

Xenia Matschke

Juan Rene Rojas Rodriguez

**Domestic Trade and its Measurement:
A Simulation Analysis**

**Research Papers in Economics
No. 10/25**

Domestic Trade and its Measurement: A Simulation Analysis

Xenia Matschke¹ Juan René Rojas Rodríguez²

October 27, 2025

Abstract

The inclusion of domestic (intra-national) trade when estimating structural gravity models is an important topic and has been shown to solve empirical puzzles such as the missing globalization puzzle (Yotov, 2022). Despite efforts to construct intra-national trade data, its measurement still remains a challenge, in particular for historical trade data. Campos et al. (2021a) recently claimed that the exact definition of domestic trade flows (gross vs. net production) is not important for the estimation of free trade agreement (FTA) effects in a gravity framework: a surprising finding, since the size differences between these measures are large, and their correlation of about 80%, while high, is not perfect. In a simulation framework, we revisit the question of the (non)-importance of whether to include and how to measure domestic trade and find some support for the conclusion of Campos et al. (2021a) concerning the estimation of trade agreement or tariff effects. However, with regard to the point estimates, marked differences arise depending on how domestic trade flows are calculated: the GDP-based domestic trade flows clearly distort the coefficient estimates. Interestingly, using only international trade flows yields results that are unbiased and as precisely estimated as those obtained when correctly measured (GO-based) domestic trade flows are included. Depending on what effects researchers are interested in, the basic gravity model without inclusion of domestic trade flows may thus be the preferred alternative after all.

JEL Classification Codes: F14, F15

Keywords: Intra-national Trade, Gravity Model

¹Universität Trier, Fachbereich IV, 54296 Trier, Germany, matschke@uni-trier.de.

²Universität Trier, Fachbereich IV, 54296 Trier, Germany; rojas@uni-trier.de

1 Introduction

The structural gravity model is the workhorse empirical model to estimate the effects of trade barriers and trade liberalization efforts on trade flows and to calculate their welfare consequences. While earlier specifications only included international trade flows, more recently, the importance of also including domestic (intra-national) flows has been stressed (Yotov, 2022). Whereas newest aggregate trade data sets construct such domestic trade flows from IO tables, using gross production and export values, many gravity specifications use proxies based on GDP, a net production value, due to limited data availability, in particular with regard to historical or developing country's gross production values or IO tables. In this article, a simulation study is conducted to evaluate in how far the inclusion of such proxies, or the omission of domestic trade flows, alters the coefficient estimates in the gravity framework with a special focus on international trade policy variables, such as free trade agreements (FTAs) or tariff policies.

Originally, the gravity model in economics was formulated as a direct analogue to Newton's theory of the forces attracting celestial bodies, and its empirical success in explaining trade flows appeared mysterious, and maybe even annoying to trade economists, given that the leading neoclassical trade models of comparative cost advantage, the Ricardian model and the Heckscher-Ohlin model, fared rather poorly when their predictions were put to the empirical test. However, at the same time when the first article of a series that established the New Trade Theory was published (Krugman, 1979), Anderson (1979) and later Eaton and Kortum (2002) and Anderson and van Wincoop (2003) were able to show that the gravity equation can be derived from trade-cost inclusive economic models both from the demand and from the supply side. From these theoretical advances and the following literature building on them, new results of what should and should not be included in a gravity equation emerged, often resulting in a solution for previously obtained empirical puzzles.

One such empirical puzzle was the so-called missing globalization puzzle: the estimates of the negative effect of geographical distance on trade flows seemed to be remarkably stable over time, although surely, globalization in the 20th century should have lowered the trade costs. However, Yotov (2012) showed that the missing globalization puzzle disappears once domestic trade enters the picture: trade even within a region or country is not frictionless, and in order to obtain the correct estimates for the impact of geographical distance, domestic trade needs to be accounted for. If domestic trade costs decrease at the same pace as international trade costs, then one should not expect the estimated coefficient on geographical distance to decrease, because international trade costs relative to domestic trade costs have not fallen. International trade costs may even appear to increase over time if domestic trade costs decline relatively more. In fact, there are many more reasons to include domestic (or in general intra-regional) trade in the gravity model: Yotov (2022) lists 15 (!) reasons for its inclusion. For example, without the inclusion of domestic trade flows, it would not be possible to evaluate, e.g., the impact of non-

discriminatory national import policies, since all international trade flows to the destination country would be affected by these policies. On the theoretical side, the main reason to include domestic trade flows is that the gravity model would otherwise not be closed (i.e., market clearing would not be ensured) if within-country trade were omitted. Therefore, domestic trade flows should be included in the gravity model if these data are available.

However, in contrast to import flows, domestic trade flows are not directly recorded at e.g., customs and hence need to be constructed from other data. Three methods are common in empirical work (Campos et al., 2021a):

1. Domestic trade is calculated as the difference of a country's GDP and its exports to the rest of the world. This proxy is not exact for two reasons: International exports are gross outputs, whereas GDP is calculated as value-added, i.e., a net output. Moreover, GDP includes services, whereas the recording of service trade has historically been spotty: most trade data sets thus still only include trade in goods, not in services.¹
2. Domestic trade is calculated as the difference of a country's gross production value and the sum of its exports to the rest of the world.
3. Domestic trade is calculated based on input-output tables and calculated as gross output sold by domestic industries to other domestic industries and final consumers.

Campos et al. (2021a) use GDP data for alternative 1, the OECD Trade in Value Added database (TIVA) for alternative 2, and the World Input Output database (WIOD) for alternative 3 and create a data set that covers 39 exporting countries and 63 importing countries for the period 1995-2011. The correlations of the log-growth rates of the obtained domestic trade flows are highest for WIOD and TIVA (88 %) and lowest for GDP and WIOD (78 %). What these really high correlations mask, is the considerable size difference in the domestic trade flow measures, however. Moreover, domestic trade flows dwarf international trade flows, regardless of which exact measure is used for the domestic trade flows: In the data set used by Campos et al. (2021a), the average international trade flow equals US\$ 2,040 millions, whereas the average domestic trade flow equals US\$ 855,445 millions when using the GDP data, US\$ 498,274 millions when using the TIVA data and US\$ 475,010 millions when using the WIOD data². The inclusion of domestic trade flows, despite the fact that only 663 of the 41,582 trade flows are domestic trade flows, dramatically alters the estimated effect of trade agreements, as shown

¹This begins to change, however: The WTO now offers a data set for service trade starting from 2005 onwards, see https://yamaraja.work/english/res_e/statis_e/services_trade_data_hub_e.htm.

²We would like to thank Jacopo Timini for sharing the data from his article with us. The cited averages of domestic and international trade flows can be found in Table A1 in Matschke and Rojas Rodríguez (2025).

in the upper part of table 2 of Campos et al. (2021a). Without inclusion of domestic trade, trade agreements do not seem to have a positive impact on trade flows, whereas with domestic trade flows included, the estimated effect of trade agreements is positive and significant at the 1 % level.³ However, when the trade agreement variable is subdivided into different kinds of trade arrangements, clear differences between GDP, TIVA, and WIOD based domestic trade flows become apparent. Moreover, when introducing a globalization time trend, the effects of even the aggregated trade agreement dummy differ in size: for example, the estimate based on GDP equals $e^{0.240} - 1 = 27.1\%$, whereas the TIVA based estimate equals $e^{0.186} - 1 = 20.4\%$ (Campos et al., 2021a, p. 3). While it is true that the estimates appear qualitatively similar as long as domestic trade flows are included, they are clearly not identical.

In our paper, we complement the existing literature by choosing a different route to investigate the importance of domestic trade flows for gravity estimation, and in particular for the estimation of the effects of international trade policy variables, such as tariffs and FTAs. Instead of using real-world international trade data and adding imputed domestic trade data based on international trade and national production data, we simulate the truth, i.e., we create explanatory variables and coefficients based on a meta-analysis of gravity models, and then calculate the trade flows, including domestic flows, based on these data with an included error component, constructing 1000 simulated trade data sets which we then use to estimate different versions of the gravity equation. Since we know the true coefficients, we can then compare the obtained estimates with the truth and see how different specifications and different ways to include or not include domestic trade data change the results. We find that while not introducing statistically significant bias, the usage of GDP-based domestic trade flows leads to lower absolute point estimates for international trade policy variables, such as free trade agreements or import tariffs. Perhaps more surprising, however, is another finding: in our simulation exercises, we cannot detect any advantage of introducing even correctly measured domestic trade flows compared to just omitting domestic trade altogether. Moreover, when omitting or mismeasuring explanatory variables, this often introduces bias, but this bias is not necessarily greater in the model that only uses international trade flows. While these results do not in any way invalidate the arguments of Yotov (2022) for using domestic trade in gravity estimation, they do present a caveat that mismeasurement of domestic trade flows may result in non-reliable estimates and that depending on the exact research question, gravity estimation with only international trade flows may lead to more reliable estimates than a domestic-trade inclusive model with questionable domestic trade data.

The remainder of this article is structured as follows. In section 2, we briefly review the structural gravity equation and how trade and production values enter it. In section 3, we explain how we construct the explanatory data and simulate the trade flows. Then, we show the estimation results for different variations of

³Matschke and Rojas Rodríguez (2025) provide an explanation for this surprising result.

the structural gravity equation based on these simulated data in section 4 and draw conclusions in section 5.

