction in trade cost increases the probability of missing observations from closer country-pairs to turn into positive trade links more than from distant ones. Eventually, all missing trade links have turned into positive ones at sufficiently low trade cost γ. We depict this in Figure 3, which isolates the sample selection example from Figure 2, by turning the hollow circles before globalization into filled ones afterwards. At the same time, a reduction of γ increases m_{ij} $(\partial m_{ij}/\partial\gamma < 0)$, but more so for further distant counry pairs $\left(\partial^2 m_{ij}/\partial\gamma\partial d_{ij} < 0\right)$. This is indicated in the top panel by the arrows that shift each observation m_{ij} upward, but more so the more distant observations. Hence, the true line fitting the data after globalization becomes flatter. But since there are no missing trade links left, there is no gap anymore between the OLS estimator and the true one. Hence, the bias from sample selection must have disappeared and the conditional expectation function of the error u_{ij}, $E[u_{ij}(\gamma^a)|d_{ij}]$, must have become horizontal after globalization, which indicates that the sample selection bias has shrunk to zero over time.

Turning to the change of the omitted variable bias over time, we need to understand again how the slope of the conditional expectation function $E[\omega_{ij}|d_{ij}]$ changes with a reduction of trade cost γ. For this purpose it is sufficient to look at how ω_{ij} changes for each observation when γ falls. From (2.15) and (2.13) we obtain immediately

$$\frac{\partial\omega_{ij}}{\partial\gamma} = -d_{ij}\tilde{\delta}\frac{e^{\tilde{\delta}z_{ij}}}{e^{\tilde{\delta}z_{ij}}-1} < 0 \tag{2.20}$$

for all ω_{ij} nonmissing. Hence, the share of exporting firms of a country j exporting to country i is increasing for each country pair when trade cost γ fall. More importantly, this share increases less for further distant trade partners:

$$\frac{\partial^2\omega_{ij}}{\partial\gamma\partial d_{ij}} = -\tilde{\delta}\frac{e^{\tilde{\delta}z_{ij}}}{e^{\tilde{\delta}z_{ij}}-1} - d_{ij}\gamma\tilde{\delta}^2 e^{-\tilde{\delta}z_{ij}}\left(1-e^{-\tilde{\delta}z_{ij}}\right)^{-2} < 0 \tag{2.21}$$

for all ω_{ij} nonmissing. This shift of ω_{ij} is depicted in Figure 4. In the bottom panel, a reduction of trade γ increases the export share ω_{ij} of each observation from the black crosses before globalization to the blue crosses after globalization. But then $E[\omega_{ij}|d_{ij}]$ must shift upward, too, since it is the average over all ω_{ij} at each level of distance d_{ij}. Because of (2.21), $E[\omega_{ij}|d_{ij}]$ must become flatter after a reduction of trade cost, which implies that the bias becomes smaller. This shift in the expected omitted variable can then be mapped into corresponding shifts of the regression line of imports on distance in the top panel of Figure 4.

Summing up, we cannot tell whether the difference in slopes between the HMR-line and the OLS-line will increase or decrease. Since we cannot tell how the bias of OLS will behave under globalization, the OLS estimate of the distance coefficient may also increase or decrease over time. This again depends on the relative change of the omitting variable bias from the number of exporting firms and the sample selection bias due to omission of zero trade flows. Both biases decrease over time with increased globalization, but the overall bias of OLS will increase or decrease depending on which of the two biases decreases faster. If the selection bias due to omission of zero trade flows decreases faster than the omitting variable bias from the number of exporting firms, the bias of OLS will increase over time, given that we observe a downward bias of the OLS distance coefficient in the first place.

Note that the sample selection bias alone is not able to solve the distance puzzle, as stated in Felbermayr and Kohler (2005). As the sample selection bias leads to an upward bias, the importance of distance would be underestimated. Hence, the level can not be captured correctly accounting for sample selection alone. However, to capture the change of the bias for the distance coefficient, we need a larger decrease in the zero-trade flows bias as compared to the omitted variable bias due to firm heterogeneity. A first glance at the data and anecdotal evidence copes with these facts. Whereas zero-trade flows decreased dramatically over the last two decades, firm sizes and productivities are still heavily dispersed.

3. Econometric analysis

3.1. Base-line estimation equation and alternative estimators

Our baseline estimation equation is the HMR gravity equation (2.17). Since our main interest rests on the coefficient of the distance variable $-\gamma$ and how it evolves over time, we will estimate this equation by year and industry separately. We use the following augmented specification:

$$\begin{aligned} m_{ij} &= \beta_0 - \gamma d_{ij} + \alpha_1 \mathrm{BORDER}_{ij} + \alpha_2 \mathrm{RTA}_{ij} + \alpha_3 \mathrm{REXCH}_{ij} + a_4 \mathrm{GDPPCDIF}_{ij} \\ &\quad + \lambda_j + \chi_i + \ln \left\{ \exp \left[\delta \left(\hat{z}^*_{ij} + \hat{\bar{\eta}}^*_{ij} \right) - 1 \right] \right\} + \beta_{u\eta} \hat{\bar{\eta}}^*_{ij} + e_{ij}, \end{aligned} \tag{3.1}$$

where we explain additional variables below. Again, note that $\ln \left\{ \exp \left[\delta \left(\hat{z}^*_{ij} + \hat{\bar{\eta}}^*_{ij} \right) - 1 \right] \right\}$ captures the omitted variable bias due to firm-level heterogeneity in the presence of sample selection, whereas $\hat{\bar{\eta}}^*_{ij}$ captures the sample selection bias of the error term from estimating (3.1) for non-zero trade. To estimate these correction terms, we add a first-stage equation in order to estimate (2.16), where:

$$\begin{aligned} z^*_{ij} &= \varphi^*_0 + \varphi^*_1 \mathrm{BORDER}_{ij} + \varphi^*_2 \mathrm{RTA}_{ij} + \varphi^*_3 \mathrm{REXCH}_{ij} + \varphi^*_4 \mathrm{GDPPCDIF}_{ij} \\ &\quad \varphi^*_5 COMM\_REL_{ij} + \varphi^*_6 COMM\_BORDER_{ij} + \\ &\quad \varphi^*_0 + \xi^*_j + \zeta^*_i - \gamma^* d_{ij} + \eta_{ij} \end{aligned} \tag{3.2}$$

3.1.1. Other estimators

We have shown that the distance puzzle can be studied by systematically comparing the estimates from HMR with corresponding estimates obtained with OLS. The OLS estimator estimates equation (3.1) omitting the correction terms for firm-level heterogeneity and sample selection, i.e., excluding $\ln \left\{ \exp \left[\delta \left(\hat{z}^*_{ij} + \hat{\bar{\eta}}^*_{ij} \right) - 1 \right] \right\}$ and $\hat{\bar{\eta}}^*_{ij}$. By comparing the HMR and OLS estimators we can evaluate how the bias of OLS evolves over time as predicted by Propositions 1 and 3. We will also compare our estimates with HMR with a number of other estimators.

Heckman The usual Heckman estimator estimates equation (3.1) omitting the correction terms for firm-level heterogeneity but including the one for sample selection, i.e., excluding $\ln\left\{\exp\left[\delta\left(\hat{z}^*_{ij}+\hat{\bar{\eta}}^*_{ij}\right)-1\right]\right\}$ but including $\hat{\bar{\eta}}^*_{ij}$.

Linear approximation of HMR As δ enters the estimation equation non-linearly, we first estimate equation (3.1) via non-linear least squares, as proposed by HMR. However, as discussed in Santos Silva and Tenreyro (2008), this derivative may be not entirely consistent with the underlying model. However, for a wide range of $\hat{z}^*_{ij}+\hat{\bar{\eta}}^*_{ij}$ the term $\ln\left\{\exp\left[\delta\left(\hat{z}^*_{ij}+\hat{\bar{\eta}}^*_{ij}\right)-1\right]\right\}$ may be well approximated by $\overline{\delta}\left(\hat{z}^*_{ij}+\hat{\bar{\eta}}^*_{ij}\right)$ for some appropriate parameter $\overline{\delta}$, which can be estimated by OLS (see our discussion in section 2.1). Hence, we also estimate the model via OLS and include $\varpi_{ij}=\overline{\delta}\left(\hat{z}^*_{ij}+\hat{\bar{\eta}}^*_{ij}\right)$ instead of $\ln\left\{\exp\left[\delta\left(\hat{z}^*_{ij}+\hat{\bar{\eta}}^*_{ij}\right)-1\right]\right\}$.

GPML In a final comparison, we will estimate (3.1) directly, that is we have a multiplicative model, given by:

$$\begin{aligned}\exp\left(m_{ij}\right) &= g\Big(\beta_0-\gamma d_{ij}+\alpha_1\mathrm{BORDER}_{ij}+\alpha_2\mathrm{RTA}_{ij}+\alpha_3\mathrm{REXCH}_{ij}+a_4\mathrm{GDPPCDIF}_{ij}\\ &\quad +\lambda_j+\chi_i+\delta\varpi_{ij}+\beta_{u\eta}\hat{\bar{\eta}}^*_{ij}+u_{ij}\Big),\end{aligned}\tag{3.3}$$

where $g(.)$ is some distribution function belonging to the family of generalized linear models.[13] As discussed in Santos Silva and Tenreyro (2006), the Poisson distribution is a prime candidate, leading to the Poisson pseudo-maximum-likelihood ($PPML$) estimator. However, if the variance is a function of higher powers of the mean, a more efficient estimator may be obtained by downweighting observations with larger conditional mean. An example for such an estimator is the Gamma pseudo-maximum-likelihood ($GPML$).

3.2. Data

3.2.1. Dependent variable

The dependent variable m_{ij} in (3.1) is the natural logarithm of bilateral imports of country i from country j in a given industry l at a given year t measured in million US$ converted by the Penn World Tables 2.0 purchasing power parity exchange rate (PPP) and deflated by the U.S. consumer price index. Data on imports are taken from Nicita and Olarreaga (2001)[14], who have compiled an industry dataset corresponding to the 3-digit ISIC, revision 2, level and contains 28 manufacturing industries for up to 100 countries during 1976-2004. Because there is a large number of missing values in the early years and we were lacking a control variable in the last year, we have restricted the sample to 1978-2003. The dataset is available for download from the World Bank (www.worldbank.org\trade). This dataset draws in turn its bilateral industry import data from COMTRADE of the UN which is based on the Standard International Trade Classification (SITC) and then transformed into ISIC. Production data are taken from UNIDO (International Yearbook of Industrial Statistics).

[13] Generalized linear models (GLMs) are based on a distribution function of the linearized exponential family which includes, for instance, the Gaussian, Poisson, negative binomial, or Gamma distributions.

[14] An update became available including data from the years 1974-2004 after we had accomplished our research.

