Int Adv Econ Res (2024) 30:299–314
https://doi.org/10.1007/s11294-024-09907-z

Check for updates

# Multilevel Modelling in Economics: A Way to Test the Natural-Partner Hypothesis

**Alejandra Martínez-Martínez**[1,2] · **Silviano Esteve-Pérez**[1] · **Salvador Gil-Pareja**[1] · **Rafael Llorca-Vivero**[1]

Accepted: 3 July 2024 / Published online: 29 July 2024


**Abstract** This study applies multilevel modelling to the analysis of trade flows. This methodology can complement the use of the standard gravity equation in addressing the influence of geographical distance on the impact of trade agreements on bilateral trade (the natural-partner hypothesis). The analysis suggests that the simplest, most obvious, and best way to nest country pairs with international trade data is by distance. To this end, a sample of 69 countries with data on domestic flows over the period 1986–2016 was used. This sample accounts for around 85% of world trade data. The estimations suggest that the pro-trade effects of preferential trade agreements decrease with distance to trade partners, while this does not occur with the multilateral trading system. What is more interesting is that the former result only holds for the preferential trade agreements involving goods, which supports the natural-trading partner hypothesis, but not for those that include services. In the latter case, the impact increases with distance. This suggests that deeper preferential trade agreements reduce the cost of distance as a barrier to trade, which probably reflects cultural (regulation) distance.

**Keywords** Domestic flows · International flows · Multilevel modelling · Preferential Trade Agreements · Multilateral trading system

**JEL Classification** C1 · F14 · F15

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s11294-024-09907-z.

✉ Alejandra Martínez-Martínez
Alejandra.martinez@uv.es

1 INTECO Research Group, University of Valencia, Valencia, Spain

2 Departamento de Estructura Económica, Facultad de Economía, University of Valencia, Avda, de los Naranjos S/N. 46022, Valencia, Spain

## Introduction

Multilevel modelling is very rarely used in economics (Oshchepkov & Shirokanova, 2022), with only a handful of exceptions (e.g., Novak & Marko, 2017). The advantage of the multilevel perspective is that it considers the existence and importance of hierarchical relationships. Disregarding these relationships could lead to mistaken conclusions relating to the impact of an independent variable on a dependent one. This is because individuals belonging to a specific group share certain characteristics that differ from those of other groups. Therefore, observations are not independent of each other, which violates a basic assumption of traditional models. If the dependence across units in a group is not considered, incorrect conclusions will be reached in the estimation of confidence intervals and erroneous statistical inferences will be made.

Only two papers (Drzewoszewska, 2014; Hikari, 2017) have tried to apply the multilevel methodology to international trade.[1] In the first case, the study is focused on the bilateral trade of European Union (EU) countries, whereas in the second case the linkage between free trade agreements and firms´ foreign direct investment in the Asia Pacific Region is the aim of the research. In this type of models, the relevant issue is the nested variable. In Drzewoszewska (2014), a three-level model was established, with time, the specific pair of countries, and the European Monetary Union (EMU) agreement as the respective levels. However, there was no clear rationale for this approach. Hikari (2017) made a clearer explanation: firms were categorized based on their sales volume, but only three groups were formed.

In longitudinal data, the nested factor is time. This applies when using country-pair fixed effects to account for unobserved heterogeneity between country pairs in trade models. However, the time groups formed this way are not sufficiently heterogeneous. This article addresses whether multilevel models can yield suitable results by using distance as the factor that accounts for the heterogeneity present across groups of bilateral trade relationships. The hypothesis is that the simplest, most obvious, and best way to nest country pairs with international trade data is by distance. Therefore, a multilevel analysis on a sample of 69 countries during the period 1986–2016 was used to estimate the incidence of distance on the impact of preferential trade agreements (PTAs) and the General Agreement of Tariff and Trade and World Trade Organization (GATT/WTO) system on bilateral trade flows.

To preview the results, it is shown that the impact of PTAs on bilateral trade is negatively shaped by distance, whereas this effect is not observed for the GATT/WTO system. These findings hold for PTAs focused on goods but not for those comprising also services. In the latter case, the impact of deeper PTAs increases with distance, that is, widening the ambit of lowering barriers to trade makes distance an irrelevant or even a positive factor for trade creation. This is an

[1] The empirical workhorse in the analysis of bilateral trade flows is the well-known gravity model with sound theoretical support (e.g., Anderson, 1979; Anderson & van Wincoop, 2003; Bergstrand, 1985). Gravity equations have been extensively used to analyse economic integration agreements (Baier & Bergstrand, 2007; Rose, 2000) and other determinants of trade barriers (Anderson & Marcouiller, 2002).

additional argument in favor of, on the one hand, multilateralism compared to regionalism and, on the other hand, deepening integration.

This paper adds to the literature by introducing a complementary tool for analysing bilateral trade flows, demonstrating that multilevel modelling is a new approach suitable for this research area. To this end, domestic trade flows, alongside international ones, were also considered, following the recommendations for trade analysis (Yotov, 2012).