2 Structural Gravity Equation and the Inclusion of Domestic Trade

To close the general equilibrium model from which e.g. the cross-sectional structural gravity equation is derived, market clearing is essential. This means that exporter (origin country) o 's exports $\sum_d X_{od}$ to all countries, including country o itself, must add up to country o 's total gross output Y_o . In turn, the trade flow from exporter o to importer (destination country) d can be written as

$$X_{od} = \left(\frac{\alpha_o p_o t_{od}}{P_d} \right)^{1-\sigma} E_d, \quad (1)$$

where E_d denotes importer d 's expenditure, α_o is a utility function parameter that measures the popularity of a good originating from country o , p_o is the factory-gate price of the good in exporter country o , t_{od} denotes the bilateral trade costs between o and d and P_d is importer d 's inward multilateral resistance term (Anderson and van Wincoop, 2003), which can be roughly interpreted as d 's consumer ideal price index (Yotov et al., 2016, p. 14). After some transformations using market clearing, the central equation of the structural gravity model can be derived⁴, namely

$$X_{od} = \frac{Y_o E_d}{Y} \left(\frac{t_{od}}{\Pi_o P_d} \right)^{1-\sigma}, \quad (2)$$

where Y denotes world gross output ($\sum_o Y_o$) and Π_o denotes exporter o 's outward multilateral resistance term (MRT). Both MRTs are recursively defined, namely

$$\Pi_o^{1-\sigma} = \sum_d \left(\frac{t_{od}}{P_d} \right)^{1-\sigma} \frac{E_d}{Y}, \quad (3)$$

$$P_d^{1-\sigma} = \sum_o \left(\frac{t_{od}}{\Pi_o} \right)^{1-\sigma} \frac{Y_o}{Y}. \quad (4)$$

A few things are noteworthy: First of all, we notice that the trade flow from o to d is linear in Y_o , exporter o 's gross production value, and E_d , importer d 's expenditure. The trade flow also depends on the bilateral trade costs t_{od} which are taken to the power of $1 - \sigma$, the trade elasticity, where σ denotes the elasticity of substitution between goods which is assumed > 1 . This means that higher bilateral trade costs ceteris paribus reduce the trade flow. Moreover, the multilateral resistance terms, which are not directly measurable and need to be calculated recursively, both increase the bilateral trade flow. The reasoning behind this is that ceteris paribus,

⁴See Yotov et al. (2016, p. 14ff.).

higher multilateral resistances lower trade flows to or from all other countries and thus relatively increase the trade flow between any two given countries (Anderson and van Wincoop, 2003).

What about the domestic trade flows? Notice that (2) also holds if $o = d$, hence no structural difference between domestic and international trade flows exists. The one important difference is, of course, that the domestic trade costs t_{oo} are typically different from t_{od} where $o \neq d$. From this insight, it is clear that when constructing our simulated data set, the only difference between the construction of domestic and international trade flows is the exact specification of the bilateral trade cost function.

The domestic trade flow X_{oo} is not recorded directly, contrary to international trade flows. Instead, it is derived from the exporter production values, as was described in the introductory section. The production value, and in particular the gross production value, is typically by a magnitude bigger than the sum of international exports which need to be subtracted from the gross production value to calculate the domestic trade flow X_{oo} for country o . The production value as the dominant determinant of X_{oo} , on the other hand, is not only equal to Y_o which appears as an explanatory variable on the right-hand side, but is also closely linked to the exporter's expenditure E_o which also appears as a regressor on the right-hand side of the gravity equation. This latter fact means that there may exist a multicollinearity problem on the RHS between Y_o and E_d whenever $o = d$.⁵ A modern structural gravity equation, however, is not estimated by using Y_o and E_d as regressors. Instead, importer and exporter fixed effects are used to absorb all variables that are importer or exporter specific. In a panel setting, Y and E , but also the non-observable multilateral resistance terms Π and P , are exporter- or importer-time-dependent and would be controlled for by importer-time and exporter-time fixed effects.

However, the fact remains that on both sides of (2), the gross production value Y_o is the dominant variable whenever a domestic trade flow is considered. Hence, the question is what effect the inclusion of domestic trade flows will have for the estimates of the trade cost coefficients as the estimates of primary interest.

First of all, since the domestic trade flows tend to be much bigger than the international trade flows, it stands to reason that the fixed effect estimates⁶ will be bigger than would be otherwise the case. Since these fixed effect estimates are often not reported, this is not something that one would see directly in the regression output.⁷

⁵As a remainder, Y_o is also defined as the sum of domestic and external goods sales by domestic producers, $X_{oo} + X_o$. Moreover, ignoring services, E_o would be the sum of domestic and external purchases by domestic consumers, $X_{oo} + M_o$.

⁶Assuming that fixed effects are not differenced out.

⁷In some cases, the importer and exporter fixed effects are of interest. For example, the MRTs can be recovered from the fixed effects (Fally, 2015) when the model is estimated using the Poisson Pseudo Maximum Likelihood (PPML) method. These MRTs can then be used to compute the general equilibrium effect of a hypothetical change in bilateral trade costs (Yotov et al., 2016). Consequently, including domestic trade flows (or their proxies) is likely to alter both the estimated importer- and exporter-time fixed effects and the derived general equilibrium effects. Furthermore, in an earlier

Repercussions on the estimates of the bilateral trade cost components included in t_{od} are possible, however, namely if these are correlated with the importer and exporter fixed effects. Additional repercussions could be also expected on t_{od} if the variables used to measure international bilateral trade costs are defined with respect to the domestic trade costs⁸. In the following, using a simulation framework, we investigate the effects of including domestic trade flows on the gravity parameter estimates with a specific focus on the bilateral international trade cost components (i.e. import tariffs and FTAs).

3 Gravity model simulation

3.1 Simulation strategy

In order to simulate our synthetic trade flows, we rely on random variables that correspond to the explanatory variables of the structural gravity model depicted in section 2. The variables are generated to resemble their empirical counterparts, or at least their theoretical representation, in the case of unobservable variables. In some cases, our variables try to mimic the characteristics of the data used by Campos et al. (2021a). The data consists of a “structural” part, based on the relationship given by equation (2), and a “stochastic” part that corresponds to the error term of the model.

The “world” in our data consists of $N = 30$ countries or geographical entities that trade with each other during $T = 10$ periods. Our panel data thus consists of $30 \times 30 \times 10 = 9,000$ observations per simulation, and a total of 1000 simulations are run. Zero trade flows are not modelled in the data in order to concentrate on the question of domestic trade flows, so our trade flow variables will always be greater than zero.

3.1.1 General formulation of the simulations

The panel approach that we use here differs from the commonly performed simulations of the gravity model that rely only on cross-sectional data (see Head and Mayer (2014) and Egger and Staub (2016) for two examples of gravity simulation). While most of the gravity theory does not go beyond one period and early gravity regressions often include only cross-sectional data (see Anderson and van Wincoop (2003) and Eaton and Kortum (2002) as two influential examples), it is

version of Campos et al. (2021a), namely Campos et al. (2021b), different proxies for domestic trade flows also lead to different general equilibrium effects, probably since the size of these flows has a direct impact on the welfare calculation when it is computed using Arkolakis et al. (2012)’s formula.

⁸Two empirical examples can be found in Yotov (2012) and Bergstrand et al. (2015). The first article estimates the effect of international distance over time and compares it with the estimates of internal distance. The second article uses yearly international border dummies to capture factors that influence international compared to domestic trade flows. Different definitions of domestic trade flows are likely to have an influence on the estimates of such effects.

now common to use panel data and to include fixed effects to control for the multilateral resistance terms, various unobservable factors and endogeneity (Head and Mayer, 2014; Yotov et al., 2016). Some newer theoretical models also extend the gravity model to the multi-period case (Olivero and Yotov, 2012). In addition, using only cross-sectional data does not allow for the inclusion of pair-fixed effects in the regressions, which are often used to limit the endogeneity problems arising from the fact that the choice of trade policy instruments, like FTAs, may not only influence trade flows, but also may depend on them (Baier and Bergstrand, 2007). Therefore, it is reasonable to simulate trade data for several periods instead of only one period, and to include trade costs and production levels that vary over time. Based on equation (2), our simulated trade flows can be expressed as

$$X_{odt} = x_{odt} \eta_{odt} = \exp [\ln Y_{ot} + \ln E_{dt} - \ln Y_t - (1 - \sigma) \ln \Pi_{ot}] \times \exp [-(1 - \sigma) \ln P_{dt} + (1 - \sigma) \ln t_{odt}] \times \eta_{odt}. \quad (5)$$

The error term, η_{odt} , enters the model multiplicatively⁹ and is drawn from a log-normal distribution with mean one and variance equal to the inverse of x_{odt} . The other variables were already defined in section 2.

For the effect of the variables in (5) on trade flows (X_{odt}), we use the theoretical elasticities predicted by the gravity model presented in section 2: unitary positive elasticities for Y_{ot} and E_{dt} ,¹⁰ unitary negative elasticity for Y_t , and $(1 - \sigma)$, also known as the trade elasticity, for the multilateral resistance terms and trade costs. The value of $(1 - \sigma)$ is set equal to -5 , close to the mean trade elasticity reported in the meta-analysis of Head and Mayer (2014, p. 164) in Table 3.5 for the structural gravity models.

In addition, and to better reflect the practices used for estimating the gravity model, we divide our trade cost variable, t_{odt} , into five other variables that represent different factors influencing international and domestic trade costs. We define our trade costs as follows:

$$t_{odt} = (Int_Dist_{od})^{\delta_1} \times (Dom_Dist_{od})^{\delta_2} \times \exp(\delta_3 DB_{od} + \delta_4 FTA_{odt}) \times (1 + \tau_{odt})^{\delta_5}, \quad (6)$$

where Int_Dist_{od} is defined as a continuous variable, representing the geographical bilateral distance between partner countries. It is set equal to one in case of domestic trade ($o = d$). Dom_Dist_{od} is analogous to Int_Dist_{od} , but is used to represent the domestic distance when $o = d$. It is set equal to one in case of international trade ($o \neq d$). DB_{od} is a dummy variable equal to one in case of domestic trade flows ($o = d$). FTA_{odt} is a dummy variable that is equal to 1 if two countries $o \neq d$ have signed a free trade agreement. Finally, τ_{odt} is a continuous variable used to reflect the tariff charged by d on imports from o . When plugged

⁹As shown in Santos Silva and Tenreyro (2006, p. 643f), the error term of the model can be expressed multiplicatively or additively without affecting the estimation results.