3.2.2. Explanatory variables

The trade dataset is merged into a balanced geography dataset covering 170 countries. These geography variables appear in (3.1)-(3.3) as follows: d_{ij} is the log of the distance in kilometers between capitals of countries i and j. λ_j and χ_i are a full set of exporter and importer dummy variables, respectively, which control for, among others, the multilateral resistance terms pointed out by Anderson and van Wincoop (2003). Additional variables used are BORDER_{ij} which is a dummy variable indicating a common border between i and j, RTA_{ij} indicates whether there is a common trade agreement between exporter i and importer j (EU, MERCOSUR, NAFTA, ASEAN, CARICOM, PATCRA, ANZD, CACM, SPARTECA, CEFTA, USIS), the real exchange rate REXCH_{ij} is measured as ratio of exporter to importer purchasing power parity from the Penn World Tables 2.0 with a value larger 1 indicating a real overvaluation of the exporter relative to the importer currency based on the prices of the Penn World Table 2.0 consumption baskets. To control for differences in trade structure, we include also a rough measure of difference in development stage, GDPPCDIF_{ij}, which measures the log difference in per capita real GDP per worker of the exporter from the one of the importer country according to the Penn World Tables 2.0.

3.2.3. Instruments

To estimate the probability to export, we use the same control variables as in (3.1) in addition to two excluded selection variables following HMR. The first excluded variable is a measure to what extent importer and exporter share a common religion in the population according to data from the Christian Research Association for the year 2003. In particular, the measure takes the sum over the set of all existing religions summing up a population's share of the importer country confessing a religion multiplied with the same share of the exporter country. This measure is bounded between 0 and 1, with large numbers indicating a large degree of overlap in the religious structure of importer and exporter country. The second excluded variable is an indicator variable that indicates whether importer and exporter share a common language.

3.3. Results

To explore the distance puzzle we thus estimate (3.1) by industry and year. With data for 28 industries over 26 years, 1978-2003 and with five specifications, this amounts to estimating $28 \times 26 = 728$ first-stage regressions and $28 \times 26 \times 5 = 3640$ second stage regressions. For expositional reasons, we therefore show our results graphically. To avoid confusion over the evolution of a negative distance coefficient, we multiply all estimated distance coefficients with -1. This facilitates an easy discussion of whether distance coefficients grow or decrease over time.

The top panel of Figure 5 depicts distance coefficients estimated with OLS and the non-linear method from HMR. For each year the mean distance coefficient is calculated over all 28 industries, which is then plotted over the full time period 1978-2003. To indicate the time pattern for each estimator, we have added a quadratic trend with an associated 95 percent confidence interval. In the lower panel we also add the linear approximation of HMR which is again compared to OLS. Several interesting features are present in Figure 5.

Note that the trend of the distance coefficient, when estimated by OLS, $\hat{\gamma}^{OLS}$, is slightly *increasing* over time. This confirms the puzzling result in previous studies that the negative impact of distance of trade seems to increase and not decrease over time, which would be expected from the globalization process. Turning to the HMR distance coefficient $\hat{\gamma}^{HMR}$, we note that irrespective of whether estimated through the full non-linear terms or through the linear approximation, the $\hat{\gamma}^{HMR}$ is indeed decreasing over time. Examining the bias of OLS, $\hat{\gamma}^{OLS} - \hat{\gamma}^{HMR}$, we note that this is positive and it is increasing over time. From Propositions 1 and 3, this is consistent with the downward bias from omitting the number of exporters dominating the selection bias from omitting zero trade flows. However, note that the bias grows over time. From theory this suggests that globalization and reduced trade costs seem to decrease the upward bias from selection bias more than it reduces the downward bias from the number of exporters. Effectively, while the omitted variable bias seems to dominate the selection bias, over time the selection bias becomes relatively more important over time.

In Figure 5(ii), we compare OLS with the linear approximation of HMR. We note that results are qualitatively the same as in Figure 5(i): the HMR distance coefficient is decreasing over time, whereas the OLS coefficient increases with the associated bias of OLS increasing. Figure 6 compares OLS with the GPML estimator using a gamma function as suggested by Santos Silva and Tenreyro (2006, 2008). Comparing Figures 5 and 6 we note that the linear approximation of HMR gives very similar results to the non-linear version of HMR and the GPML estimate. That the linear approximation of the HMR works satisfactory is useful information for future application of the HMR methodology, given the cumbersome estimation of the non-linear version of HMR. Finally, we make a comparison with results obtained with the usual Heckman procedure. Again, results are qualitatively similar.

3.3.1. Industry estimates

Figure 7 shows changes over time of level of distance coefficient for the average over all 28 industries from HMR and OLS. Most industries show a similar pattern, where the distance coefficient with OLS is increasing over time, the HMR distance coefficient is decreasing over time, producing an increasing bias. In particular, these patterns are present in industries that are characterized by intra-industry trade, whereas patterns seem weaker in industries where the pattern of trade is more explained by comparative advantage. This is also what should be expected since trade in the HMR model generates intra-industry trade.

3.3.2. Globalization and transport costs

We have so far assumed that the data-generating process is the HMR-model and hence that (3.1) will consistently estimate the distance coefficient. We can informally test this assumption by relating the estimated distance coefficient $-\hat{\gamma}^{HMR}$ to actual trade costs. This is done in Figure 8, where Figure 8(i) shows that the estimated distance coefficients are strongly positively correlated with shipping costs in data recently published by Hummels. Figure 8(ii) shows the $-\hat{\gamma}^{HMR}$ is also positively correlated with oil prices, which should be an important determinant of transport costs. Finally, we note that the OLS estimate of the distance coefficient is negatively correlated with these data on transport costs. Again, this non-intuitive correlation can be explained because OLS does neither control for the omitted variable of the number of exporters nor for the omission of zero trade flows.

4. Conclusions

Globalization rapidly advanced during the last two decades: Many free trade and investment agreements were concluded, tariffs were reduced heavily, goods and services trade as well as international capital flows strongly grew. However, the influence of distance in empirical estimates of bilateral trade flows remained high and did not show a decreasing trend over time ("distance puzzle").

While there are many empirical and some theoretical attempts to solve these puzzles, we show that employing the most recent developments in empirical trade flow estimations accounting for zero trade flows and firm heterogeneity based on the HMR framework, the puzzles can be solved. We first show theoretically that the OLS estimate of the distance coefficient may be up- or downward biased, depending on whether the omitted variable bias from the number of exporting firms or the sample selection bias due to omission of zero trade flows dominates, respectively. Further, relating globalization to a fall of the distance coefficient, both the upward bias from sample selection from omitting zero trade flows and the downward bias from omitting the number of exporting decreases from globalization in absolute value. As these two biases work in opposite directions, the overall change of the bias from OLS estimates is ambiguous. Hence, it is important to control for both, the extensive margin and the intensive margin, as well as for firm level heterogeneity in order to understand the evolution of the distance coefficient over time.

We use industry level data over 28 industries for a time span ranging from 1978 to 2003 and confirm earlier results that the level of the distance coefficient is larger using OLS and that the distance coefficient is increasing over time using OLS. Hence, the bias from not controlling for firm heterogeneity is larger than the sample selection bias from omitting zero trade flows. However, the decrease of the bias over time is larger for the latter.

Employing the non-linear estimation of HMR leads indeed to a declining distance coefficient over time as expected. This can also be related to trade costs, since the estimated distance coefficients are strongly correlated with the time pattern in freight costs.

5. Appendix

5.1. Proof of Proposition 1

Proof. Rewrite (2.10) as $m_{ij} = \boldsymbol{\beta}'\mathbf{X}_{ij} + \omega_{ij} + u_{ij}$ and z_{ij}^* in (2.16) as $z_{ij}^* = \boldsymbol{\varphi}^{*\prime}\mathbf{X}_{ij} - \kappa\phi_{ij} + \eta_{ij}$ where $\boldsymbol{\beta} \equiv (\beta_0, \lambda_j, \chi_i, -\gamma)'$, and $\boldsymbol{\varphi}^* = \left(\gamma_0^*, \xi_j^*, \zeta_i^*, -\gamma^*\right)'$. Let, $\hat{\boldsymbol{\beta}}^{OLS}$ denote the OLS estimator of $\boldsymbol{\beta}$. We then obtain:

$$\text{plim}\left(\hat{\boldsymbol{\beta}}^{OLS}\right) = \boldsymbol{\beta} + \left[\mathbf{X}_{ij}'\mathbf{X}_{ij}\right]^{-1}\mathbf{X}_{ij}'\text{E}\left[\omega_{ij} + u_{ij}\left|z_{ij}^* > 0\right.\right], \tag{5.1}$$

where we have exploited that the $\mathbf{X}_{ij}$ variables contain only geography information and are therefore deterministic.

To evaluate (5.1), examine the conditional expectations $\text{E}\left[\omega_{ij}\left|z_{ij}^* > 0\right.\right]$ and $\text{E}\left[u_{ijl}\left|z_{ij}^* > 0\right.\right]$. Using formula (16.36) on p. 549 in Cameron and Trivedi (2005), we first obtain:

$$\begin{aligned} & E\left[u_{ij}\left|z_{ij}^* > 0\right.\right] = cov\left(u_{ij}, \eta_{ij}^*\right)\text{E}\left[\eta_{ij}^*|\eta_{ij}^* > \boldsymbol{\varphi}^{*\prime}\mathbf{X}_{ij} - \kappa^*\phi_{ij}\right] \\ = \; & corr\left(u_{ij}, u_{ij} + \nu_{ij}\right)\frac{\sigma_u}{\sigma_\eta}\frac{\phi\left(\boldsymbol{\varphi}^{*\prime}\mathbf{X}_{ij} - \kappa^*\phi_{ij}\right)}{\Phi\left(\boldsymbol{\varphi}^{*\prime}\mathbf{X}_{ij} - \kappa^*\phi_{ij}\right)} \\ \equiv \; & \beta_{u\eta}\overline{\eta}_{ij} > 0, \end{aligned} \tag{5.2}$$

where $\beta_{u\eta} = corr\left(u_{ij}, u_{ij} + \nu_{ij}\right)\sigma_u/\sigma_\eta$ and $\overline{\eta}_{ij} = \frac{\phi\left(\boldsymbol{\varphi}^{*\prime}\mathbf{X}_{ij} - \kappa^*\phi_{ij}\right)}{\Phi\left(\boldsymbol{\varphi}^{*\prime}\mathbf{X}_{ij} - \kappa^*\phi_{ij}\right)}$. Further, we have assumed that u_{ij} and η_{ij}^* are bivariate normally distributed. Note that this implies that $u_{ij} = cov(u_{ij}, \eta_{ij}^*)$ $\eta_{ijl}^*/\sigma_\eta^2 + \varrho_{ij}$, where ϱ_{ij} is independent of η_{ijl}^*. Hence, $\text{E}\left[u_{ij}\left|\eta_{ij}^* > \boldsymbol{\varphi}^{*\prime}\mathbf{X}_{ij} - \kappa^*\phi_{ij}\right.\right] = cov(u_{ij}, \eta_{ij}^*)$ $1/\sigma_\eta^2\text{E}\left[\eta_{ij}^*\left|\eta_{ij}^* > \boldsymbol{\varphi}^{*\prime}\mathbf{X}_{ij} - \kappa^*\phi_{ij}\right.\right]$ and $cov\left(u_{ij}, \eta_{ij}^*\right) = corr\left(u_{ij}, u_{ij} + \nu_{ij}\right)\sigma_u\sigma_\eta$.