## Methodology

The basic assumption of multilevel models is the existence of hierarchy in the data. The first-level observations, which are the focus of the current analysis, are nested into the second-level groups. More importantly, the multilevel model provides information about the incidence of the variables of interest in the explanation of these two-level differences. Multilevel regressions can decompose the total variance at the first level into the variance between the groups and the variance within the groups. What is of particular interest is that these models can provide information about the effects of the independent variables (at the first or second levels) on the aforementioned variances. Furthermore, with this method, it is possible to assess whether the effect of a specific independent variable differs between the groups. Multilevel modelling can also be useful in international economics.

Let $X_{ijR}$ denote the exports of country $i$ to country $j$ (if $j$ coincides with $i$, these are domestic flows) in group R. The simplest version of the multilevel model is the one known as the empty model, which is expressed as follows:

$$X_{ijR} = \beta_{0R} + \varepsilon_{ijR} \tag{1}$$

where $\beta_{0R}$ is the average group level and $\varepsilon_{ijR}$ is the deviation with respect to the average group's level. The average group's level can be expressed as follows:

$$\beta_{0R} = \gamma_{00} + u_{0R}. \tag{2}$$

In this case, $\gamma_{00}$ is the average of the sample and $u_{0R}$ is a random variable that measures the distance of each group with respect to the general average. Therefore:

$$X_{ijR} = \gamma_{00} + u_{0R} + \varepsilon_{ijR}. \tag{3}$$

Hence, in its simplest version, multilevel analysis entails explaining the variance in bilateral trade flows by the divergence of each formed country group's' bilateral trade compared to the general average (the variance of the intercept) plus the divergence of each unit (the pair of countries) with regard to the respective group. In other words, the total variance of bilateral trade flows is explained by the between-group variance [$var(u_{0R})$] plus the within-group variance [$var(\varepsilon_{ijR})$].

In this sense, the intraclass correlation coefficient (ICC) is the percentage that the between-group variance represents compared to the total variance. It measures the extent to which the divergences among the established groups are relevant in explaining the dependent variable.

The crucial point is the choice of the nested variable, which establishes the hierarchy of the data. It is generally accepted in the literature that the number of groups formed should be around 30 (Maas & Hox, 2005; Stegmueller, 2013), although there is a debate about this point (Bryan & Jenkins, 2016). In this study, the selected nested variable is distance, which is likely the most important factor in explaining trade flows according to the literature. In an analysis covering 1,467 distance-impact estimates from 103 papers, Disdier and Head (2008) found persistence over time of the negative impact of distance on trade. This result was independent of samples and methods. Grossman (1998) stated that the negative impact of distance on trade is not only due to transport costs (a small part of total costs) but also due to cultural differences and confidence. Additionally, distance is a variable that uniquely defines groups over time (this is not the case for the product of bilateral incomes, for instance), which makes things simpler. Therefore, in this research, 30 distance groups at equally distanced intervals were formed.[2] Obviously, each pair of countries is encapsulated in only one group for the entire period. The estimations performed suggest that the hierarchy established in this way is appropriate based on the information reported by the corresponding Lagrange test, whose null hypothesis is the validity of the respective pooled ordinary least squares (OLS) estimation.

These types of models are enriched with the inclusion of independent variables at the aggregate (second level) or individual (first level) levels. When the respective coefficients are not allowed to vary by group, these regressors act as fixed effects. The alternative is to consider random slopes; in this case, the impact of the variables of interest differs by group. Moreover, this permits one to draw conclusions regarding the influence of including additional regressors on both between-group and within-group variance. In this research, regressors at the first level were included.

Assume, for simplicity, the inclusion of two regressors; one of them acts as a fixed effect, and the other incorporates a random component. Therefore,

$$X_{ijR} = \gamma_{00} + u_{0R} + \beta_1 w_{ijR} + \beta_{2j} z_{ijR} + \varepsilon_{ijR} \tag{4}$$

with $w_{ijR}$ as the fixed effect and $z_{ijR}$ as the random one with $\beta_{2R} = \gamma_{20} + u_{2R}$, where $\gamma_{20}$ is the common coefficient for all the groups and $u_{2R}$ is the random component.

Then, by substituting:

$$X_{ijR} = \gamma_{00} + u_{0R} + \beta_1 w_{ijR} + \gamma_{20} z_{ijR} + u_{2j} z_{ijR} + \varepsilon_{ijR}. \tag{5}$$

Hence, in this type of model, the estimation of the variance of the random components of the slope is also obtained. Additionally, it is possible to allow for the estimation of the covariance of the random component of the slope and the random

[2] Distance is measured as the great-circle distance between the countries weighted by population and large-city location in each country.

component of the intercept. This analysis allowed for random slopes in the coefficients of the dummies controlling for PTAs and the GATT/WTO system. The rest of the (gravity) variables (*incomes* –$Y_{i(j)}$ -, *contiguity*, common *language*, and the fact that one country was a *colony* of the other) act as fixed effects.