¹⁰Our expenditure variable corresponds to the final expenditure (value added) on goods and services ($y_{dt} + z_{dt}$).

into equation (5), each δ_k will be multiplied by the trade elasticity. To simplify notation, these products are represented by λ_k (i.e., the coefficient on any of these trade cost variables is then $\lambda_k = (1 - \sigma) \times \delta_k$).

Substituting (6) into (5) and including the beta and lambda coefficients for clarity, the trade flows are given by:

$$X_{odt} = \exp [\beta_1 \ln Y_{ot} + \beta_2 \ln E_{dt} + \beta_3 \ln Y_t + \beta_4 \ln \Pi_{ot}] \times \exp [\beta_5 \ln P_{dt} + \lambda_1 \ln Int_Dist_{od} + \lambda_2 Dom_Dist_{od}] \times \exp [\lambda_3 DB_{od} + \lambda_4 FTA_{odt} + \lambda_5 \ln(1 + \tau_{odt})] \times \eta_{odt}, \quad (7)$$

where $\beta_1 = \beta_2 = 1$, $\beta_3 = -1$, $\beta_4 = \beta_5 = 5$. For our trade cost variables, we mainly base our coefficients on the values reported in Tables 3.4 and 3.5 of Head and Mayer (2014) for the Structural Gravity. We set our lambdas as follows: $\lambda_1 = -1$ (Head and Mayer (2014, p. 26) report a mean of -1.1 for distance) and $\lambda_2 = -0.5$ to represent the smaller internal trade costs; $\lambda_3 = 2$ (mean of 1.9 for home trade); $\lambda_4 = 0.35$ (mean of 0.36 for trade agreements); $\lambda_5 = -5$ (mean of -5.13 for trade elasticity). Table 2 summarizes the information about the coefficients that were used for the construction of the simulation data.

[Insert Table 2]

3.1.2 FTAs and trade diversion

In the context of the gravity estimation, λ_4 represents the change in the trade flow X_{odt} after signing an FTA. For domestic trade flows, the FTA variable is set equal to zero. An FTA reduces the bilateral trade costs of the signatory countries o and d . This effect occurs regardless of the size of domestic trade flows compared to international trade flows, and even when domestic trade flows are omitted from the estimation sample.

3.1.3 Domestic trade flow proxies

In the following construction of variables, we simplify the description and consider goods as tradable and services as non-tradable, notwithstanding the fact that in reality, non-tradable goods and tradable services certainly exist. Exports are then based on the gross production value of (tradable) goods, whereas GDP is the net value of (tradable) goods and (non-tradable) services production. Three types of domestic trade flows are considered for the simulations:

1. The true domestic flows generated by the gravity model in equation (7).
2. Domestic flows computed as the difference between the GDP (value added) of o and its total gross exports during a given year: $X_{oot} = E_{ot} - X_{ot}$. This is one of the methods used in Campos et al. (2021a), which is also very common in the literature since GDP values are easy to obtain. This

approach has two main problems, however: 1) it mixes gross value and value added observations (gross exports vs. GDP), and 2) it mixes (non-tradable) services and (tradable) goods from the GDP with the tradable goods of gross exports.¹¹

3. The difference between the gross output of o and its total gross exports during a given year: $X_{oot} = Y_{ot} - X_{ot}$, where $X_{ot} = \sum_{d \neq o} X_{odt}$. This approach is also used in Campos et al. (2021a). In our data, it should correspond to the true domestic trade flows generated by the model, with a small deviation coming from the error term η_{odt} when summing all X_{odt} exports of o .

3.2 Data Construction

The random variables in (7) are generated as follows:

Y_{ot} : In period $t = 1$, country o 's gross (tradable) goods output Y_{o1} is drawn from a gamma distribution with shape parameter $\alpha = \frac{420,000^2}{800,000^2} = 0.275625$ and scale parameter $\theta = \frac{800,000^2}{420,000} = 1,523,809.524$. We also add a constant equal to 2,000 to avoid having numbers too close to zero and having a fixed lower bound. The shape and scale parameters are chosen to approximate the average and standard deviation of the gross production variable (TIVA) of Campos et al. (2021a) during the year 1995 (mean 435,720.7 and standard deviation of 814,979, both in millions of US \$, see Table 1).¹² For $t > 1$, Y_{ot} grows with respect to Y_{o1} by an annual factor such that $Y_{ot} = Y_{o1} \times (\prod_{k=2}^t (1 + \kappa_{ok}))$, where κ_{ok} is a random number drawn from a normal distribution with a mean of 0.025 and a standard deviation of 0.15. The parameters of the normal distribution are chosen to mimic the mean and standard deviation values of the log-difference of the gross production (TIVA) in Campos et al. (2021a) (0.052 and 0.15, respectively, see Table 1).

Y_t : It is simply the sum of all countries' annual gross production values in a given year t : $(Y_t = \sum_{o=1}^N Y_{ot})$.

ω_{ot} : It represents the share of value added (y_{ot}) in gross output (Y_{ot}) in goods and is derived from a gamma distribution with shape parameter $\alpha = \frac{0.45^2}{0.1^2} = 20.25$ and scale parameter $\theta = \frac{0.1^2}{0.45} = 0.02\bar{2}$, reflecting that value added is smaller than the gross value of production. The mean and standard deviation used for ω_{ot} are based on the share of GDP to GO for a sample of 30

¹¹Most data sets about bilateral international trade still do not include information about service trade.

¹²The mean of the gamma distribution is given by $\alpha\theta$ and its variance by $\alpha\theta^2$. For a given variable, α is calculated as the square of the mean divided by the variance, and θ as the variance divided by the mean.

countries in 1995 (mean = 53.22% and standard deviation = 0.09953).¹³

Z_{ot} : It represents the gross production value of services and is generated as a proportion of the gross production value of goods for the first period of simulated data: $Z_{o1} = \iota_{o1} \times Y_{o1}$. ι_{o1} is generated using a gamma distribution with shape parameter $\alpha = \frac{0.45^2}{0.1^2} = 20.25$ and scale parameter $\theta = \frac{0.1^2}{0.45} = 0.022$.¹⁴ For the following years, $Z_{ot>1}$ grows relative to Z_{o1} by an annual factor such that $Z_{ot} = Z_{o1} \times (\prod_{k=2}^t (1 + \nu_{ok}))$, where ν_{ok} is a random number drawn from a normal distribution with a mean of 0.08 and a standard deviation of 0.1.

z_{ot} : It represents the value added in service production. Similarly to the value added in goods production, it is generated as $z_{ot} = \psi_{ot} \times Z_{ot}$. ψ_{ot} represents the share of value added (z_{ot}) in gross output (Z_{ot}) in services and is derived from a gamma distribution with shape parameter $\alpha = \frac{0.55^2}{0.1^2}$ and scale parameter $\theta = \frac{0.1^2}{0.55}$. These parameters mirror the observed ratio in the cited data of the World Bank and the UN Statistics Division for the same sample of countries.

E_{dt} : It is defined as the sum of value added in goods production and value added in services production of country d : $E_{dt} = \omega_{dt} \times Y_{dt} + \psi_{dt} \times Z_{dt} = y_{dt} + z_{dt}$.

Π_{ot} and P_{dt} : The multilateral resistance terms are solved iteratively/recursively and simultaneously using the values generated for Y_{ot} , E_{dt} , Y_t and t_{odt} , following equations (3) and (4). As Π_{ot} and P_{dt} can be only solved up to a scalar (Anderson and Yotov, 2010), we choose P_{1t} (value for country 1) as the reference value for each period.¹⁵

Int_Dist_{od} : The bilateral distance between o and d is drawn from a gamma distribution with a shape parameter of $\frac{8,519.015^2}{4,681.06^2} = 3.312$ and a scale parameter of $\frac{4,681.06^2}{8,519.015} = 2,572.17$. The shape and scale parameters are chosen to approximate the average international bilateral distance (“distw”), according to the GeoDist dataset of CEPII (Mayer and Zignago, 2011). The variable is set to one in case of domestic trade.

Dom_Dist_{od} : The distance between o and d when $o = d$ is also drawn from a gamma distribution. The shape and scale parameters are changed to $\frac{182.8028^2}{229.932^2} =$

¹³We calculate these shares using GDP data from the World Bank (“World Development Indicators”) and data from the UN Statistics Division (“Output, gross value added and fixed assets by industries at current prices (ISIC Rev. 4)”) for the gross output. The sample includes Austria, Belgium, Bulgaria, Canada, Croatia, Cyprus, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, Romania, Spain, Sweden, Switzerland, United Kingdom, and the United States.

¹⁴These parameters mirror the ratio of services to goods in the gross output data provided by the World Bank and the UN Statistics Division for the same sample of countries.

¹⁵More details about the iterative process are given in Appendix 6.12.

0.6321 and $\frac{229.932^2}{182.8028} = 289.21$, respectively, to reflect the smaller internal distance, as in the GeoDist dataset of CEPII. The variable is set to one in case of international trade.

FTA_{odt} : To represent trade agreements, a dummy variable equal to one if country pair od is part of an FTA during period t and zero otherwise is used. In period $t = 1$, the variable is drawn from a binomial distribution with a probability of 0.2 that each symmetric pair (od and do with $d \neq o$) is part of the same FTA, close to the average value of the general trade agreement dummy variable used in Campos et al. (2021a) in its first year (1995), 0.3573. In the following periods, the country pairs that are already in an FTA remain in it, and the pairs currently not in an FTA can sign an FTA with probability 0.1 (binomial distribution). The FTA variable is always equal to zero in case of domestic trade.