To proceed, use a linear approximation of ω_{ij} in noting that $\omega_{ij} = \ln\left[\left(Z_{ij}^*\right)^\delta - 1\right]$ for $z_{ij}^* > 0$. We can then write $\omega_{ij} = \ln\left[\left(Z_{ij}^*\right)^\delta - 1\right] = \ln\left[\exp\left(\delta z_{ijl}^*\right) - 1\right] \approx \delta z_{ij}^* > 0$, where $\delta = \sigma_\eta\frac{k-\varepsilon+1}{\varepsilon-1}$ is defined as above. [15] We then obtain:

$$\begin{aligned} & \text{E}\left[\omega_{ij}\left|z_{ij}^* > 0\right.\right], \\ = \; & \text{E}\left[\delta z_{ij}^*\left|z_{ij}^* > 0\right.\right] = \delta\text{E}\left[\left\{\text{E}\left[z_{ij}^*\left|d_{ij,}\mathbf{X}_{ij}\right.\right] + \eta_{ij}^*\right\}\left|z_{ij}^* > 0\right.\right] \\ = \; & \delta\text{E}\left[z_{ij}^*\left|d_{ij,}\mathbf{X}_{ij}\right.\right] + \delta\text{E}\left[\eta_{ij}^*\left|z_{ij}^* > 0\right.\right], \\ = \; & \delta\text{E}\left[\gamma_0^* + \xi_j^* + \zeta_i^* - \gamma^* d_{ij} - \kappa^*\phi_{ij}\left|d_{ij,}\mathbf{X}_{ij}\right.\right] + \delta\text{E}\left[\eta_{ij}^*\left|z_{ij}^* > 0\right.\right], \\ = \; & \delta\left[\gamma_0^* + \xi_j^* + \zeta_i^* - \gamma d_{ij}^* + \overline{\eta}_{ij}\right], \\ = \; & \delta\boldsymbol{\varphi}^{*\prime}\mathbf{X}_{ij}' + \delta\overline{\eta}_{ij}, \end{aligned} \tag{5.3}$$

where we used the assumption that ϕ_{ij} is independent of $\mathbf{X}_{ij}$ in the second last row. Noting that $\left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\varphi}^*\delta = \boldsymbol{\varphi}^*\delta$, we obtain:

$$\text{plim}\left(\hat{\boldsymbol{\beta}}^{OLS}\right) = \boldsymbol{\beta} + \boldsymbol{\gamma}\delta + \delta\overline{\eta}_{ijl} + \left[\mathbf{X}_{ij}'\mathbf{X}_{ij}\right]^{-1}\mathbf{X}_{ij}'\beta_{u\eta}\overline{\eta}_{ij} \gtreqless 0. \tag{5.4}$$

[15] It can be shown that this approximation works very well in the range of ω_{ij} from $[0.5, \infty]$ and estimated values of δ around 1.

Since country dummies in $\mathbf{X}_{ij}$ are not correlated by construction and distance is hardly correlated with country dummies the matrix $\mathbf{X}'\mathbf{X}$ can be viewed as diagonal. But then:

$$\text{plim}\left(-\hat{\gamma}^{OLS}\right) = -\gamma - \gamma\delta + \left[\delta + \beta_{u\eta}\frac{\sum_i \sum_j d_{ij}}{\sum_i \sum_j (d_{ij})^2}\right]\bar{\eta}_{ij}, \tag{5.5}$$

and hence

$$\text{Bias}(\ -\hat{\gamma}^{OLS}) = -\gamma\delta + \left[\delta + \beta_{u\eta}\frac{\sum_i \sum_j d_{ij}}{\sum_i \sum_j (d_{ij})^2}\right]\bar{\eta}_{ij}.$$

■

5.2. Proof of Proposition 2

Proof: Suppose the true data generating process is (HMR model)

$$m_{ij} = x\beta + \omega_{\iota j} + u_{ij}, \tag{5.6}$$

where x is a row vector that contains the regression constant, country exporter- and importer dummies and d_{ij}, and β is the corresponding coefficient vector (β_0 $\boldsymbol{\lambda}$ $\boldsymbol{\chi}$ $-\gamma$)'. Suppose now that these data are estimated with a Heckman 2-stage estimator, which estimates in the second stage:

$$m_{ij} = xb + \hat{\rho}\lambda\left(z_{ij}\right) + e_{ij}, \tag{5.7}$$

for $z_{ij} > 0$, where e_{ij} is an error term not necessarily white noise, and

$$\lambda\left(z_{ij}\right) = \phi\left(z_{ij}\right)/\Phi\left(z_{ij}\right) = E[u_{ij}|x, \phi_{ij}, z^*_{ij} > 0] \tag{5.8}$$

is the inverse Mills ratio obtained from a first stage probit estimation of T_{ij} on z_{ij}, i.e.

$$P(T_{ij} = 1|x, \phi_{ij}, z^*_{ij} > 0) = \Phi\left(z^*_{ij} > 0\right) \tag{5.9}$$

with

$$z^*_{ij} = x\boldsymbol{\gamma}^* - \kappa^*\phi_{ij} + u_{ij} + \nu_{ij} \tag{5.10}$$

and $\boldsymbol{\gamma}^*$ is a vector (γ^*_0 $\boldsymbol{\varsigma}^*$ $\boldsymbol{\xi}^*$ $-\gamma^*$)', and a star is a convenient scaling of a coefficient. Recall that u_{ij} and ν_{ij} are correlated, causing sample selection. Note also that

$$\omega_{\iota j} = \ln\left[e^{\delta z^*_{ij}} - 1\right] \approx \delta z^*_{ij} = \delta\left(x\boldsymbol{\gamma}^* - \kappa^*\phi_{ij} + u_{ij} + \nu_{ij}\right) \tag{5.11}$$

Now, apply the Frisch-Waugh theorem to (5.7):

$$m_{ij} - \lambda a_0 = \left(x - \lambda a_1\right)b + e_{ij}, \tag{5.12}$$

where

$$\begin{aligned} a_0 &= \left(\lambda'\lambda\right)^{-1}\lambda' m_{ij} && (5.13) \\ a_1 &= \left(\lambda'\lambda\right)^{-1}\lambda' x && (5.14) \end{aligned}$$

and analogously

$$m_{ij} - \lambda a_0 = (x - \lambda a_1)\,\beta + (\omega_{\iota j} - \lambda a_2) + u_{ij}, \tag{5.15}$$

where

$$a_2 = \left(\lambda'\lambda\right)^{-1} \lambda'\omega_{\iota j}. \tag{5.16}$$

Note that the OLS estimate b of (5.12) is given by

$$b_{Heck} = \left[(x - \lambda a_1)'\,(x - \lambda a_1)\right]^{-1} (x - \lambda a_1)'\,(m_{ij} - \lambda a_0)\,, \tag{5.17}$$

which is the two-stage Heckman estimator. The asymptotic bias of the Heckman estimator is thus for i and j with $z_{ij}^* > 0$

$$\begin{aligned}
& p\lim b_{Heck} - \beta \\
= \;& p\lim \left[(x - \lambda a_1)'\,(x - \lambda a_1)\right]^{-1} (x - \lambda a_1)'\,[(x - \lambda a_1)\,\beta + (\omega_{\iota j} - \lambda a_2) + u_{ij}] - \beta \\
= \;& p\lim \left[(x - \lambda a_1)'\,(x - \lambda a_1)\right]^{-1} (x - \lambda a_1)'\,(\omega_{\iota j} - \lambda a_2 + u_{ij}) \\
\approx \;& p\lim \left[(x - \lambda a_1)'\,(x - \lambda a_1)\right]^{-1} (x - \lambda a_1)'\left[\delta z_{ij}^* - \lambda a_2 + u_{ij}\right] \\
= \;& p\lim \left[(x - \lambda a_1)'\,(x - \lambda a_1)\right]^{-1} \left[p\lim\,(x - \lambda a_1)'\,\delta z_{ij}^*\right] \\
& -p\lim \left[(x - \lambda a_1)'\,(x - \lambda a_1)\right]^{-1} \left[p\lim\,(x - \lambda a_1)'\,(\lambda a_2)\right] \\
& +p\lim \left[(x - \lambda a_1)'\,(x - \lambda a_1)\right]^{-1} \left[p\lim\,(x - \lambda a_1)'\,u_{ij}\right] \\
\equiv \;& p\lim \left[(x - \lambda a_1)'\,(x - \lambda a_1)\right]^{-1} [A + B + C]
\end{aligned}$$

where we inserted (5.15) and (5.17) in line 2, and the approximation (5.11) in line 4. Note first that

$$\begin{aligned}
& A \\
= \;& E_{x,z_{ij}}[(x - \lambda a_1)'\left(\delta z_{ij}^*\right) |z_{ij}^* > 0] \\
= \;& E_{x,z_{ij}}\left[\left(x - E\left[x|\lambda\left(z_{ij}^*\right)\right]\right)'\,(\delta z_{ij})\,|z_{ij}^* > 0\right] \\
= \;& E_{z_{ij}}\left[E_x\left[\left(x - E\left[x|\lambda\left(z_{ij}^*\right)\right]\right)'\left(\delta z_{ij}^*\right)|z_{ij}^*\right]|z_{ij}^* > 0\right] \\
= \;& \delta E_{z_{ij}}\left[E_x\left[\left(x - E\left[x|\lambda\left(z_{ij}^*\right)\right]\right)|z_{ij}^*\right]' z_{ij}^*\,|z_{ij}^* > 0\right] \\
= \;& \delta E_{z_{ij}}\left[\left(E\left[x|z_{ij}^*\right] - E\left[x\,|\lambda\left(z_{ij}^*\right),\; z_{ij}^*\right]\right)' z_{ij}^*\,|z_{ij}^* > 0\right] \\
= \;& \delta E_{z_{ij}}\left[\left(E\left[x|z_{ij}^*\right] - E\left[x\,|z_{ij}^*\right]\right)' z_{ij}^*\,|z_{ij}^* > 0\right] \\
= \;& 0,
\end{aligned}$$

where we use in line 2 a law of large numbers assuming the usual regularity conditions, in line 3 the fact that $a_1\lambda = E\left[x|\lambda\left(z_{ij}^*\right)\right]$, i.e. it is the score of a regression of x on λ according to (5.14), line 4 uses the law of iterated expectations, in line 5 z_{ij}^* can be taken out of the inner conditional expectation operator, since it is conditioned on z_{ij}^*, and line 6 uses the fact that $E\left[E\left[x|\lambda\left(z_{ij}^*\right)\right]|z_{ij}^*\right] = E\left[E\left[x|z_{ij}^*\right]|z_{ij}^*\right] = E\left[x|z_{ij}^*\right]$ (e.g. Wooldrdige, 2002, p. 30, property CE.4) together with the law of iterated expectations.

The term B can be developed as follows for i and j with $z_{ij}^* > 0$:

$$\begin{aligned}
& B \\
\equiv\ & p\lim\left(x-\lambda a_1\right)'\left(\lambda a_2\right) \\
=\ & p\lim\left(x-\lambda\left(\lambda'\lambda\right)^{-1}\lambda'x\right)'\left(\lambda\left(\lambda'\lambda\right)^{-1}\lambda'\omega_{\iota j}\right) \\
=\ & p\lim x'\left(\lambda\left(\lambda'\lambda\right)^{-1}\lambda'\omega_{\iota j}\right) \\
& -p\lim\left(\lambda\left(\lambda'\lambda\right)^{-1}\lambda'x\right)'\left(\lambda\left(\lambda'\lambda\right)^{-1}\lambda'\omega_{\iota j}\right) \\
=\ & p\lim x'\left(\lambda\left(\lambda'\lambda\right)^{-1}\lambda'\omega_{\iota j}\right) \\
& -p\lim\left(x'\lambda\left(\lambda'\lambda\right)^{-1}\lambda'\lambda\left(\lambda'\lambda\right)^{-1}\lambda'\omega_{\iota j}\right) \\
=\ & 0,
\end{aligned}$$

where we inserted (5.14) and (5.16).