Therefore, the following (linear) multilevel model in its more complete version was estimated:

$$\begin{aligned} lntrade_{ijRt} = {} & \gamma_{00} + u_{0R} + \beta_1 Y_{iRt} + \beta_2 Y_{jRt} + \beta_3 contiguity_{ijR} \\ & + \beta_4 language_{ijR} + \beta_5 colony_{ijR} \\ & + \gamma_1 PTA_{ijRt} + u_{1R} PTA_{ijRt} + \gamma_2 GATT/WTO_{ijRt} \\ & + u_{2R} GATT/WTO_{ijRt} + \gamma_t + \mu_i + \alpha_j + \varepsilon_{ijR} \end{aligned}$$

where incomes (*Y*) are the nominal gross domestic product (GDP) (in United States (U.S.) dollars) for each country, *contiguity* is a dummy for countries sharing a common border, *language* is a dummy for countries sharing at least one common language, *colony* is a dummy variable equal to 1 if one country was a colony of the other, *PTA* is a dummy variable that takes the value of 1 if the countries are engaged in a PTA of any type and *GATT/WTO* is a dummy variable equal to 1 if the countries belong to the GATT/WTO system in year *t*. $\gamma_t$ are the time effects, $\mu_i$ is the exporter fixed effect, and $\alpha_j$ is the importer fixed effects.[3]

The defenders of multilevel models have powerful reasons for not using traditional one-level estimations. The general argument is that, if hierarchy matters, to overlook this fact will lead to many mistaken conclusions regarding the impact of an independent variable on a dependent one. The point is that individuals belonging to a specific group share certain characteristics that may differ from those of other groups. These observations are more similar to each other than to others randomly drawn from the entire population. This implies that observations are not independent from each other, which violates a basic assumption of traditional regressions (Osborne, 2000). If the dependence across units in a group is disregarded, incorrect conclusions are generally reached both in the estimation of confidence intervals and statistical inferences. Consequently, this kind of model controls for unspecified heterogeneity by allowing the intercept and some slope coefficients to have a stochastic element in their variation.

Additionally, it could be interesting to learn what lies behind between-group variation, which is not considered in classical regression models when estimating longitudinal data by fixed effects, where only within-group variation is assessed. In this sense, multilevel models are an example of partial pooling. In other words, they are an intermediate step between complete pooling and no pooling. Therefore, with such models, it is possible to estimate not only group averages but also group-level effects, thus establishing a link between the within-group estimate and the regression

[3] When the slopes are random, the variables must be centered in order to correct interpretations of the coefficients. However, in this case, the variables of interest are dummies, and no problem occurs in this sense.

estimator that overlooks group indicators. Finally, multilevel methods usually lead to consistent and efficient estimations.

Another option for the analysis performed herein is the inclusion of interactive dummies for distance groups in traditional gravity equations. However, this fixed-effects estimation confounds the effects of group-level predictors with the effects of the group characteristics. Additionally, Oshchepkov and Shirokanova (2022, p. 13) clearly demonstrated that, in this case, multilevel modelling is "more convenient, parsimonious, and efficient" than the aforementioned alternative.[4]

## Data

The selected data for the gravity and trade agreement variables (incomes, distance, contiguity, common language, colony, PTAs, and GATT/WTO system) were taken from the Dynamic Gravity Dataset (DGD) developed by the United States International Trade Commission, 2019; Gurevich & Herman, 2018a). This database includes a total of 286 countries over the period 1948–2019. The database includes a large number of variables regarding macroeconomic, geographic, and institutional indicators. Concerning PTAs, the database follows the WTO definition. Hence, these agreements are those that "grant more favorable conditions to the agreement members than those faced by other WTO members" (Gurevich & Herman, 2018b, p. 19). In this case, the aggregate variable includes customs unions, economic-integration agreements, free-trade agreements, and partial-scope agreements. As Gurevich and Herman (2018a) stated, this database was designed to be easily combined with the usual sources of bilateral trade data. Trade data comes from the structural gravity database (structural gravity database, World Trade Organization, 2019). The advantage of these data is that they incorporate both domestic-trade and international-trade flows for manufacturing (Larch et al., 2019). A selected sample of 69 countries over the period 1986–2016 was used.[5] The sample of countries accounts for around 85% of world trade.

## Empirical Analysis and Results

Table 1 presents an initial approximation of the multilevel analysis based on the explanations provided in the methodology section. The assumption is that bilateral flows can be nested depending on the distance between the partners. Heterogeneity would emerge if there was a hierarchy between the country pairs in their bilateral trade as a consequence of their physical distance, which may also be a reflection of cultural distance or institutional distance, for example. Therefore, 30 equal

[4] For a detailed discussion of the statements in this section, see Gelman and Hill (2007) and Oshchepkov and Shirokanova (2022).

[5] Online Supplemental Appendix Table 1 displays the list of countries. The last available year of this data source is 2016.