τ_{odt} : It represents the ad valorem tariff imposed by d on imports from o in period t and is set equal to zero in the case of intra-national trade. In period $t = 1$, each country d sets τ_{odt} with each partner o following a gamma distribution with a shape parameter of $\frac{0.083^2}{0.03^2} = 7.6544$ and a scale parameter of $\frac{0.03^2}{0.083} = 0.0108$. The shape and scale parameters are chosen to approximate an average tariff rate of 8%, as in the Campos et al. (2021a) data for 1995 (see Table 1 where the importing countries' MFN tariff average equals 8.3%). For the periods $t > 1$, d changes τ_{od1} by an annual factor such that $\tau_{odt} = \tau_{od1} \times (\prod_{k=2}^t (1 + h_{odk}))$, where h_{odk} is normally distributed with a mean of -0.03 and a standard deviation of 0.09, so that the tariff rates decrease on average by 3% each period. In the case that d and o have signed an FTA in period t , τ_{odt} is set equal to zero to represent the trade liberalizing effects of trade agreements. By construction, the simulated trade agreements are all bilateral (only include o and d).

$DOM(X)_{MODEL_{oot}}$: represents the “true” domestic trade flows generated from (2) as $X_{oot} = \frac{Y_{ot}E_{ot}}{Y_t} \left(\frac{t_{oot}}{\Pi_{ot}P_{ot}} \right)^{(1-\sigma)}$.

$DOM(X)_{GDP_{oot}}$: represents the domestic trade flows calculated from GDP, namely generated as $E_{ot} - \sum_{d \neq o} X_{odt}$.

$DOM(X)_{GO_{oot}}$: represents the domestic trade flows calculated from gross goods output, generated as $Y_{ot} - \sum_{d \neq o} X_{odt}$. These trade flows are approximately equal, but not identical to the model trade flows $DOM(X)_{MODEL_{oot}}$.

Tables 3, 12 and 13 show the summary statistics and the correlation matrix of our full simulated data set, respectively. The values shown in the tables are calculated using the total data, i.e., 9000 observations per simulation times 1000 simulation runs. The correlation coefficients are typically very small, with the notable exception of the correlation between the MRTs (approximately -0.5) and the FTA and

tariff variable (approximately -0.8). The latter strong correlation is caused by the fact that whenever the FTA variable is generated as 1, the tariff rate is automatically set to 0. In practice, the correlation is probably weaker because FTAs tend to be formed between countries which already have low tariff rates directed against each other since the trade liberalization process is gradual.

[Insert Table 3]

[Insert Table 12]

[Insert Table 13]

4 Simulation Results

In this section, we examine different scenarios for which the omission or mismeasurement of domestic trade flows, sometimes in conjunction with other problems, such as e.g. the omission of the multilateral resistance terms, may cause problems when estimating the coefficients of the gravity model. All models are estimated using the Poisson Pseudo Maximum Likelihood (PPML) method, given that the model, as predicted by theory, was generated in its multiplicative form, with an error term directly proportional to the expected value of the dependent variable.¹⁶ The coefficients and their standard errors presented in this section are the mean values obtained from the regressions of 1,000 replications of the data generation procedure described in section 3. We also include the average p-value of each coefficient in square brackets, testing if the estimated coefficient of each variable is statistically different from its true value (see Table 2).

We have divided this part into several subsections to explore different approaches used in the literature for the estimation of the gravity model of trade. We start with the canonical model that includes all variables that were used to simulate the trade flows without the inclusion of fixed effects. Then we move towards the naive gravity model that omits the multilateral resistance terms (the so-called gold medal error of gravity estimation) and also investigate the impact of using GDP both for the importer expenditure and exporter output variable as well as the effect of using only one distance variable instead of distinguishing between international and intra-national distance.

In a next step, we switch to a model with exporter-time and importer-time fixed effects which absorb all country-time specific variables, i.e., multilateral resistance terms and exporter output and importer expenditure. Also in this setting, we investigate what happens if only one distance variable is used. Finally, we additionally introduce pair-fixed effects to obtain the gravity specification that is typically used in the literature if the focus lies on the impact of international trade policy.

¹⁶As shown in Santos Silva and Tenreyro (2006), the log-linear transformation commonly used to estimate the gravity model by OLS can lead to significant bias in the obtained coefficient estimates.

In all model specifications, we investigate how the handling of domestic trade flows alters the coefficient estimates and whether in case of incorrect model specification, it adds to or lessens any present estimation bias. When referring to bias, we distinguish between two forms: a clearly discernible difference between the size of the known true coefficient and the obtained estimate and a statistically significant difference between estimate and truth, manifesting itself as a p-value of less than 0.1 (statistical significance at least at the 10 % level). Obviously, with high standard errors, an apparent bias in the sense of a clear size difference between estimate and truth may not constitute a statistically significant bias, whereas on the other hand, with small standard errors, a statistically significant bias may arise even though estimated and true coefficient appear close.

The way how domestic trade flows are measured is indicated at the bottom of each table. The true domestic trade flows, created as part of the model simulation, are the “model” domestic trade flows. These domestic trade flows are very close, but not identical to those labelled “GO”, which are calculated as the difference between an exporter’s tradable share of domestic gross output minus the sum of the exporter’s international exports. The clearly incorrect measure of domestic trade flows (labelled “GDP”) uses exporter GDP as production measure and then subtracts all international exports. Since the difference between GDP and total gross exports may be negative¹⁷, we deal with this problem by simply censoring the negative observations at zero. Finally, the label “NO” designates the case where domestic trade flows are simply excluded from the regressions, as used to be the standard procedure in traditional gravity regressions.

Table 3 provides an overview of the simulated data. Of the 9 million observations, only 300,000 (3.3 %) are domestic trade flow observations. For these observations, a comparison between model trade flows and GO-based trade flows shows very little difference, although it was not possible in the simulation to actually “close” the model and force the model-generated domestic trade flows to be exactly equal to the difference of tradable gross output and all international sales exporter by exporter, as should be theoretically the case. The GDP-based domestic trade flow average exceeds the model and GO-based trade flows by far and also features a much higher standard deviation. Comparing the international trade flows (INT_X_{odt}) to the domestic trade flows, we see that all domestic trade flow versions dwarf the international trade flows by an average factor of at least 230. With regard to the other country-time varying variables, we see that the exporter output measure Y_{ot} is on average considerably lower than the importer expenditure measure E_{dt} because the latter includes services: it is also true that the exporter measure is a gross value and the importer measure a net value, but this difference is secondary compared to the services difference. Finally, the international distance exceeds the domestic one by a factor of more than 1,000.

¹⁷Remember that we have generated the GDP observations as a share of the total gross output of each country, which could result in total exports (measured in gross values) greater than our simulated GDP observations, see section 3.

4.1 Canonical Gravity Form

We start with the simulation results of the true model, i.e., we include all the data-generating variables in the regressions in order to verify that the PPML estimation can recover the true model coefficients. From this starting point, we can identify any potential issues before proceeding with the other simulations. The results for the model-generated trade flows, shown in column M1 of Table 4, are as expected: The true coefficients are estimated without any problems and with small standard errors, indicating precise coefficient estimates. For none of the coefficients is the null hypothesis of no statistical difference between the estimated and the true coefficient rejected (in fact, all p-values lie above 0.97). The same is true for column M3 where domestic trade flows are calculated as tradable gross output minus the sum of all exports. Since the domestic trade flows in M3 are very similar to the model-generated ones in M1 and should be theoretically identical, this result is not surprising and shows that the utilization of the gross output-based measure for domestic trade flows delivers precise and unbiased estimates.

In column M2 of Table 4, we use the proxy of domestic trade flows generated by using GDP instead of tradable gross output. All estimated coefficients clearly differ from the true values, and the standard errors are, in comparison to column M1, relatively larger. With regard to the sign of the deviation from the truth, we see that the coefficient on exporter output is smaller and the coefficient on importer expenditure is higher. In contrast, most other estimates are smaller in absolute value. For example, the point estimate on the world production value is less than half the true size. Moreover, both multilateral resistance term coefficients and also the distance coefficients are (absolutely) smaller than their true counterparts. The same holds for the international trade policy (FTAs, tariffs) variable coefficients, whereas the estimate on the domestic trade dummy exceeds its true value, picking up the inflated domestic trade values. A separate question, however, is in how far these apparent estimation differences are statistically significant. Checking out the p-values in square brackets, we see that for some variables, the true values do lie within the confidence bands. In particular, no statistically significant bias is present for the coefficients on the international trade policy variables FTA and $\ln(1 + \tau)$. The simulation exercise thus confirms the finding of Campos et al. (2021a) that the exact calculation of the domestic trade flow proxy does not alter the trade policy effect estimates in a statistically significant way. Yet, the point estimates are different, and the inclusion of the GDP-based domestic trade flows seems to lead to an underestimation of the effect of free trade agreements and tariffs. For the FTA effect, this result appears reasonable since the FTA variable for the high domestic trade flows is set to 0. Several other coefficient estimates, however, are also statistically different from their true value when the GDP-based domestic trade flow measures are employed. In particular, the coefficients of the variables $\ln(Y_{ot})$, Π_{ot} , $\ln(wY_t)$ and even the variable $\ln(Dom_Dist_{od})$, that was especially introduced to account for domestic trade flows, exhibit a statistically significant downward bias, whereas $\ln(E)_{dt}$ and DB_{od} are upward-biased.

We now exclude all domestic trade flows from the model and re-estimate it to mimic traditional/common gravity estimations, which lack domestic trade observations. As shown in column M4 of Table 4, the results of excluding these observations from the regressions are equivalent to those of column M1: The true coefficients are estimated without problem and with relatively small standard errors. This is not surprising, given that only a few observations (3.3 %) are omitted from the data set during each iteration/repetition, and all variables that explain international trade remain in the estimation.¹⁸ The variables FTA_{odt} and $\ln(1 + \tau_{odt})$ are equal to zero in case of domestic trade, and $\ln(Y_{ot})$, $\ln(E_{dt})$ and $\ln(wY_t)$ have all the same effect on domestic and international trade, so the omission of domestic trade flows does not affect the estimation of the “international” variable coefficients. The standard errors of the coefficients are comparable to the versions with the “right” domestic trade flows (M1 and M3), in contrast to the model where the GDP-based domestic trade flow proxy is included, which results in higher standard errors and lower p-values.