The term C captures the potential sample selection bias in presence of the correction term for it given i and j with $z_{ij}^* > 0$:

$$\begin{aligned}
C &= p\lim\left(x-a_1\lambda\left(z_{ij}^*\right)\right)'u_{ij} \\
&= E[\left(x-a_1\lambda\left(z_{ij}^*\right)\right)'u_{ij}|z_{ij}^*>0] \\
&= E[\left(x-a_1\lambda\left(z_{ij}^*\right)\right)'\left[E[u_{ij}|x,\phi_{ij},z_{ij}^*>0]\right]|z_{ij}^*>0] \\
&= E\left[\left(x-a_1\lambda\left(z_{ij}^*\right)\right)'\lambda\left(z_{ij}^*\right)|z_{ij}^*>0\right] \\
&= E\left[\left(x-E\left[x|\lambda\left(z_{ij}^*\right)\right]\right)'\lambda\left(z_{ij}^*\right)|z_{ij}^*>0\right] \\
&= E\left[E\left[x-E\left[x|\lambda\left(z_{ij}^*\right)\right]|z_{ij}^*\right]'\lambda\left(z_{ij}^*\right)|z_{ij}^*>0\right] \\
&= E\left[\left(E\left[(x)|z_{ij}^*\right]-E\left[E\left[x|\lambda\left(z_{ij}^*\right)\right]|z_{ij}^*\right]\right)'\lambda\left(z_{ij}^*\right)|z_{ij}^*>0\right] \\
&= 0,
\end{aligned}$$

where we use in line 2 a law of large numbers assuming the usual regularity conditions, in line 3 we apply the law of iterated expectations, in line 4 we insert (5.8), in line 5 we use the fact that $a_1\lambda = E\left[x|\lambda\left(z_{ij}^*\right)\right]$, in line 6 we apply the law of iterated expectations and place $\lambda\left(z_{ij}^*\right)$ outside the inner conditional expectation, because z_{ij}^* is conditioned on, in line 8 $E\left[E\left[x|\lambda\left(z_{ij}^*\right)\right]|z_{ij}^*\right] = E\left[E\left[x|z_{ij}^*\right]|z_{ij}^*\right] = E\left[x|z_{ij}^*\right]$ (e.g. Wooldrdige, 2002, p. 30, property CE.4) together with the law of iterated expectations.

Summing up, we have for the asymptotic bias of the distance coefficient:

$$p\lim -\gamma_{Heck} + \gamma = 0.$$

■

5.3. Proof of Proposition 3

Proof: From (2.19), we have $\text{Bias}(-\hat{\gamma}^{OLS}) = -\delta\gamma + \left[\delta + \Xi\beta_{u\eta}\right]\bar{\eta}_{ij}$. Thus, it follows that $\frac{\partial \text{Bias}(-\hat{\gamma}^{OLS})}{\partial t} = -\delta\frac{\partial(\gamma)}{\partial t} + \left[\delta + \Xi\beta_{u\eta}\right]\frac{\partial \bar{\eta}_{ij}}{\partial t}$. The change of the omitted variable bias over time is simply given by:

$$\frac{\partial\left(-\delta\gamma\right)}{\partial t} = -\delta\frac{\partial\gamma}{\partial t} > 0$$

The sign of the change in sample selection bias depends on the sign of

$$\begin{aligned}
\frac{\partial\bar{\eta}_{ij}}{\partial t} &= \frac{\partial\left(\frac{\phi(z_{ij})}{\Phi(z)}\right)}{\partial t} \\
&= \frac{1}{\Phi(z_{ij})^2}\left[\left(\phi'\left(z_{ij}\right)\cdot\Phi\left(z_{ij}\right) - \phi\left(z_{ij}\right)^2\right)\right]\frac{\partial z_{ij}}{\partial t} \\
&= \left[\frac{-z_{ijl}\phi\left(z_{ij}\right)}{\Phi\left(z_{ij}\right)} - \left(\frac{\phi\left(z_{ij}\right)}{\Phi\left(z_{ij}\right)}\right)^2\right]\frac{\partial z_{ij}}{\partial t} \\
&= \left[-z_{ij}\bar{\eta}_{ijl} - \left(\bar{\eta}_{ijl}\right)^2\right]\frac{\partial z_{ij}}{\partial t} \\
&= -\bar{\eta}_{ijl}\left[z_{ij} + \bar{\eta}_{ijl}\right]\frac{\partial z_{ij}}{\partial t}.
\end{aligned} \tag{5.18}$$

Note that

$$\frac{\partial z_{ij}}{\partial t} = -d_{ij}\frac{\partial\gamma\left(t\right)}{\partial t} > 0.$$

The derivative of the mills ratio $\frac{\partial\bar{\eta}_{ij}}{\partial z_{ij}} = -\bar{\eta}_{ij}\left[z_{ij} + \bar{\eta}_{ij}\right]$ is negative. This can be shown by noting that

$$E\left[\eta^*_{ij}\left|\eta^*_{ij} > -\boldsymbol{\varphi}'\mathbf{X}\right.\right] = \frac{\phi\left(\boldsymbol{\varphi}'\mathbf{X}\right)}{\Phi\left(\boldsymbol{\varphi}'\mathbf{X}\right)} = \frac{\phi\left(-\boldsymbol{\varphi}'\mathbf{X}\right)}{1-\Phi\left(-\boldsymbol{\varphi}'\mathbf{X}\right)}, \tag{5.19}$$

and using the result derived in Sampford (1953) and also given in Theorem 24.2 on page 866 in Greene (2008), that for $\phi\left(x\right)/\left(1-\Phi\left(x\right)\right)$ the derivative with respect to x is given by

$$\frac{\phi\left(x\right)}{1-\Phi\left(x\right)}\left[\frac{\phi\left(x\right)}{1-\Phi\left(x\right)} - x\right], \tag{5.20}$$

and bounded between zero and one. Using the equality given in equation (5.19), we may write this as:

$$\frac{\phi\left(\boldsymbol{\varphi}'\mathbf{X}\right)}{\Phi\left(\boldsymbol{\varphi}'\mathbf{X}\right)}\left[\frac{\phi\left(\boldsymbol{\varphi}'\mathbf{X}\right)}{\Phi\left(\boldsymbol{\varphi}'\mathbf{X}\right)} + \boldsymbol{\varphi}'\mathbf{X}\right] = \bar{\eta}_{ij}\left[z_{ij} + \bar{\eta}_{ij}\right]. \tag{5.21}$$

Hence, this expression differs from our derivative of $\bar{\eta}_{ij}$ only by the multiplication with -1. Hence, the derivative of $\bar{\eta}_{ij}$ with respect to z_{ij} is bounded between -1 and 0. But then $\frac{\partial\bar{\eta}_{ij}}{\partial t} = \partial\left[\phi\left(z_{ij}\right)/\Phi\left(z_{ij}\right)\right]/\partial t < 0$. The change in the bias for OLS is therefore ambiguous, depending on whether the change in the sample selection bias or the change in the omitted variable bias is larger:

$$\frac{\partial\text{Bias}(-\hat{\gamma}^{OLS})}{\partial t} = -\delta\frac{\partial\left(\gamma\right)}{\partial t} + \left[\delta + \Xi\beta_{u\eta}\right]\frac{\partial\bar{\eta}_{ij}}{\partial t} \gtreqless 0. \tag{5.22}$$

■

References

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

Anderson, J.E. and E. van Wincoop (2004), "Trade Costs," *Journal of Economic Literature* **42**(3), pp. 691-751.

Baldwin, R. and J. Harrigan (2007), "Zeros, Quality and Space: Trade Theory and Trade Evidence," *NBER Working Paper Nr. 13214*.

Bernard, A.B., J.B. Jensen, and P.K. Schott (2006), "Trade Costs, Firms and Productivity," *Journal of Monetary Economics* **53**(1), pp. 917–937.

Berthelon, M. and C. Freund (2008), "On the Conservation of Distance in International Trade," *Journal of International Economics* **75**(2), pp. 310-320.

Blum, B. and A. Goldfarb (2006), "Does the Internet Defy the Law of Gravity?", *Journal of International Economics* **70**(2), pp. 384–405.

Brun, J.-F., C. Carrère, P. Guillaumont and J. de Melo (2005), "Has Distance Died? Evidence from a Panel Gravity Model," *World Bank Economic Review* **19**(1), pp., 99-120.

Cameron, A.C. and P.K. Trivedi (2005), *Microeconometrics - Methods and Applications*, Cambridge University Press, Cambridge, United Kingdom.

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

Disdier, A.-C. and K. Head (2008), "The Puzzling Persistence of the Distance Effect on Bilateral Trade," *Review of Economics and Statistics* **90**(1), pp. 37-48.

Dixit, A.K. and J.E. Stiglitz (1977), "Monopolistic Competition and Optimum Product Diversity," *American Economic Review* **67**(3), pp. 297-308.

Dreher, Axel, Noel Gaston and Pim Martens (2008), Measuring Globalization – Gauging its Consequences (New York: Springer).

Duranton, G. and M. Storper (2008), "Rising Trade Costs? Agglomeration and Trade with Endogenous Transaction Costs," *Canadian Journal of Economics* **41**(1), pp. 292-319.

Felbermayr, G.J. and W. Kohler (2006), "Exploring the Intensive and Extensive Margins of World Trade," *Review of World Economics* **142**(4), pp. 642-674.

Flam, H. and H. Nordström (2008), "Gravity Estimation of the Intensive and Extensive Margin: An Alternative Procedure and Alternative Data," unpublished manuscript, available at: http://www-2.iies.su.se/∼flamh/workingpapers.html.

Glaeser, E.L. and J.E. Kohlhase (2004), "Cities, Regions and the Decline of Transport Costs," *Papers in Regional Science* **83**(1), pp. 197–228.

Greene, W. (2008), *Econometric Analysis*, 6th edition, Pearson, Prentice Hall.

Grossman, G.M. (1998), "Comment" (pp. 29–31), in J. A. Frankel (Ed.), *The Regionalization of the World Economy*, Chicago, University of Chicago Press, National Bureau of Economic Research project report, 1998.

Heckman, J.J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica* **47**(1), pp. 153-161.

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

Hillberry, R. and D. Hummels (2008), "Trade responses to geographic frictions: A decomposition using micro-data," *European Economic Review* **52**(3), pp. 527-550.

Hummels, D. (2007), "Transportation Costs and International Trade in the Second Era of Globalization," *Journal of Political Economy* **21**(3), pp. 131-154.

Hummels, D. and P.J. Klenow (2005), "The Variety and Quality of a Nation's Exports," *American Economic Review* **95**(3), pp. 704–723.

Krautheim, S. (2008), "Gravity and Information: Heterogeneous Firms, Exporter Networks and the 'Distance Puzzle'," unpublished manuscript, available at: http://ces.univ-paris1.fr /membre/Krautheim/Index.htm.