**Table 1** Multilevel (lineal) analysis of the logarithm of bilateral trade flows. Period 1986–2016 (69 countries)

| Variable | Empty model (1) | Empty model (2) | Empty model (3) | Model with incomes (4) | Model with incomes and time (5) | Model with incomes and time (6) |
|---|---|---|---|---|---|---|
| Intercept | 2.883 (0.267)*** | 2.826 (0.234)*** | 3.213 (0.127)*** | -6,490 (0.241)*** | -6,251 (0.241)*** | -2,137 (0.309)*** |
| lnYi | - | - | - | 1.164 (0.003)*** | 1.218 (0.003)*** | 0.572 (0.017)*** |
| lnYj | - | - | - | 0.840 (0.003)*** | 0.896 (0.003)*** | 0.623 (0.016)*** |
| Time | - | - | - | - | -0.047 (0.001)*** | 0.017 (0.002)*** |
| Exp. and imp. dummies | No | No | No | No | No | Yes |
| Var (intercept) | 2.139*** | 1.641*** | 0.501*** | 1.735*** | 1.714*** | 2.470*** |
| ICC | 0.157 | 0.126 | 0.037 | 0.312 | 0.317 | 0.491 |
| Type of data | All | Internat | All | All | All | All |
| Number of observations | 130,969 | 128,877 | 130,969 | 124,252 | 124,252 | 124,252 |
| Second level variable | Distance (30) | Distance (30) | Year (31) | Distance (30) | Distance (30) | Distance (30) |
| LR-test Pr(chibar2) | 22,107 (0.000) | 15,669 (0.000) | 4,591 (0.000) | 44,187 (0.000) | 44,863 (0.000) | 56,963 (0.000) |

*significant at 10% level, ** significant at 5% level, ***significant at 1% level. Data source: Dynamic Gravity Dataset (United States International Trade Commission, 2019) interval-distance groups with all the bilateral trade combinations in the sample were formed. The first interval includes those country pairs closer to each other based on a specific range of kilometers, whereas the last group comprises those country pairs further away from each other given the same range of kilometers. There is some consensus in the literature that 30 groups in the second level is an adequate number for multilevel analysis (Maas & Hox, 2005; Stegmueller, 2013). In this case, each group has a range of 654 km.[6]

Column (1) presents the empty model. As explained earlier, the estimated intercept is simply the general mean of the sample. This value is 2.883, with a variance of 2.139 and an ICC of 0.157. Thus, the heterogeneity across the (distance) groups is remarkable in relation to the total heterogeneity of bilateral flows, as expected. The corresponding Lagrange-ratio (LR) test confirms the adequacy of the multilevel approximation. As already mentioned, the recent literature outlines the necessity to incorporate domestic flows to adequately estimate the effect of the variables of interest on trade flows. Therefore, it is interesting to examine whether there is an important difference when omitting domestic flows and considering only international flows. Column (2) presents these results. There is almost no change in the estimated intercept. However, the inter-group variance, although still high, is substantially reduced to a value of 1.641, implying a reduction in the ICC to 0.126. Thus, the group heterogeneity driven by the inclusion of domestic flows remains quite notable.

In a multilevel model, longitudinal data are a special case in which time is the second-level (nested) variable. Given the novelty of the current analysis, it is useful to verify the result obtained when time serves as the variable determining the hierarchy of data (column 3). As expected, trade does not change considerably across the years. Thus, the variance of the intercept is small as is that of the corresponding ICC, which has a value of 0.037. This casts some doubts about the convenience of using this type of analysis in this context, even though the corresponding test shows that the multilevel analysis is appropriate.

Next, the variables affecting bilateral trade flows based on the gravity literature were included. These variables act at the first level. Initially, the incomes of the exporter and the importer country were introduced. The values of the estimated parameters are as expected and highly significant ($p < 0.01$). The between-group variance decreases (compared to column 1). However, more interestingly, the ICC is now considerably higher (0.312). This means that, although the inclusion in the regression of countries' incomes lessens the heterogeneity across the groups in absolute terms, this variability increases in relative terms because the within-group heterogeneity is reduced much more. The time dimension is incorporated in column (5) by adding a time variable as an independent variable, as is usually done in multilevel models. Finally, exporter and importer fixed effects are included to account for multilateral resistance (Anderson & van Wincoop, 2003), invariable with time. The result obtained (column 6) is quite surprising. The between-group heterogeneity increases considerably in both absolute and relative terms. This means that accounting for relative barriers to trade (i.e., bilateral barriers between the countries in a

[6] Estimations with 15 groups were also performed. No relevant changes were found.

**Table 2** Multilevel analysis of the logarithm of bilateral-trade flows (linear model; period: 1986–2016; 69 countries)