In conclusion, the results confirm that the PPML estimator is appropriate to estimate the model. Moreover, we find that the GDP-based proxies for domestic trade lead to statistically significant bias in many of the estimated coefficients, even when including all the relevant variables. For the international trade policy variables, however, the bias in the GDP-based model is not significant, although the estimated coefficients clearly differ from the true values. In contrast, the model with only international trade flows (NO, M4) performs just as well as the “right” domestic trade flow inclusive variants (M1 and M3).

[Insert Table 4]

4.2 The Gold Medal Error

We continue by evaluating the effects of including domestic trade flows in conjunction with the so-called “Gold Medal Error” of gravity estimation, so named by Baldwin and Taglioni (2006), which consists of omitting the multilateral resistance terms when estimating the gravity model. As by construction, the MRTs are correlated with the other explanatory variables in the model¹⁹, their exclusion should induce a bias in the estimation of all coefficients, as pointed out in Anderson and van Wincoop (2003), regardless of whether or not domestic trade flows are included in the estimation. Using the same simulated data, we repeat the regression specifications shown in Table 4, with the only difference that now the multilateral

¹⁸It is not possible to estimate the coefficients of the domestic trade cost proxies, as they are always equal to zero when domestic trade observations are omitted. The MRT variables are the correct ones, i.e. their calculation employs the true domestic trade flows.

¹⁹See equations (3) and (4). With regard to our simulated data set, the correlation between the explanatory variables appears quite small, with the exception of the correlation between the exporter and importer MRT, which takes a value of about -0.5, and the correlation between the FTA and the tariff variable (-0.8) as can be seen from Tables 12 and 13.

resistance terms are omitted. The results are shown in Table 5, following the same order as in the previous table.

The first three columns of Table 5 show the (severe) bias that the MRT omission induces: Most of the estimated coefficients are biased compared to their true value. Curiously, the exclusion of domestic trade flows from the regression (last column M8) appears to help mitigate the omitted variable bias of many of the coefficients compared to the domestic-trade inclusive versions in the other columns of Table 5. The reduced bias in M8 is for example apparent in the point estimate for importer expenditure, and even when the coefficient estimate bias is significant, the point estimates are closer to the truth than is the case in the other columns.

However, when checking out the coefficients for those variables that trade economists are most interested in, namely the international trade policy variables FTA_{odt} , and $\ln(1 + \tau)_{odt}$, the picture changes. While bias is apparent when looking at the point estimates, the difference to the true value is not statistically significant in any of the models. For the FTA variable, we see that the coefficient estimate in all estimations equals about 0.16, less than half the true value of 0.35. But the difference is statistically insignificant. With regard to the point estimate difference to the truth, we note that the models with the “right” domestic trade flows (M5 and M7) perform very marginally better than the model with GDP-based domestic trade flows (M6). The “best model” for estimating the FTA coefficient appears to be the one with only international trade flows (M8), but the quality difference is really small. For the import tariff, M5 and M7 exhibit a slight, but insignificant downward bias in the absolute value of the coefficient estimate, whereas the estimate for the GDP-based trade flow (M6) is absolutely smaller and further away from the truth. This time, the model with only international trade flows (M8) performs the worst: the size of the bias in M8 when domestic trade flows are omitted appears quite substantial (estimate -3.02 vs. truth -5), but even there, the difference is not statistically significant.

We thus conclude that the omission of the multilateral resistance terms biases the coefficient estimates, but this bias, while often substantial in size, is not statistically significant in any of the estimated models, as far as international trade policy effects are concerned. We confirm, however, the conjecture in Yotov (2022) that the inclusion of domestic trade helps obtain more accurate point estimates of these variable coefficients, with the GO-based domestic trade flow inclusion yielding closer-to-the-truth estimates.

[Insert Table 5]

4.3 GDP instead of gross output

Early gravity models used exporter and importer GDP in the regression. In theory, however, exporter gross output and importer expenditure, a net value, should be included. In the model simulation, Y_{ot} is defined as exporter tradable gross output

and E_{dt} as importer GDP. In the next exercise in Table 6, the effect of replacing Y_{ot} by exporter GDP is investigated.

Checking out the results for the regressions with the “right” domestic trade flows (M9 and M11), it appears that the coefficients on the exporter and importer “mass” variables (namely $\ln(Y)$, where Y has now been replaced by exporter GDP, and $\ln(E)$) are not biased significantly: even the point estimates are close to the true value 1. A clear bias emerges for the time dependent world output variable $\ln(wY)$, whose true coefficient -1 is now estimated as approximately -1.5, and the exporter multilateral resistance term $\ln(\Pi)$ that now picks up the effects of the inclusion of wrong values for the other exporter-time dependent variable²⁰. Whereas the true coefficient equals 5, the estimate is 5.35. For the other estimates, the equality with the true values cannot be rejected in columns M9 and M11. The point estimate of the FTA variable lies close to the truth (0.352 vs. 0.35), whereas the absolute value of the import tariff variable lies above the true value (-5.21 vs. -5).

This result changes when GDP-based domestic trade flows are used. In column M10, we see that except for the international trade policy variables FTA and $\ln(1 + \tau)$ and the international distance variable, which has a p-value of 0.105, all estimates are statistically different from the true coefficient values at least at a significance level of 10%. Whereas the true coefficients of the international trade policy variables lie within the calculated confidence intervals, the point estimate of the FTA effect is higher (0.36 > 0.35) and the tariff effect is absolutely smaller (-4.72 vs. -5). For the import tariffs, the bias direction when GDP-based domestic trade flows are used is thus opposite that when the model/GO based domestic trade flows are employed.

The comparison with the model using only international trade flows (M12) also yields interesting results. The replacement of gross output Y by GDP significantly biases the multilateral resistance term coefficient estimates and also those for the international distance. The FTA estimate is statistically unbiased, but the point estimate is higher than the truth, and also higher than the estimates from the other specifications. In contrast, the tariff estimate is, of all 4 specifications, the closest to the truth and also statistically unbiased.

Just as with the gold medal mistake, we thus find that using GDP instead of gross output to measure exporter mass introduces bias, but this bias is not statistically significant with respect to the international trade policy variables. The inclusion of the “right” domestic trade flows yields FTA effect estimates closest to the truth, whereas the pure international trade flow model performs best for the estimation of the tariff impact.

[Insert Table 6]

²⁰ $\ln(Y)$, now replaced by $\ln(GDP)$.

4.4 Single distance variable

When simulating the model, we distinguished between the impact of international distance and domestic distance by setting the coefficient on international distance equal to -1 and for domestic distance to -0.5, because it appears reasonable to assume that distance-related international trade barriers exist that are not present for domestic trade. In Table 7, we investigate what happens if the estimation does not distinguish between international and domestic distance.

Starting with the “right” domestic trade flow-inclusive models M13 and M15, we see that the coefficients for the multilateral resistance terms are significantly downward biased, whereas the coefficient on the domestic trade dummy is upward-biased, as could be expected since this dummy now partly picks up the missing distinction between international and domestic distance. Looking at the coefficient of the distance variable, we see that the coefficient is basically equal to -0.5, i.e., the true coefficient on domestic distance. Despite the fact that most trade flows are international, the estimate is determined by the domestic trade flows because they exceed the international trade flows by far. The international trade policy variable coefficients are statistically unbiased, but the point values both lie somewhat below the true coefficient values.

Turning to the GDP-based domestic trade flows, we see that except for the international trade policy variables, all coefficient estimates are significantly biased. The estimate for the effect of the distance variable -0.466 lies, in absolute value, even below the true coefficient for domestic distance. The FTA effect, although not significantly biased, lies at 0.32 and thus below the true value; and the estimated absolute value for the tariff coefficient is also clearly lower than the true coefficient, although the difference is not statistically significant.

A very different picture arises for the pure international trade flow model (M16). This model now delivers by far the most precise estimates, because domestic trade flows are not included and hence distance is always the international distance. The point estimate for the coefficient on $\ln(Dist)$ thus accurately measures the true coefficient (-1). The same can be observed for the coefficients of the other variables. This exercise thus leads us to conclude that if, in reality, the coefficients on international and national distance are different, one should either include both distances as regressors or just use international trade flows. Our result also points to possible bias if distances are not measured correctly. Given that the literature’s take on the measurement of gravity distances is not unanimous (e.g. with respect to population weighting or whether one uses multiple locations in a country to construct a distance measure), this seems a potential problem that is often overlooked.

[Insert Table 7]

4.5 Gravity with Country-Time Fixed Effects

We now evaluate the impact of replacing the multilateral resistance terms (Π_{ot} and P_{dt}), gross output (Y_{ot}), expenditure (E_{dt}) and “world” output (wY_t) variables by

exporter- and importer-time fixed effects. Since these two sets of fixed effects will control for the effects of country-time specific variables, the coefficients of the country-pair variables should be consistently estimated without omitted variable bias. As before, we will repeat the basic estimation specifications presented in Table 4 with differently constructed domestic and only international trade flows.

In general, Table 8 confirms that using exporter- and importer-time fixed effects, which is nowadays a common practice to control for the unobservable multilateral resistance terms, leads to unbiased estimates for the remaining variable coefficients. The omission of the domestic trade flows (column M20) or the use of the simulated or GO-based domestic trade flow observations (columns M17 and M19) lead to similar, very precisely estimated coefficients basically equal to the true ones.

If the domestic trade flows are calculated as the difference between the GDP and the (gross) total exports (column M18), the point estimates are somewhat different from the true coefficients, although the bias remains statistically insignificant, with the exception of the domestic dummy DB whose coefficient is overestimated ($2.42 > 2$). In absolute value, both the coefficient on FTA and on import tariffs is underestimated, although the difference is not big and statistically insignificant.