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

Lawless, M. and K. Whelan (2007), "A Note on Trade Costs and Distance," *Research Technical Paper of the Central Bank and Financial Services Authority of Ireland*.

Nicita, A. and M. Olarreaga (2001), "Trade and Production: 1976-1999," *World Bank Policy Research Working Paper No. 2701*, available at: http://www.worldbank.org \trade.

Sampford, M.R. (1953), "Some Inequalities on Mill's Ratio and Related Functions," *The Annals of Mathematical Statistics* 24(1), pp. 130-132.

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

Santos Silva, J.M.C. and S. Tenreyro (2008), "Trading Partners and Trading Volumes: Implementing the Helpman-Melitz-Rubinstein Model Empirically," unpublished manuscript, available at: http://privatewww.essex.ac.uk/∼jmcss/research.html?.

World Trade Report (2008), *Trade in a Globalizing World*, World Trade Organization.

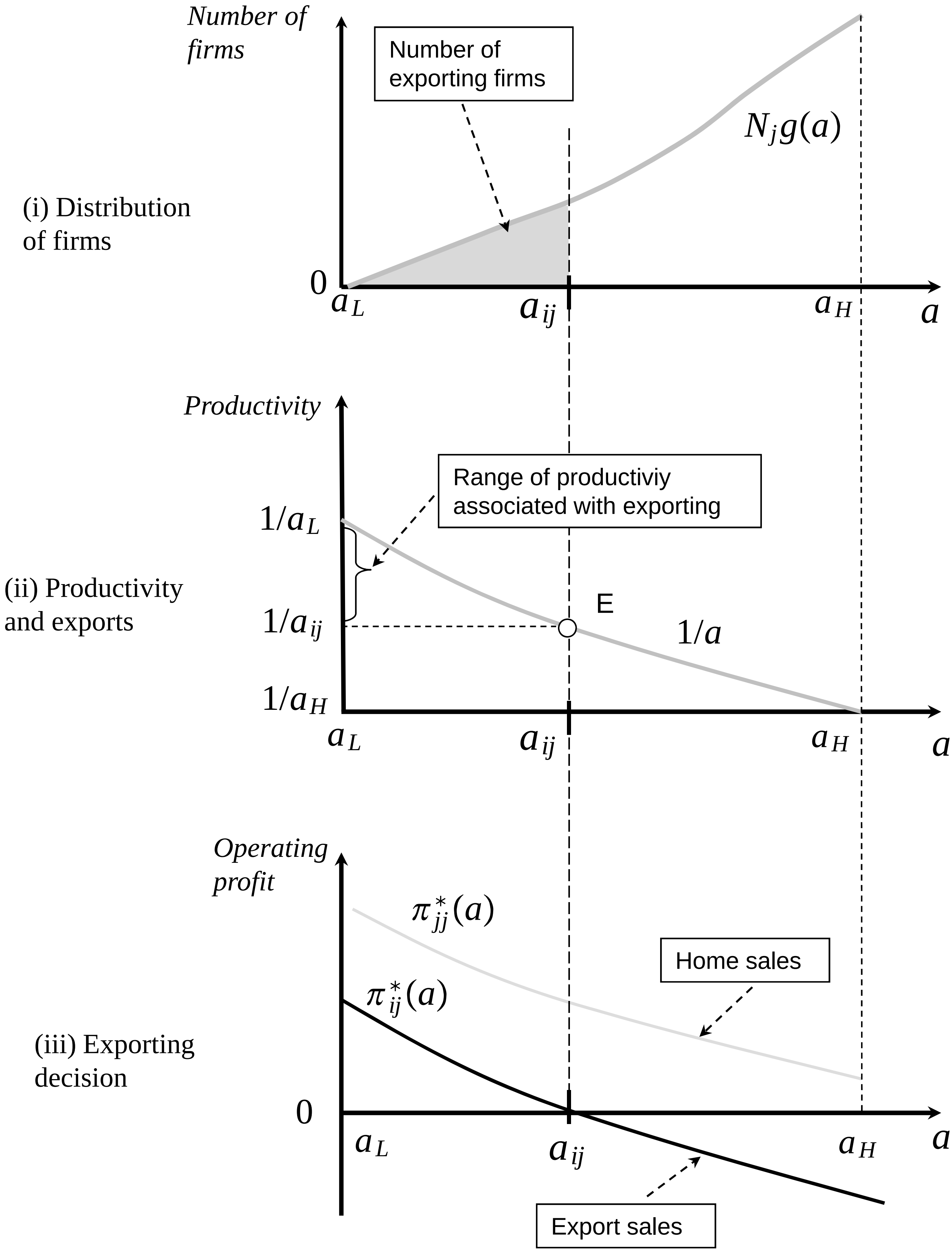

Figure 1: The HMR model

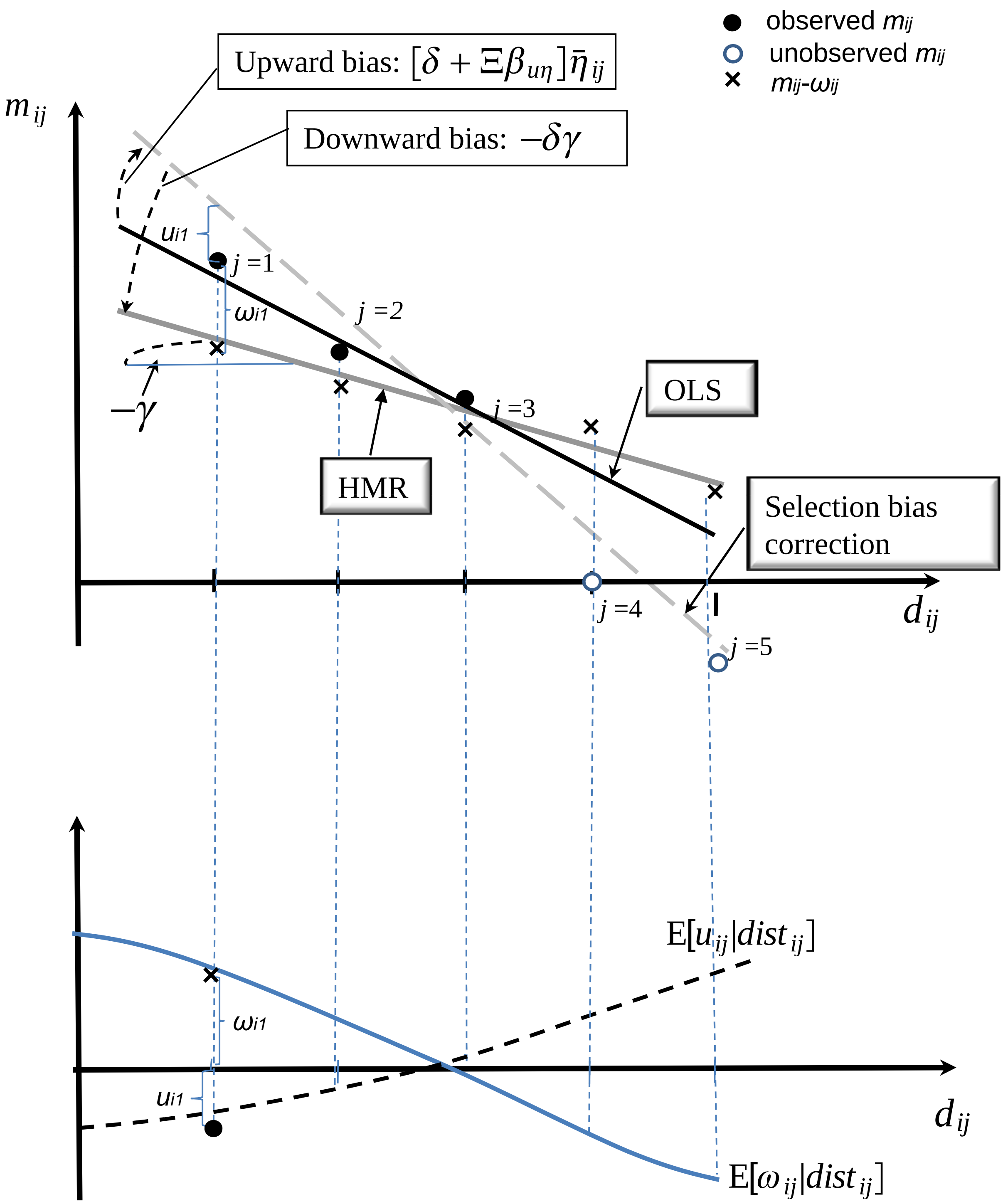

Figure 2: Illustrating the bias of OLS

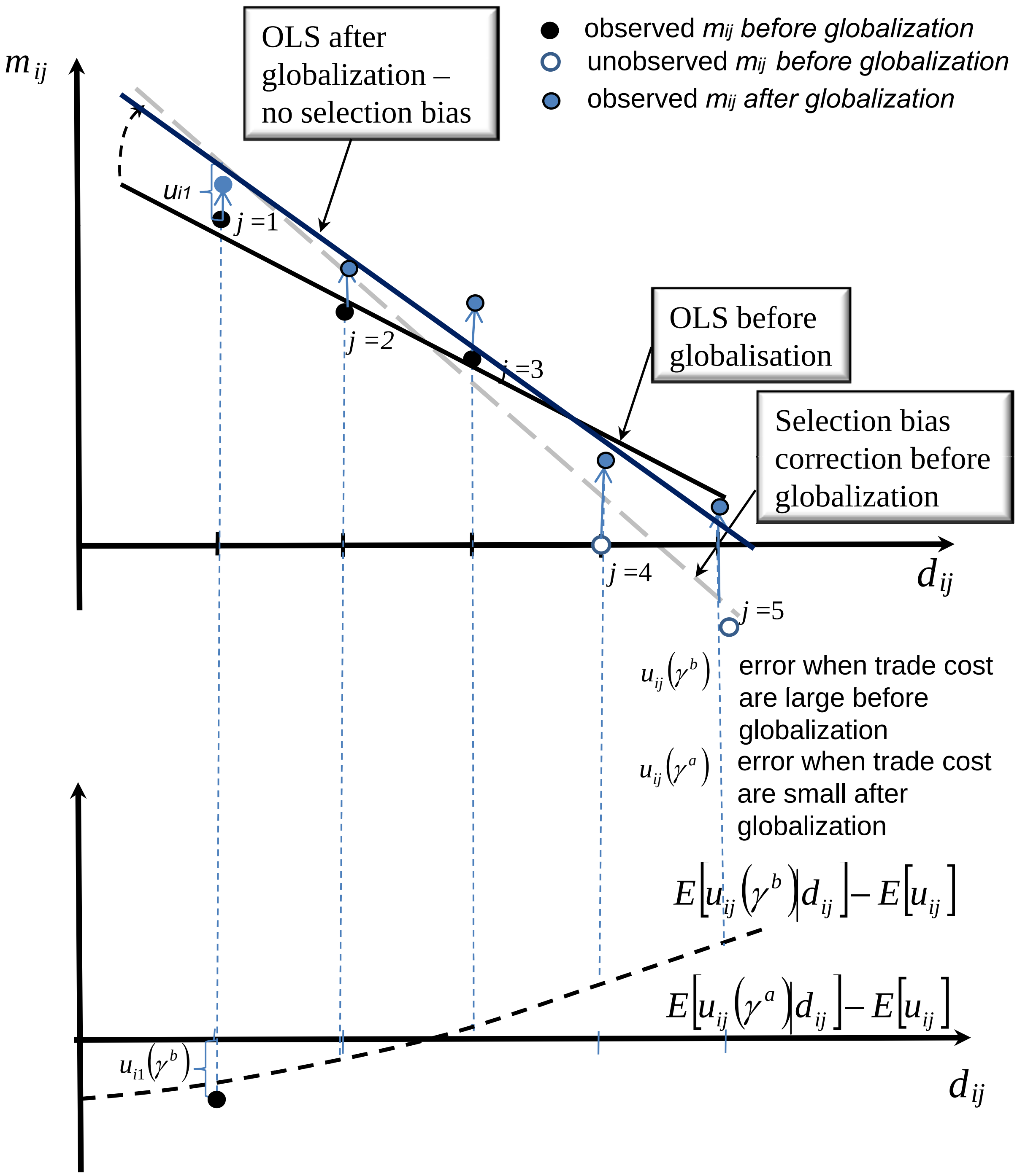

Figure 3: Illustrating the change in sample selection bias of OLS

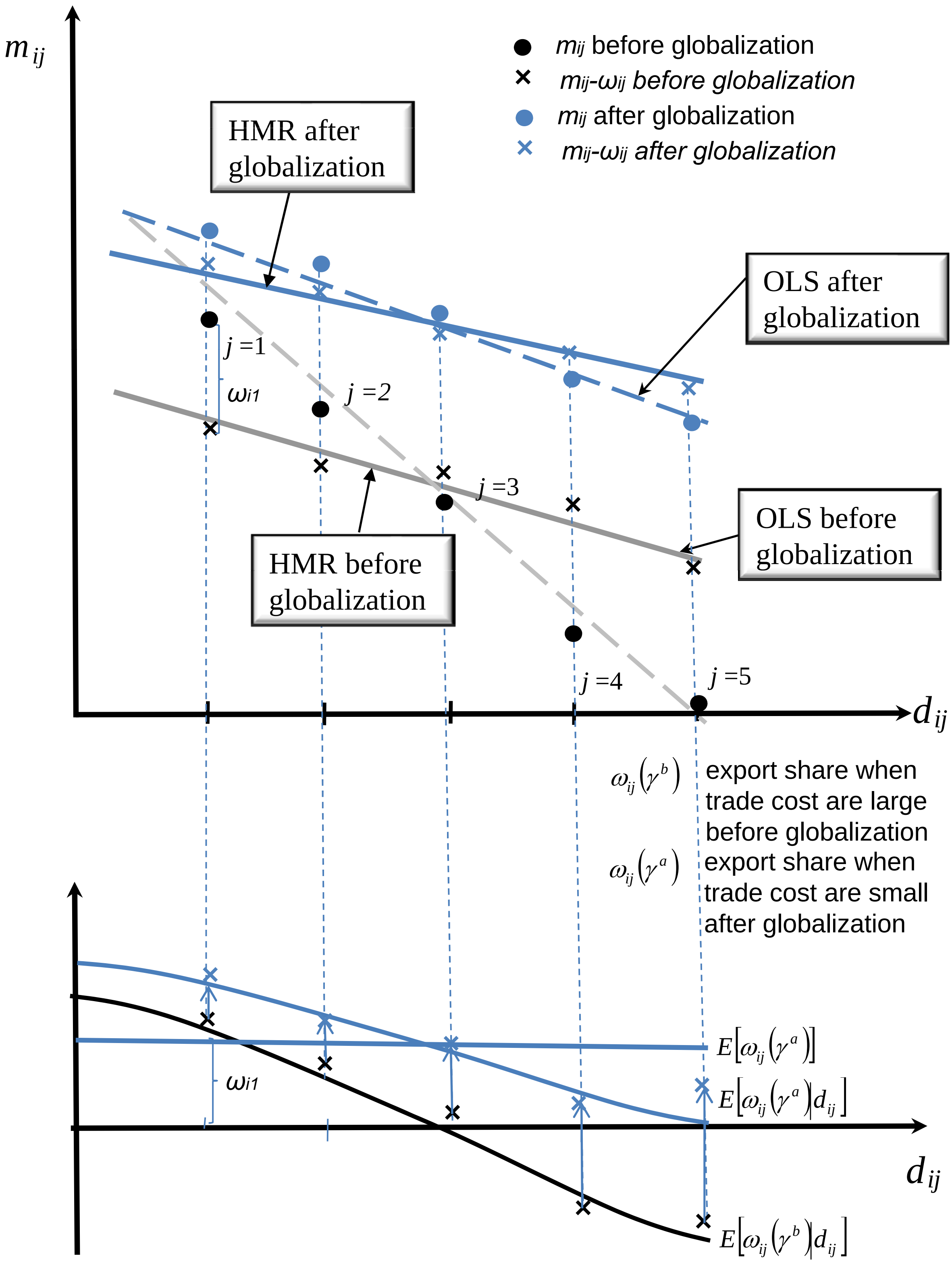

Figure 4: Illustrating the change in omitted variable bias of OLS

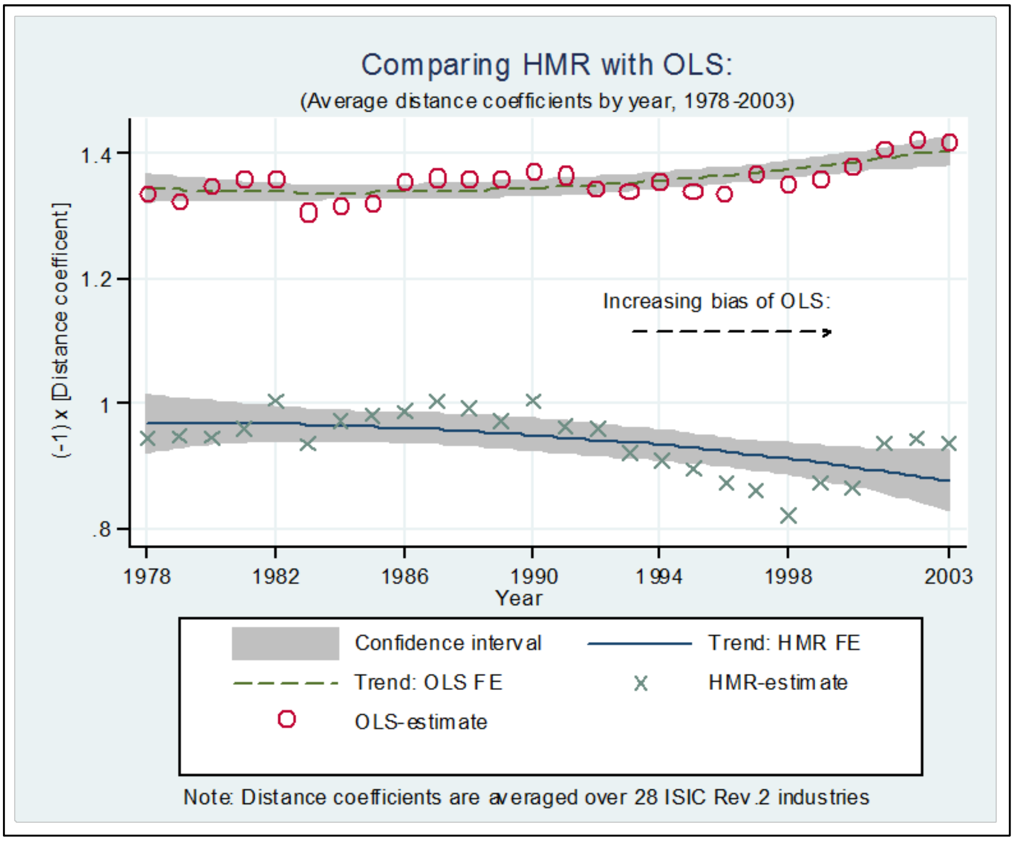

Figure 3 (i)

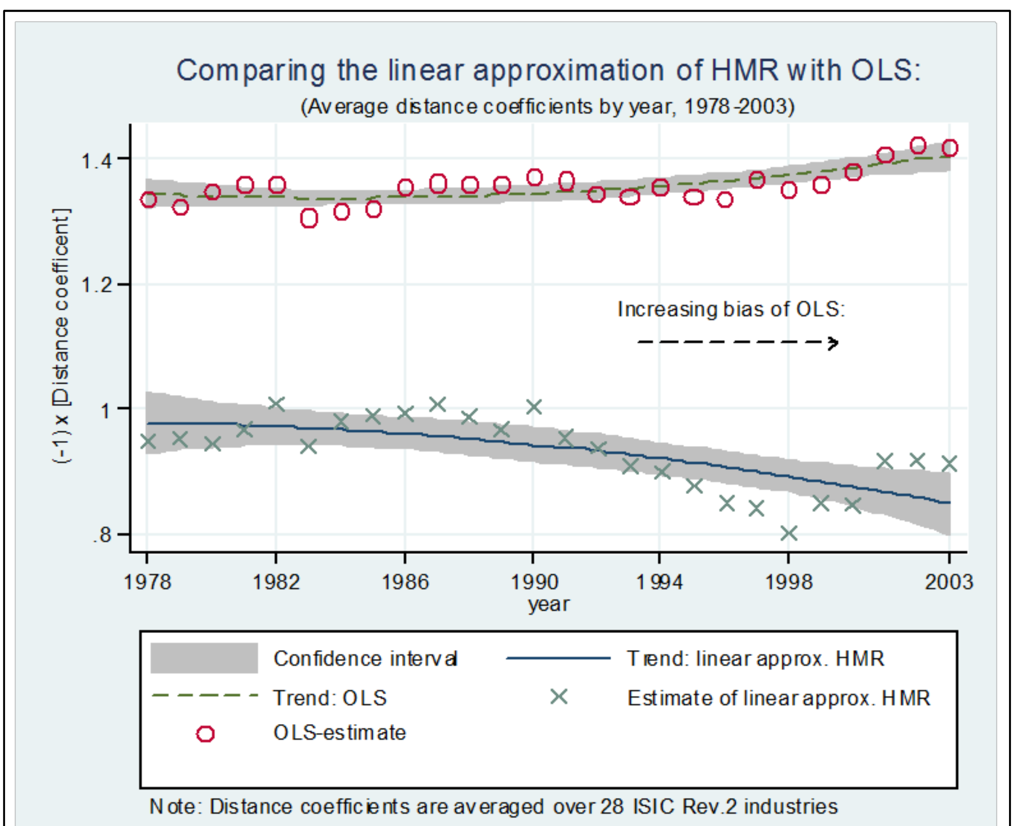

Figure 3 (ii)