| Variable | Basic Model (1) | Extended model (2) | Extended model Random slopes (3) | Extended model Cov Intercept (4) | Extended model Random slopes Internat. Flows (5) | Extended model Cov Intercept Internat Flows (6) | Extended model Cov Intercept PTAs Disaggreg. (7) |
|---|---|---|---|---|---|---|---|
| Intercept | 12.294 (0.300)$^{***}$ | 11.961 (0.300)$^{***}$ | 11.918 (0.338)$^{***}$ | 11.914 (0.339)$^{***}$ | 11.785 (0.256)$^{***}$ | 11.781 (0.257)$^{***}$ | 11.965 (0.339)$^{***}$ |
| lnYi | 0.512 (0.016)$^{***}$ | 0.501 (0.016)$^{***}$ | 0.500 (0.016)$^{***}$ | 0.501 (0.016)$^{***}$ | 0.508 (0.015)$^{***}$ | 0.508 (0.015)$^{***}$ | 0.494 (0.016)$^{***}$ |
| lnYj | 0.546 (0.016)$^{***}$ | 0.535 (0.016)$^{***}$ | 0.535 (0.016)$^{***}$ | 0.535 (0.016)$^{***}$ | 0.545 (0.015)$^{***}$ | 0.545 (0.015)$^{***}$ | 0.529 (0.016)$^{***}$ |
| Time | 0.016 (0.002)$^{***}$ | 0.016 (0.002)$^{***}$ | 0.016 (0.001)$^{***}$ | 0.016 (0.001)$^{***}$ | 0.016 (0.001)$^{***}$ | 0.016 (0.001)$^{***}$ | 0.017 (0.001)$^{***}$ |
| Contiguity | −0.508 (0.033)$^{***}$ | −0.503 (0.033)$^{***}$ | 0.202 (0.033)$^{***}$ | 0.202 (0.034)$^{***}$ | 0.525 (0.033)$^{***}$ | 0.524 (0.033)$^{***}$ | 0.179 (0.034)$^{***}$ |
| Language | 0.438 (0.016)$^{***}$ | 0.436 (0.016)$^{***}$ | 0.555 (0.016)$^{***}$ | 0.556 (0.016)$^{***}$ | 0.658 (0.016)$^{***}$ | 0.657 (0.016)$^{***}$ | 0.549 (0.016)$^{***}$ |
| Colony | 0.716 (0.032)$^{***}$ | 0.716 (0.032)$^{***}$ | 0.657 (0.031)$^{***}$ | 0.657 (0.031)$^{***}$ | 0.629 (0.030)$^{***}$ | 0.630 (0.030)$^{***}$ | 0.634 (0.031)$^{***}$ |
| PTAs | - | 0.016 (0.013) | 0.064 (0.143) | 0.069 (0.143) | 0.200 (0.049)$^{***}$ | 0.198 (0.048)$^{***}$ | - |
| PTAs_goods_only | - | - | - | - | - | - | 0.053 (0.145) |
| PTAs_g_services | - | - | - | - | - | - | 0.232 (0.089)$^{***}$ |
| GATT/WTO | - | 0.455 (0.031)$^{***}$ | 0.309 (0.076)$^{***}$ | 0.311 (0.076)$^{***}$ | 0.134 (0.085)$^{***}$ | 0.140 (0.086)$^{*}$ | 0.309 (0.074)$^{***}$ |
| Var (intercept) | 2.350$^{**}$ | 2.337$^{**}$ | 3.083$^{***}$ | 3.115$^{***}$ | 1.625$^{***}$ | 1.646$^{***}$ | 3.108$^{***}$ |
| Var(PTAs) | - | - | 0.606$^{***}$ | 0.607$^{***}$ | 0.064$^{***}$ | 0.063$^{***}$ | - |
| Var(PTAs_g_only) | - | - | - | - | - | - | 0.619$^{***}$ |

**Table 2** (continued)

| Variable | Basic Model (1) | Extended model (2) | Extended model Random slopes (3) | Extended model Cov Intercept (4) | Extended model Random slopes Internat. Flows (5) | Extended model Cov Intercept Internat Flows (6) | Extended model Cov Intercept PTAs Disaggreg. (7) |
|---|---|---|---|---|---|---|---|
| Var(PTAs_g_ser) | - | - | - | - | - | - | 0.185*** |
| Var(GATT) | - | - | 0.141*** | 0.143*** | 0.186*** | 0.190*** | 0.133*** |
| Cov(PTAsGATT) | - | - | - | 0.089 | - | 0.021 | - |
| Cov(PTAgGATT) | - | - | - | - | - | - | 0.077 |
| Cov(PTAgsGATT) | - | - | - | - | - | - | 0.032 |
| Cov(PTAgPTAgs) | - | - | - | - | - | - | −0.031 |
| Cov(PTAsIntcp) | - | - | - | −0.952*** | | 0.036 | |
| Cov(PTAgIntcp) | - | - | - | - | - | | −0.888*** |
| Cov(PTAgsIntcp) | - | - | - | - | - | - | −0.330*** |
| Cov(GATT,intcp) | - | - | - | −0.298** | - | −0.330** | −0.289*** |
| Type of data | All | All | All | All | Internat | Internat | All |
| Exp. and imp. dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of observations | 125,191 | 125,191 | 125,191 | 125,191 | 123,317 | 122,250 | 125,191 |
| LR test Pr(chibar2) | 47,032 (0.000) | 44,206 (0.000) | 51,134 (0.000) | 51,160 (0.000) | 22,095 (0.000) | 19,813 (0.000) | 51,136 (0.000) |

*significant at 10% level, **significant at 5% level, ***significant at 1% level. “All” for the type of data means that the sample includes international and domestic flows. “Internat” means that only international flows are considered. “Cov.” means covariance and “Disaggreg” means disaggregated. In all the cases, the second-level variable is distance. Data sources: Structural Gravity database (World Trade Organization, 2019); Dynamic Gravity Dataset (United States International Trade Commission, 2019) pair in relation to the barriers of each country with the rest of their partners) makes the established distance hierarchy particularly relevant.

The next step was to perform the analysis with the variables of interest (Table 2). First (column 1), some commonly used gravity variables, such as contiguity (sharing a common border), language, and colony, were added. The coefficients of language and colony are as expected and highly significant ($p < 0.01$), but this is not the case for contiguity, which has a highly significant ($p < 0.01$), though counter-intuitive, coefficient. Note that, although distance is a key variable in gravity models, it is usual to include a dummy variable for countries sharing a common border because this is assumed to be an additional factor that facilitates trade. In any case, although the aforementioned variables have certain effects on bilateral trade, their influence in explaining the between-group and within-group variance is negligible.