The most important take-away from this exercise is that the models with the “right” domestic trade flows and the model without domestic trade flows basically yield the same, precise estimates of the true coefficients, whereas the inclusion of GDP-based domestic trade flows produces slightly inferior estimates. From this, we conclude that there is no advantage of including domestic trade flows: the model with just international trade flows delivers estimates that are just as good.

[Insert Table 8]

If only a single distance variable is included jointly with the sets of fixed effects mentioned, excluding the domestic trade flows (M24) outperforms using any type of domestic trade flow (Table 9), mirroring the results found in Table 7. In general, using the “right” domestic flows (M21 and M23), or the GDP-based proxy, lead to similar results although the GDP-based domestic trade flows increase the difference of the estimates from the truth. The absolute values of the international trade policy variable coefficients are downward-biased whenever domestic trade flows are used. The domestic trade-exclusive model delivers the best results.

[Insert Table 9]

4.6 Three-Way Fixed Effects

We now explore the impact of using three-way fixed effects in the gravity estimation. The inclusion of exporter-importer pair-fixed effects will capture the impact of all time-invariant bilateral factors affecting bilateral trade flows between

two countries.²¹ These fixed effects are also used to mitigate the possible endogeneity when analyzing the effects of trade policies between countries, like trade agreements.²² Consequently, only time-variant variables can be included in the regressions (in our case: FTA_{odt} and $\ln(1 + \tau)_{odt}$).

[Insert Table 10]

After the inclusion of fixed effects, only FTA_{odt} and $\ln(1 + \tau)_{odt}$ remain as regressors. The regression results are shown in Table 10. In general, the use of three-way fixed effects leads to practically identical results regardless of whether the “right” domestic trade flows (M25, M27) or just international trade flows are used (M28), as reported in Table 10. The inclusion of GDP-based domestic trade flows does not lead to statistically significant bias, but an absolute downward bias for both estimated international trade policy coefficients is noticeable. In principle, the results of Table 10 thus confirm the main results of Campos et al. (2021a) that using domestic trade observations based on the difference between GDP and total gross exports does not lead to significantly biased estimates with regard to the FTA effects.

If only the FTA variable is left in the regression and the tariff variable is omitted, all estimates for the FTA coefficient are upward biased by about a factor 2, as Table 11 shows. The bias is slightly less in size when the GDP-based domestic trade flow is used, mirroring the lower coefficient estimate of M26 in Table 10. The take-away of this table is that the omission of tariffs as the other important international trade policy variable leads to a clear upward bias in the FTA estimate, which is not surprising because of the strong negative correlation (-0.8) between the two trade policy variables.²³ As was already pointed out in the discussion of the correlation tables in Section 3, we expect the correlation between these two variables to be less strong in real-world data, because FTAs tend to be formed between countries whose tariff barriers against each other are already lower, hence the upward bias on the FTA variable should subsequently then also be lower. Yet, the results should serve as a warning that the omission of trade policy variables, for example because they are hard to measure or tedious to collect, certainly comes with the risk of biasing the estimates for the included trade policy variables. This bias may be substantial if the included and the omitted variables are highly correlated.

[Insert Table 11]

5 Conclusion

The inclusion of domestic (intra-national) trade when estimating structural gravity models is an important topic and has been shown to solve empirical puzzles such

²¹Symmetric or asymmetric, depending on the nature of the pair-fixed effects used.

²²See Baier and Bergstrand (2007) and Yotov et al. (2016, p. 111).

²³Additionally, our tariff variable is set equal to zero by construction when o and d form an FTA (i.e., if $FTA = 1$, then $\tau = 0$).

as the missing globalization puzzle (Yotov, 2022). Despite efforts to construct domestic trade data, its measurement remains a challenge, in particular for historical data. Campos et al. (2021a) have recently claimed that the way how domestic trade flows are calculated (in particular the correct gross output (GO) based trade flows vs. GDP based trade flows as proxy) is not important for the estimation of free trade agreement (FTA) effects in a gravity framework: a surprising finding, since the size differences between these measures are large, and their correlation of about 80%, while substantial, is not perfect.

In this article, we have revisited the question of the (non)-importance of whether to include and how to measure domestic trade in a simulation framework. We find that the conclusion of Campos et al. (2021a), that the exact way how domestic trade is measured does not matter for the estimation of the effect of FTAs is basically correct if we evaluate the results based on statistically significant diversion from the true parameters. However, with regard to the point estimates, marked differences arise depending on how domestic trade flows are calculated: the GDP-based domestic trade flows clearly distort the coefficient estimates. Interestingly, if only international trade flows are used, the results are unbiased and just as precisely estimated as if one includes the correctly measured (GO-based) domestic trade flows. Depending on what effects researchers are interested in, the basic gravity model without inclusion of domestic trade flows may thus be the preferred alternative after all.

Moreover, our simulation results also underline a basic insight from statistics that omitted variable bias is more severe the higher the absolute correlation between omitted and included variables. In our context, the high absolute correlation between the two international trade policy variables (import tariffs and FTAs) in our simulated data set implies a strong (in our case upward) bias for the included FTA variable when the import tariff variable is omitted. This bias appears regardless of whether or not domestic trade flows are included in the regression. Accordingly, the omission of other relevant trade policy variables, for example because they are hard to measure or tedious to collect, may present a substantially more compelling reason to doubt the validity of the estimation results than the omission or inclusion of domestic trade flows as additional observations in the structural gravity estimation.

References

- ANDERSON, J. E. (1979): “A theoretical foundation for the gravity equation,” *American Economic Review*, 69, 106–116.
- ANDERSON, J. E. AND E. VAN WINCOOP (2003): “Gravity with Gravitas: A Solution to the Border Puzzle,” *American Economic Review*, 93, 170–192.
- ANDERSON, J. E. AND Y. V. YOTOV (2010): “The changing incidence of geography,” *American Economic Review*, 100, 2157–2186.

- ARKOLAKIS, C., A. COSTINOT, AND A. RODRÍGUEZ-CLARE (2012): “New trade models, same old gains?” *American Economic Review*, 102, 94–130.
- BAIER, S. L. AND J. H. BERGSTRAND (2007): “Do free trade agreements actually increase members’ international trade?” *Journal of International Economics*, 71, 72–95.
- BALDWIN, R. AND D. TAGLIONI (2006): “Gravity for Dummies and Dummies for Gravity Equations,” Working Paper 12516, National Bureau of Economic Research.
- BERGSTRAND, J. H., M. LARCH, AND Y. V. YOTOV (2015): “Economic integration agreements, border effects, and distance elasticities in the gravity equation,” *European Economic Review*, 78, 307–327.
- CAMPOS, R. G., J. TIMINI, AND E. VIDAL (2021a): “Structural gravity and trade agreements: does the measurement of domestic trade matter?” *Economics Letters*, 208.
- (2021b): “Structural gravity and trade agreements: Does the measurement of domestic trade matter?” Documentos de trabajo del Banco de España 2117, Bank of Spain.
- EATON, J. AND S. KORTUM (2002): “Technology, Geography, and Trade,” *Econometrica*, 70, 1741–1779.
- EGGER, P. H. AND K. E. STAUB (2016): “GLM estimation of trade gravity models with fixed effects,” *Empirical Economics*, 50, 137–175.
- FALLY, T. (2015): “Structural gravity and fixed effects,” *Journal of International Economics*, 97, 76–85.
- HEAD, K. AND T. MAYER (2014): “Gravity Equations: Workhorse, Toolkit, and Cookbook,” in *Handbook of International Economics*, ed. by G. Gopinath, E. Helpman, and K. Rogoff, Elsevier, vol. 4 of *Handbook of International Economics*, 131–195.
- KRUGMAN, P. R. (1979): “Increasing returns, monopolistic competition, and international trade,” *Journal of International Economics*, 9, 469–479.
- MATSCHKE, X. AND J. R. ROJAS RODRÍGUEZ (2025): “Do We Really Need Domestic Trade Flows to Evaluate the Impact of Trade Agreements? A Comment,” Research Papers in Economics 11/25, Universität Trier.
- MAYER, T. AND S. ZIGNAGO (2011): “Notes on CEPII’s distances measures: The GeoDist database,” Working Papers 2011-25, CEPII research center.
- OLIVERO, M. P. AND Y. V. YOTOV (2012): “Dynamic gravity: endogenous country size and asset accumulation,” *Canadian Journal of Economics*, 45, 64–92.

- SANTOS SILVA, J. M. C. AND S. TENREYRO (2006): “The Log of Gravity,” *The Review of Economics and Statistics*, 88, 641–658.
- YOTOV, Y. V. (2012): “A simple solution to the distance puzzle in international trade,” *Economics Letters*, 117, 794–798.
- (2022): “On the role of domestic trade flows for estimating the gravity model of trade,” *Contemporary Economic Policy*, 40, 794—798.
- YOTOV, Y. V., R. PIERMARTINI, J.-A. MONTEIRO, AND M. LARCH (2016): *An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model*, Geneva: World Trade Organization.