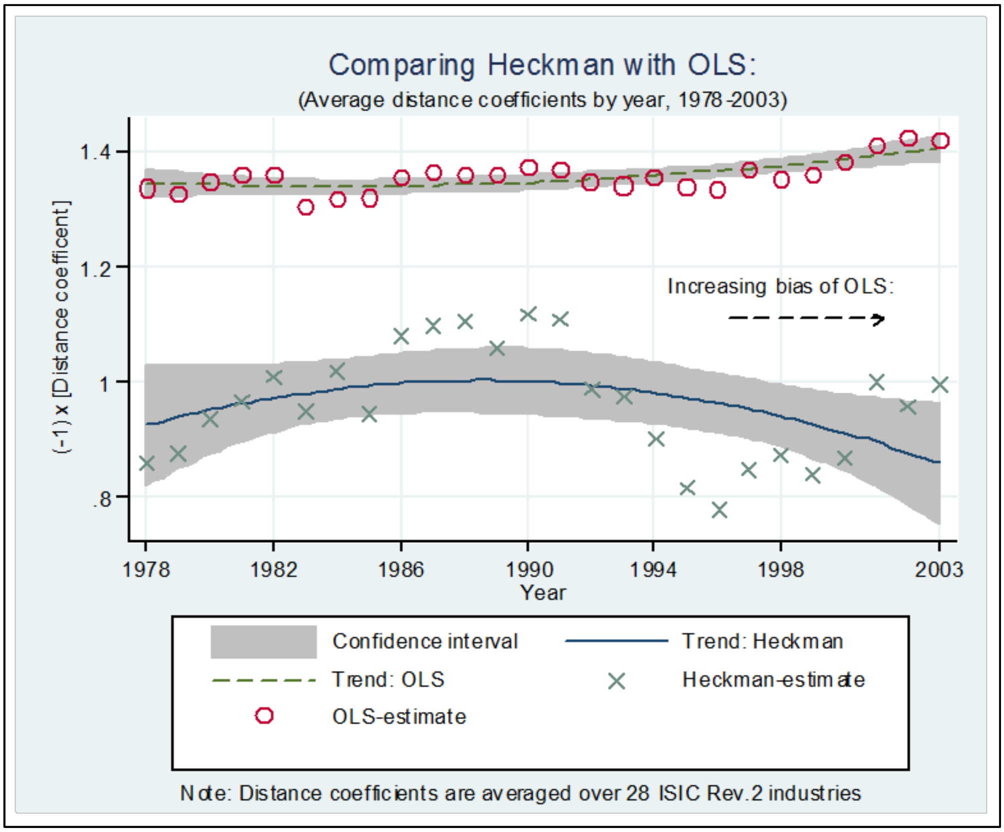

Figure 4 (i)

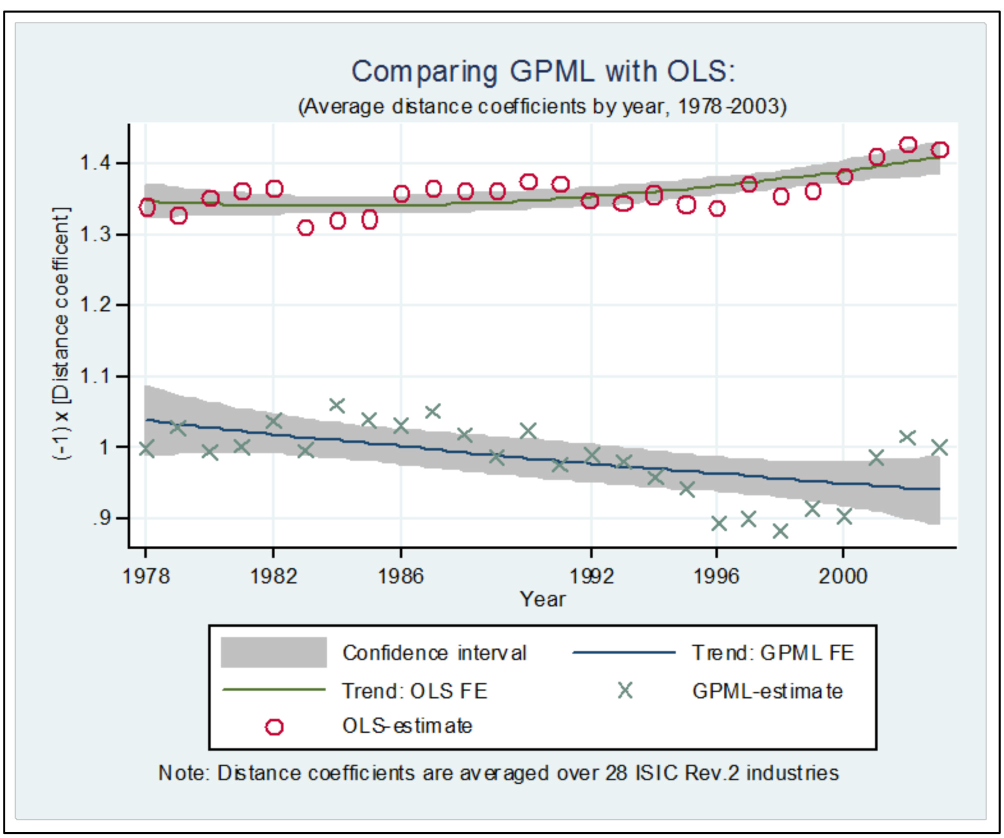

Figure 4 (ii)

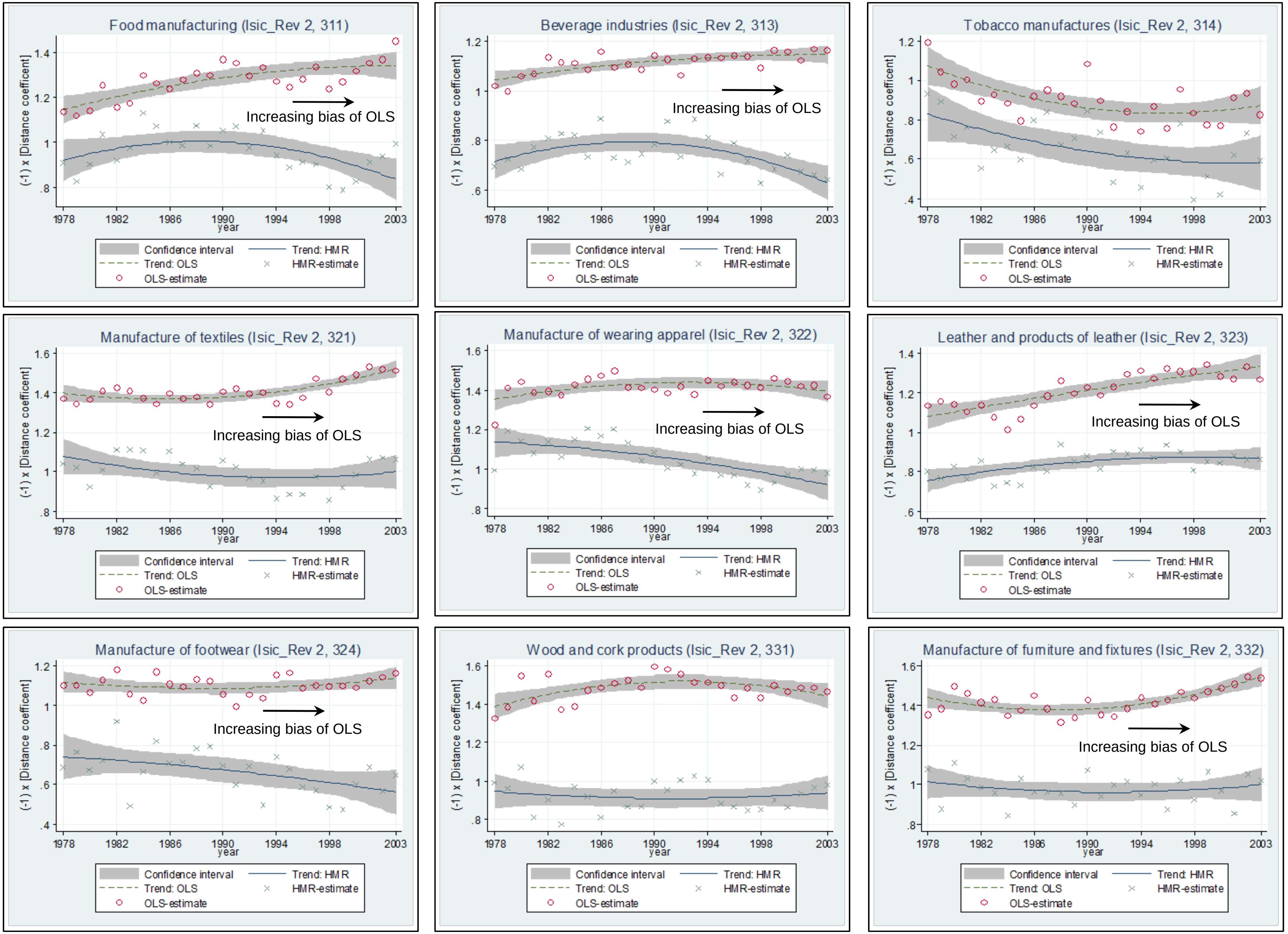

Figure 5a: Industry estimates for HMR and OLS , 1978-2003.

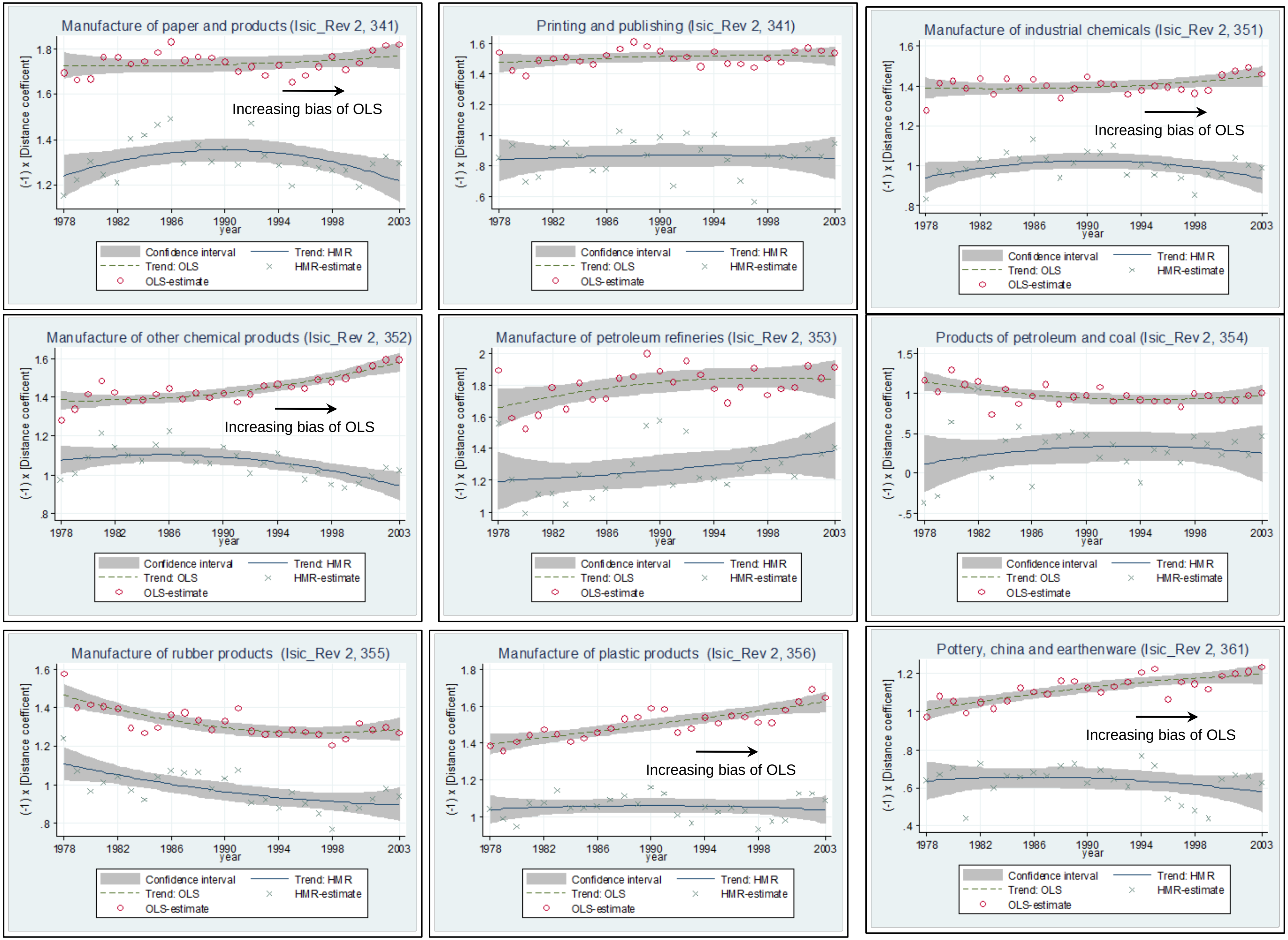

Figure 5b: Industry estimates for HMR and OLS, 1978-2003.