In column 2 of Table 2, the dummies for PTAs and GATT/WTO (the variables of interest) are included. Again, despite their effects on trade flows, the reduction in between- and within-group differences is almost zero. On average, PTAs have no effect on international trade flows. Furthermore, membership in the GATT/WTO system positively influences trade, with a considerable effect of around 60% [exp (0.045) -1].

However, the analysis in column 2 can be refined by considering that the estimated coefficients for PTAs and the GATT/WTO vary by group. In other words, the respective slopes are allowed to be random (in multilevel terminology). This is presented in column (3) of Table 2. The between-group heterogeneity undergoes a relevant increase, which suggests that distance may be a factor with a strong role in the impact of preferential and/or multilateral agreements on trade. It is also observed that the coefficient for PTAs (the common group's impact) rises, though it continues to be far from significant ($p = 0.65$). However, there is a high and significant variance for this slope ($p < 0.01$). The estimated coefficient of GATT/WTO diminishes, but it maintains a strong impact. The variance of the slope is also significant ($p < 0.01$) but not as high. In other words, the impact of preferential agreements exhibits an important variability depending on the distance group of the country pair. This variability is considerably less relevant for multilateral agreements.

In the next column (5), relationships between the random coefficients of the intercept and the slopes of the trade agreements are allowed. The result is a highly negative relationship between the intercept and the impact of PTAs. For countries with historically strong relationships of bilateral trade, the incidence of trade agreements is lower. The same occurs for the impact of the GATT/WTO, though to a lesser extent. Columns (5) and (6) reflect the same regressions as columns (3) and (4), but only with international trade flows. In this case, in line with the recent literature, some relevant changes occur with respect to the variables of interest, which may suggest that the coefficients are biased. Both coefficients (for PTAs and the GATT/WTO) are positive and significant ($p < 0.01$), with similar effects of between 15 and 20%. This finding differs from the one that considered also domestic flows, where a non-significant coefficient was found ($p = 0.65$) in the first case and a higher impact in the second one. By allowing for random slopes and covariance between them, it is noted that the variances are significant ($p < 0.01$). However, in contrast to what occurred before, the heterogeneity in relation to distance is clearly higher for the GATT/WTO than for PTAs. There is no negative relationship between the intercept

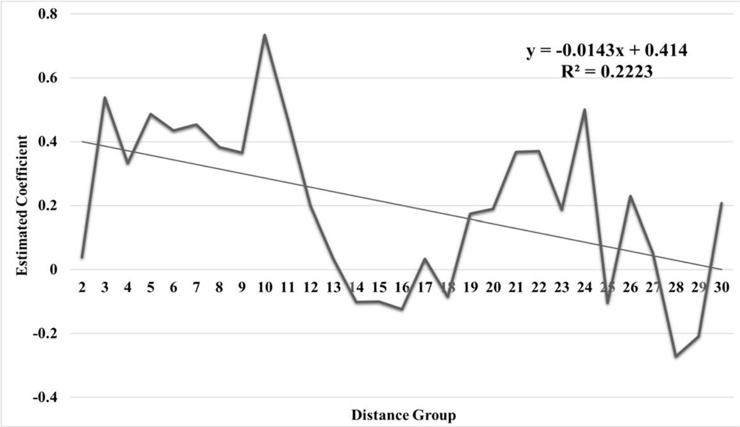


**Fig. 1** Estimated coefficients for PTAs by distance groups. Period 1986–2016. Data sources: Structural Gravity Database, (World Trade Organization, 2019); Dynamic Gravity Dataset (United States International Trade Commission, 2019)

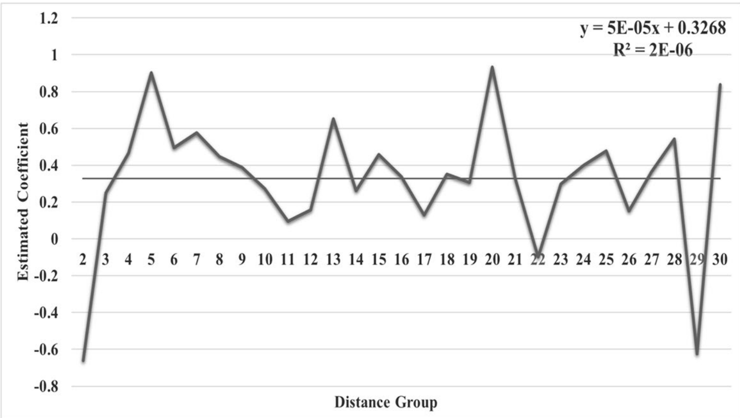


**Fig. 2** Estimated coefficients for GATT/WTO by distance groups. Period 1986–2016. Data sources: Structural Gravity Database (World Trade Organization, 2019); Dynamic Gravity Dataset (United States International Trade Commission, 2019)

and the impact of PTAs. These results confirm the bias encountered when considering only international flows.

At this point, it is interesting to verify which slopes are obtained for PTAs and the GATT/WTO depending on the distance group the country pair belongs to. Figures 1 and 2 display values for these slopes as a function of the 30 distance groups formed. Initially, there seemed to be a negative relationship between the impact of PTAs and the distance between the trading partners, whereas this is clearly not the case for the GATT/WTO. After running OLS regressions for both cases,[7] a negative and significant ($p=0.01$) coefficient

[7] The value of the first distance group was eliminated because it was an outlier that influenced the regressions.