6 Appendix

6.1 Summary of Campos et al. (2021) data

Table 1: Data Summary of Campos et al. (2021a) for 1995
(39 domestic trade flow observations)

(1)					
	Mean	SD	Min	Max	N
Y_GDP	692,003.943	1,484,969.30	3,418.83	7,586,226.00	39
Y_PROD	435,720.741	814,978.99	2,811.50	3,676,203.75	39
Y_WIOD	423,222.710	812,195.20	2,993.24	3,803,668.00	39
E_GDP	691,617.696	1,494,802.73	3,990.23	7,725,554.00	39
E_PROD	435,334.507	821,171.37	3,382.90	3,815,531.75	39
E_WIOD	422,836.476	819,426.83	3,601.34	3,942,996.00	39
MFN_tariff	0.083	0.03	0.00	0.15	39
Dom_trade_GDP	617,770.058	1,394,515.77	1,807.33	7,154,965.50	39
Dom_trade_PROD	361,486.859	726,394.19	1,200.00	3,244,943.25	39
Dom_trade_WIOD	348,988.828	724,902.22	1,238.80	3,372,407.50	39

6.2 True Coefficients for the Simulations

Table 2: Coefficients used in (7)

Coefficient	Value	Variable
β_1	1	Y_{it}
β_2	1	E_{jt}
β_3	-1	Y_t
β_4	5	Π_{it}
β_5	5	P_{jt}
λ_1	-1	Int_Dist_{ij}
λ_2	-0.5	Dom_Dist_{ij}
λ_3	2	DB_{ii}
λ_4	0.35	FTA_{ijt}
λ_5	-5	τ_{ijt}

6.3 Summary of simulated data

Table 3: Data Summary

	Mean	SD	Min	Max	N
X_{odt}	15,690.046	173,792.241	0.000	18,489,742.000	9,000,000
t_{odt}	5.835	1.203	0.161	10.280	9,000,000
η_{odt}	1.000	2.831	0.000	5,078.818	9,000,000
Y_t	14,121,000.837	5,301,656.606	3,193,998.500	40,198,912.000	9,000,000
Y_{ot}	470,700.028	965,522.893	368.931	20,660,794.000	9,000,000
E_{dt}	679,779.542	1,373,290.191	1,132.108	35,955,976.000	9,000,000
Int_Dist_{odt}	8,241.960	4,856.190	1.000	55,788.777	9,000,000
Dom_Dist_{odt}	7.034	53.103	0.000	2,927.135	9,000,000
FTA_{odt}	0.462	0.499	0.000	1.000	9,000,000
τ_{odt}	0.037	0.043	0.000	0.364	9,000,000
Π_{ot}	2.963	1.471	0.102	9.061	9,000,000
P_{dt}	1.202	0.778	0.059	8.428	9,000,000
X_{ot}	52,362.220	191,818.576	0.006	7,466,556.500	9,000,000
M_{dt}	52,362.220	174,293.914	0.005	6,140,284.500	9,000,000
INT_X_{odt}	1,805.594	19,677.562	0.000	3,989,915.750	8,700,000
$DOM(X)_{MODEL}$	418,339.166	852,741.904	93.509	18,489,742.000	300,000
$DOM(X)_{GO}$	418,337.808	852,744.430	63.042	18,487,972.000	300,000
$DOM(X)_{GDP}$	627,431.386	1,300,621.429	0.000	35,955,900.000	300,000

6.4 Simulation results - True Model

Table 4: True Model

	M1	M2	M3	M4
$\ln(Y)$	1.00000 (0.00020) [0.98516]	0.82091 (0.02771) [0.00000]	1.00000 (0.00019) [0.98451]	1.00000 (0.00026) [0.99653]
$\ln(E)$	1.00000 (0.00021) [0.98255]	1.11539 (0.02258) [0.00000]	1.00000 (0.00020) [0.99068]	1.00000 (0.00026) [0.98790]
$\ln(wY)$	-0.99899 (0.00126) [0.97675]	-0.39363 (0.22427) [0.00044]	-0.99899 (0.00064) [0.99837]	-0.99878 (0.00435) [0.98421]
$\ln(\Pi)$	5.00003 (0.00114) [0.98987]	4.53762 (0.10700) [0.00001]	5.00000 (0.00110) [0.99547]	4.99994 (0.00166) [0.97137]
$\ln(P)$	5.00004 (0.00114) [0.98411]	4.88774 (0.10730) [0.44291]	5.00001 (0.00110) [0.99526]	4.99999 (0.00163) [0.99551]
$\ln(\text{Int_Dist})$	-0.99999 (0.00049) [0.98346]	-0.95069 (0.03136) [0.15898]	-0.99999 (0.00049) [0.97827]	-0.99998 (0.00051) [0.97660]
$\ln(\text{Dom_Dist})$	-0.50000 (0.00013) [0.98896]	-0.46750 (0.01529) [0.03786]	-0.50000 (0.00011) [0.99904]	0.00000 (0.00000) [.]
DB	2.00017 (0.00424) [0.97551]	2.42717 (0.25183) [0.07553]	2.00016 (0.00422) [0.97732]	0.00000 (0.00000) [.]
FTA	0.35003 (0.00130) [0.98427]	0.33621 (0.06805) [0.86490]	0.35003 (0.00129) [0.98370]	0.35002 (0.00131) [0.98852]
$\ln(1 + \tau)$	-4.99963 (0.01716) [0.98103]	-4.35286 (0.86153) [0.41771]	-4.99965 (0.01713) [0.98174]	-4.99961 (0.01743) [0.98498]
Domestic Trade	model	GDP	GO	NO

Notes: Coefficients are the mean of each variable based on 1000 repetitions. The mean standard error is in parentheses. P-value, H_0 : the estimated coefficient is statistically equal to the true coefficient, in square brackets.

6.5 Simulation results - Gold Medal Mistake

Table 5: Gold Medal Mistake

	M5	M6	M7	M8
$\ln(Y)$	0.60007 (0.01967) [0.00000]	0.41769 (0.03160) [0.00000]	0.60007 (0.01967) [0.00000]	1.16049 (0.06788) [0.08119]
$\ln(E)$	0.52000 (0.02063) [0.00000]	0.65438 (0.02538) [0.00000]	0.52000 (0.02063) [0.00000]	1.04831 (0.06355) [0.94686]
$\ln(wY)$	-0.37139 (0.22453) [0.00012]	0.07800 (0.30613) [0.00002]	-0.37139 (0.22452) [0.00012]	-0.23559 (1.09419) [0.38191]
$\ln(Int_Dist)$	-0.44340 (0.12324) [0.00000]	-0.44195 (0.12441) [0.00000]	-0.44340 (0.12324) [0.00000]	-0.49759 (0.11823) [0.00001]
$\ln(Dom_Dist)$	-0.00533 (0.00984) [0.00000]	-0.00267 (0.01499) [0.00000]	-0.00533 (0.00984) [0.00000]	0.00000 (0.00000) [.]
DB	0.78744 (1.12464) [0.20116]	1.28442 (1.13610) [0.42806]	0.78743 (1.12464) [0.20115]	0.00000 (0.00000) [.]
FTA	0.16204 (0.34241) [0.64865]	0.16185 (0.34297) [0.65221]	0.16205 (0.34241) [0.64866]	0.16539 (0.31347) [0.55380]
$\ln(1 + \tau)$	-4.86099 (4.17429) [0.82053]	-4.34853 (4.16694) [0.71041]	-4.86098 (4.17429) [0.82052]	-3.02122 (3.92404) [0.51584]
Domestic Trade	model	GDP	GO	NO

Notes: Coefficients are the mean of each variable based on 1000 repetitions. The mean standard error is in parentheses. P-value, H_0 : the estimated coefficient is statistically equal to the true coefficient, in square brackets.

6.6 Simulation results - Using GDP instead of Y

Table 6: GDP instead of Y

	M9	M10	M11	M12
$\ln(GDP)$	1.00702 (0.01028) [0.43789]	1.03369 (0.00765) [0.00000]	1.00702 (0.01028) [0.43804]	1.00316 (0.00662) [0.60075]
$\ln(E)$	1.00594 (0.01099) [0.52705]	1.02924 (0.00805) [0.00014]	1.00594 (0.01099) [0.52718]	1.00235 (0.00705) [0.73056]
$\ln(wY)$	-1.49863 (0.23862) [0.00386]	-1.05215 (0.03884) [0.08522]	-1.49863 (0.23862) [0.00386]	-1.04161 (0.15258) [0.54937]
$\ln(II)$	5.34579 (0.09094) [0.00002]	5.35509 (0.05597) [0.00000]	5.34576 (0.09095) [0.00002]	5.71049 (0.05094) [0.00000]
$\ln(P)$	4.93195 (0.08366) [0.38814]	5.33449 (0.05697) [0.00000]	4.93191 (0.08366) [0.38791]	5.29729 (0.04989) [0.00000]
$\ln(Int_Dist)$	-1.00128 (0.02461) [0.93550]	-1.03546 (0.02289) [0.10507]	-1.00127 (0.02462) [0.93568]	-1.04472 (0.01687) [0.00695]
$\ln(Dom_Dist)$	-0.51334 (0.01424) [0.30613]	-0.53345 (0.00589) [0.00000]	-0.51334 (0.01424) [0.30623]	0.00000 (0.00000) [.]
DB	1.88130 (0.20098) [0.53804]	2.31257 (0.19216) [0.07469]	1.88130 (0.20098) [0.53805]	0.00000 (0.00000) [.]
FTA	0.35217 (0.05566) [0.92428]	0.36155 (0.05449) [0.84387]	0.35217 (0.05566) [0.92427]	0.36761 (0.04059) [0.64457]
$\ln(1 + \tau)$	-5.21188 (0.68403) [0.75317]	-4.71506 (0.65409) [0.64738]	-5.21189 (0.68405) [0.75317]	-5.03750 (0.49944) [0.92060]
Domestic Trade	model	GDP	GO	NO

Notes: Coefficients are the mean of each variable based on 1000 repetitions. The mean standard error is in parentheses. P-value, H_0 : the estimated coefficient is statistically equal to the true coefficient, in square brackets.