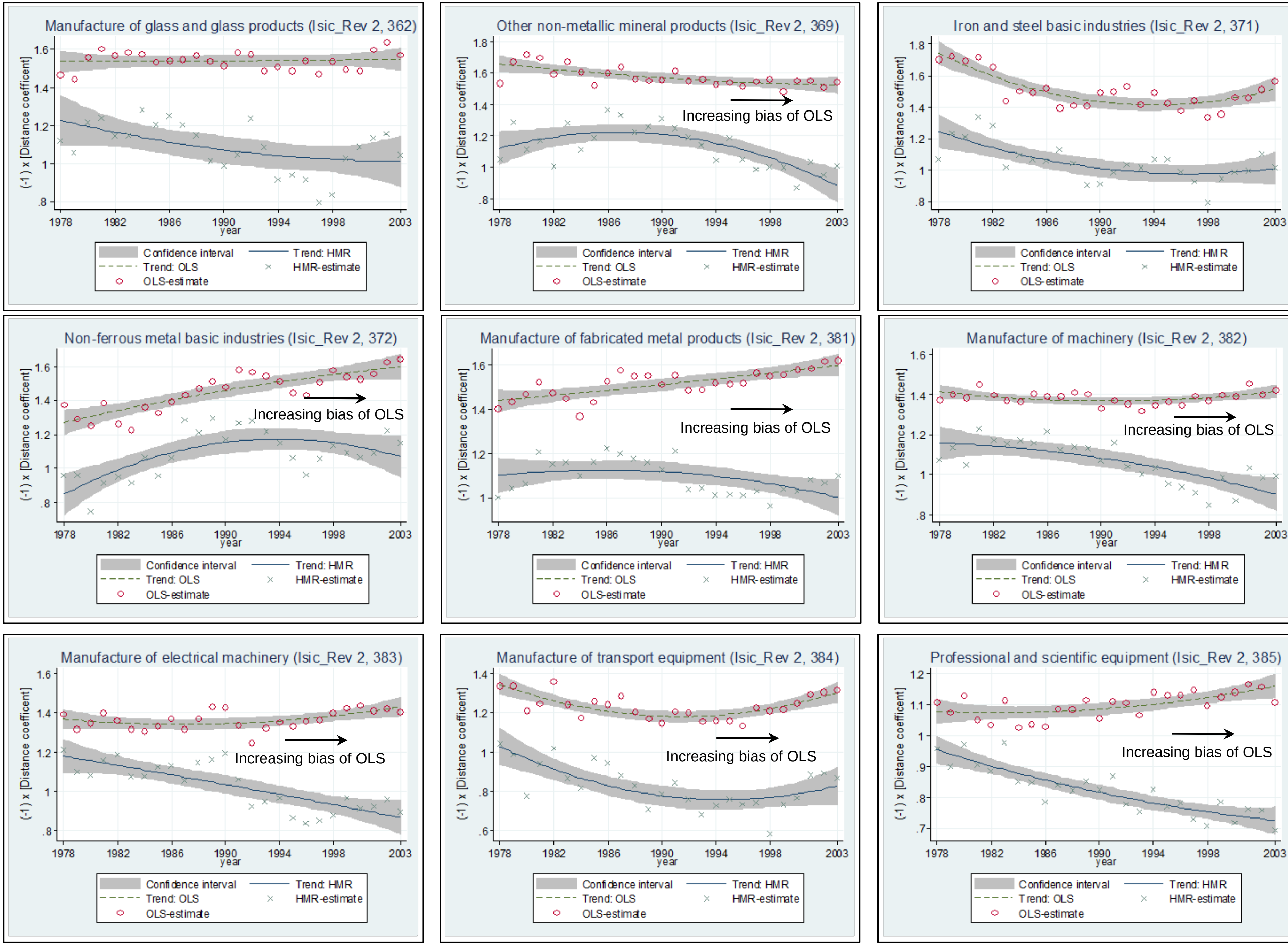

Figure 5c: Industry estimates for HMR and OLS, 1978-2003.

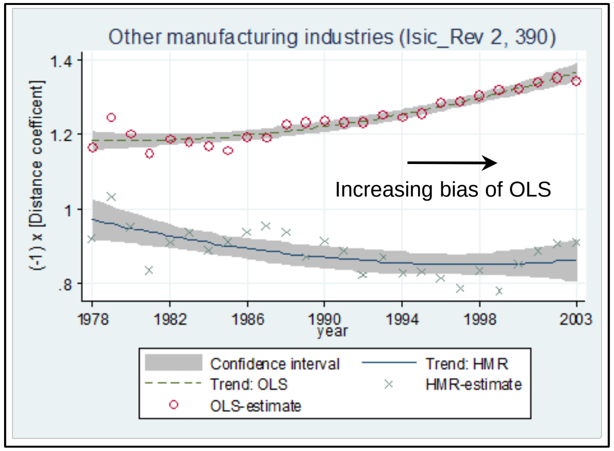

Figure 5d: Industry estimates for HMR and OLS, 1978-2003.

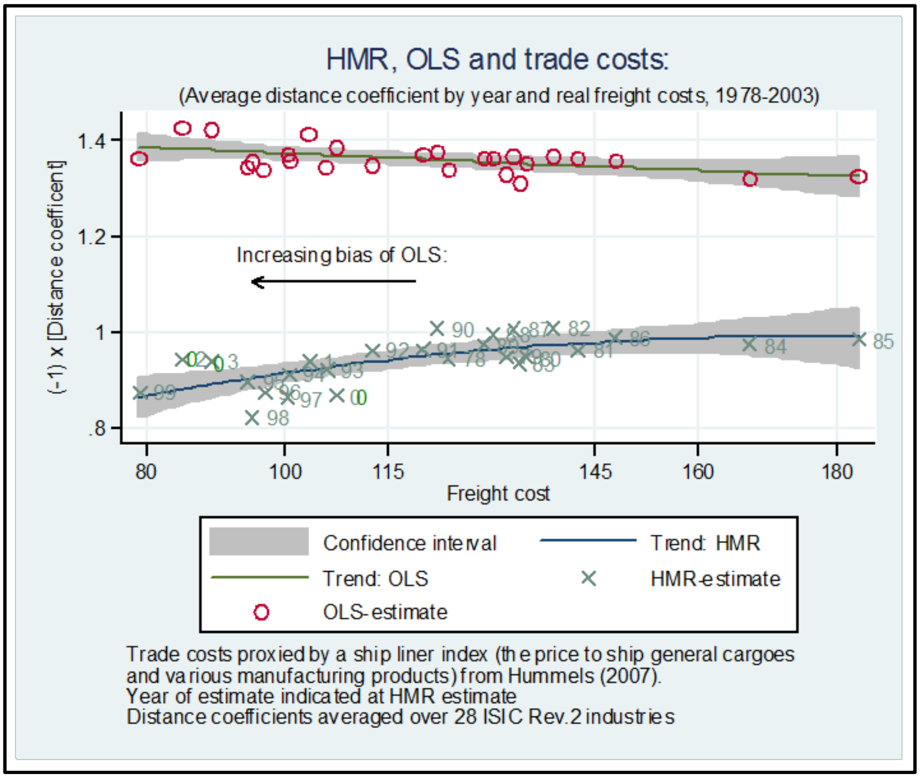

Figure 6 (i)

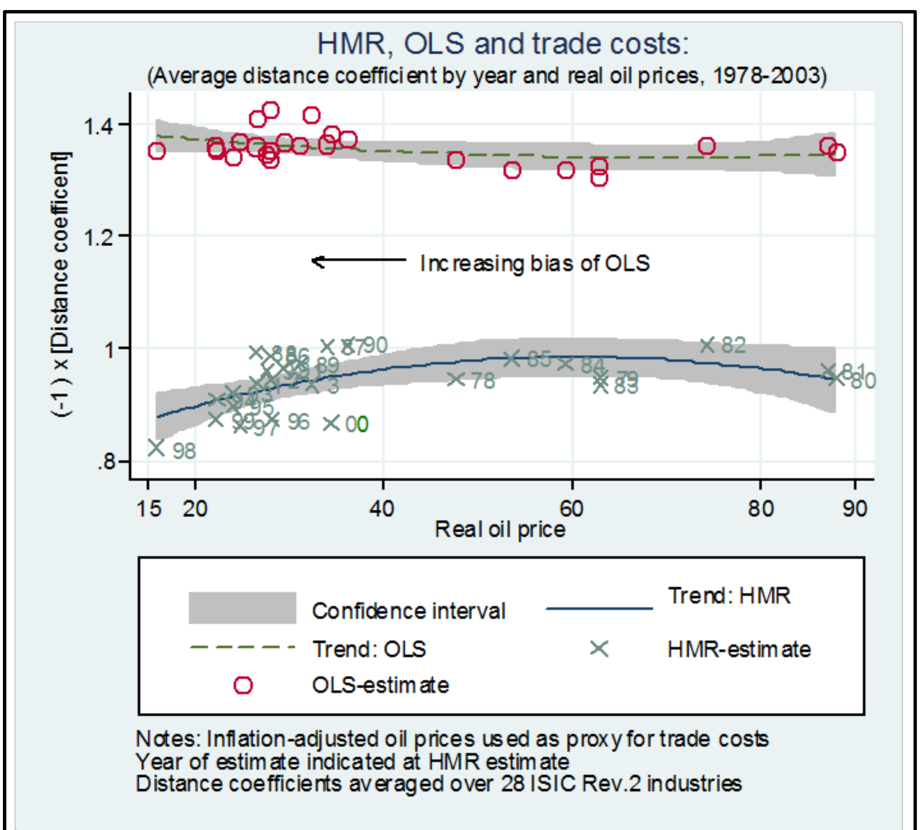

Figure 6 (ii)