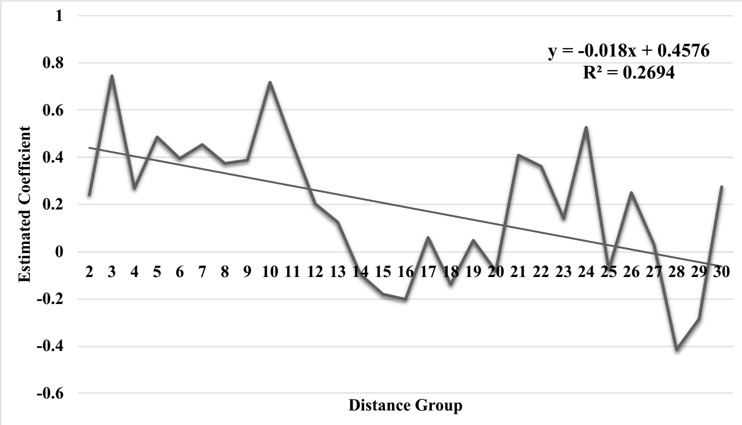


**Fig. 3** Estimated coefficients for PTAs' goods only by distance groups. Period 1986–2016. Data sources: Structural Gravity Database (World Trade Organization, 2019); Dynamic Gravity Dataset (United States International Trade Commission, 2019)

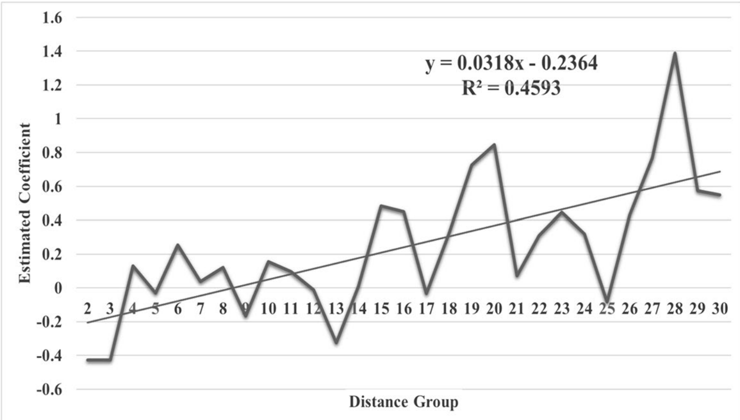


**Fig. 4** Estimated coefficients for PTAs' goods and services by distance groups. Period 1986–2016. Data sources: Structural Gravity Database (World Trade Organization, 2019); Dynamic Gravity Dataset (United States International Trade Commission, 2019)

for PTAs was found (the fit is shown in Fig. 1). The value of this coefficient is −0.0143. This implies that, on average, the move from one distance group to the next one reduces the slope of the respective PTA to around 0.014 points. Put differently, for a given PTA, a pair of countries located, for example, in distance group 9 will face an impact on their trade in this PTA of approximately 33% (slope = 0.285), whereas for a pair located in group 18 (slope = 0.142), this impact is reduced to 15.2%. As already mentioned, no relationship was found between the distance groups and the GATT/WTO agreement (Fig. 2). Therefore, the current multilevel analysis reveals that multilateral agreements have an impact on bilateral trade regardless of the distance between the pair of countries, whereas preferential agreements are better at promoting trade for closer members.

An interesting feature of the database is the distinction between the PTAs focused on the liberalization of trade in goods and those that include also services. Therefore, the generic PTA dummy is replaced with two dummies (*PTAs_goods_only* and *PTAs_g_services*) that reflect this distinction. Then, the same estimation as the one in column 4 of Table 2 is performed. The results are shown in column 7 of Table 2. The estimated parameter for the goods-only PTAs remains non-significant ($p=0.715$).[8] However, the coefficient obtained for the PTAs that include goods and services is positive and highly significant ($p=0.01$), with an impact on bilateral trade of almost 25%. Moreover, when analyzing the effect by distance groups, a clear difference emerges. The respective results are presented in Figs. 3 and 4. There is a negative trend in the impact of the PTAs exclusively focused on goods with respect to the distance among the trading partners (a linear adjustment with an estimated parameter of $-0.018$, significant at the 1% level). The natural-partner hypothesis is thus verified. However, a strong positive relationship between the PTAs that include services and distance was detected. This seems a quite reasonable result. With the development of information and communication technologies, geographical distance is less of a barrier to international trade, even for manufactured goods. The services sector is closely linked to the industrial sector, with the latter often encompassing the former. For countries that are far away from each other, the elimination of other kinds of barriers (mainly regulatory ones) seems to have a larger impact, probably because geographical distance now measures cultural (institutional) distance.

Finally, the variance across the distance groups is much higher for the goods-only PTAs than it is for the PTAs that include services, which suggests a considerably higher heterogeneity in the first case. Similarly, historically strong trade relationships among partners condition the impacts of the goods-only PTAs to a higher extent than they do for the PTAs incorporating services.