6.7 Simulation results - Using a single DIST variable

Table 7: Single DIST variable

	M13	M14	M15	M16
$\ln(Y)$	0.99032 (0.00997) [0.24088]	0.81436 (0.02856) [0.00000]	0.99032 (0.00997) [0.24085]	1.00000 (0.00026) [0.99653]
$\ln(E)$	0.98751 (0.01085) [0.16696]	1.10395 (0.02515) [0.00008]	0.98751 (0.01085) [0.16689]	1.00000 (0.00026) [0.98790]
$\ln(wY)$	-0.98134 (0.03980) [0.53851]	-0.37955 (0.22537) [0.00035]	-0.98135 (0.03977) [0.53857]	-0.99878 (0.00435) [0.98421]
$\ln(\Pi)$	4.88706 (0.06640) [0.04278]	4.44643 (0.11994) [0.00000]	4.88703 (0.06639) [0.04274]	4.99994 (0.00166) [0.97137]
$\ln(P)$	4.88702 (0.06639) [0.04292]	4.79643 (0.12004) [0.09042]	4.88699 (0.06638) [0.04286]	4.99999 (0.00163) [0.99551]
$\ln(\text{Dist})$	-0.50045 (0.00713) [0.00000]	-0.46623 (0.01644) [0.00000]	-0.50045 (0.00713) [0.00000]	-0.99998 (0.00051) [0.97660]
DB	6.25817 (0.08530) [0.00000]	6.56819 (0.12347) [0.00000]	6.25815 (0.08530) [0.00000]	0.00000 (0.00000) [.]
FTA	0.32810 (0.08434) [0.85362]	0.31628 (0.10742) [0.80460]	0.32810 (0.08434) [0.85363]	0.35002 (0.00131) [0.98852]
$\ln(1 + \tau)$	-4.96725 (1.03430) [0.93885]	-4.34131 (1.32842) [0.55627]	-4.96724 (1.03429) [0.93887]	-4.99961 (0.01743) [0.98498]
Domestic Trade	model	GDP	GO	NO

Notes: Coefficients are the mean of each variable based on 1000 repetitions. The mean standard error is in parentheses. P-value, H_0 : the estimated coefficient is statistically equal to the true coefficient, in square brackets.

6.8 Simulation results - Exp- and Imp-Time FE

Table 8: Exp- and Imp-Time FE

	M17	M18	M19	M20
$\ln(Int_Dist)$	-0.99999 (0.00061) [0.97871]	-0.99792 (0.00267) [0.52012]	-0.99998 (0.00061) [0.97371]	-0.99997 (0.00064) [0.96590]
$\ln(Dom_Dist)$	-0.50007 (0.00118) [0.93785]	-0.49511 (0.00756) [0.52810]	-0.50008 (0.00119) [0.93327]	0.00000 (0.00000) [.]
DB	2.00056 (0.00803) [0.93304]	2.42243 (0.04262) [0.00000]	2.00065 (0.00806) [0.92524]	0.00000 (0.00000) [.]
FTA	0.35006 (0.00154) [0.97376]	0.34734 (0.00539) [0.59922]	0.35006 (0.00154) [0.97372]	0.35008 (0.00156) [0.96120]
$\ln(1 + \tau)$	-4.99940 (0.02056) [0.97754]	-4.95136 (0.06914) [0.49873]	-4.99941 (0.02058) [0.97787]	-4.99914 (0.02079) [0.96759]
Domestic Trade	model	GDP	GO	NO

Notes: Coefficients are the mean of each variable based on 1000 repetitions. The mean standard error is in parentheses. P-value, H_0 : the estimated coefficient is statistically equal to the true coefficient, in square brackets.

6.9 Simulation results - Exp- and Imp-Time FE and single DIST variable

Table 9: Exp- and Imp-Time FE

	M21	M22	M23	M24
$\ln(Dist)$	-0.89470 (0.00851) [0.00000]	-0.89123 (0.00918) [0.00000]	-0.89470 (0.00851) [0.00000]	-0.99997 (0.00064) [0.96590]
DB	4.92413 (0.03459) [0.00000]	5.36710 (0.03925) [0.00000]	4.92413 (0.03459) [0.00000]	0.00000 (0.00000) [.]
FTA	0.34434 (0.01502) [0.68934]	0.34167 (0.01621) [0.58633]	0.34434 (0.01502) [0.68935]	0.35008 (0.00156) [0.96120]
$\ln(1 + \tau)$	-4.96789 (0.19137) [0.84892]	-4.91697 (0.20569) [0.67633]	-4.96791 (0.19137) [0.84899]	-4.99914 (0.02079) [0.96759]
Domestic Trade	model	GDP	GO	NO

Notes: Coefficients are the mean of each variable based on 1000 repetitions. The mean standard error is in parentheses. P-value, H_0 : the estimated coefficient is statistically equal to the true coefficient, in square brackets.

6.10 Simulation results - Three-way FE

Table 10: Three-way FE

	M25	M26	M27	M28
<i>FTA</i>	0.34987 (0.00359) [0.98826]	0.33967 (0.01375) [0.53533]	0.34986 (0.00360) [0.98591]	0.34991 (0.00367) [0.99920]
$\ln(1 + \tau)$	-5.00130 (0.04670) [0.99500]	-4.75953 (0.17939) [0.18568]	-5.00149 (0.04676) [0.99165]	-5.00088 (0.04778) [0.99889]
Domestic Trade	model	GDP	GO	NO

Notes: Coefficients are the mean of each variable based on 1000 repetitions. The mean standard error is in parentheses. P-value, H_0 : the estimated coefficient is statistically equal to the true coefficient, in square brackets.

6.11 Simulation results - Omitting tariff variable

Table 11: Omitting tariff variable

	M29	M30	M31	M32
<i>FTA</i>	0.69424 (0.01134) [0.00000]	0.66853 (0.01315) [0.00000]	0.69425 (0.01134) [0.00000]	0.69245 (0.01200) [0.00000]
Domestic Trade	model	GDP	GO	NO

Notes: Coefficients are the mean of each variable based on 1000 repetitions. The mean standard error is in parentheses. P-value, H_0 : the estimated coefficient is statistically equal to the true coefficient, in square brackets.

6.12 Iterative process of the MRTs

For the construction of the multilateral resistance terms, we based our loop on the one presented in the appendix and codes of Head and Mayer (2014). The loop considers each year independently, such that all calculations are performed only for the evaluated year in each iteration of the loop.

1. The multilateral resistance terms for all countries during period t are set equal to one as a starting point for the calculations.
2. Then, “new” MRTs are calculated for each country, using the generated output, expenditure, and trade costs, and the “old” MRT values of the evaluated year t , based on 3 and 4.
3. The inward MRTs are normalized by the value of the inward MRTs of country “1” during year t .
4. The process is repeated from step two onwards until the variation between the current and the previous iteration of the loop converges (the change between the old and new values is smaller than the threshold).
5. When the loop converges for period t of the data, it moves to period $t + 1$ and repeats steps one to four until the last period is reached.

6.13 Correlation tables

Table 12: Correlation matrix of the simulated data: part 1

	trade	t	error	wY	Y	E	GDP
<i>trade</i>	1						
<i>t</i>	-0.3210	1.0000					
<i>error</i>	-0.000004	0.0002	1.0000				
<i>wY</i>	0.0339	-0.0216	-0.0002	1.0000			
<i>Y_{ot}</i>	0.1850	-0.0048	0.0006	0.1830	1.0000		
<i>E_{dt}</i>	0.1760	-0.0106	0.00001	0.1830	0.0334	1.0000	
<i>GDP</i>	0.1730	-0.0105	0.0006	0.1830	0.9350	0.0404	1.0000
<i>int_dist</i>	-0.1380	0.7450	0.0002	0.0012	-0.0006	-0.0004	-0.0004
<i>dom_dist</i>	0.2600	-0.4370	-0.00001	0.0009	-0.0005	-0.0003	-0.0003
<i>ln_int_dist</i>	-0.4050	0.8880	0.0001	0.0004	-0.0002	-0.0002	-0.0002
<i>ln_dom_dist</i>	0.3930	-0.6660	-0.00001	0.0003	-0.00001	0.0001	0.0001
<i>FTA</i>	-0.0715	-0.2130	-0.0003	0.0554	0.0111	0.0259	0.0259
<i>tariff</i>	-0.0721	0.4670	0.0003	-0.0612	-0.0120	-0.0281	-0.0279
<i>MRT_{ot}</i>	-0.0216	0.0090	-0.0003	0.0385	-0.1160	-0.0016	-0.2210
<i>MRT_{dt}</i>	-0.0236	-0.0001	-0.0003	0.0019	0.0004	-0.1210	0.0032
<i>X_{ot}</i>	0.1210	-0.0123	-0.00002	0.1160	0.6530	0.0255	0.4390
<i>M_{dt}</i>	0.0947	-0.0125	-0.00001	0.1280	0.0234	0.5610	0.0280
<i>ln_trade</i>	0.2600	-0.4330	0.0507	0.0769	0.2720	0.2350	0.1800
<i>ln_trade_GDP</i>	0.2650	-0.4430	0.0503	0.0765	0.2700	0.2330	0.1790

Table 13: Correlation matrix of the simulated data: part 2

	int_dist	dom_dist	ln_int_dist	ln_dom_dist	FTA	tariff	MRT_ot	MRT_dt	X_ot	M_dt	ln_trade	ln_trade_GDP
<i>int_dist</i>	1											
<i>dom_dist</i>	-0.1930	1.0000										
<i>ln_int_dist</i>	0.5980	-0.5750	1.0000									
<i>ln_dom_dist</i>	-0.2880	0.7780	-0.8590	1.0000								
<i>FTA</i>	0.0537	-0.1050	0.1610	-0.1570	1.0000							
<i>tariff</i>	0.0519	-0.0988	0.1520	-0.1480	-0.8070	1.0000						
<i>MRT_{ot}</i>	0.0016	0.0143	0.0008	0.0098	-0.0180	0.0198	1.0000					
<i>MRT_{dt}</i>	0.0008	0.0102	0.0006	0.0068	0.0030	-0.0046	-0.4960	1.0000				
<i>X_{ot}</i>	-0.0013	0.0005	-0.0004	0.0004	0.0288	-0.0318	0.1430	0.0012	1.0000			
<i>M_{dt}</i>	-0.0002	0.0016	-0.0001	0.0008	0.0307	-0.0339	-0.0019	0.1210	0.0302	1.0000		
<i>ln_trade</i>	-0.2760	0.2480	-0.4400	0.3720	0.0278	-0.1650	0.1680	0.1360	0.3300	0.3420	1.0000	
<i>ln_trade_GDP</i>	-0.2800	0.2580	-0.4540	0.3860	0.0245	-0.1660	0.1660	0.1360	0.3270	0.3410	1.0000	1.0000