## Conclusions

This article represents an attempt to apply multilevel modelling to analyze trade flows, an area where the traditional estimation of gravity equations has been widely adopted. The starting point was the consideration that country pairs are nested by geographical distance, an assumption that held up well. The analysis focused on the impact of PTAs and the GATT/WTO system on bilateral trade flows. To this end, a sample of 69 countries over the period 1986–2016 was used.

The most striking result is that the impact of preferential agreements centered on goods weakens with the distance among country pairs, whereas this does not occur when considering also the international liberalization of services, in which case the opposite occurs. This finding probably suggests the relevance of cultural distance. Furthermore, this result indicates that the natural-partner hypothesis is valid in the case of trade for goods, but not for services. When taking into account the

[8] As noted, no relevant changes occurred with respect to the estimation (and variance) of the parameter for the GATT/WTO. The focus is on the two new variables.

multilateral system, distance is not a relevant factor. Deepening the integration process and multilateralism is the way forward. This application of multilevel modelling to international trade flows has proven not only effective but also highly valuable.

**Funding** Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. Silviano Esteve, Salvador Gil, and Rafael Llorca acknowledge financial support from the PROMETEO project (CIPROM/2022/50).



## References

Anderson, J. E. (1979). A Theoretical Foundation for the Gravity Equation. *American Economic Review, 69*(1), 106–116.

Anderson, J. E., & Marcouiller, D. (2002). Insecurity and the Pattern of Trade: An Empirical Investigation. *The Review of Economics and Statistics, 84*(2), 342–352.

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

Baier, S. L., & Bergstrand, J. H. (2007). Do Free Trade Agreements Actually Increase Members' International Trade? *Journal of International Economics, 71*(1), 72–95.

Bergstrand, J. H. (1985). The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence. *The Review of Economics and Statistics, 67*(3), 474–481.

Bryan, M. L., & Jenkins, S. P. (2016). Multilevel Modelling of Country Effects: A Cautionary Tale. *European Sociological Review, 32*(1), 3–22.

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

Drzewoszewska, N. (2014). Multilevel Modelling of Bilateral Trade Flows between European Union Countries. *Przeglad Statystyczny, 61*(3), 283–306.

Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models* (p. 648). Cambridge: Cambridge University Press.

Grossman, G. (1998). Comment on Deardorff. In J. A. Frankel (Ed.), *The Regionalization of the World Economy* (pp. 29–31). University of Chicago Press.

Gurevich, T. & Herman, P. (2018a). *The Dynamic Gravity Dataset: 1948–2016*. United States International Trade Commission Working Paper 2018–02-A. Washington, D.C. February, 1–24. https://www.usitc.gov/publications/332/working_papers/gurevich_herman_2018_dynamic_gravity_dataset_201802a.pdf. Accessed 01/07/2021

Gurevich, T. & Herman, P. (2018b). The Dynamic Gravity Dataset: Technical Documentation, United States International Trade Commission, Washington, D.C, February, 1-54. https://www.usitc.gov/data/gravity/dynamic_gravity_technical_documentation_v1_00_1.pdf

Hikari, I. (2017). Multilevel Analysis of Free Trade Agreements and Foreign Direct Investment in the Asia Pacific Region. *RIETI Discussion Paper Series* 17-E-012, February, 1–37. https://www.rieti.go.jp/jp/publications/dp/17e012.pdf

Larch, M., Monteiro, J.A., Piermartini, R. & Yotov, Y. (2019). On the Effects of GATT/WTO Membership on Trade: They are Positive and Large After All, *WTO Staff Working Papers* ERSD-2019–09, World Trade Organization (WTO), Economic Research and Statistics Division. https://www.wto.org/english/res_e/reser_e/ersd201909_e.pdf

Maas, C. J. M., & Hox, J. (2005). Sufficient Sample Sizes for Multilevel Modelling. *Methodology European Journal of Research Methods for the Behavioral and Social Sciences, 1*(3), 86–92.

Novak, M., & Marko, P. (2017). Using a Multilevel Modelling Approach to Explain the Influence of Economic Development on the Subjective Well-Being of Individuals. *Economic Research-Ekonomska Istraživanja, 30*(1), 705–720.

Osborne, J. W. (2000). Advantages of Hierarchical Linear Modelling. *Practical Assessment, Research, and Evaluation, 7*(1), 1–3.

Oshchepkov, A., & Shirokanova, A. (2022). Bridging the Gap between Multilevel Modelling and Economic Methods. *Social Science Research, 104*(102689), 1–19.

Rose, A. K. (2000). One Money, One Market: The Effect of Common Currencies on Trade. *Economic Policy, 15*(30), 7–46.

Stegmueller, D. (2013). How Many Countries for Multilevel Modelling? A Comparison of Frequentist and Bayesian Approaches. *American Journal of Political Science, 57*(3), 748–761.

United States International Trade Commission (2019). Dynamic Gravity Dataset. Available at https://sites.google.com/view/peterherman/dynamic-gravity-dataset. Last accession 01/07/2021.

World Trade Organization (2019). Structural Gravity Database. Available at https://www.wto.org/english/res_e/reser_e/structural_gravity_e.htm. Last accession 01/07/2021.

Yotov, Y. V. (2012). A Simple Solution to the Distance Puzzle in International Trade. *Economic Letters, 117*(3), 794–798.